Achilleas Zapranis ·
Antonios Alexandridis

Weather Derivatives

Modeling and Pricing Weather-Related Risk

Springer

Weather Derivatives

Antonis K. Alexandridis • Achilleas D. Zapranis

Weather Derivatives

Modeling and Pricing Weather-Related Risk

Springer

Antonis K. Alexandridis
School of Mathematics,
Statistics and Actuarial Science
University of Kent
Canterbury, UK

Achilleas D. Zapranis
Department of Accounting and Finance
University of Macedonia
Thessaloniki, Greece

ISBN 978-1-4614-6070-1 ISBN 978-1-4614-6071-8 (eBook)
DOI 10.1007/978-1-4614-6071-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012952584

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

In the last decade, a new category of financial derivatives was developed, namely, weather derivatives. These financial products were developed by energy and utility companies as an effective tool for hedging the volumetric risk of the energy units sold, rather than the price risk of each unit. Weather derivatives are financial instruments that can be used in order to reduce risk associated with adverse or unexpected weather conditions. The payoff of a weather derivative depends upon the index of an underlying weather measure such as rainfall, temperature, humidity, or snowfall. In contrast to the classical financial markets, the weather market is incomplete in the sense that the underlying assets cannot be stored or traded.

The weather market is one of the fastest developing markets. However, many investors are hesitant to actively enter the market. The main reasons are the difficulty in pricing the traded financial products as well as the difficulty in accurately modeling the underlying weather variables. In addition, the number of publications arising from the academic community is limited. Finally, most investors lack the knowledge of the existence of the weather market and the benefits that it can provide.

The aim of this book is to provide a concise and rigorous treatment of the stochastic modeling of weather market as well as to significantly contribute to the existing literature. Presenting a pricing and modeling approach for weather derivatives written on various underlying weather variables will help investors and companies to accurately price weather derivatives and will help them to effectively hedge against weather risk.

First, the basic aspects of the weather market are discussed. More precisely, we discuss about the purpose of weather derivatives and the history of the weather market, we present the investors of this market, and we point out the differences between weather derivatives and insurance as well as the main problems of the market. Next, we discuss methods for data preprocessing. More analytically, methods for cleaning the data as well as identifying and model trends, jumps, and discontinuities are presented. Third, we present the available modeling and pricing

methodologies of weather derivatives. Understanding the weaknesses and advantages of each method will help us build a model that accurately describes the dynamics of the weather variables. Fourth, linear and nonlinear, nonparametric methods are presented in modeling temperature wind and precipitation. Fifth, the models that developed for the weather variables are used in order to provide analytical pricing formulas for the various securities that are traded in the weather market. Next, we discuss how our modeling and pricing methodologies can be improved using meteorological forecasts. Finally, the notion of basis risk is explained and discussed. In the sense of weather derivatives, basis risk has two components arising from both the choice of weather station where a derivative contract is written, as well as the relationship between the hedged volume and the underlying weather index.

We tried to make the material accessible and readable without excessive mathematical requirements, for example, at a level of advanced MBA or Ph.D. students and industry professionals with a background in financial econometrics. Moreover, there is an introduction – tutorial – to data preprocessing to acquaint non-statisticians to the basic principles and a similar but more extensive introduction to stochastic calculus for scientists of non-finance area, firstly introducing the basic theorems and gradually building up to more complex frameworks. This introductory chapter to stochastic calculus will present all the necessary material to the reader to help him understand and follow the sections about the pricing of the weather derivatives.

Familiarity with the operations of the capital markets will help (e.g., to understand the mechanics of the various weather securities), but it is not a prerequisite. The book will take the reader to the level where he is expected to be able to apply the proposed methodologies in modeling and pricing weather derivatives in any desired location on different weather variables.

Through extensive examples and case studies, this book provides a step-by-step guide for modeling and pricing various weather derivatives written on different weather variables. The content is written in an easy to understand way, and the methodologies are separated in various stages for a better understanding, constituting the book a guidebook for both investors and students and helping them to apply the methodologies in order to achieve a successful weather risk management strategy and accurate weather derivative pricing.

This book is aimed to be used (*a*) by weather-sensitive companies in sectors like energy, electricity, agriculture, transportation, retail, construction, and entertainment, whose revenues are exposed to weather risk and are interested in weather risk management, and (*b*) by students in advanced postgraduate programs in finance, MBA, and mathematical modeling courses as well as in agriculture, energy sector, mathematical, meteorological, and engineering sciences that are seeking employment in the mathematical modeling and financial services industry. In addition, the book is aimed to be used (*c*) by researchers in incomplete financial

markets, in alternative investments, electricity and CO_2 markets, and in relevant Ph.D. programs and finally, (*d*) by investment professionals in investment institutions such as banks, insurance companies, security houses, fund managers, institutional investors, companies with intensive international activities, and financial consultancy firms.

During the writing stages of this book, the help of Christina Ioannidou was significant, and we would like to thank her.

Canterbury, UK Antonis K. Alexandridis
Thessaloniki, Greece Achilleas D. Zapranis

Contents

List of Abbreviations

AccHDD	Cumulative heating degree day
AccCDD	Cumulative cooling degree day
ACF	Autocorrelation function
ADF	Augmented Dickey–Fuller
AIC	Akaike's information criterion
AR	Autoregressive
ARCH	Autoregressive conditional heteroskedasticity
ARFIMA	Autoregressive fractional integrated moving average
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
AROMA	Autoregressive on moving average
BP	Back-propagation
BE	Backward elimination
BIC	Bayesian information criterion
BM	Brownian motion
BS	Bootstrap
CAR	Continuous autoregressive
CAT	Cumulative average temperature
CAWS	Cumulative average wind speed
CET	Central European time
CDD	Cooling degree day
CDF	Cumulative density function
CME	Chicago mercantile exchange
CR	Cumulative rainfall
CV	Cross-validation
DAT	Daily average temperature
DAWS	Daily average wind speed
ECAD	European climate assessment & dataset
ECMWF	European centre for medium-range weather forecasts

EEX	European energy exchange
ENSO	El Niño–Southern oscillation
FBM	Fractional brownian motion
FPE	Akaike's final prediction error
FT	Fourier transform
GARCH	Generalized autoregressive conditional heteroskedastic
GCV	Generalized cross-validation
HBA	Historical burn analysis
HDD	Heating degree day
HU	Hidden unit
IPOCID	Independent prediction of change in direction
JB	Jarque–Bera
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
LLF	Log likelihood function
MAE	Mean absolute error
MAPE	Mean absolute percentage error
Max AE	Maximum absolute error
MC	Monte Carlo
MLE	Maximum likelihood estimation
MPR	Minimum prediction risk
MRBM	Mean reverting Brownian motion
MSE	Mean square error
NCDC	National climatic data center
NCEF	National center of environmental forecasts
NESDIS	National environmental satellite, data and information service
NIG	Normal inverse Gaussian
NMSE	Normalized mean square error
NN	Neural network
NOAA	National oceanic and atmospheric administration
NYMEX	New York mercantile exchange
OPEC	Organization of petroleum exporting countries
ORBS	Orthogonalized residual based selection
OTC	Over-the-counter
OU	Ornstein–Uhlenbeck
PAC	Pacific rim
PACF	Partial autocorrelation function
PCA	Principal component analysis
PDF	Probability density function

POCID	Prediction of change in direction
POS	Position of sign
RBFN	Radial basis functions networks
RBS	Residual based selection
RMSE	Root mean square error
SAROMA	Seasonal autoregressive on moving average
SBP	Sensitivity based pruning
SSO	Stepwise selection by orthogonalization
WA	Wavelet analysis
WDC	World data center for meteorology
WN	Wavelet network
WRT	With respect to
WTI	West Texas intermediate

Chapter 1
The Weather Derivatives Market

1.1 Introduction

Weather derivatives are financial instruments that can be used by organizations or individuals as part of a risk management strategy to reduce risk associated with adverse or unexpected weather conditions. Just as traditional contingent claims, whose payoffs depend upon the price of some fundamental, a weather derivative has an underlying measure such as rainfall, temperature, humidity, or snowfall. The difference from other derivatives is that the underlying asset has no value and it cannot be stored or traded while at the same time the weather should be quantified in order to be introduced in the weather derivative. To do so, temperature, rainfall, precipitation, or snowfall indices are introduced as underlying assets.

Today, weather derivatives are being used for hedging purposes by companies and industries, whose profits can be adversely affected by unseasonal weather or, for speculative purposes, by hedge funds and others interested in capitalizing on those volatile markets.

The purpose of this book is to develop a model that explains the dynamics of the various weather variables. A model that describes accurately the evolution of these variables can be later used to derive closed form solutions for the pricing of weather derivatives on various indices. This majority of the material in this book is focused on temperature since the majority of the traded weather derivatives are written on temperature indices. Our findings and proposals can be very useful not only to researchers but also to traders, hedging companies, and new investors.

The rest of the chapter is organized as follows. In Sect. 1.2, the basic aspects of the weather market are discussed. More precisely, in Sect. 1.2.1, the purpose of weather derivatives is presented. The history of the weather marker is presented in Sect. 1.2.2, while in Sect. 1.2.3, the investors that are actively involved in the weather market are shown. In Sect. 1.2.4, various weather securities are described. The differences between weather derivatives and insurance are presented in Sect. 1.2.5. The concept of the basis risk is introduced in Sect. 1.2.6. In Sect. 1.3, the related markets to the weather market are described. More precisely in Sect. 1.3.1 the electricity market is

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_1,

described while in Sect. 1.3.2 the oil and gas markets are described. In Sect. 1.4, the main problems of the weather market are presented. The purpose and the usefulness of this book are analytically described in Sect. 1.5. Finally, in Sect. 1.6, an outline of the book and an overview of each chapter are presented.

1.2 The Weather Market

In this section, the basic aspects of the weather market are discussed. More precisely, the purpose of weather derivatives, the history of the weather marker, the investors that are actively involved in the weather market, the weather securities, the differences between weather derivatives and insurance, the basis risk, and finally the common approaches for pricing temperature derivatives are described.

1.2.1 The Purpose of Weather Derivatives

Weather derivatives are financial instruments whose payoffs depend upon the value of some underlying weather index. The underlying weather index can be rainfall, temperature, humidity or snowfall, or any other weather variable. Weather derivatives are used by organizations or individuals as part of a risk management strategy to reduce risk associated with adverse or unexpected weather conditions.

In general, weather derivatives are designed to cover non-catastrophic weather events. Rainy or dry and warm or cold periods which are expected to occur frequently can cause large fluctuation on the revenues of a particular company. A company that uses weather derivatives as a part of its hedging strategy can eliminate the risk related to weather. As a result, the volatility of the year-to-year profits will be significantly reduced. Jewson et al. (2005) present various reasons why this is important. First, low volatility in revenues reduces the risk of great losses and bankruptcy. Second, it decreases the volatility in the share price of the company while it increases the share price. Finally, the interest rate that the company can borrow money is reduced.

Government organizations can also use weather derivatives, in local or national level, in order to avoid unexpected raise in their running costs.

In Jewson et al. (2005) and Cao and Wei (2003), various examples of weather hedging are presented. Weather can affect the revenues of a company directly by affecting the volume of sales. An amusement park that wants to hedge against rainy days in which fewer visitors will be attracted can enter a weather contract written on rainfall. Similarly, an electricity company that wants to avoid a reduced demand in electricity due to a warm winter can use a temperature derivative. A ski resort could use weather derivatives to hedge against a reduced snowfall which will attract fewer visitors. On the other hand, government organization can use weather derivatives in

Table 1.1 Industries with weather exposure and the type of risk they face

Hedger	Weather type	Risk
Agricultural industry	Temperature/ precipitation	Significant crop losses due to extreme temperatures or rainfall
Air companies	Wind	Cancellation of flights during windy days
Airports	Frost days	Higher operational costs
Amusement parks	Temperature/ precipitation	Fewer visitors during cold or rainy days
Beverage producers	Temperature	Lower sales during cool summers
Building material companies	Temperature/ snowfall	Lower sales during severe winters (construction sites shut down)
Construction companies	Temperature/ snowfall/rainfall	Delays in meeting schedules during periods of poor weather
Energy consumers	Temperature	Higher heating/cooling costs during cold winters and hot summers
Energy industry	Temperature	Lower sales during warm winters or cool summers
Hotels	Temperature/ precipitation	Fewer visitors during rainy or cold periods
Hydroelectric power generation	Precipitation	Lower revenue during periods of drought
Municipal governments	Snowfall	Higher snow removal costs during winters with above-average snowfall
Road salt companies	Snowfall	Lower revenues during low snowfall winters
Ski resorts	Snowfall	Lower revenue during winters with below-average snowfall
Transportation	Wind/snowfall	Cancellation of ship services due to wind or buses due to blocked roads

order to avoid an increase in the costs of cleaning roads in case of snowfall or icy days.

Weather can also affect the revenues or induce costs to the company indirectly, for example, a construction company that experiences delays when constructors cannot work due to weather. Similarly, cancellation of flights due to weather conditions can cause large costs to airlines. In Table 1.1, various industries that are exposed to weather risk and the particular risk that they face are presented.

Trading strategies vary from company to company, and weather derivatives can be used to create profitable investment portfolios in a number of ways (Jewson 2004). High possible returns while keeping the risk very low can be obtained by a portfolio that contains weather derivatives and commodity trades because of the correlation between the weather and commodity prices. Alternatively, adding weather derivatives on a stock portfolio will reduce its risk because of the lack of correlation between the weather derivatives and the wider financial markets. Finally, a diversified portfolio of weather derivatives can give good return for very low risk because of the many different and uncorrelated weather indices on which weather derivatives are based, (Jewson 2004).

1.2.2 The Weather Market History

The necessity of weather products resulted to the creation of a weather market which developed very quickly. Since their inception in 1996, weather derivatives have known a substantial growth. The first parties to arrange for, and issue weather derivatives in 1996, were energy companies, which after the deregulation of energy markets were exposed to weather risk.

Energy and utility companies already had tools for hedging the price of the energy unit. However, as the competition was increasing, the demand in energy was uncertain. Weather affects both short-term demand and long-term supply of energy. A particular pattern of weather conditions, like a warming trend, can affect the long-term supply (Cao and Wei 2003). In addition, weather anomalies could result to severe changes in the price of energy and gas. Therefore, weather derivatives were developed as an effective tool for hedging the volumetric risk, rather than the price risk (Muller and Grandi 2000).

The effects of unpredictable seasonal weather patterns had previously been absorbed and managed within a regulated, monopoly environment. The deregulated environment together with the close association between the short-term demand for energy and the weather conditions created a fertile environment for weather derivatives and the development of the weather market (Cao and Wei 2003).

The first transaction in the weather derivatives market took place between 1996 and 1997. The weather transaction was executed by Aquila Energy as a weather option embedded in a power contract (Considine 2000). The first public weather derivative transaction was between Koch Energy and Enron in 1997 in order to transfer the risks of adverse weather. The deal was concerning a temperature index for Milwaukee for the winter of 1997–1998. Since then, the weather market has quickly expanded. In the following years, transaction in Europe, Asia, and Australia took place.

In September 1999, the Chicago Mercantile Exchange (CME) launched the first exchange-traded weather derivatives. In Fig. 1.1, a categorization of the financial derivatives traded in the CME is presented. CME's contracts represent the first exchange-traded, temperature-based weather derivatives (Cao and Wei 2003). The CME offered new weather derivatives in various cities in the USA, attracting more participants. Initially, weather derivatives were offered in 10 cities which were chosen based on population, the variability in their seasonal temperatures, and the activities seen in over-the-counter (OTC) markets. The regulatory system offered by the CME helped the market to evolve. The CME eliminated the default risk. Moreover, the transparency on the transactions was increased since the prices of the contracts were public. Consequently, the weather market attracted new participants.

In 2004, the national value of CME weather derivatives was $2.2 billion and grew tenfold to $22 billion through September 2005, with open interest exceeding 300,000 and volume surpassing 630,000 contracts traded. However, the OTC market was still more active than the exchange, so the bid-ask spreads were quite large.

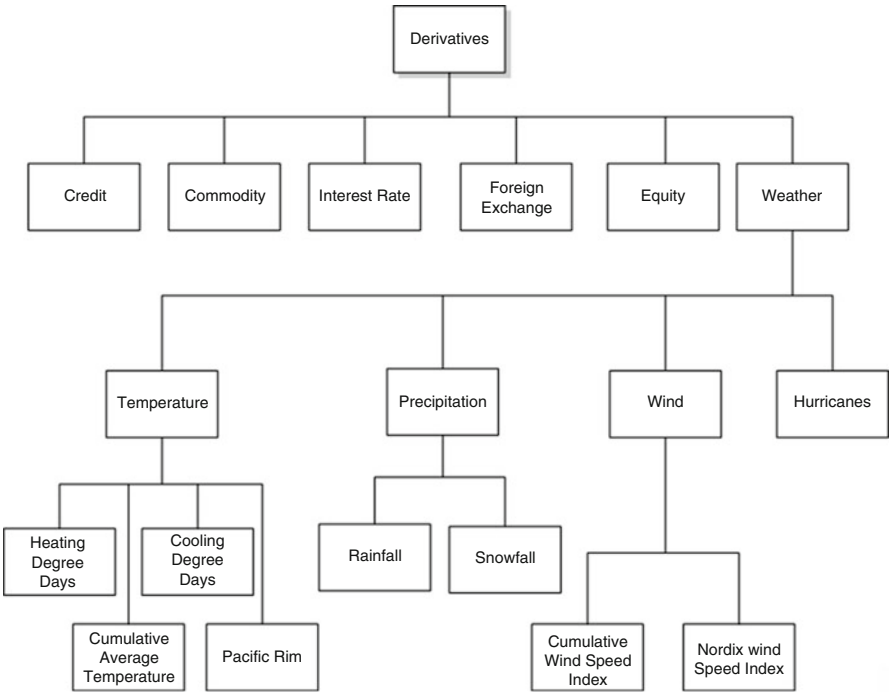

Fig. 1.1 Categorization of financial derivatives

According to the annual survey by the Weather Risk Management Association (WRMA 2009), the estimated national value of weather derivatives – OTC and exchange-traded – traded in 2008/2009 was $15 billion, compared to $32 billion the previous year and $45 billion in 2005–2006. However, there was a significant growth compared to 2005 and 2004 (Ceniceros 2006). According to CME, the recent decline reflected a shift from seasonal to monthly contracts.

Although the overall number of contracts decreased, following the general decline in financial markets, the weather market continues to develop, broadening its scope in terms of geography, client base, and interrelationship with other financial and insurance markets. In Asia, the number of contracts in 2009 rose to 250% compared to the period of 2007–2008. In Europe, there were 34,068 contracts traded in 2008–2009 compared to the previous year's 25,290 (WRMA 2010).

The weather derivatives market is organized as any other financial market. Hedgers and speculators are involved on transactions. Transaction between hedgers and speculators takes place in the primary market. In the secondary market, speculators trade between themselves.

The group of hedgers consists of companies who buy weather derivatives to hedge the weather risk in their businesses, while the group of speculators consists of banks, insurance companies, reinsurance companies, and hedge funds. Speculators

are involved in trading weather derivatives in order to make a profit rather than to hedge their risks.

Today, weather derivatives can be structured in order to cover almost any weather variable for various periods ranging from a week to several years.

1.2.3 Market Participants

According to Challis (1999) and Hanley (1999), nearly $1 trillion of the US economy is directly exposed to weather risk. It is estimated that nearly 30% of the US economy and 70% of the US companies are affected by weather (CME 2005). The electricity sector is especially sensitive to the temperature. According to Li and Sailor (1995) and Sailor and Munoz (1997), temperature is the most significant weather factor explaining electricity and gas demand in the United States. The impact of temperature in both electricity demand and price has been considered in many papers, including Henley and Peirson (1998), Peirson and Henley (1994), Gabbi and Zanotti (2005), Zanotti et al. (2003), Pirrong and Jermakyan (2008), and Engle et al. (1992). Hence, it is logical that energy companies are the main investors of the weather market. In 2004, the 69% of the weather market was consisting of energy companies. As more participants were entering the market, the energy companies were corresponding to 46% of the weather market in 2005.

Agricultural companies are greatly affected by weather conditions. However, only recently, companies from the agricultural sector started to participate in the weather market. The willingness to pay for climate derivatives is measured in Edwards and Simmons (2004) and Simmons et al. (2007). Under a general class of mean–variance utility functions with constant absolute risk aversion, they conclude that there is a demand for climatic hedging tools by wheat farmers. In Asseldonk (2003), Dubrovsky et al. (2004), Edwards and Simmons (2004), Harrington and Niehaus (2003), Hess et al. (2002), Lee and Oren (2007), Myers et al. (2005), Simmons et al. (2007), and Turvey (2001), the impact of the weather risk management for agricultural and agri-business is discussed.

Transportation, public utilities, retail sales, amusement and recreation services, and construction sectors are also very sensitive to weather (Dutton 2002). Figure 1.2 presents the participation of various industry sectors in the weather derivatives market. It is clear that until 2005, the weather derivatives markets were dominated by energy companies. However, as weather derivatives gain popularity, new players enter the market especially from agriculture and retail sectors.

The development of the weather market draws new members whose profits do not depend on weather conditions, like insurers and reinsurers, investment banks, and hedge funds.

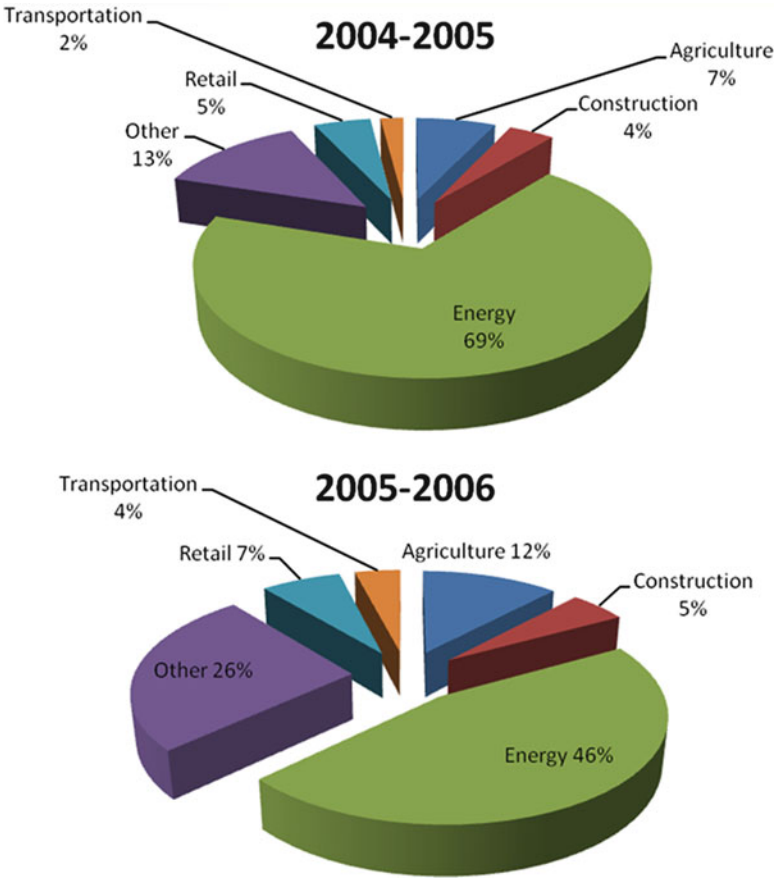

Fig. 1.2 Weather derivative potential by sector in 2004–2005 and 2005–2006 (Data obtained from WRMA. www.wrma.org)

Investment banks understood the potential of weather derivatives as a financial risk management product that they could cross sell along with other financial products for hedging interest rate or currency risks.[1] Finally, some commodity traders and hedge funds saw opportunities to trade weather on a speculative basis, or to take advantage of arbitrage opportunities relative to other energy or agricultural commodities.[2]

[1] Climetrix, http://www.climetrix.com/WeatherMarket/MarketOverview/default.asp

[2] Climetrix, http://www.climetrix.com/WeatherMarket/MarketOverview/default.asp

1.2.4 Weather Securities

The list of traded contracts on the weather derivatives market is extensive and constantly evolving. The CME offers various weather futures and option contracts. They are index-based products geared to average seasonal and monthly weather in 46 cities[3] around the world – 24 in the USA, 10 in Europe, 6 in Canada, 3 Australian, and 3 in Japan. At the end of 2009, the CME trades weather products written on the following 10 European cities: Amsterdam, Barcelona, Berlin, Essen, London, Madrid, Oslo, Paris, Rome, and Stockholm. In the USA, there are contracts for the following 24 cities: Atlanta, Baltimore, Boston, Chicago, Cincinnati, Colorado Springs, Dallas, Des Moines, Detroit, Houston, Jacksonville, Kansas City, Las Vegas, Little Rock, Los Angeles, Minneapolis–St. Paul, New York, Philadelphia, Portland, Raleigh, Sacramento, Salt Lake City, Tucson, and Washington D.C. Also, there are 6 Canadian cities, Calgary, Edmonton, Montreal, Toronto, Vancouver, and Winnipeg; 3 Australian cities, Brisbane, Melbourne, and Sydney; and finally, there are 3 Japanese cities, Hiroshima, Tokyo, and Osaka.

However, over 95% of the contracts are written on temperature heating degree days (HDD), cooling degree days (CDD), pacific rim, and cumulative average temperature (CAT) indices.

In Europe, CME weather contracts for the summer months are based on an index of CAT. The CAT index is the sum of the daily average temperatures (DATs) over the contract period. The average temperature is measured as the simple average of the minimum and maximum temperature over 1 day.

In the USA, Canada, and Australia, CME weather derivatives are based on the HDD or CDD index. A HDD is the number of degrees by which the daily temperature is below a base temperature, and a CDD is the number of degrees by which the daily temperature is above the base temperature. The base temperature is usually 65 °F in the USA and 18 °C in Europe and Japan. HDDs and CDDs cannot be negative, and usually, they are accumulated over a month or over a season. The CME also trades HDD contracts for the European cities.

For the three Japanese cities, weather derivatives are based on the Pacific Rim Index. The Pacific Rim Index is simply the average of the CAT index over the specific time period.

A weather contract is specified by the following parameters (Alaton et al. 2002):

- The contract type
- The strike or future price
- The tick size
- The maximum payout
- The contract period
- The underlying index (CAT, HDDs, rainfall, snowfall)
- Weather station from which the underlying variable data are obtained
- A premium paid from the buyer to the seller (negotiable)

[3] The number of cities that the CME trades weather contracts at the end of 2009

Weather derivatives are based on standard derivative structures such as puts, calls, swaps, collars, straddles, and strangles. As in the classical financial derivatives, the payout of these contracts depends on the strike price (the value at which the underlying index may be bought or sold) and the tick size (the smallest increment of the index that leads to a payout amount). Usually, the payout of the contract is capped. A cap in the payout is added in order to protect the two parties against extreme adverse weather conditions. In option derivatives, a premium must be given from the buyer to the seller. The premium is the price of the option.

All contracts have a defined start date and end date that constrains the period over which the underlying index is calculated. The period of the contract can range from 1 week to several years. In CME monthly and seasonal contracts are traded. Some contracts have more specific periods such as the measurements of the underlying index are considered only in working days and not at weekends.

In the contract, the underlying index must be specified. The underlying index is based on a weather variable and defines the payoff of the contract. Usually, contracts are written on CDDs, HDDs, or CAT over a specified period. Some derivatives are based on event indexes which count the number of times that temperature exceeds or falls below a defined threshold over the contract period. Similar indexes are also used for other variables, for example, cumulative rainfall or the number of days on which snowfall exceeds a defined level.

All weather contracts are based on the actual observations of weather at one specific weather station. A backup station is used in the case the main station fails. Most transactions are based on a single station, although some contracts are based on a weighted combination of readings from multiple stations and others on the difference in observations at two stations.

1.2.5 Weather Derivatives and Insurance

In the past, insurance contracts and catastrophe bonds were widely used by companies in weather-sensitive industry sectors. Like insurance contracts, the purpose of weather derivatives is to protect the buyer of the contract against adverse weather conditions. In other words, weather derivatives also provide insurance against fluctuations of the weather conditions. However, a closer inspection of these two products reveals many differences.

The first difference is the weather events that each tool covers. Insurance contracts are written on rare weather events such as extreme cold or heat and hurricanes or floods. These events are highly liked to create great catastrophes with huge impact on the revenues of the company. In contrast, weather derivatives can protect a company from recurrent weather conditions with large probability of occurrence. Unlike insurance and catastrophe-linked instruments, which cover high-risk and low-probability events, weather derivatives usually shield revenues against low-risk and high-probability events (e.g., mild or cold winters).

Claiming compensation from an insurance company usually is time consuming and expensive. The insured party must first prove that the weather had catastrophic

effects on his company while the outcome depends on the subjective opinion of each regulator. On the other hand, in the case of weather derivatives, the company receives the profit of the contract immediately. In addition, there is no need for a catastrophe to occur on the company in order to receive the compensation. Weather derivatives are based on objective criteria like the index of the temperature, the rainfall, or any other underlying index which is accurately measured on a predefined weather station.

Another advantage of weather derivatives is the additional freedom that they offer to the buyer in contrast to the insurance contracts. Hedging the impact of the weather on the competitive companies using weather derivatives is possible. For example, an agricultural company on area A can hedge against weather effects in a different area B where a competitive company is established. Favorable weather conditions in area B will result to the increase of the quantity and quality of a particular agricultural product in area B. Consequently, the demand and price for this particular product from the company in area A will decrease.

Finally, since weather derivatives are financial instruments, a weather derivative can be later sold in a third party, for speculative reasons, before the expiration day of the contract.

Companies, especially, on the agriculture and energy sector can significantly benefit from the advantages that weather derivatives offer as a weather risk management tool, (Hess et al. 2002; Pirrong and Jermakyan 2008; Simmons et al. 2007; Turvey 2001).

1.2.6 Basis Risk

Weather risk is unique in that it is highly localized, and despite great advances in meteorological science, it still cannot be predicted precisely and consistently. Risk managers often face unique basis risks arising from both the choice of weather station where a derivatives contract is written as well as the relationship between the hedged volume and the underlying weather index (Manfredo and Richards 2009). We will refer to the first as spatial or geographical basis risk, while to the second as basis risk.

The exchange-traded weather derivatives eliminated the default risk while at the same time the liquidity and the transparency increased. On the other hand, investors who wish to trade weather derivatives outside the list of the traded cities in CME face a spatial risk.

Geographical basis risk results from the distance between the hedging company and the site at which the weather measurement takes place. Geographical basis risk can reach critical levels in some cases (Rohrer 2004). As the distance between a hedging company and the measurement weather station of the weather derivative increases, the demand for weather derivative decreases (East 2005; Edwards and Simmons 2004).

It is expected that spatial risk will always be positive. However, Woodard and Garcia (2008) show that weather derivatives from a variety of stations around the hedging company can improve the hedging effectiveness. Using nonlocal derivatives for a weather variable that is highly spatially correlated, the hedging strategy obtained may be as good as the one obtained using locally derived contracts (Woodard and Garcia 2008).

In many studies, energy and weather are considered highly correlated. Hence, companies from the energy sector are extensively using weather derivatives to hedge both the price and volumetric risk of energy demand (Gabbi and Zanotti 2005; Henley and Peirson 1998; Pirrong and Jermakyan 2008). Moreover, weather derivatives are used for the valuation of gas and CO_2 emissions contracts (Bataller et al. 2006; Zanotti et al. 2003; Geman 1999).

However, these two variables, energy and temperature (or any other weather variable), are not perfectly correlated. The payoff of the weather derivative depends on the weather index, and it is unlikely that the payoff will compensate exactly for the money lost due to weather (Jewson et al. 2005). As a result, a risk is induced on the hedging strategy, called basis risk. As the correlation between the weather index and the financial loss increases, it is expected of the basis risk to decrease.

The study and understanding of spatial and basis risk will draw new participants to the weather market.

1.3 Weather Derivatives and Related Markets

In this section, we will examine the markets that are closely related to the weather market. The energy market, for example, electricity, oil, and gas, has similar characteristics to the weather market. Many power companies trade also weather derivatives due to the high correlation of the underlying variables between the two markets. Energy companies are among the most active and sophisticated users of derivatives (Hull 2003). Many energy products trade in both the OTC and on exchanges.

1.3.1 The Electricity Market

As it was mentioned earlier, the weather market emerged by contracts developed by energy companies. At the same time, the majority of the participants in the weather market consist of companies from the energy sector. On the other hand, temperature and electricity consumption and prices are highly correlated.

After the deregulation of the energy market in the early 1990s, the energy companies developed financial derivatives on electricity price in order to hedge themselves against excess production and limited consumption of electricity. Electricity is the same as any weather variable in the sense that it cannot be stored.

Table 1.2 Wholesale electricity markets

Location	Power exchange	Website
Scandinavia	Nord Pool Spot	http://www.nordpoolspot.com/
France	Powernext	http://www.powernext.fr/
Germany	EEX	http://www.eex.com/
Great Britain	Elexon	http://www.elexon.co.uk/
India	PXIL	http://www.powerexindia.com/
USA	PJM	http://www.pjm.com/
USA	NYMEX	http://www.cmegroup.com/company/nymex.html
USA	New York market	http://www.nyiso.com/

Furthermore, it has to be consumed exactly at the same time as it is produced. As a result, the electricity market cannot be considered complete, and the arbitrage-free pricing approach cannot be applied. Hence, there is a need in energy market for real time balancing between supply and demand. The excess power is sold to another control area through a transmission line. Although supply and demand are the two key factors determining the price, transmission capacity and costs also play a role.

Each area has to secure its supply of electricity and the stability of its network. This is the role of a transmission system operator. The system operator in a deregulated market has to be a noncommercial organization and independent with regard to the market participants. The operator should not own any generating assets that could benefit from its decisions. In addition, the independent system operator has the superior, physical ruling and control of the energy system in his area.

In Table 1.2, the electricity exchanges in the world are presented. The larger are the Nord Pool that covers the area of the Scandinavia, the European Energy Exchange (EEX) located in Germany, and the NYMEX and PJM in the USA which is world's largest wholesale electricity market. Each power exchange has its own pricing mechanism, products, and settlement principals.

The contracts traded in the weather market are divided in two categories. In the first category, physical delivery of electricity is needed, while in the second one, there is cash settlement of the contracts.

The day-ahead market is the main market where contracts between buyer and seller have physical delivery of energy the next 24 h. Usually, a utility company has the role of the buyer. Each day, the buyer bids for price and the amount of electricity that he needs for the following day, hour by hour. Similarly, the seller, usually an electricity production unit, announces the price and the amount of electricity he can deliver the next day, hour by hour. At the Nord Pool, at 12:00 CET is the time of gate closure for bids. The offer orders are aggregated to form the supply curve, while bid orders are aggregated to form the demand curve. The price is set where the curves for sell and buy intersect for each hour and the price is announced to the market at 13:00 CET, and after that, the contracts are settled. From 00:00 CET the next day, the settled contracts are delivered physically hour by hour.

Each market is divided in different control areas. Hence, in case of congestion of power flow, area prices are calculated as mechanism to relieve the power grid. More specifically, the price in area with the excess electricity is lowered in order to lead to

an increase in purchase and a decrease in sale. On the other hand, the price is increasing in the area with lower-electricity production, so the participants in this area will sell more and purchase less.

There is also the so-called intraday market. In Nord Pool, this market is called Elbas and helps to secure the necessary balance between supply and demand in the power market for northern Europe. Balancing between supply and demand is also secured in the day-ahead market. However, there is a possibility that some incident between the closing time and the delivery at the next day may break this balance. In the intraday market, participants can trade electricity close to real time prices in order to bring the market back to balance.

Elbas is a continuous market, and participants can trade 24 h a day until 1 h before delivery. Prices are set based on the best prices. The prices are ranked where highest-buy price and lowest-sell price are the best prices.

The participants in the electricity market are energy producers and retailers, traders, brokers, financial analysts, and clearing companies. Also, large end users trade at the exchange and in order to obtain electricity directly from the market instead of doing it through a supplier.

On the other hand, contracts with financial cash settlement are used for price hedging and risk management. Usually, these contracts have duration from a day to up to year and have a time horizon up to 6 years (at Nord Pool). In these contracts, there is no physical delivery of electricity. The contracts are settled with cash, while cash settlements take place daily.

1.3.2 The Oil and Gas Market

Crude oil is considered one of the most important commodities since it constitutes a decisive factor in the configuration of prices of all the other commodities while its price fluctuation is an indication and also a cause of important changes in global economies. The rise, the stability, or the decline of crude oil prices have a direct impact in the economies of various states but also in the more general international economy. Both oil and natural gas are widely used for heating and electricity production. As any commodity, the prices of oil and gas are determined by supply and demand.

The oil and gas market is sharing a lot of characteristics with the electricity market. However, unlike energy and weather, both oil and gas can be stored and traded. The market is split between the financial market and the physical market, the price paid for actual deliveries of natural gas and individual delivery points around the United States. In general, market mechanisms, although they vary between the USA and Europe, are similar.

Both oil and gas can be traded through New York Mercantile Exchange (NYMEX). Henry Hub and Pine Prairie Energy Center is the pricing point for natural gas future contracts traded on the NYMEX. On the other hand, oil is traded

in various West Texas Intermediate (WTI), Brent, and Russian Export Blend Crude Oil indices.

Similar to the electricity market, both oil and gas demands are highly correlated to temperature. As a result, both their prices and volume are fluctuating, depending on the temperature. For example, a mild winter will result to a significant lower demand for gas and oil for heating. As a result, the consumption of oil and gas will decrease which will lead to a decrease in prices.

However, many other factors affect the final price of oil and gas. For example, global demand for petroleum products is highly seasonal, and it is higher during the winter months, when countries increase their use of distillate heating oil and residual fuels. Supply of crude oil, including both production and net imports, also shows a similar seasonal variation. Finally, OPEC's (Organization of Petroleum Exporting Countries) decisions have had considerable influence on international oil prices, for example, in the 1973 oil embargo which resulted to a great increase in prices and an economic recession throughout the world. OPEC is an intergovernmental organization of 12 oil-producing countries.[4] OPEC's objective is to coordinate and unify petroleum policies among member countries in order to secure fair and stable prices for petroleum producers; an efficient, economic, and regular supply of petroleum to consuming nations; and a fair return on capital to those investing in the industry (OPEC 2012).

1.4 Weather Derivatives Pricing and Other Issues

It is clear that the weather market is developing rapidly as more investors and participants are actively involved. Nevertheless, there are still some issues that are hampering the further development of the market. A generally accepted pricing model, like the Black–Scholes model, does not exist. Also, many companies have to deal with spatial and basis risk. Finally, the market is still relatively illiquid while practitioners and risk management companies keep weather market data private and do not publish their models.

Solving the first two problems would attract new participants in the market, and the liquidity would increase. By extending the existing list of weather indices, companies would be able to match the weather effects to their loss of revenues, and by expanding the list of cities that the CME trades, weather derivatives would reduce the spatial risk.

As we have already mentioned, in this book we focus on temperature, wind, and precipitation derivatives. A generally accepted framework for pricing temperature (or in general weather) derivatives does not exist. Most investors use the historical burn analysis (HBA) pricing methodology, (Dorfleitner and Wimmer 2010),

[4] Algeria, Angola, Ecuador, Iran, Iraq, Kuwait, Libya, Nigeria, Qatar, Saudi Arabia, the United Arab Emirates, and Venezuela

which is very easy to understand and to replicate. However, HBA is bound to be biased and inaccurate. In fact, HBA is considered to be the simplest pricing method in terms of implementation and the most prone to large pricing errors (Jewson et al. 2005).

More recent studies utilize dynamic models which directly simulate the future behavior of temperature. The estimated dynamic models can be used to derive the corresponding indices and price various temperature derivatives. Using models for daily temperatures can, in principle, lead to more accurate pricing than other alternatives. Daily models very often show great potential accuracy since they make a complete use of the available historical data. Finally, it is easy to incorporate meteorological forecasts. However, deriving an accurate model for the daily temperature is not a straightforward process. The risk with daily modeling is that small misspecifications in the models can lead to large mispricing in the contracts.

Building a dynamic model for the temperature requires the selection of a stochastic differential equation. Temperature shows seasonality in the mean and variance, so the seasonality component must be accurately modeled. It is important that a precise estimate of the speed of mean reversion is obtained and that the distribution of the residuals is correctly selected. Finally, the appropriate length of the historical data should be chosen in order to estimate the various parameters of the discretized version of the stochastic model.

Building an algorithm that would correctly define the basic features of temperature would lead to an accurate pricing of weather derivatives.

1.5 Purpose of the Book and Readership Level

The purpose of this book is to study analytically and in depth the financial products that are traded in the weather market. This book provides a concise and rigorous treatment of the stochastic modeling of weather market. Presenting a pricing and modeling approach for weather derivatives written on various underlying weather variables will help investors and companies to accurately price weather derivatives and will help them to effectively hedge against weather risk. In addition, this book aims to significantly contribute to the existing literature since only few papers have been published so far considering the modeling issues of other weather variables except temperature in the context of weather derivatives pricing. Finally, the study of the basis risk and the market price of risk expect to draw new participants to the market while it will provide additional accuracy on pricing decreasing the bid-ask spreads.

This book is trying to link the mathematical aspects of the modeling procedure of the weather variables to the financial markets and the pricing of the weather derivatives. Through extensive examples and case studies, this book will provide a step-by-step guide for modeling and pricing various weather derivatives written on different weather variables. The content is written in an easy-to-understand way,

and the methodologies are separated in various stages for a better understanding, constituting the book a guidebook for both investors and students.

The book is innovative in making the material accessible and readable without excessive math requirements, for example, at a level of advanced MBA or Ph.D. students and industry professionals with a background in financial econometrics. There is an introduction/tutorial to data preprocessing to acquaint non-statisticians to the basic principles and a similar but more extensive introduction to stochastic calculus for scientists of non-finance area, firstly introducing the basic theorems and gradually building up to more complex frameworks. This introductory chapter to stochastic calculus will present all the necessary material to the reader to help him understand and follow the sections about the pricing of the weather derivatives.

Familiarity with the operations of the capital markets will help (e.g., to understand the mechanics of the various weather securities), but it is not a prerequisite. The book will take the reader to the level where he/she is expected to be able to apply the proposed methodologies in modeling and pricing weather derivatives in any desired location on different weather variables.

The objective of this book is twofold: First, to expand the framework that was developed in (Alexandridis 2010) for modeling and pricing temperature derivatives; second, to develop a framework for modeling and pricing weather derivatives on the remaining underlying weather variables like wind, rainfall, and snowfall. In order to accurately price weather derivatives based on weather indices, a model that will describe the evolution of the underlying weather variable should be developed. In this book, various methods for modeling the dynamics of temperature, wind, and precipitation are described. For this purpose, the daily modeling approach is pursued. First, classic linear models are described in order to develop a more advance nonlinear model. We intend to do so using wavelet analysis and wavelet networks, two state-of-the-art tools previously successfully used in various applications (Zapranis and Alexandridis 2009).

It is also important to study the so-called basis risk. Weather risk is unique in that it is highly localized, and despite great advances in meteorological science, it still cannot be predicted precisely and consistently. Risk managers often face unique basis risks arising from both the choice of weather station where a derivative contract is written and the relationship between the hedged volume and the underlying weather index. We will refer to the first as spatial or geographical basis risk, while to the second as basis risk.

It is crucial for the accurate pricing of the various weather derivatives the understanding of the market price of risk. The weather derivatives market is a classical incomplete market. The basis risk and the market price of risk are very important for the accurate pricing of weather derivatives and hedging of the company's weather risk. However, since today an extended study on basis risk and market price of risk does not exist. The study and understanding of these two sources of risk will draw new participants to the weather market and will help investors to accurately price weather derivatives.

This book sheds light to various topics regarding the modeling of the weather variables like temperature, wind, rainfall, and snowfall. This book, with its findings

and proposals, can be very useful not only to researches but also to traders, hedging companies, and new investors.

Furthermore, our study may prove to be useful to market participants that do not have the theoretical background to model weather derivatives. Our approach is described step by step and is presented in an algorithm style. Hence, anyone can apply or repeat our approach in any length of datasets in any city of preference since our approach is not restricted to the cities presented here. We hope that our analysis will help new investors to understand the weather market as well as to attract new participants to the market.

1.6 Overview of the Book

This book is separated in three parts. Part I consists of the introductory chapters. The second part, from Chaps. 4, 5, 6, 7, and 8, covers the material for modeling and pricing temperature derivatives and basis risk. The final part and in the remaining Chaps. 9 and 10, cover the modeling and pricing issues of the remaining weather variables.

Chapter 2 offers an introduction to stochastic calculus for scientists of non-finance area, firstly introducing the basic theorems and gradually building up to more complex frameworks. This introduction to stochastic calculus chapter will present all the necessary material to the reader to help him understand and follow the sections about the pricing of the weather derivatives.

In Chap. 3, the necessary data preprocessing procedures are described. More precisely, methods for cleaning the data, identifying trends, patterns, and seasonalities are presented. This chapter further examines the impact of El Niño and La Niña in the values of the DATs. Finally, a novel method for selecting the length of the historical data for analysis is analytically described.

In Chap. 4, we focus on pricing approaches of temperature derivatives. This chapter reviews in detail the most important and more often cited models proposed in literature to represent the temperature-driving process. In this chapter, the strengths and weaknesses of prior studies will be analyzed in order to develop an appropriate model that describes the temperature dynamics and that it can be used in pricing of various temperature derivatives. More precisely, the following approaches will be analyzed: the actuarial methods, the historical burn analysis, index modeling, and daily modeling. Finally, some alternative methods proposed in literature will be presented and discussed.

Chapter 5 proposes a model for the DAT. More precisely, the purpose of this chapter is to develop a model that accurately describes the dynamics of the DAT. The statistical properties of the DATs will be examined in order to propose a process that exhibits the same behavior. The proposed model will be evaluated and compared in-sample and out-of-sample in various locations (with active weather market) against models previously proposed in literature. More precisely, a statistical model will be built step by step using two state-of-the-art tools: wavelet

analysis and wavelet networks. In addition, the various dynamics that govern the temperature process will be analytically studied and discussed. In order to obtain a better understanding of the distributions of the residuals, we expand our analysis by fitting additional distributions. More precisely, a Lévy family distribution is fitted to the residuals. The Lévy family contains many known distributions as subclasses.

Based on the temperature model, we derive pricing formulas for weather derivatives on various temperature indices in Chap. 6. The model that was developed in the previous chapter described the daily dynamics of the temperature. Hence, it can be applied in order to estimate the various indices. In this chapter, the pricing formulas of various temperature derivatives will be presented first under the assumption of normal distribution and then under the assumption of a Lévy motion noise. More precisely, the pricing formulas for the following indices will be derived: CAT, Accumulated HDDs (AccHDD), Accumulated CDDs (AccCDD), and the Pacific Rim. Finally, the market price of risk will be analytically examined.

Chapter 7 describes how meteorological forecasts can be used in order to improve the forecasting ability of the temperature models. More precisely, methods for incorporating the ensemble or the probabilistic forecasts will be presented. Finally, the impact of the introduction of meteorological forecast in the pricing procedures will be discussed.

In Chap. 8, the effects of the geographical and basis risk are discussed. Risk managers often face unique basis risks arising from both the choice of weather station where a derivatives contract is written as well as the relationship between the hedged volume and the underlying weather index. In this chapter, these two sources of risk will be studied. This analysis will be very helpful for selecting an appropriate weather risk management and hedging strategy.

Chapter 9 presents methodologies for modeling and pricing the various wind indices. There are further presented the pricing approaches and the pricing formulas for the wind derivatives traded in the market.

In Chap. 10, rainfall and snowfall derivatives are discussed. Since the dynamics of rainfall and snowfall are very similar, both variables are described as precipitation. Precipitation derivatives and precipitation modeling are presented. An indifference pricing framework is presented while hedging precipitation exposure is also discussed.

References

Alaton P, Djehince B, Stillberg D (2002) On modelling and pricing weather derivatives. Appl Math Finance 9:1–20

Alexandridis A (2010) Modelling and pricing temperature derivatives using wavelet networks and wavelet analysis. University of Macedonia, Thessaloniki

van Asseldonk MAPM (2003) Insurance against weather risk: use of heating degree-days from non-local stations for weather derivatives. Theor Appl Climatol 74:137–144

Bataller MM, Tornero AT, Micó EV (2006) CO_2 prices, energy and weather http://papers.ssrn.com/sol3/papers.cfm?abstract_id=913964. Accessed 8 July 2006

Cao M, Wei J (2003) Weather derivatives: a new class of financial instruments. University of Toronto. http://www.rotman.utoronto.ca/~wei/research/JAI.pdf. Accessed 5 Jan 2006
Ceniceros R (2006) Weather derivatives running hot. Bus Insur 40(32), p. 11
Challis S (1999) Bright forecast for profits. Reactions June edition
CME (2005) An introduction to CME weather products. http://www.cme.com/edu/res/bro/cmeweather. Accessed 15 Jan 2007
Considine G (2000) Introduction to weather derivatives. http://www.cme.com/files/weather.pdf. Accessed 10 Jan 2006
Dorfleitner G, Wimmer M (2010) The pricing of temperature futures at the Chicago Mercantile Exchange. J Bank Finance. doi:10.1016/j.bankfin.2009.12.004
Dubrovsky M, Buchtele J, Zalud Z (2004) High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modelling. Climate Change 63:145–179
Dutton JA (2002) Opportunities and priorities in a new era for weather derivatives. Bull Am Meteorol Soc 83:1303–1311
East M (2005) Issues of geographical basis risk in weather derivatives for Australian wheat farmers. In: 49th annual conference of the Australian Agricultural and Resource Economics Society, Coffs Harbour, Feb 2005
Edwards M, Simmons P (2004) Preliminary results for the measurement of willingness to pay for climate derivatives. In: 48th annual conference of the Australian Agricultural and Resource Economics Society, Melbourne, Feb 2004
Engle RF, Mustafa C, Rice J (1992) Modelling peak electricity demand. J Forecasting 11:241–251
Gabbi G, Zanotti G (2005) Climate variables and weather derivatives: gas demand, temperature and the cost of weather for a gas supplier. http://www.efmaefm.org/efma2005/papers/285-zanotti_paper.pdf. Accessed 12 Aug 2008
Geman H (1999) Insurance and weather derivatives. RISK Books, London
Hanley M (1999) Hedging the force of nature. Risk Professional 1:21–25
Harrington S, Niehaus G (2003) United grain growers: enterprise risk management and weather risk. Risk Manage Insur Rev 6(2):193–217
Henley A, Peirson J (1998) Residential energy demand and the interaction of price and temperature: British experimental evidence. Energy Econ 20:157–171
Hess U, Richter K, Stoppa A (2002) Weather risk management for agriculture and agri-business in developing countries. In: Dischel RS (ed) Climate risk and the weather market: financial risk management with weather hedges. Risk Books, London
Hull CJ (2003) Option, futures and other derivatives, 5th edn. Prentice Hall, New Jersey
Jewson S (2004) Introduction to weather derivative pricing. http://ssrn.com/abstract=557831. Accessed 21 June 2004
Jewson S, Brix A, Ziehmann C (2005) Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations. Cambridge University Press, Cambridge, UK
Lee Y, Oren SS (2007) An equilibrium pricing model for weather derivatives in a multi-commodity setting. Paper presented at the decision and risk analysis, Univeristy of Texas, Dallas, 21–22 May
Li X, Sailor DJ (1995) Electricity use sensitivity and climate and climate change. World Resour Rev 3:334–346
Manfredo MR, Richards TJ (2009) Hedging with weather derivative: a role for options in reducing basis risk. Appl Financ Econ 19(2):87–97
Muller A, Grandi M (2000) Weather derivatives: a risk management tool for weather-sensitive industries. Geneva Pap Risk Insur 25(1):273–287
Myers RJ, Liu Y, Hanson S (2005) How should we value agricultural insurance contracts. In: American agricultural economics associastion annual meeting, Rhode Island, 24–27 July 2005
OPEC (2012) Brief history. http://www.opec.org/opec_web/en/about_us/24.htm. Accessed 7 May 2012

Peirson J, Henley A (1994) Electricity load and temperature issues in dynamic specification. Energy Econ 16:235–243

Pirrong C, Jermakyan M (2008) The price of power: the valuation of power and weather derivatives. J Bank Finance 32(12):2520–2529

Rohrer MB (2004) The relevance of basis risk in the weather derivatives market. In: Constantino M, Brebbia CA (eds) Computational finance and its applications. WIT Press, Southampton, pp 67–76

Sailor DJ, Munoz R (1997) Sensitivity of electricity and natural gas consumption to climate in the USA – meteorology and results for eight states. Energy Intern J 22:987–998

Simmons P, Edwards M, Byrne J (2007) Willingness to pay for weather derivatives by Australian wheat farmers. In: European association of agricultural economists, 101st Seminar, Berlin, 5–6 July 2007

Turvey CG (2001) Weather derivatives for specific event risks in agriculture. Rev Agricul Econ 23 (2):333–351

Woodard JD, Garcia P (2008) Basis risk and weather hedging effectiveness. Agricul Finance Rev 68:111–124

WRMA (2009) Celebrating 10 years of weather risk industry growth. http://www.wrma.org/pdf/WRMA_Booklet_%20FINAL.pdf. Accessed 22 Aug 2009

WRMA (2010) Weather derivatives volume plummets. www.wrma.org/pdf/weatherderivatives-volumeplummets.pdf. Accessed 22 Jan 2010

Zanotti G, Gabbi G, Laboratore D (2003) Climate variables and weather derivatives: gas demand, temperature and seasonality effects in the Italian case. http://ssrn.com/abstract=488745. Accessed 6 Jan 2003

Zapranis A, Alexandridis A (2009) Model identification in wavelet neural networks framework. In: Iliadis L, Vlahavas I, Bramer M (eds) Artificial intelligence applications and innovations III, vol IFIP 296, Springer, New York, pp 267–277

Chapter 2
Introduction to Stochastic Calculus

2.1 Introduction

Any variable whose value changes over time in an uncertain way is said to follow a stochastic process. Instead of dealing with only one possible way the process might develop over time, in a stochastic process, there is some indeterminacy described by probability distributions. This means that even if the initial condition is known, there are many possibilities the process might follow. However, some paths may be more possible.

Stochastic process can be classified in time as discrete and continuous. When the values of a stochastic process change at certain fixed points in time, then this process is called a discrete time stochastic process. On the other hand, when changes can occur at any time, this process is called a continuous time stochastic process.

In addition, stochastic processes can be classified depending on their variable. If the underlying variable can take only discrete values, then this process is called a discrete variable stochastic process. On the other hand, when the variable can take any value (within a range), this process is called continuous variable stochastic process.

Hence, four categories of stochastic processes exist:

- Discrete time–discrete variable
- Discrete time–continuous variable
- Continuous time–discrete variable
- Continuous time–continuous variable

As an example of stochastic process, we mention the following examples. The closing price of oil in the exchange is a discrete time–discrete variable process. The price of a stock between 13:00 and 14:00 is a continuous time–discrete variable process. The amount of water in a lake is a continuous time–continuous variable process.

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_2,

The purpose of this chapter is to give the necessary background in stochastic calculus. It is not meant to provide a complete background in stochastic theory but rather present all the necessary theorems and results that will be used later on in order to derive the prices of various weather derivatives on different weather indexes. The basic concepts and a collection various theorems that will provide the reader an adequate background in order to follow the rest of the book are presented. We refrain from providing proofs for most theorems since a higher stochastic and mathematical level is needed which is beyond the scope of this book. The interested reader can refer to Mikosch (1998), Shiryaev (1999), Wilmott et al. (1995), Shreve (2005), Karatzas and Shreve (1991), and Baxter and Rennie (1996).

The rest of the chapter is organized as follows. In Sect. 2.2 some basic families of stochastic process are presented. In Sect. 2.3 the Itô integral is defined, and its properties are presented. The various forms of the Itô formula are presented in Sect. 2.4, while in Sect. 2.5 some applications of the Itô formula are described. In Sect. 2.6 the Girsanov's theorem is presented, while an extension of the Girsanov's theorem, namely, the Esscher transform, is presented in Sect. 2.7. Finally, in Sect. 2.8 we conclude.

2.2 Some Stochastic Processes

During the years different stochastic processes were proposed in order to model stock movements. All four types of stochastic processes that were presented in the introduction can be used in order to model stock prices. However, the continuous time–continuous variable types of processes have been proved very useful in stock price modeling and as a result are commonly used.

A first attempt was by a Markov process. A Markov process is a type of stochastic process where the future values of the variable depend only on the present value of a variable and not on how the variable reached its present value. In other words, the history of the variable is irrelevant for the future behavior of the variable.

If we focus on the stock market, the Markov property means that any forecasts for the future values of the stock are irrelevant to the past prices of the stock. The only relevant information is the present price of the stock. The Markov property of stock prices is consistent with the weak form of market efficiency, (Hull 2003). In other words, the present price of a stock includes all the information contained in the historical prices of the stock.

A particular type of Markov stochastic process with mean rate of zero and variance rate of t per year is called a Wiener process or Brownian motion.

Definition 2.1 *A stochastic process* $B = (B(t), t \in [0, \infty))$ *is called* (*standard*) *Brownian motion or a Wiener process if the following conditions are satisfied:*

- It starts at zero: $B(0) = 0$.
- The function $t \to B(t)$ is almost surely continuous.

- It has independent increments with distribution $B(t) - B(s) \sim N(0, t - s)$, that is, the random variable $B(t) - B(s)$ is independent of the random variable $B(u) - B(v)$ when $t > s \geq u > v \geq 0$.

The definition of the Brownian motion (BM) indicates that $B(t)$ has zero mean (drift rate) and variance of t. In other words, $\mathrm{E}[B(t)] = 0$ and $\mathrm{var}(B(t)) = t$. The above definition can be expanded in order to retrieve a stochastic process with any drift rate a and variance expressed in terms of $dB(t)$.

A process with drift and variance expressed in terms of $dB(t)$ is called a generalized Wiener process or a Brownian motion and can be defined as

$$dS(t) = adt + bdB(t). \tag{2.1}$$

Hence, S has drift rate equal to a and variance b^2 per unit time. Note that in (2.1) the parameters a and b are considered constants. This is not very useful if, for example, we want to model stock price changes. Equation (2.1) implies that an investor is just as uncertain of the percentage return when the stock price is €50 as when it is €1, (Hull 2003). More precisely, for a stock price we can speculate that the expected proportional change in a short period of time remains constant, but not the expected absolute change in a short period. In addition, we can also speculate that our uncertainty regarding the magnitude of future changes in share price is proportional to the share price. Hence, an expansion of model (2.1) can be constructed:

$$dS(t) = aS(t)dt + bS(t)dB(t). \tag{2.2}$$

Model (2.2) is called geometric Brownian motion, and it is the most widely used model of stock price changes. A closer inspection of (2.2) reveals that both the drift rate and variance rate are not constants but rather functions of S that change over time. This is known as the Itô process. This is a generalized Wiener process where a and b are functions of the underlying variable x and time t. The Itô process can be written as

$$dS(t) = a(S,t)dt + b(S,t)dB(t). \tag{2.3}$$

Finally, another important category of stochastic processes is the martingale processes.

Definition 2.2 *A stochastic process X_t is called martingale if the following conditions are satisfied*:

- $\mathrm{E}[|X(t)|] < \infty$.
- $\mathrm{E}[X(t)|X(0), X(1), \ldots, X(s)] = X(s), \quad 0 \leq s < t$.

In the second condition, the expectation of the stochastic variable is conditioned in the history of $X(t)$ up to s. Similar this can be written as $\mathrm{E}[X(t)|\mathbf{F}_s] = X(s)$.

We will not dig further into martingales. The interested reader is pointed to the books of Shiryaev (1999), Mikosch (1998), and Karatzas and Shreve (1991). This section will be closed with the following theorem:

Theorem 2.1 *A Brownian motion is a martingale.*

Proof The first condition is straightforward since $\mathrm{E}[B(t)] = 0$. We focus on the second condition when $0 \leq s < t$.

$$\begin{aligned}\mathrm{E}[B(t)|\mathbf{F}_s] &= \mathrm{E}[B(t) - B(s) + B(s)|\mathbf{F}_s] \\ &= \mathrm{E}[B(t) - B(s)|\mathbf{F}_s] + \mathrm{E}[B(s)|\mathbf{F}_s]. \\ &= \mathrm{E}[B(t) - B(s)] + B(s) \\ &= B(s)\end{aligned}$$

□

2.3 Itô Integral

Before we move on presenting the Itô formula, a notion of integration of stochastic differential equation is needed. In this section some necessary introductory definitions are given, and then a definition of the Itô integral is presented. Finally, some useful properties of the Itô integral are presented. These properties will be used later on deriving weather derivative prices.

Definition 2.3 *A σ-field $\mathbf{F}$ on Ω is a collection of subsets of Ω satisfying the following conditions:*

- It is not empty: $\emptyset \in \mathbf{F}$ and $\Omega \in \mathbf{F}$.
- If $A \in \mathbf{F}$, then $A^c \in \mathbf{F}$.
- If $A_1, A_1, \ldots \in \mathbf{F}$, then

$$\bigcup_{i=1}^{\infty} A_i \in \mathbf{F} \text{ and } \bigcap_{i=1}^{\infty} A_i \in \mathbf{F}.$$

Definition 2.4 *A random variable X is called $\mathbf{F}_s$ adapted if X can be written as (a limit of a sequence of) function of $B(\tau)$ for one or more $\tau \leq s$, but not as a function of any $B(u)$ with $u>s$. A stochastic process $X(s)$ is called adapted if for each time $s \in [0, t]$ the random variable $X(s)$ is $\mathbf{F}_s$ adapted.*

The *integrator* of the Itô integral is a Brownian motion $B(t)$ with an associated filtration $\mathbf{F}_t$ and let an *integrand* process X have the following properties:

- X is adapted to the Brownian motion on $[0, T]$.
- The integral $\int_0^T \mathrm{E}[X^2(s)]ds$ is finite.

The above conditions mean that $X(t)$ is a function of $B(s)$ for $0 \leq s<t$ and that $\int_0^T \mathrm{E}[X^2(s)]ds<\infty$. Now, the Itô integral can be defined as follows:

Definition 2.5 *A stochastic process $X(s)$ is called Itô integrable on the interval $[0,t]$ if $X(s)$ is adapted for $s \in [0,t]$ and $\int_0^t \mathrm{E}[X^2(s)]ds<\infty$.*

The Itô integral is defined as the random variable:

$$I_t(X) = \int_0^t X(s)dB(s) = \lim_{n\to\infty} \sum_{i=1}^{n-1} X(s_i)(B(s_{i+1}) - B(s_i)).$$

The Itô integral has the following appealing properties. First, the Itô integral has the *adaptedness* property. For each t, I_t is $\mathbf{F}_t$-measurable. Second, it has the *linearity* property. For constants c_1, c_2, and processes $X^{(1)}$ and $X^{(2)}$ on $[0,T]$, we have that

$$\int_0^t \left[c_1X^{(1)}(s) + c_2X^{(2)}(s)\right]dB(s) = c_1\int_0^t X^{(1)}(s)dB(s) + c_2\int_0^t X^{(2)}(s)dB(s).$$

Third, $I_t(X)$ is a *martingale*. Fourth, $I_t(X)$ *has expectations zero*, $\mathrm{E}\left[\int_0^t X(s)dB(s)\right] = 0$. Fifth, it has the *continuity* property. In a more mathematical manner, this means that $I_t(X)$ is a continuous function of the upper limit of the integration t. Finally, *the Itô isometry* holds:

$$\mathrm{E}\left[\int_0^t X(s)dB(s)\right]^2 = \int_0^t \mathrm{E}\left[X^2(s)\right]ds.$$

All of these properties will be extensively used in order to solve the stochastic differential equation that describe the weather variables and then derive a price of the weather derivatives on various weather indexes.

2.4 Itô Formula

The objective of this section is to derive the Itô formula. The Itô formula derives the dynamics of an Itô process. More precisely if S is an Itô process, then applying the Itô formula, we can derive the dynamics of $f(S)$. The method presented in (Wilmott et al. 1995) will be followed.

Suppose $f(S)$ is an ordinary function which is a smooth function of S and not a stochastic one. If S varies by a small amount dS, then f varies by a small amount. If we use Taylor approximation on ordinary functions, we can write that

$$df = \frac{df}{dS}dS + \frac{1}{2}\frac{d^2f}{dS^2}dS^2 + \cdots. \tag{2.4}$$

Although dS is given by (2.3), here is simply a number. Hence, dS^2 is given by

$$\begin{aligned} dS^2 &= (a(S,t)dt + b(S,t)dB)^2 \\ &= a^2(S,t)(dt)^2 + 2a(S,t)b(S,t)dBdt + b^2(S,t)(dB)^2. \end{aligned} \tag{2.5}$$

Note that the following calculation rules apply:

$$\begin{aligned} &dB_i(t)dB_i(t) = dt \\ &dB_i(t)dB_j(t) = 0, \ i \neq j \\ &(dt)^2 = 0 \\ &dB_i(t)dt = dtdB_i(t) = 0. \end{aligned} \tag{2.6}$$

Hence, (2.5) becomes

$$dS^2 = b^2(S,t)dt. \tag{2.7}$$

Finally, replacing (2.3) and (2.7) to (2.4), we get

$$\begin{aligned} df &= \frac{df}{dS}(a(S,t)dt + b(S,t)dB) + \frac{1}{2}\frac{d^2f}{dS^2}b^2(S,t)dt \\ &= \left(a(S,t)\frac{df}{dS} + \frac{1}{2}\frac{d^2f}{dS^2}b^2(S,t)\right)dt + b(S,t)\frac{df}{dS}dB. \end{aligned} \tag{2.8}$$

This is the Itô formula in differential form. By integrating (2.8), we get the Itô formula in integral form. Now, we will give a more rigorous definition of the Itô formula.

Theorem 2.2 *Itô Formula for Brownian Motion. Let $f(x)$ be a twice differentiable function. Then,*

$$f(B(t)) = f(B(s)) + \int_s^t f'(B(u))dB(u) + \frac{1}{2}\int_s^t f''(B(u))du, \quad s<t \tag{2.9}$$

is a simple form of the Itô formula.

Theorem 2.3 *Extension I of Itô Formula. Let $f(t,x)$ be a function whose second-order partial derivatives are continuous. Then,*

$$\begin{aligned} f(t,B(t)) &= f(s,B(s)) \\ &+ \int_s^t \left[\frac{\partial f(u,B(u))}{\partial t} + \frac{1}{2}\frac{\partial f^2(u,B(u))}{\partial u^2}\right]du + \int_s^t \frac{\partial f(u,B(u))}{\partial u}dB(u), \quad s<t. \end{aligned} \tag{2.10}$$

In the previous section, a notion of the Itô process in differential form was presented. Here, we give a more rigorous definition of the Itô process in integral from.

Definition 2.6 *The stochastic process $X(t)$ is called Itô process if there exist two Itô integrable processes $a(X,t)$ and $b(X,t)$ such that*

$$X(t) = X(0) + \int_0^t a(X,s)ds + \int_0^t b(X,s)dB(s). \tag{2.11}$$

Theorem 2.4 *General Itô Formula. Assume that $f(t,x)$ is a function which is once differentiable in t and twice in x, and $X(t)$ is an Itô process. Then,*

$$\begin{aligned} f(t,X(t)) = f(s,X(s)) &+ \int_s^t b(X,u)\frac{\partial f(u,X(u))}{\partial x}dB(u) \\ &+ \int_s^t \left[\frac{\partial f(u,X(u))}{\partial t} + a(X,u)\frac{\partial f(u,X(u))}{\partial x} + \frac{1}{2}b^2(X,u)\frac{\partial f^2(u,X(u))}{\partial x^2}\right]ds, \end{aligned} \tag{2.12}$$

or equivalently

$$\begin{aligned} f(t,X(t)) = f(s,X(s)) &+ \int_s^t \left[\frac{\partial f(u,X(u))}{\partial t} + \frac{1}{2}b^2(X,u)\frac{\partial f^2(u,X(u))}{\partial x^2}\right]du \\ &+ \int_s^t \frac{\partial f(u,X(u))}{\partial x}dX(u), \end{aligned} \tag{2.13}$$

where

$$dX(u) = a(X,u)du + b(X,u)dB(u). \tag{2.14}$$

Example 2.1 *Let $f(x) = x^2$ and suppose a Brownian motion $B(t)$. Applying the Itô formula (2.9) in the interval $[0,t]$ with $f'(x) = 2x$ and $f''(x) = 2$, we have that*

$$\begin{aligned} f(B(t)) &= f(B(0)) + \int_0^t f'(B(s))dB(s) + \frac{1}{2}\int_0^t f''(B(s))ds \\ B(t) &= f(0) + \int_0^t 2B(s)dB(s) + \frac{1}{2}\int_0^t 2ds \\ B(t) &= 0 + \int_0^t 2B(s)dB(s) + \int_0^t ds \\ B(t) &= \int_0^t 2B(s)dB(s) + t. \end{aligned}$$

Rearranging the above equation, we have that

$$\int_0^t B(s)dB(s) = \frac{1}{2}B^2(t) - \frac{t}{2}.$$

The reader can notice that there is an extra term $-\frac{t}{2}$ in contrast to the classical integral. In order to verify that the above equation is correct, we take expectations on both sides. We have that by definition

$$\mathrm{E}\left[\int_0^t B(s)dB(s)\right] = 0$$

and

$$\mathrm{E}\left[\frac{1}{2}B^2(t) - \frac{t}{2}\right] = \frac{1}{2}\mathrm{E}\left[B^2(t)\right] - \frac{t}{2} = \frac{t}{2} - \frac{t}{2} = 0.$$

□

Very often the *differential form* of the Itô formula is adopted which provides a more compact notation. Hence, (2.12) can be rewritten as

$$\begin{aligned} df(t,X(t)) = &\left[\frac{\partial f(t,X(t))}{\partial t} + a(X,t)\frac{\partial f(t,X(t))}{\partial x} + \frac{1}{2}b^2(X,t)\frac{\partial f^2(t,X(t))}{\partial x^2}\right]dt \\ &+\mathrm{b}(X,t)\frac{\partial f(t,X(t))}{\partial x}dB(t), \end{aligned} \tag{2.15}$$

or equivalently from (2.13), we get

$$df(t,X(t)) = \left[\frac{\partial f(t,X(t))}{\partial t} + \frac{1}{2}b^2(X,t)\frac{\partial f^2(t,X(t))}{\partial x^2}\right]dt + \frac{\partial f(t,X(t))}{\partial x}dX(t). \tag{2.16}$$

As it is mentioned in (Shreve 2005), the mathematical meaningful form of Itô formula is in integral form. This is because solid definitions exist for both integrals in the right-hand side which are a summation of an Itô integral and a Riemann integral. However, for computations, it is more convenient to use the differential form. The differential formula becomes mathematically respectable only after we integrate it (Shreve 2005).

Next, we will present a multidimensional Itô formula. First, we define a multidimensional Brownian motion.

Definition 2.7 *Multidimensional Brownian Motion. A d-dimensional Brownian motion is a process*

$$B(t) = (B_1(t), \ldots, B_d(t))$$

with the following properties:

- Each $B_i(t)$ is one-dimensional Brownian motion.
- If $i \neq j$, then the processes $B_i(t)$ and $B_j(t)$ are independent.

Associated with a d-dimensional Brownian motion, we have filtration $\{\mathbf{F}_t\}$ such that:

- For each t, the random vector $B(t)$ is $\mathbf{F}_t$-measurable.
- For each $t \leq t_1 \leq \ldots \leq t_n$, the vector increments

$$B(t_1) - B(t), \ldots, B(t_n) - B(t_{n-1})$$

are independent of $\mathbf{F}_t$.

Theorem 2.5 *Multidimensional Itô Formula. Let* $X_i(t)$ *be Itô processes with dynamics*

$$dX_i(t) = a_i(X,t)dt + b_{i1}(X,t)dB_1(t) + \cdots + b_{1m}(X,t)dB_m(t) \tag{2.17}$$

and $i = 1 \ldots n$. *If* $\mathbf{X}(t)$ *stands for the vector* $(X_1(t), \ldots, X_n(t))'$ *and* $\mathbf{g}(t,\mathbf{x}) = \left(g_1(t,\mathbf{x}), \ldots, g_p(t,\mathbf{x})\right)'$ *is a vector-valued function of* t *and* $\mathbf{x} \in \mathbb{R}^n$, *then the stochastic dynamics of* $\mathbf{g}(t,\mathbf{X}(t))$ *are given from*

$$\begin{aligned} dg_k(t,\mathbf{X}(t)) = \frac{\partial g_k(t,\mathbf{X}(t))}{\partial t} + \sum_{i=1}^{n} \frac{\partial g_k(t,\mathbf{X}(t))}{\partial x_i} dX_i(t) \\ + \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 g_k(t,\mathbf{X}(t))}{\partial x_i \partial x_j} dX_i(t) dX_j(t). \end{aligned} \tag{2.18}$$

2.5 Applications of Itô Formula

In this section we will write the Itô Lemma in a differential form but in a more compact form. Then few examples will be presented in order to allow the reader to become familiar with the use of Itô formula. The Itô formula is probably the most important theorem in derivative pricing, and it will be used in later on in order to derive the prices of various weather derivatives.

Itô Lemma. Let x be variable that follows an Itô process:

$$dx = a(x,t)dt + b(x,t)dz \tag{2.19}$$

where dz is a Wiener process. The variable x has drift rate a and variance b^2. Then, the process $G(x,t)$ follows also an Itô process:

$$dG = \left(\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right)dt + \frac{\partial G}{\partial x}bdz \tag{2.20}$$

with drift rate

$$\left(\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right) \tag{2.21}$$

and variance

$$\left(\frac{\partial G}{\partial x}b\right)^2. \tag{2.22}$$

Example 2.2 *The Geometric Brownian Motion. Suppose that the changes of a stock are described by the following dynamics*:

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(t)$$

where μ and σ *are constants. Our objective is to solve the above stochastic differential equation and find the stochastic process that defines the dynamics of* $S(t)$. *According to notation in* (2.19), *we have that* $a = \mu S(t)$ *and* $b = \sigma S(t)$. *We define a process* $G(S,t) = \ln S(t)$. *Note that the processes* G *satisfy the initial condition,* $G(0) = \ln S(0)$.

In order to apply the Itô Lemma and find the Itô process of dG, first the partial derivatives of G must be calculated. We have that

$$G_t = \frac{\partial G(S,t)}{\partial t} = 0,$$

$$G_S = \frac{\partial G(S,t)}{\partial S} = \frac{1}{S},$$

$$G_{SS} = \frac{\partial^2 G(S,t)}{\partial S^2} = -\frac{1}{S^2}.$$

Applying formula (2.20), we have that

$$\begin{aligned} dG(S,t) &= \left(\mu S(t)\frac{1}{S(t)} - \frac{1}{2S^2(t)}\sigma^2 S^2(t)\right)dt + \sigma S(t)\frac{1}{S(t)}dB(t) \\ &= \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dB(t). \end{aligned}$$

We can integrate the above equation in the interval $[0, t]$, and we get

$$\begin{aligned} G(S,t) - G(S,0) &= \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma \int_0^t dB(s) \\ \ln S(t) - \ln S(0) &= \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma dB(t) \\ \ln \frac{S(t)}{S(0)} &= \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma dB(t) \\ \frac{S(t)}{S(0)} &= e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma dB(t)} \\ S(t) &= S(0)e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma dB(t)}. \end{aligned}$$

□

Next, a more complex problem is considered.

Example 2.3 *The Ornstein–Uhlenbeck (O–U) Process. Suppose that $X(t)$ follows the stochastic differential equation*:

$$dX(t) = cX(t)dt + \sigma dB(t),$$

where c and σ *are constants. Our objective is to solve the above stochastic differential equation. We will do that applying the Itô Lemma. According to notation in (2.19), we have that $a = cX(t)$ and $b = \sigma$. We define a process $G(x,t) = e^{-ct}X(t)$. Note that the processes G and X satisfy the same initial condition, $G(0) = X(0)$.*

In order to apply the Itô Lemma and find the Itô process of dG, first the partial derivatives of G must be calculated. We have that

$$G_t = \frac{\partial G(X,t)}{\partial t} = -ce^{ct}X,$$

$$G_X = \frac{\partial G(X,t)}{\partial X} = e^{-ct},$$

$$G_{XX} = \frac{\partial^2 G(X,t)}{\partial X^2} = 0.$$

Applying formula (2.20), we have that

$$\begin{aligned} dG(X,t) &= (-ce^{-ct}X(t) + ce^{-ct}X(t))dt + \sigma e^{-ct}dB(t) \\ &= \sigma e^{-ct}dB(t). \end{aligned}$$

We can integrate the above equation in the interval $[0, t]$, and we get

$$G(X,t) - G(X,0) = \sigma \int_0^t e^{-cs} dB(s)$$
$$e^{-ct}X(t) - X(0) = \sigma \int_0^t e^{-cs} dB(s)$$
$$X(t) = e^{ct}X(0) + \sigma e^{ct} \int_0^t e^{-cs} dB(s)$$

which is the solution of the Ornstein–Uhlenbeck process.

□

2.6 Girsanov's Theorem

In probability theory, the Girsanov's theorem (named after Igor Vladimirovich Girsanov) describes how the dynamics of stochastic processes change when the original measure is changed to an equivalent probability measure. The theorem is especially important in the theory of financial mathematics as it tells how to convert from the physical measure which describes the probability that an underlying instrument (such as a share price or interest rate) will take a particular value or values to the risk-neutral measure which is a very useful tool for pricing derivatives on the underlying.

Theorem 2.6 *Girsanov's Theorem. The stochastic process*

$$M(t) = \exp\left\{\theta B(t) - \frac{1}{2}\theta^2 t\right\},\ t \in [0,T] \tag{2.23}$$

is a martingale with respect to the natural Brownian filtration $\mathbf{F}_t = \sigma(B(s), s \leq t)$ *for* $t \in [0,T]$ *under the probability measure* P.

The relation

$$Q(A) = \int_A M(T;\omega) dP(\omega),\ A \in \mathbf{F} \tag{2.24}$$

defines a probability measure Q on $\mathbf{F}$ which is equivalent to P. Under the probability measure Q, the process $W(t)$ defined by

$$W(t) = B(t) - \theta t,\ t \in [0,T] \tag{2.25}$$

is a standard Brownian motion. The process $W(t)$ is adapted to the filtration $\mathbf{F}_t = \sigma(B(s), s \leq t)$.

Note, that the Girsanov's theorem is used in order to change the measure and eliminate the drift term in a stochastic differential equation. We will use this theorem in order to price temperature, wind, and rainfall derivatives.

Example 2.4 *Geometric Brownian Motion. Consider the stochastic differential equation*

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(s).$$

If we define

$$W(t) = B(t) + \frac{\mu}{\sigma}t,$$

we have that $dW(t) = dB(t) + \frac{\mu}{\sigma}dt$. By replacing this in to the original equation, we have that

$$dS(t) = \mu S(t)dt + \sigma S(t)\left(dW(t) - \frac{\mu}{\sigma}dt\right) = \sigma S(t)dW(t). \tag{2.26}$$

By Girsanov's theorem, $W(t)$ is a standard Brownian motion under the equivalent probability measure Q. Note that the drift rate from (2.26) was eliminated. Also the solution of (2.26) is a martingale under Q, but not under P.

2.7 Esscher Transform

In the previous section, we defined a risk-neutral probability measure $Q \sim P$ such that all tradable assets in the market are martingales after discounting. However, in this book we are interested in the weather market. Weather variables are not tradable assets. Temperature, for example, cannot be stored or be traded. As a result, we find that an infinite number of equivalent probabilities that $Q \sim P$ is a risk-neutral probability measure exist. In addition, Girsanov's theorem transform a Brownian motion under P to a new Brownian motion under Q. However, the Brownian motion implies a normal distribution which is preserved through the transformation. As it presented in the next chapters, weather variables often are not driven by a normal distribution. Rather, more complicated jump processes are often suggested which provide a better fit to the residuals of the weather variables.

In this section an extension of the Girsanov theorem is presented, namely, the Esscher transform. More precisely, the Esscher transform transforms a probability density $f(x)$ to a new probability density $f(x, \theta)$ with a parameter θ (Gerber and Shiu 1994). The transform first was introduced by Esscher (1932) and used in derivative pricing by Gerber and Shiu (1994).

Theorem 2.7 *The Esscher Transform. Let $f(x)$ be a probability density; then, its Esscher transform is defined as*

$$f(x,\theta) = \frac{e^{\theta x} f(x)}{\int_{-\infty}^{\infty} e^{\theta x} f(x) dx} \tag{2.27}$$

More generally, if P is a probability measure, then the Esscher transform of P is a new probability measure Q_θ which has density

$$\frac{e^{\theta x} f(x)}{\int_{-\infty}^{\infty} e^{\theta x} f(x) dP(x)} \tag{2.28}$$

with respect to P.

The Esscher transform is a generalization of the Girsanov's theorem. While the Girsanov's transform preserves the normality of the distribution of the Brownian motion, the Esscher transform, on the other hand, is preserving the distributional properties of a jump process. In other words, the characteristics of the jump process are also now known under the risk-neutral measure (Benth et al. 2008). Although the parameter θ changes the characteristics of the jump process, the independent increment property is still preserved (Benth et al. 2008). The Esscher transform will be used later on in order to derive temperature derivative prices when the noise process of the temperature is driven by a Lévy process.

2.8 Conclusions

In this chapter the basic notions of stochastic calculus were presented. The purpose of this chapter was to give the necessary background in stochastic calculus. It is not meant to provide a complete background in stochastic theory but rather present all the necessary theorems and results that will be used later on in order to derive the prices of various weather derivatives on different weather indexes. The basic concepts and a collection various theorems that will provide the reader an adequate background in order to follow the rest of the book were presented.

In this sense, various families of stochastic processes were presented. In addition, the Itô integral was defined in order to present the Itô formula. The Itô formula is probably the most important theorem that will be used in weather derivative pricing. Another important theorem that was presented is the Girsanov's theorem and its extension, namely, the Esscher transform.

We refrained from providing proofs for most theorems since a higher stochastic and mathematical level is needed which is beyond the scope of this book. The interested reader can refer to Mikosch (1998), Shiryaev (1999), Wilmott et al. (1995), Shreve (2005), Karatzas and Shreve (1991), and Baxter and Rennie (1996).

References

Baxter M, Rennie A (1996) Financial calculus: an introduction to derivative pricing. Cambridge University Press, Cambridge

Benth FE, Saltyte-Benth J, Koekebakker S (2008) Stochastic modelling of electricity and related markets, vol 11, Advance series on statistical science & applied probability. World Scientific, Singapore

Esscher F (1932) On the probability function in the collective theory of risk. Skandinavisk Aktuarietidskrift 15:175–195

Gerber HU, Shiu ESW (1994) Option pricing by Esscher transforms. Trans Soc Actuaries 46:99–191

Hull CJ (2003) Option, futures and other derivatives, 5th edn. Prentice Hall, New Jersey

Karatzas I, Shreve SE (1991) Brownian motion and stochastic calculus, 2nd edn. Springer, New York

Mikosch T (1998) Elementary stochastic calculus with finance in view, Advanced series on statistical science & applied probability. World Scientific, Singapore

Shiryaev AN (1999) Essentials of stochastic finance: facts, models, theory, Advanced series on statistical science & applied probability. World Scientific, Singapore

Shreve SE (2005) Stochastic calculus for finance, vol I, The binomial asset pricing models. Springer, New York

Wilmott P, Howison S, Dewynne J (1995) The mathematics of financial derivatives: a student introduction. Repr. with corrections. edn. Cambridge University Press, Cambridge

Chapter 3
Handling the Data

3.1 Introduction

Since the underlying index of weather derivatives is a weather variable like temperature, rainfall, precipitation, or snowfall, historical weather data is important for pricing these derivatives. Not only an adequate amount of data is needed but also it has to be of high quality for an appropriate pricing and risk management of the weather risk (Dunis and Karalis 2003).

High-quality weather data are used for weather risk management, weather derivative pricing, marking to market, and settlement of weather contracts. Hence, data from meteorological stations must be reported continuously and accurately. However, surprisingly, meteorological data not only offered with great limitations but at high cost too. As a result, the unavailability and the cost of meteorological data is one of the major reasons that hinder the weather market to further expand (Boissonnade et al. 2002).

In the USA the main source of weather data is the National Oceanic and Atmospheric Administration (NOAA).[1] One of the line offices of NOAA is the National Environmental Satellite, Data, and Information Service (NESDIS)[2] which runs the National Climatic Data Center (NCDC).[3] NCDC is the world's largest active archive of weather data. NCDC also operates the World Data Center for Meteorology (WDC).[4]

In Europe a reliable source of weather data is the European Climate Assessment & Dataset (ECAD).[5] ECAD which offers freely daily datasets from 12 climate variables observed at 4,641 meteorological stations in 62 countries.

[1] http://www.noaa.gov/

[2] http://www.nesdis.noaa.gov/

[3] http://www.ncdc.noaa.gov/oa/ncdc.html

[4] http://www.ncdc.noaa.gov/oa/wdc/index.php

[5] http://eca.knmi.nl/

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_3,

However, obtaining weather data from other locations around the world, like Asia, Africa, or Australia, is often costly (Banks 2002).

Easy access to high-quality weather data for long periods and for various stations would help the market to evolve and would offer liquidity. Unfortunately it is still very hard and costly to obtain this type of data. Moreover, the datasets available to researchers have many flaws, like missing data, gaps, and errors (Nelken 2000).

Some stations had to be moved during the years or to be replaced by more modern equipment; as a result, jumps will occur on the data. Another aspect is the range of the data. Previous studies use datasets containing historical data from 5 to 230 years to fit various models. However, if a very long period is considered, then the datasets will be affected by trends like urban effects. On the other hand, when studying very small datasets, there is a possibility that important dynamics of the temperature process will not be revealed which will result to an incorrect model and to mispricing of the corresponding weather contracts. Some areas exhibit urban effects due to industrial activity or pollution that result to warming trends. Finally, there are areas that are affected by extreme weather patterns like the El Niño and La Niña that must be accounted when pricing a weather derivative.

Both the NCDC and the ECAD offer raw data. In order to clean the data, a series of techniques must be applied in order to make the data appropriate for weather derivative pricing and risk management.

Concluding, data inhomogeneities can be separated in two categories. The first one includes artificial discontinuities. This category includes missing or error values due to changes in station location and changes or malfunctions of measurement instrumentation. The second category includes climatic inhomogeneities. This category includes long-term (global warming), regional (El Niño), or local trends. Both categories will be analytically discussed in the next sections.

The rest of the chapter is organized as follows. In Sect. 3.2 data cleaning and preprocessing methods are discussed. In Sect. 3.2.1 methods for filling the missing values are presented, while in Sect. 3.2.2 methods for correcting erroneous values are described. Methods for detecting and correcting jumps and discontinuities in the data are presented in Sect. 3.2.3. In Sect. 3.3 the identification and modeling of trends is analyzed. The reasons that trends appear in meteorological data are discussed in Sect. 3.3.1. Urbanization effects are explained in Sect. 3.3.1.1, while the effects of the random and predicted variability are discussed in Sects. 3.3.1.2 and 3.3.1.3, respectively. In Sect. 3.3.2 the various structures of trends are presented. More precisely, the polynomial trends are described in Sect. 3.3.2.1, while the filtering methods like the moving average and the Loess and Lowess methods are described in Sects. 3.3.2.2 and 3.3.2.3, respectively. In Sect. 3.3.2.4 an example of indentifying and removing a trend from real data is presented. The procedure of identifying and removing seasonalities from meteorological data is presented in Sect. 3.4. In Sect. 3.5 how El Nino and La Nina affect the weather is discussed. In Sect. 3.6 a procedure for selecting the length of historical data is discussed. Finally, in Sect. 3.7 we conclude.

3.2 Data Cleaning and Preprocessing

For an expansion of the weather market, providing of continuously and accurate weather data is crucial. Hence, automated systems must report continuously from meteorological stations. The stations should also have a long history and not be prone to relocation (Boissonnade et al. 2002).

A common problem on large dataset is the existence of incorrect values of the weather variables, for example, temperature. These values must be corrected before modeling the dynamics of the weather variable. Elsewise, these values can cause large mispricing errors.

For example, the daily average temperature in CME is measured as the average of the daily minimum and maximum temperature. Hence, while cleaning a temperature time series, one must check for the following:

- Missing values.
- The minimum value is larger than the maximum value.
- There are no unreasonable values of the weather variable for the particular location for the time of the year.
- Differences between nearby weather stations are not questionably large.
- Values noted as error or untrusting measurements by the data provider.
- Leap years.

3.2.1 Missing Values

One of the major problems of the data is the missing values. Missing values may occur because of a broken equipment, a break in the transmission of the weather observation, or because a weather value is lost after it is recorded (Boissonnade et al. 2002).

In Dunis and Karalis (2003) different methods for filling, the missing data were described. In the naïve approach the missing value is replaced by the temperature at the same day the previous year. This method is highly likely to produce large jumps in the temperature time series. Another approach is to fill the missing data using nearby weather stations to the one in interest. Dunis and Karalis (2003) propose and test more complex methods like the expectation maximization algorithm or the data augmentation algorithm, state space models and Kalman filter, neural networks (NNs), and principal component analysis (PCA) with the latter to outperform all other methods. However, PCA requires additional correlated cleaned temperature data (Dunis and Karalis 2003). This means that high-quality temperature data from neighboring meteorological stations is required.

In this book the procedure described below is followed in order to fill the missing values. Let T_t be the temperature at day t which value is missing. First, the average temperature of that particular day across the years is calculated denoted by Avy. Next, the average temperature of 7 days ago and 7 days after the missing value is

calculated denoted by *Avd*. Then the missing value, $T_{t,miss}$, is replaced by the average of these two parameters.

$$T_{t,miss} = \frac{\left(T_{avy,t} + T_{avd,t}\right)}{2}. \tag{3.1}$$

$$T_{avy,t} = \frac{1}{N}\sum_{yr=1}^{N} T_{t,yr}. \tag{3.2}$$

$$T_{avd,t} = \frac{\sum_{i=1}^{7} T_{t-i} + \sum_{i=1}^{7} T_{t+i}}{14}. \tag{3.3}$$

The above procedure is very easy in implementation and very efficient. More precisely a normal average is obtained by (3.2) which is balanced by the temporal temperature conditions around the missing values by (3.3). However, in some cities there are consecutive missing values. In this case, the missing values are filled using only (3.2).

In cases were a large set of consecutive weather values are missing, spatial interpolation can be used. The missing values are filled by interpolation between observation across several stations (Boissonnade et al. 2002). However, as in the case of PCA, more high-quality data are required. In addition, the selection of related station is not always a straightforward process. Nearby meteorological stations are not always the higher correlated ones. Existence of microclimates in some areas is a common issue. Hence, both the correlation and the distance between meteorological stations must be accounted for in spatial modeling.

Furthermore, each weather variable governed by different correlation dynamics. As it will be described in Chap. 8, the correlation of both rainfall and temperature changes differently as the distance increases while the correlation of each variable changes over time.

3.2.2 Erroneous Values

A common mistake on the data is when the minimum value is greater than the maximum value in a particular day. The first and simpler test is to check if the minimum of the daily values is indeed smaller than the maximum values and that their average is indeed the daily average temperature.

A second test is to check the magnitude of the daily measurements according to the location and the time of the year.

There is a possibility that a block of data is significantly different than the values of the nearby weather stations. In such cases the block of data must be removed of the data or adjusted by a weighted average of the nearby stations.

Sometimes, data providers note values as error values or not trustworthy. This means that although there is a measurement for the particular day at the weather station, this value maybe is not correct due to equipment malfunction, poorly calibrated equipments, or any other problems.

In order to clean the data for these errors, usually two approaches are followed. The first one is the meteorological approach. In this case, values from nearby stations are used. The correlation and the distances between the meteorological stations are calculated and then a weighted average for the error value is estimated. The second one is the time-series approach. In this case a model for the dataset is build (excluding the error values). Then new values are generated and replace the error values by the estimated model.

However, replacing these erroneous values is not always a straightforward approach. Different countries report weather data with different time conventions. For example, in the USA the daily average temperature is measured as the average of the daily maximum and the daily minimum as they measured from midnight to midnight. On the other hand, in UK temperature is reported daily from 09:00 a.m. to 09:00 a.m.

One more common problem but easy to solve is the existence of leap years. The usual and simpler approach is to remove the extra value, 29th of February, in order to have equal datasets each year. Another approach is to adjust the daily models in order to include the 29th of February every 4 years.

3.2.3 Jump and Discontinuities Detection

After removing or correcting incorrect values and filling any missing values, one has to identify any possible jumps in the weather time series. The most common cause of a jump in weather measurements is the change of the location of the meteorological stations. Over the years, a meteorological station may be placed in a different nearby location. For example, a meteorological station may be moved from the center of the city to the airport or the instrumentation can be lifted few meters above the ground, that is, to a different floor. Another reason that causes jumps in the data is the change of the equipment of the meteorological station. Old measurement equipments in a station may be replaced by new, more sensitive, or more modern with greater accuracy instruments. These changes result to jumps in the values of the data that can vary from small changes to several degrees. Another reason may be the changes of the surroundings of the meteorological station. A station that was in the shade may now be on sun or the construction of large building or a parking lot next to the meteorological station, etc.

Identifying jumps is essential for accurate weather pricing. The impact of the jump or the discontinuities in the pricing of weather contracts can be significant. Old measurements can have a significant impact on the estimated parameters of a model. Usually, free weather data are not corrected for such jumps. However, usually, there is a note on the time series when changes on a meteorological

station occur. On the other hand, there is possibility of an unknown jump in the time series. These jumps can be identified either visually using various graphs or using statistical test, like test for structural breaks. Another approach is to analyze data from nearby and surrounding meteorological stations (Jewson et al. 2005).

In order to identify and remove the jumps and discontinuities from the data, the following procedure must be followed. First, the potential dates that a jump occurs must be identified. Second, the possible jump must be verified. Finally, the verified jump must be quantified.

It is clear that a historical analysis of a meteorological station will significantly help the data cleaning and enhancement procedure. This information gathered by the historical analysis is generally referred to as "metadata." An accurate history analysis will provide the exact days and the physical causes of possible discontinuities in the data. Although, information about the changes in the location or the equipments of a station is reported in the metadata, information about the changes in the surrounding of the stations usually is not reported. Hence, additional statistical test must be performed in order to identify this kind of discontinuities, like examining a moving average of the temperature or examine each month of the historical data separately.

There are several approaches in order to verify a discontinuity in the weather data. These methods can be separated in two categories: methods that use data from the station in question and methods that used data from several related stations. An analytic overview of these methods can be found in Boissonnade et al. (2002). In the first category, quintiles of the distribution can be analyzed (DeGaetano and Allen 1999), or filters can be applied to the data (Rhoades and Salinger 1993), to detect large discontinuities. As it is mentioned in Boissonnade et al. (2002), these methods were applied in literature in order to identify large discontinuities, over 2°. Alternatively, in Boissonnade et al. (2002) a 12-month moving average is proposed to indentify smaller jumps.

When multiple stations are used, a weighted average of the neighboring related stations is estimated. Then the difference between this weighted average and the values of the meteorological station is used in order to identify the jumps. However, again more high-quality data from the related station are needed. Also, it is rare for the weather time series of the related stations not to contain discontinuities.

After the dates of discontinuities have been identified and the jumps have been verified, the step of quantification of the jumps is estimated. The most common approaches to adjust the data in order to remove the jumps are the following. If the jump is very small and does not have a statistical significant effect on the values, then it can be ignored. If the dataset after the jump is large, over 10 years, then data before the jump can be disregarded. Finally, if the jump is significant and past data cannot be removed, then the data prior to the jump must be adjusted according to the size of the jump. The magnitude of the jump can be estimated as the difference of the data before and after the jump. For example, yearly means before and after the discontinuity can be calculated, their difference is an estimation of the magnitude of the jump.

3.3 Identifying and Removing Trends

After cleaning the weather data, one can proceed in modeling the dynamics of the weather variable. In order to do so, one common approach is to deconstruct the data in various parts. The first choice that one had to do is to estimate the possible trend on the data. The researcher must analyze both the historical values of the weather variable as well as the metadata of the meteorological station. This analysis will provide valuable insight in understanding the reasons of the existence of the trend and its structure.

In the next sections, a discussion of the most common reasons that trends in data are evident is examined. Next, various methods that can be applied to model a trend in the data are presented together with various examples for better understanding.

3.3.1 Reasons of Trends

The last years' extreme temperatures both in summer and winter are observed in various places around the world. The cause of these extreme usually attributed to global warming. In fact it is believed that global warming affects temperature with an upward trend while at the same time the remaining weather variables are affected in various ways. For example, due to global warming, reduced rainfall in summer and excessive rainfall in winter are observed in some places, like Australia. However, global warming is not the only reason that a trend might appear in meteorological data.

Before we proceed to temperature modeling and temperature derivative pricing, we must understand the reasons that trends appear in data and how we can model them. Almost all temperature time series appear to incorporate trends in the long run, and there are a number of possible explanations for such trends. In the next section, the main reasons that trends in DATs appear will be discussed.

3.3.1.1 Urbanization

The most common reason of trends in temperature measurements is the effect of urbanization. Temperature is affected by global warming and urban effects. In areas under development, the surface temperature rises as more people and buildings concentrate. This is due to the sun's energy absorbed by the urban buildings and the emissions of vehicles, release of CO_2, industrial buildings, and cooling units. Hence, urbanization around a weather station results to an increment in the observed measurements of temperature.

3.3.1.2 Random Variability

Identifying a trend is very important. However, extra care must be taken when dealing with trend identification. Before any action is taken, one must be sure that increase/decrease of the measurements is real and not a part of a random variability.

3.3.1.3 Predicted Variability

An observed trend may be part of larger cycle. Studying 5 or 10 years of data may result to a linear trend; however, a larger historical dataset like 50 years may indicate a larger cycle where a slight increase in temperature occurs in the first 25 years and a slight decrease occurs in the last 25 years.

3.3.2 Structures of Trends

As it was already mentioned, weather is highly localized. As a result, there is a possibility that different dynamics govern the same weather variable in different locations. For example, different type of trends may characterize the change of dynamics in temperature in different cities depending on the development and urbanization effects in each city. Moreover, a different type of trend may be appropriate for a weather index and a weather variable, for example, the HDDs index and temperature.

3.3.2.1 Polynomial Trends

The usual approach is to fit a polynomial to the data in order to remove the trend. The polynomial trend is given by

$$Trend_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n. \tag{3.4}$$

The simplest polynomial is the linear trend, where $n = 1$. In this case, a simple line is fitted to the data.

$$Trend_t = a_0 + a_1 t. \tag{3.5}$$

Although the linear trend is the simplest method of detrending , it is also the most common approach. Sometimes a second-order polynomial is used, while linear trend of higher order is rarely used.

$$Trend_t = a_0 + a_1 t + a_2 t^2. \tag{3.6}$$

When a second- or higher-order polynomial is used, there is possibility that the observed trend may be part of larger cycle.

In addition, when a jump is identified in the data, a piecewise linear trend can be used:

$$Trend_t = \begin{cases} a_1 + b_1 t & t \le t_0 \\ a_2 + b_2 t & t \ge t_0 \end{cases}. \tag{3.7}$$

The polynomial trend is used in daily data as well as in detrending index series.

3.3.2.2 Moving Average

Another common approach for detrending time series and is often used in index detrending is the moving average approach. More precisely the moving average is actually a smoothing method rather than a method for identifying trends. In contrast to the polynomial method, the moving average is a nonparametric approach. Hence, a formula for the possible trend is not provided. The moving average is given by

$$Trend_t = \frac{1}{p} \sum_{i=-m}^{m} Y(t+i). \tag{3.8}$$

where $p = 2m + 1$. In this case the trend is an average of past and future data.

The moving average is a rolling window that uses p values of the data to create a new smoother time series.

3.3.2.3 Loess and Lowess Filtering Methods

A more sophisticated method is to use a filter in order to remove trends and smooth the data. One approach is the Loess and Lowess methods originally developed by Cleveland (1979) and Cleveland and Devlin (1988). Loess is built on linear and nonlinear least squares regression. The Loess is a nonparametric smoothing method. It fits a low-degree polynomial to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point. The polynomial is fitted using weighted least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away. In this book we associate Lowess with a 1st-degree polynomial model, while the Loess with a 2nd-degree polynomial.

Although Loess has many advantages, Loess requires fairly large, densely sampled datasets in order to produce good models. In addition, Loess is a computationally intensive method. Finally, the Lowess algorithm is sensitive and prone to

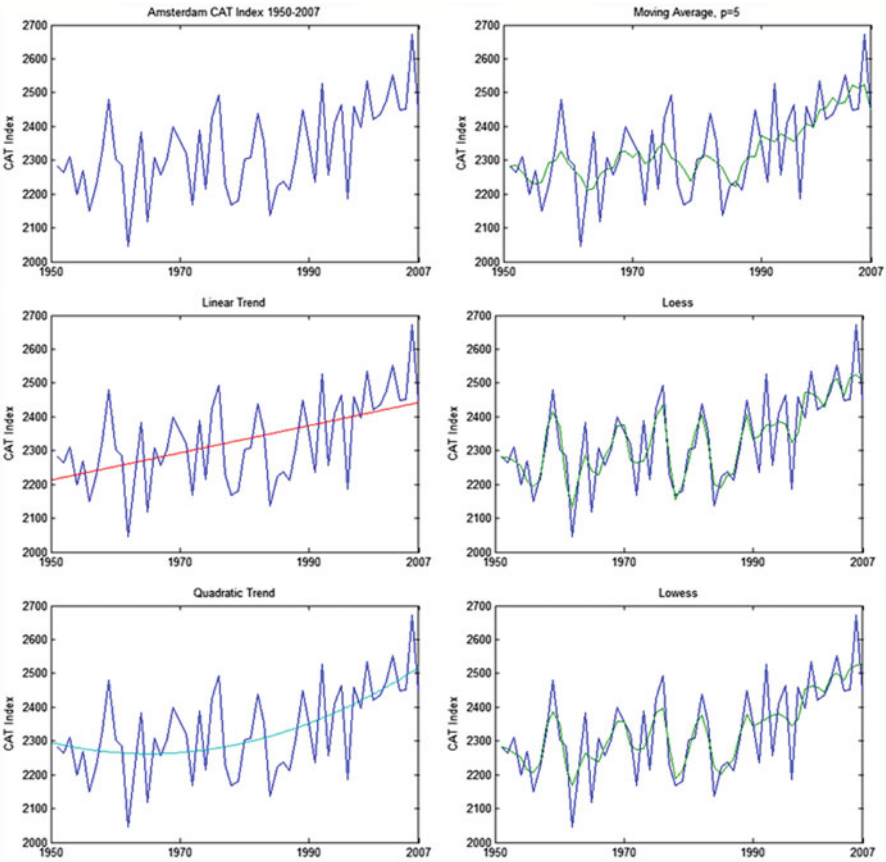

Fig. 3.1 The CAT index and various fitted trends at Amsterdam Schiphol Airport

the effect of outliers which are common in weather time series. These reasons make Loess or Lowess an unfavorable option for detrending.

3.3.2.4 An Example of Fitting a Trend in Temperature

In this section we will apply the previous detrending methods on real temperature data. Figure 3.1 shows different trends fitted to Amsterdam Schiphol Airport. The data corresponds to historical CAT indices from 1950 to 2007 for the period of May to September each year. We chose this periods since the CAT index in the CME for the European cities is measured from May to September.

In Table 3.1 the mean and standard deviation of the various detrending methods at the CAT index at the Amsterdam Schiphol are calculated: no trend, linear, quadratic, moving average, Loess, and Lowess.

Table 3.1 The mean and standard deviation of the CAT Index at Amsterdam Schiphol Airport estimated using different trends

	Trend		Residuals	
	Mean	St.D.	Mean	St.D.
No trend	2329.60	126.58	–	–
Linear trend	2329.60	66.84	0.00	107.49
Quadratic trend	2329.60	75.65	0.00	101.49
Moving average	2328.90	81.492	0.71	91.58
Loess	2329.00	101.22	0.58	67.17
Lowess	2329.50	93.96	0.13	72.95

As Jewson et al. (2005) mention, it is only worthy to model the trend if a significant number of historical data in years are used. Modeling the trend with only a few years of data can do more harm than good (Jewson et al. 2005).

3.4 Identifying and Removing Seasonalities

It is well known that weather variables and especially temperature exhibit strong seasonalities. We expect that the temperature each year will move around a seasonal mean. In order to develop a proper model that fits the dynamics of a temperature process, the seasonality must be accurately modeled. Removing the seasonal cycle, we expect to transform our temperature data to a stationary series.

In order to remove the seasonal cycle, one could average the temperature measurements in daily values. In other words, first we average all the measurements for the 1st of January, then we move on 2nd of January and so on. If the result is not smooth enough, a moving average can be used to remove any remaining noise. This approach is very fast and easy in implementation and can be used on large datasets.

Alternatively, to model the seasonality in temperature, a single sinusoid can be fitted, $S(t) = a\sin(2\pi(t+\varphi)/365)$, adjusted for the amplitude, a, and the phase, φ. The amplitude points out the difference between the highest and the mean temperature. On the other hand, the phase specifies where in the cycle the oscillation begins.

A more sophisticated approach is to model the seasonal mean by a truncated Fourier series:

$$S(t) = \sum_{i=1}^{I_1} a_i \sin(2i\pi(t-f_i)/365) + \sum_{g=1}^{J_1} b_j \cos(2j\pi(t-g_i)/365). \tag{3.9}$$

Usually, only a few harmonics are sufficient to capture the seasonal mean. In Fig. 3.2 the DAT and the fitted seasonality and a linear trend are presented for the period 1/1/1991 to 31/12/2000. The data are taken from the Schiphol airport at Amsterdam. Note that the fitted seasonality is presented together with a linear trend

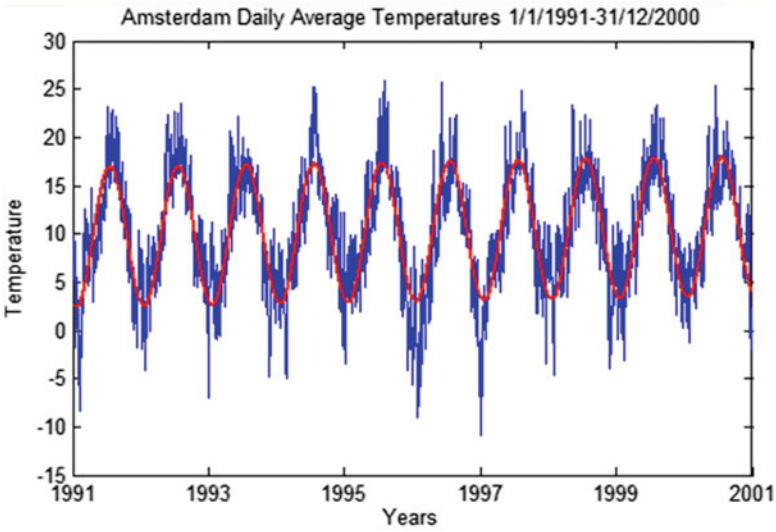

Fig. 3.2 Daily average temperature and fitted seasonality in Amsterdam Schiphol Airport

given by (3.5). The fitted seasonality is modeled by (3.9) with only one sine and only one cosine, that is, $I_1 = J_1 = 1$.

The previous method is very simple and can be easily applied in very large datasets. On the other hand, Lau and Weng (1995) confirmed periodicities in the temperature series with a period significantly greater than 1 year. Lau and Weng (1995) examined the monthly Northern Hemisphere surface temperature for the period January 1854–July 1993 using wavelet analysis (WA). They reported that the temperature has three main frequency branches: interannual (2–5 years), interdecadal (10–12 years, 20–25years, and 40–60 years), and century (~180 years) scales.

This conclusion was also reached in Zapranis and Alexandridis (2006, 2008, 2009, 2011). More precisely Zapranis and Alexandridis (2006) used the Daubechies 11 wavelet at level 11 to decompose 100 years of the average daily temperature time series of Paris. Specifically, in their paper, WA captured dynamics of temperature such as an upward trend, and periodicities expanding up to 13 years. This method will be analytically discussed in the next chapters.

3.5 El Niño and La Niña Effects

El Niño-Southern Oscillation (ENSO) is defined by prolonged differences in Pacific Ocean surface temperatures when compared with the average value. The accepted definition is a warming or cooling of at least 0.5 °C (0.9 °F) averaged over the east-central tropical Pacific Ocean. Typically, this anomaly happens at irregular

Table 3.2 Historical El Niño and La Niña episodes since 1950

El Niño	La Niña
1957–1958	1950–1951
1965–1966	1954–1957
1968–1969	1965–1965
1972–1973	1970–1971
1982–1983	1973–1974
1986–1987	1975–1976
1991–1992	1988–1989
1997–1998	1995–1996
2002–2003	1998–2000
2004–2005	2000–2001
2006–2007	2007–2008
2009–2010	2010–2011

Source: National Weather Service Prediction Center
http://www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/enso.shtml#history

intervals of 2–7 years and lasts from 9 months to 2 years. The average period length is 5 years. When this warming or cooling occurs for only 7–9 months, it is classified as El Niño/La Niña "conditions"; when it occurs for more than that period, it is classified as El Niño/La Niña "episodes."

The first signs of an El Niño are:

- Rise in surface pressure over the Indian Ocean, Indonesia, and Australia.
- Fall in air pressure over Tahiti and the rest of the central and eastern Pacific Ocean.
- Trade winds in the south Pacific weaken or head east.
- Warm air rises near Peru, causing rain in the northern Peruvian deserts.
- Warm water spreads from the west Pacific and the Indian Ocean to the east Pacific. It takes the rain with it, causing extensive drought in the western Pacific and rainfall in the normally dry eastern Pacific.[6]

The Southern Oscillation is the atmospheric component of El Niño. This component is an oscillation in surface air pressure between the tropical eastern and the western Pacific Ocean waters. In Table 3.2 the historical occurrences of El Niño and La Niña since 1950 are presented.

El Niño episodes are defined as sustained warming of the central and eastern tropical Pacific Ocean. This results in a decrease in the strength of the Pacific trade winds and a reduction in rainfall over eastern and northern Australia. It creates increased rainfall across the east-central and eastern Pacific Ocean including several portions of the South American west coast. Winters, during the El Niño effect, are warmer and drier than average in the Northwest, Northmidwest, and

[6] National Climatic Data Center (June 2009). "El Niño / Southern Oscillation (ENSO) June 2009." National Oceanic and Atmospheric Administration

Northmideast United States, and therefore those regions experience reduced snowfalls. Meanwhile, significantly wetter winters are present in northwest Mexico and the southwest United States including central and southern California, while both cooler and wetter than average winters in northeast Mexico and the southeast United States occur during the El Niño phase of the oscillation.

El Niño's effects on Europe are not entirely clear. There is some evidence that an El Niño may cause a wetter, cloudier winter in Northern Europe and a milder, drier winter in the Mediterranean Sea region.

According to the previous analysis, El Niño/La Niña not only affects various regions around the world but also affects various weather variables like temperature, wind, rainfall, and snowfall, while El Niño is credited with suppressing hurricanes. On the other hand, the 2010–2011 La Niña was one of the strongest ever observed with devastating effects on eastern Australia. Hence, when modeling the dynamics of the previous weather variables, first we must decide if the area of the meteorological station is affected by El Niño or La Niña.

If the area is affected by El Niño\La Niña, then there are two methods for estimating a weather model and predict the future evolution of the weather variable. The first, and more common, approach is to use a regime-switching model where the parameters adjusted depending on the occurrence of an ENSO year.

Alternatively, the effects of the El Niño\La Niña on the historical data can be quantified and removed. Next, on forecasting the evolution of the weather variable, when ENSO effects are predicted, the forecasted values must be adjusted according to the quantified effects of the El Niño\La Niña. However, since the impact on the weather variables of each El Niño\La Niña varies significantly, this method runs the danger of under/overestimating the effect of the El Niño\La Niña on the predicted weather variable.

Hence, first the El Niño\La Niña episodes have to be predicted and then their effects.

There are two major types of El Niño prediction models. The first category consists of simple statistical models. In these models, statistical relationships, derived from historical data, are used to predict future El Niño events. These modes are easily implemented. However, the period of the historical meteorological data is too short. Therefore, the statistical models are usually subjective. In addition, the dynamics of El Niño are not modeled. As a result, the accuracy of the prediction of the statistical models is limited.

On the other hand, hydrodynamic-coupled ocean–atmosphere models are widely used. In this method, a system of differential equation describes the behavior and the relationship between the ocean and the atmosphere. Although these models offer more accuracy, they are computationally expensive.

There are also hybrid models, where an ocean model is coupled to a statistical atmospheric model. These hybrid models try to combine the computational efficiency of the statistical models with the accuracy of the hydrodynamic models.

Given a forecast for El Niño, one has to model the impact of El Niño on the particular weather variable and meteorological station that he is interested. One approach is to use statistical models. Jewson (2004) tried to statistically relate El

Niño and US winter temperatures in New York, Chicago, Portland, and Tucson. Their results indicate that it is not possible to discern a clear relation and that it is not possible to fit a simple statistical model. Jewson (2004) studied only the impact on temperature. A study on wind, rainfall, and snowfall may reveal different results.

Alternatively, modeling the impact of El Niño is to use simulations on the hydrodynamic-coupled ocean–atmosphere models. This approach is still on early stages of developing (Jewson et al. 2005).

3.6 Selection of the Length of Historical Data

In this section the subject of selecting the length of the historical data will be discussed. Fitting a model to datasets of different size, it is expected that the estimated parameters will deviate.

If the data is of high quality, then using as many years of data usually is good choice to increase the accuracy of the estimation of the parameters. However, often this is not the case. In addition, we must ensure that we can identify any trends or seasonalities.

As it was mentioned earlier, if there is a jump on the data, it is better to remove the data prior to the jump. However, if the remaining dataset is very small, then it is better to adjust the data and use the complete set.

A first test is to plot the mean and the standard deviation of the index versus the number of years. Figure 3.3 shows the estimated mean and standard deviation of the CAT as a function of the number of years. These plots give an indication of the sensitivity of our results to the number of years but do not answer the question of how many years should be used.

In Fig. 3.4 the CAT index at Amsterdam Schiphol Airport together with the mean CAT index as a function of time is shown. Note that the clear trend was not removed from the data. As it is shown in Fig. 3.4, the historical average CAT as function of number of years is almost constant (with a slight upward slope) the last

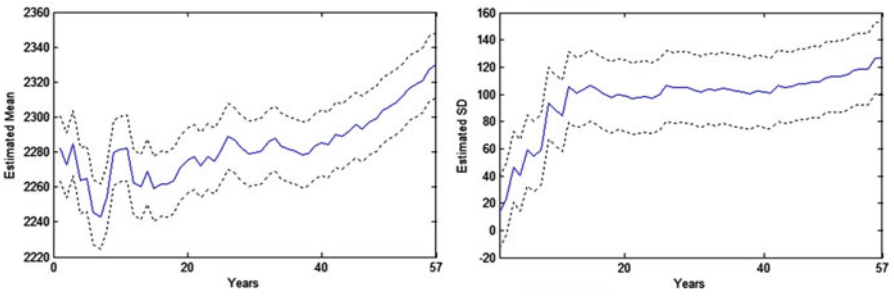

Fig. 3.3 The estimated mean CAT index at Amsterdam Schiphol Airport and the estimated standard deviation as a function of the number of years. The dashed lines show the uncertainty at +/− 1 standard deviation

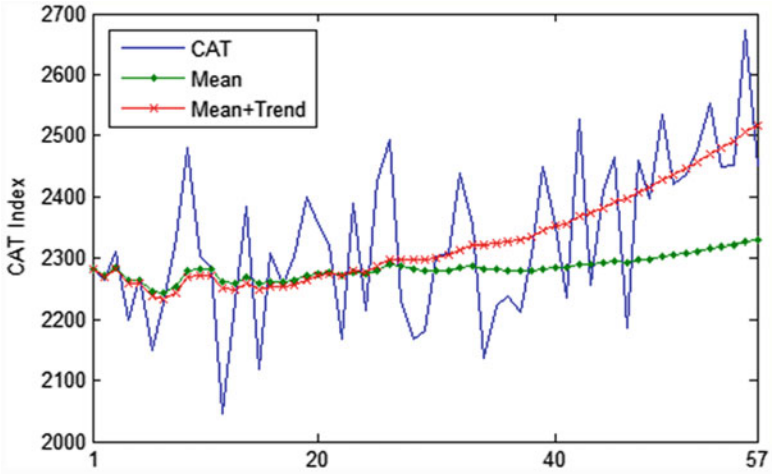

Fig. 3.4 The real CAT index at Amsterdam Schiphol Airport, the historical average CAT index versus the number of years and the historical average CAT index adjusted for a quadratic trend versus the number of years

35 years. However, it is clear that it almost always underestimates the CAT index, especially in the most recent years. In other words, the mean CAT index is heavily influenced by old and probably irrelevant data, and as a result, the future evolution of the CAT index is always underestimated. On the other hand, one can first detrend the data and then estimate the mean CAT index as a function of number of years. Then, the mean CAT index can be adjusted according to the trend to estimate the future evolution of the CAT index. In Fig. 3.4 a quadratic (second-order) polynomial is fitted to the original CAT index. Then the mean CAT index over the years is estimated. Finally, the estimated mean CAT index is adjusted according to the quadratic trend. Although this method indicates that it is best to use the last 20 years only, where a linear upward trend dominates the data, it does not answer to the question of how many years should be used.

A common approach to decide the length of the historical data is to follow the backtesting method. The backtesting method examines how good a model will work in the previous years. The length of the historical data that produces the best results is used for forecasting the weather variables and weather indices. However, this method has a main disadvantage. It assumes that past dynamics that affect a weather variable, for example, temperature, will continue to affect the temperature with the same magnitude which is not necessary true.

Market practitioners trading temperature derivatives often use only 10 years of data to estimate the temperature of index and then derive the payoff of the temperature contract. However, as it will presented later, more advanced methods produce more accurate results.

3.7 Conclusions

For a successful weather risk management, weather derivative pricing, marking to market, and settlement of weather contracts, the use of high-quality weather data is necessary. Hence, data from meteorological stations must be reported continuously and accurately. However, surprisingly, meteorological data not only offered with great limitations but at high cost too. As a result, the unavailability and the cost of meteorological data is one of the major reasons that hinder the weather market to further expand.

In this chapter various techniques for data cleaning and preprocessing were presented. First, methods for filling missing values were presented. Then, various approaches for identifying and correcting erroneous values were discussed. Finally, a framework for identifying and verifying jumps and discontinuities as well as a methodology for adjusting the data when a jump occurs was analytically presented.

On the second part of this chapter, various methods for identifying and modeling trends on temperature data were discussed. In addition to the various trends, two smoothing techniques were presented. We conclude that the choice of trend depends on the size of the dataset as well as the index that it is modeled. Finally, a weather variable (temperature) and the corresponding weather index (HDDs) on the same variable may be governed by a different trend.

Weather variables exhibit strong seasonality. We tried to model and extract the seasonal mean of daily measurements of the weather variables using truncated Fourier series.

El Niño and La Niña not only affect various regions around the world but also affect various weather variables like temperature, wind, rainfall, and snowfall, while El Niño is credited with suppressing hurricanes. First, the El Niño\La Niña episodes have to be predicted and then their effects. There are two major types of El Niño prediction models. The first category consists of simple statistical models. On the other hand, hydrodynamic-coupled ocean–atmosphere models are widely used. There are also hybrid models, where an ocean model is coupled to a statistical atmospheric model. Given a forecast for El Niño/La Niña, one has to model the impact of El Niño/La Niña on the particular weather variable and meteorological station that he is interested. If the area is affected by El Niño\La Niña, then there are two methods for estimating a weather model and predict the future evolution of the weather variable. The first is to use a regime-switching model. Alternatively, the effects of the El Niño\La Niña on the historical data can be quantified and removed.

Selecting the length of the historical data is an issue among researchers and practitioners. Fitting a model to datasets of different size, it is expected that the estimated parameters will deviate. If the data is of high quality, then using as many years of data usually is good choice to increase the accuracy of the estimation of the parameters. However, often this is not the case. A common approach to decide the length of the historical data is to follow the backtesting method. However, this method has a main disadvantage. It assumes that past dynamics that affect a

weather variable, for example, temperature, will continue to affect the temperature with the same magnitude which is not necessarily true.

Market practitioners trading temperature derivatives often use only 10 years of data to estimate the temperature of index and then derive the payoff of the temperature contract. However, as it will be presented later, more advanced methods produce more accurate results.

References

Banks E (ed) (2002) Weather risk management: markets, products and applications, Finance and capital markets. Palgrave Macmillan, Basingstoke

Boissonnade AC, Heitkemper LJ, Whitehead D (2002) Weather data: cleaning and enhancement. In: Dischel RS (ed) Climate risk and the weather market: financial risk management with weather hedges. Risk Books, London

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74(368):829–836

Cleveland WS, Devlin SJ (1988) Locally-weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc 83(403):596–610

DeGaetano AT, Allen RJ (1999) Single-station test to adjust in homogeneities in daily extreme temperature series. In: Eleventh conference on applied climatology, Dallas, Texas, 1999. American Meteorological Society, Boston, pp 193–195

Dunis CL, Karalis V (2003) Weather derivative pricing and filling analysis for missing temperature data. Derivative Use Trading Regul 9(1):61–83

Jewson S (2004) A preliminary assessment of the utility of seasonal forecasts for the pricing of U.S. temperature based weather derivatives. SSRN eLibrary

Jewson S, Brix A, Ziehmann C (2005) Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations. Cambridge University Press, Cambridge, UK

Lau KM, Weng HY (1995) Climate signal detecting using wavelet signal transform. Bull Am Meteorol Soc 76:2391–2401

Nelken I (2000) Weather derivatives – pricing and hedging. http://www.supercc.com. Accessed 3 Jan 2000

Rhoades DA, Salinger MJ (1993) Adjustment of temperature and rainfall records for site changes. Int J Clim 13:899–913

Zapranis A, Alexandridis A (2006) Wavelet analysis and weather derivatives pricing. Paper presented at the 5th Hellenic Finance and Accounting Association (HFAA), Thessaloniki, 15–16 Dec

Zapranis A, Alexandridis A (2008) Modelling temperature time dependent speed of mean reversion in the context of weather derivative pricing. Appl Math Finance 15(4):355–386

Zapranis A, Alexandridis A (2009) Weather derivatives pricing: modelling the seasonal residuals variance of an Ornstein-Uhlenbeck temperature process with neural networks. Neurocomputing 73:37–48

Zapranis A, Alexandridis A (2011) Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing. Neural Comput Appl 20(6):787–801. doi:10.1007/s00521-010-0494-1

Chapter 4
Pricing Approaches of Temperature Derivatives

4.1 Introduction

Early methods such as the actuarial method or the HBA were used to derive the price of a temperature derivative written on a temperature index without actually modeling the dynamics of the temperature. Both methods measure how a temperature derivative would perform the previous years. The average (discounted) payoff that was derived from the previous years is considered to be the payoff of the derivative.

Alternatively, one can directly model the corresponding index, namely, "index modeling," such as the HDD index, the CDD index, the CAT index, the AccHDDs index, or the AccCDDs index. A different model must be developed for each index. In literature, few papers suggest that temperature index modeling (HDD or CDD index) might be more appropriate (Davis 2001; Dorfleitner and Wimmer 2010; Geman and Leonardi 2005; Jewson et al. 2005).

Another approach to estimate the temperature driving process is to use models based on daily temperatures. Daily modeling can, in principle, lead to more accurate pricing than modeling temperature indices, (Jewson et al. 2005), as a lot of information is lost due to existing boundaries in the calculation of temperature indices by a normal or lognormal process, such as HDD being bounded by zero. On the other hand, deriving an accurate model for the daily temperature is not a straightforward process. Observed temperatures show seasonality in all of the mean, variance, distribution, and autocorrelation, and there is evidence of long memory in the autocorrelation. The risk with daily modeling is that small misspecifications in the models can lead to large mispricing of the temperature contracts (Jewson et al. 2005).

In the literature, two methods have been proposed for the modeling of the DAT, the usage of a discrete or a continuous process. Moreno (2000) argues against the use of continuous processes in the temperature modeling based on the fact that the values of temperature are in discrete form; hence, a discrete process should be used directly. Caballero and Jewson (2002), Caballero et al. (2002), Campbell and

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_4,

Diebold (2005), Cao et al. (2004), Cao and Wei (1999, 2000, 2003, 2004), Carmona (1999), Franses et al. (2001), Jewson and Caballero (2003a, b), Moreno (2000), Roustant et al. (2003a, b), Svec and Stevenson (2007), Taylor and Buizza (2002, 2004), and Tol (1996) make use of a general autoregressive moving average (ARMA) framework.

On the other hand, Alaton et al. (2002), Bellini (2005), Benth (2003), Benth and Saltyte-Benth (2005, 2007), Benth et al. (2007, 2008), Bhowan (2003), Brody et al. (2002), Dischel (1998a, b, 1999), Dornier and Queruel (2000), Geman and Leonardi (2005), Hamisultane (2006a, b, 2007, 2008), McIntyre and Doherty (1999), Oetomo and Stevenson (2005), Richards et al. (2004), Schiller et al. (2008), Torro et al. (2003), Yoo (2003), and Zapranis and Alexandridis (2006, 2007, 2008, 2009a, b) suggest a temperature diffusion stochastic differential equation. The continuous processes used for modeling daily temperatures usually take a mean-reverting form, which has to be discretized in order to estimate its various parameters. Once the parameters of the process are estimated, one can then value any contingent claim by taking expectation of the discounted future payoff. Given the complex form of the process and the path-dependent nature of most payoffs, the pricing expression usually does not have closed-form solutions. In that case, Monte Carlo (MC) simulations are used. This approach typically involves generating a large number of simulated scenarios of weather indices to determine the possible payoffs of the weather derivative. The fair price of the derivative is then the average of all simulated payoffs, appropriately discounted for the time value of money; the precision of the MC approach depends on the correct choice of the temperature process and the look back period of available weather data. In Fig. 4.1, the evolution of the weather derivatives literature using continuous stochastic differential equations is presented.

In Fig. 4.2, the main methods for estimating and modeling the temperature indices and the temperature process for weather derivative pricing can be found. In this study, we focus on daily modeling since using models for daily temperatures can, in principle, lead to more accurate pricing than modeling temperature indices since more information is obtained.

The rest of this chapter is organized as follows. In Sect. 4.2, the actuarial method that is used for the pricing of various derivatives is described. In Sect. 4.3, the HBA is discussed. In Sect. 4.4, the advantages and disadvantages of index modeling are analyzed. Approaches that modeled the temperature using daily models are presented in Sect. 4.5. More precisely, Sect. 4.5.1 presents the discrete processes used in previous studies in daily temperature modeling, while Sect. 4.5.2 presents the continuous processes. In Sect. 4.6, alternative modeling methods are presented. Finally, in Sect. 4.7, we conclude.

4.2 Actuarial Method

A pricing methodology for weather derivatives that is widely used in insurance is the actuarial (or insurance) method. In actuarial pricing, appropriate datasets of meteorological data and forecasts are used in order to derive the distribution of all

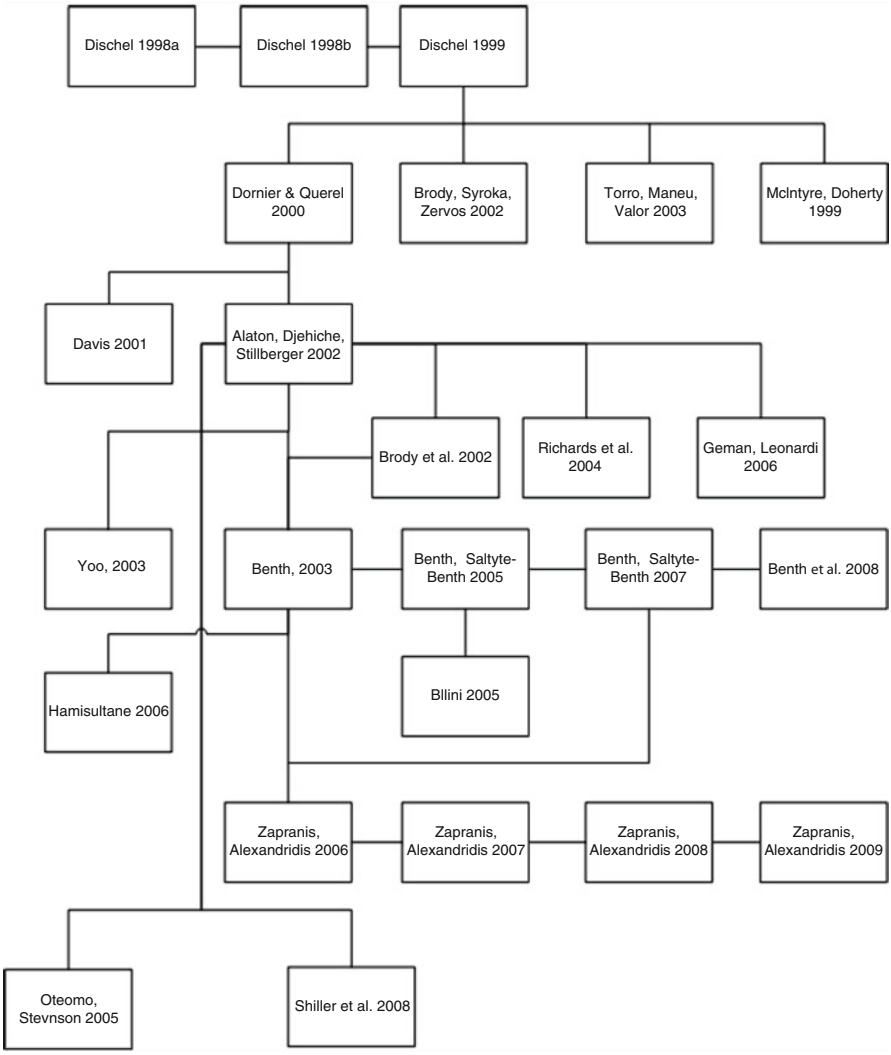

Fig. 4.1 Evolution of weather derivatives literature using continuous stochastic differential equations

possible outcomes for the settlement index (Jewson 2004), while historical data are used to calculate the expected payoff. The expected payoff is discounted at the risk-free rate to obtain the price. This method is based on statistical analysis, and it is less applicable in contracts with underlying variables that follow recurrent, predictable patterns. Since this is the case for most of the weather derivatives contracts, actuarial analysis is not considered the most appropriate pricing approach unless the contract is written on rare weather events such as extreme cold or heat. Moreover,

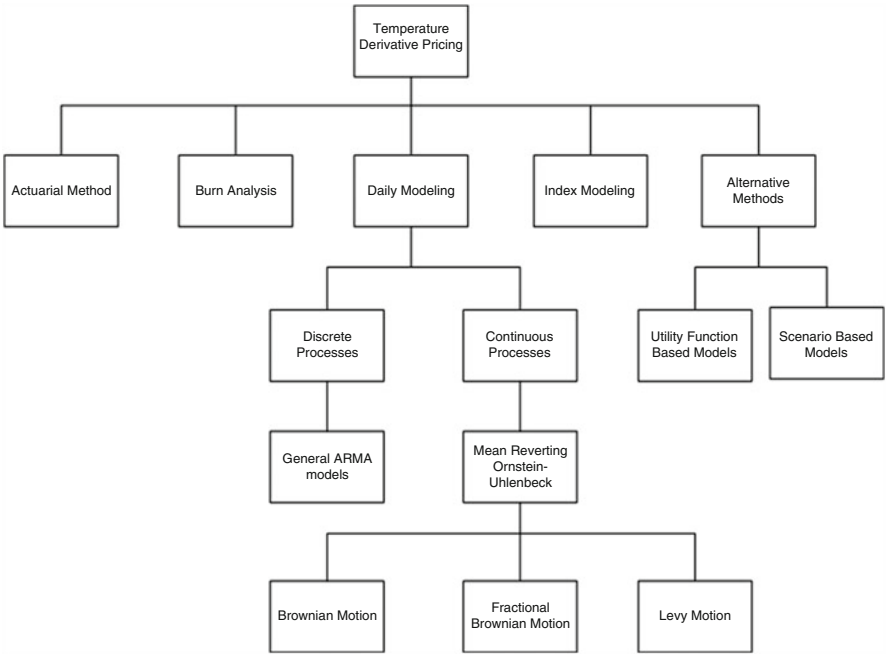

Fig. 4.2 Methods for estimating and modeling the temperature indices and the temperature process for weather derivative pricing

the estimated expected payoff is in the real world, meaning that the actuarial approach is correct only when the expected payoff from the derivative is the same in both the real and the risk-neutral world (Hull 2003, 2005).

4.3 Historical Burn Analysis

A classical approach for weather derivatives pricing is the performing of simulations based on historical data, known as HBA. More precisely, in HBA the average payoff of the weather derivatives in the past n years is computed. HBA is often considered as a benchmark approach. The main assumption of this method is that the historical record of the payoff of the weather contract gives a precise illustration of the distribution of the potential payoffs (Dischel 1999). In other words, HBA assumes that history will repeat itself with the same likelihood (Turvey 2001; Hamisultane 2008). If weather risk is calculated as the payoffs standard deviation, then the price of the contract will be $P(t) = D(t, T) \times (\mu \pm \alpha \times \sigma)$, where $D(t, T)$ is the discount factor from contract maturity T to the pricing time t, μ is the historical average payoff, σ is the historical standard deviation of the payoffs, and a is a positive number denoting risk tolerance. Often time series of length

between 10 and 30 years are used (Cao et al. 2004). HBA is very easy to calculate since there is no need to fit the distribution of the temperature or to solve any stochastic differential equations. Moreover, HBA is based in very few assumptions. Firstly, the temperature time series is assumed to be stationary. Next, the data for different years is assumed to be independent and identically distributed.

A closer inspection of a temperature time series shows that none of these assumptions is correct. It is apparent that the temperature time series contains seasonalities, jumps, and trends (Zapranis and Alexandridis 2006, 2007,2008). Moreover, there is evidence that the volatility and the average of temperature is not constant for different historical record lengths (Dischel 1999). According to the observations above and since this approach does not incorporate forecasts, this method is bound to be biased and inaccurate. Furthermore, the assumption that the data originating from different years is independent is under question (Moreno 2000). Jewson et al. (2005) suggest that these assumptions can be made if the data is cleaned and detrended, although their results indicate that pricing still remains inaccurate. Other methods, such as index and daily modeling, are considered more accurate, but still HBA is usually regarded as an acceptable first approximation of the derivative's price and is widely used by market participants. In fact, HBA is considered to be the simplest pricing method in terms of implementation and the most prone to large pricing errors (Jewson et al. 2005). Finally, similar to the actuarial method, the market price of risk related to the temperature cannot be estimated in the HBA framework (Cao et al. 2004).

4.4 Index Modeling

Early studies tried to model directly different temperature indices like the HDD, the CDD, or the CAT index. Geman and Leonardi (2005) discuss the statistical properties of both HDD and AccHDD indices. Their results from 55 years of data indicate that, in the case of the December HDDs in Paris, the hypothesis of the normal distribution is rejected, while in the case of the December AccHDD, the normality hypothesis is accepted. They conclude that modeling directly the HDDs is not appropriate. On the other hand, if someone wants to discuss the various issues related to the valuation of the contracts over a particular period, the AccHDD index can be modeled directly (Geman and Leonardi 2005).

Davis (2001) tried to model the AccHDD index using a continuous stochastic differential equation. More precisely, the AccHDD index X_t modeled by a lognormal process:

$$dX_t = \nu X_t dt + \gamma X_t dW_t \tag{4.1}$$

and at time T,

$$X_T = \exp(m_T + \gamma W_T), \tag{4.2}$$

where

$$m_T = \log X_0 + (v - \frac{1}{2}\gamma^2)T. \tag{4.3}$$

Using 11 years of data from Birmingham, England, Davis (2001) concludes that modeling the AccHDD index X_t as lognormal process is convenient but affects the pricing. Also, the choice of the initial value X_0 can significantly affect the option prices by $\pm 10\%$.

Jewson et al. (2005) tested HDD, CDD, and CAT indices in various sites in the USA. Their results indicate that almost always the normality test is accepted for seasonal contracts. However, for shorter periods, the normality hypothesis is often rejected. Moreover, Jewson et al. (2005) concludes that there is very little theory in what distribution should be used to fit the indices. As a result, there is a significant probability that a non-appropriate distribution is used. In that case, great errors will be induced in the derived estimated prices.

According to Jewson et al. (2005), most practitioners rely on the index modeling framework since it is easy to understand. Based on this observation, Dorfleitner and Wimmer (2010) recently tested the effectiveness of index modeling in forecasting temperature index and temperature futures prices. Two models were tested, one incorporating a linear trend and one without a linear trend. In both models, weather forecast was included. Their results indicate that the model with the linear trend has better forecasting ability in the case of predicting the temperature HDD and CDD indices; however, a model without a linear trend is better at forecasting the price of temperature futures. In other words, although a trend exists in the temperature HDD and CDD indices, market practitioners prefer a simpler model without a trend.

4.5 Daily Modeling

More recent studies utilize dynamic models which directly simulate the future behavior of temperature. The estimated dynamic models can be used to derive the corresponding indices and price various temperature derivatives.

Using models for daily temperatures can, in principle, lead to more accurate pricing than modeling temperature indices. Daily models very often show greater potential accuracy than the HBA (Jewson et al. 2005), since daily modeling makes a complete use of the available historical data. In the contrary, calculating the temperature index, such as HDDs, as a normal or lognormal process, a lot of information both in common and extreme events is lost (e.g., HDD is bounded by zero).

It is clear that using index modeling a different model must be estimated for each index. On the other hand, using daily modeling, only one model is fitted to the data and can be used for all available contracts on the market on the same location. Also using a daily model, an accurate representation of all indices and their distribution

can be obtained. Finally, in contrast to index modeling and HBA, it is easy to incorporate meteorological forecasts.

However, deriving an accurate model for the daily temperature is not a straightforward process. The risk with daily modeling is that small misspecifications in the models can lead to large mispricing in the contracts.

Previous studies propose that modeling DAT can be done either using a discrete or a continuous process. In the next sections, discrete and continuous daily temperature models previously proposed in the literature will be analytically discussed.

4.5.1 Discrete Process

The values of temperature are in discrete form; hence, a discrete process can be used directly (Moreno 2000). In Moreno (2000) a mean-reverting discrete process and a general AR(p) proposed by Carmona (1999) were compared. The temperature was modeled by a mean-reverting process that was given by

$$T_{i+1} = T_i + (S_{i+1} - S_i) + a(S_i - T_i) + \sigma_i \varepsilon_i, \tag{4.4}$$

where T_i is the DAT, S_i is the mean seasonal temperature, σ_i is the volatility of the temperature changes, and ε_i $\sim i.i.d$ follow the $N(0,1)$ distribution. Moreno (2000) findings from Paris-Orly and Marseille suggest that the volatility is not constant and shows seasonality. It is modeled by

$$\sigma_i = 1 + \sin^2(\theta i + \beta). \tag{4.5}$$

The temperature T_i in the second model, originally proposed by Carmona (1999), is given by

$$T_{i+1} = S_{i+1} + \mathrm{AR}(p). \tag{4.6}$$

The fitting from both models is very good with the autoregressive (AR) model (4.6) to outperform the mean-reverting model (4.4). Moreno (2000) studied the distribution of the residuals in a monthly basis and found that the distribution of the residuals for different periods is not the same. Moreno (2000) concluded that the distribution of the residuals is not constant through the year. More precisely, the residuals are independently but not identically distributed. Hence, both models cannot be used in order to simulate the temperature process.

Cao and Wei (2000) also argue about the use of diffusion processes in temperature modeling. Using one-factor diffusion processes cannot incorporate autocorrelation in the temperature, while there is a possibility that a simulated temperature path will not resemble a real one. Cao and Wei (2000) suggest a discrete process to capture the unique characteristics of DAT. Studying temperature data, Cao and Wei

(1999, 2000, 2003) and Cao et al. (2004) build their framework on the following five assumptions about DAT:

- It follows a predicted cycle.
- It moves around a seasonal mean.
- It is affected by global warming and urban effects.
- It appears to have autoregressive changes.
- Its volatility is higher in the winter than in summer.

By denoting $T_{yr,t}$ the temperature on date t in year yr, the variable $U_{yr,t}$ is the daily temperature whose mean and trend have been removed:

$$U_{yr,t} = T_{yr,t} - \hat{\overline{T}}_{yr,t}, \tag{4.7}$$

where $\hat{\overline{T}}_{yr,t}$ is the adjusted historical mean temperature and

$$\overline{T}_t = \frac{1}{m}\sum_{yr=1}^{m} T_{yr,t} \tag{4.8}$$

is the average temperature on a particular date over m years. In Cao and Wei (2000) a k-lag autocorrelation system is used for the daily temperature residuals:

$$U_{yr,t} = \sum_{i=1}^{k} \rho_i U_{yr,t-i} + \sigma_{yr,t}\varepsilon_{yr,t}, \tag{4.9}$$

$$\sigma_{yr,t} = \sigma - \sigma_1 \left|\sin\left(\frac{\pi t}{365} + \varphi\right)\right|, \tag{4.10}$$

where ρ_i is the autocorrelation coefficient for the i^{th} lag, $\varepsilon_{yr,t}$ are *i.i.d.* standard normally distributed, and $\sigma_{yr,t}$ is the volatility. The parameters can be estimated using the maximum likelihood estimation (MLE). The above model is easy to estimate and captures features of temperature such as seasonality in mean and volatility, the autocorrelation property, and uneven variations through the year (Bellini 2005). Also weather forecast can be used to enhance the predictability of the model. However, as Cao and Wei (2000) comment, the above model probably cannot be used for long time periods since long-range forecasts cannot have daily precision.

In a more recent paper, Cao and Wei (2004) adapt the framework proposed by Lucas (1978) to derive a valuation framework for temperature derivatives and to study the market price of risk. Their results indicate that the market price of risk associated to the temperature is significant and that the market price of risk affects option values much more than forward prices, mainly due to payoff specification.

Campbell and Diebold (2005) expand the model proposed by Cao and Wei (2000). They use a low-order Fourier series with autoregressive lags to model the

seasonal mean temperature. In addition, the conditional variance is allowed to exhibit seasonality in the variance as well as autoregressive effects:

$$T_t = S_t + \sum_{l=i}^{L} \rho_{t-l} T_{t-l} + \sigma_t \varepsilon_t, \tag{4.11}$$

where

$$S_t = a + \beta t + \sum_{p=1}^{P} \left(\delta_{c,p} \cos\left(\frac{2\pi p t}{365}\right) + \delta_{s,p} \sin\left(\frac{2\pi p t}{365}\right) \right), \tag{4.12}$$

$$\sigma_t^2 = \sum_{q=1}^{Q} \left(\gamma_{c,q} \cos\left(\frac{2\pi q t}{365}\right) + \gamma_{s,q} \sin\left(\frac{2\pi q t}{365}\right) \right) + \sum_{r=1}^{R} a_r \varepsilon_{t-r}^2, \tag{4.13}$$

$$\varepsilon_t \sim i.i.d.N(0,1). \tag{4.14}$$

Using Fourier series in (4.12) and (4.13), Campbell and Diebold (2005) produce a smooth seasonal pattern while reduce significantly the number of parameters that have to be estimated. Model (4.11) incorporates a linear trend that reflects dynamics such as global warming or urban effects around a meteorological station. The parameters L and P are estimated using both Akaike's and Schwarz criteria. Campbell and Diebold (2005) use temperature data from 1/1/1960 to 11/5/2001 to estimate the systems (4.11, 4.12, 4.13 and 4.14) in ten locations in the USA. The large estimated value of $L = 25$ reveals a long memory in the temperature dynamics, and the estimated value of $R = 1$ reveals autoregressive effects in the variance. Since $L = 25$, in order to obtain good estimations of the parameters, large datasets must be used. Bellini (2005) suggests that the linear trend might be part of a long-term cycle and also suggests that the quality of the trend might be deteriorate if large datasets as in Campbell and Diebold (2005) are used.

Roustant et al. (2003a, b) use a general ARMA model to calculate the price uncertainty of weather derivatives. As in Cao and Wei (1999), the temperature is modeled by a linear model with periodic variance:

$$T_t = S_t + \sigma_t Z_t, \tag{4.15}$$

where the S_t is given by (4.12) and

$$\sigma_t = a + b \cos\left(\frac{2\pi t}{365}\right) + c \sin\left(\frac{2\pi t}{365}\right) \tag{4.16}$$

and Z_t is the ARMA processes with variance 1:

$$Z_t = \varphi_1 Z_{t-1} + \ldots + \varphi_p Z_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}. \tag{4.17}$$

Their results from Paris indicate that the ARMA model reduces to a simple AR (3) model as in many studies such as in Benth et al. (2007) and Carmona (1999). The seasonal part S_t has only the two first components of the Fourier series. Their results indicate large values for the price uncertainty, especially for weather options prices. Roustant et al. (2003b) conclude that the uncertainty comes from the modeling of the trend and seasonality. Hence, a sophisticated algorithm for modeling the trend and seasonal part must be derived.

Results from Tol (1996) indicate that the volatility of temperature is not constant but shows some systematic variation. A generalized autoregressive conditional heteroskedastic (GARCH) model is used to capture this feature in temperature data from the Netherlands.

Franses et al. (2001) propose a nonlinear GARCH model for weekly temperature in the Netherlands. Their results indicate a strong asymmetry in the volatility and that the nonlinear GARCH outperforms the linear GARCH model.

Taylor and Buizza (2004, 2006) expand the works of Franses et al. (2001) and Tol (1996) by using a low-order Fourier series to model the seasonality S_t, as in Campbell and Diebold (2005). They use only 5 years of DATs from the UK to test the forecast ability of an AR–GARCH and an atmospheric model as well as the impact of the inclusion of weather ensemble forecast to the models. The AR–GARCH model is defined as

$$T_t = S_{\mu,t} + \varphi T_{t-1} + \varepsilon_t, \tag{4.18}$$

$$e_t = \sigma_t \varepsilon_t, \tag{4.19}$$

$$\sigma_t^2 = S_{\omega,t} + a\left(e_{t-1} - S_{\gamma,t}\right)^2 + \beta\sigma_{t-1}^2, \tag{4.20}$$

$$S_{\lambda,t} = \lambda_0 + \lambda_1 \sin\left(\frac{2\pi t}{365}\right) + \lambda_1 \cos\left(\frac{2\pi t}{365}\right) + \lambda_1 \sin\left(\frac{4\pi t}{365}\right) + \lambda_1 \cos\left(\frac{4\pi t}{365}\right). \tag{4.21}$$

The results from Taylor and Buizza (2004, 2006) indicate that the atmospheric model outperforms the AR–GARCH model.

Caballero et al. (2002) argue about the use of simple ARMA models since their results indicate that a slow decay in the autocorrelation function (ACF) of the temperature is evident. In order to capture the long memory in the ACF, an autoregressive fractional integrated moving average model (ARFIMA) is used. ARFIMA models are defined as the ARIMA (autoregressive integrated moving average) models proposed by Box and Jenkins (1970):

$$\Phi(L)(1-L)^d T_i = \Psi(L)\varepsilon_i, \tag{4.22}$$

where $\Phi(L)$ and $\Psi(L)$ are polynomials in the lag operator L and ε_i is a white noise process. For $0<d<\frac{1}{2}$ the process has long memory with intensity d, for $-\frac{1}{2}<d<0$ the process has short memory, while for $d \geq \frac{1}{2}$ the process is

nonstationary (Bellini 2005). An ARFIMA (p,d,q) model is equivalent to an ARMA (∞,q) model with only $p + q + 1$ parameters. Caballero et al. (2002) use DATs from the UK, and their results indicate that long-range dependence is present in their data and that ARFIMA models can accurately and parsimoniously reproduce the autocovariance structure of the observed data. However, fitting an ARFIMA model is extremely computationally expensive and prohibitively for long datasets. Moreover, the ARFIMA model proposed by Caballero et al. (2002) fails to capture the seasonality in the ACF of temperature (Jewson and Caballero 2003b). Hence, the memory is underestimated in summer and overestimated in winter (Bellini 2005).

In Jewson and Caballero (2003b), a new form of AR processes was presented, called autoregressive on moving average (AROMA) to model the DAT. In AROMA $(m_1, \ldots, m_M)$ process, the detrended and deseasonalized temperature $\tilde{T}_t$ in day t is regressed onto a number of moving averages of previous detrended and deseasonalized temperatures; all end in day $t - 1$:

$$\tilde{T}_t = \sum_{i=1}^{M} \left(a_i \sum_{j=t-1}^{t-m_i} \tilde{T}_j \right) + \varepsilon_t. \tag{4.23}$$

The AROMA process was extended to incorporate seasonality (SAROMA) by fitting a different model with different regression parameters for each day. In order to obtain good estimations of the parameters, the number of moving averages must be as small as possible (Jewson and Caballero 2003b). The regression parameter for a moving average of length m can be fitted when the length of the fitting window is significant larger than m. Jewson and Caballero (2003b) suggest the use of 4 moving averages, and for the length of the moving average, all the possible combinations with lengths up to 35 were tested. The length of each moving average was chosen according to the root mean square error (RMSE) between the real ACF and the modeled ACF. The above method runs the danger of overfitting the data. Algorithms for the optimal selection of the number of the moving averages as well as the length of each moving average must be derived. Finally, although the proposed model can capture the slow decay of the ACF, it cannot capture more rapid changes (Jewson and Caballero 2003b).

In Svec and Stevenson (2007), various models were compared in modeling and forecasting DAT. More precisely, one intraday and two daily models based on Fourier transformation of temperature as well as a wavelet reconstructed Fourier transformation were compared. The above models were a modification of the model originally proposed by Campbell and Diebold (2005). The dataset that was used covers the period from 1997 to 2005, and their results indicate that the modified models outperform the original model. Finally, Svec and Stevenson (2007) tested their data for fractionality, and their estimates indicate that Sydney DATs do not exhibit long memory.

4.5.2 Continuous Process

The continuous processes used for modeling DAT usually take a mean-reverting form, which has to be discretized in order to estimate its various parameters. Once the process is estimated, one can then value any contingent claim by taking expectation of the discounted future payoff. Given the complex form of the process and the path-dependent nature of most payoffs, the pricing expression usually does not have closed-form solutions. In that case, MC simulations are being used. This approach typically involves generating a large number of simulated scenarios of weather indices to determine the possible payoffs of the weather derivative. The fair price of the derivative is then the average of all simulated payoffs, appropriately discounted for the time value of money; the precision of the MC approach is dependent on the correct choice of the temperature process and the look back period of available weather data.

Since temperature exhibits strong and clear seasonality, most models already proposed in weather derivatives literature make use of a mean-reverting process. Most models can be written as nested forms of the following mean-reverting O–U process:

$$dT(t) = dS(t) - \kappa(T(t) - S(t))dt + \sigma(t)dB(t), \tag{4.24}$$

where T is the temperature, κ is the speed of mean reversion, $S(t)$ is a deterministic function modeling the trend and seasonality, $\sigma(t)$ is the daily volatility of temperature variations, and $B(t)$ is the driving noise process.

Dischel (1998a, b) is the first to propose a continuous stochastic model. Since temperature cannot be stored or traded, the weather market has the classical form of an incomplete market. Dischel (1998b) argues about the use of a Black–Scholes model to price weather derivatives. Weather derivatives are different from other financial derivatives in that the underlying weather index (HDD, CDD, CAT, etc.) cannot be traded. Furthermore, the corresponding market is relatively illiquid. Consequently, since weather derivatives cannot be cost-efficiently replicated by other weather derivatives, arbitrage pricing cannot directly apply to them. The weather derivatives market is a classic incomplete market, because the underlying weather variables are not tradable. When the market is incomplete, prices cannot be derived from the no-arbitrage condition, since it is not possible to replicate the payoff of a given contingent claim by a controlled portfolio of the basic securities. Consequently, the classical Black–Scholes–Merton pricing approach, which is based on no-arbitrage arguments, cannot be directly applied. In addition, market incompleteness is not the only reason; weather indices do not follow random walks (as the Black and Scholes approach assumes) and the payoffs of weather derivatives are determined by indices that are average quantities, while the Black–Scholes payoff is determined by the value of the underlying exactly at the maturity date of the contract (European options).

In Dischel (1998b), the following stochastic process is proposed to model temperature:

$$dT(t) = (\kappa S(t) + \beta T(t))dt + \gamma\tau(t)dz_1 + \delta\sigma(t)dz_2, \tag{4.25}$$

where $S(t)$ is the average seasonal historical mean, that is, $S(t) = \frac{1}{m}\sum_{i=1}^{m} T_{i,t}$. $S(t)$ represents the average temperature on a particular date across all years of the dataset. According to Dischel (1998b), $S(t)$ also represents the mean reversion parameter, while κ denotes the speed of the mean reversion which is considered constant. The random part of the process is given by the second part of equation (4.25), $\gamma\tau(t)dz_1 + \delta\sigma(t)dz_2$ where dz_1, dz_2 denote the Weiner processes corresponding to the distribution of the temperature, $T(t)$, and the distribution of the changes in temperature, $\Delta T(t)$. Dischel (1998b) makes no assumptions about the distributions dz_1, dz_2 and tries to extract them by bootstrapping the past data. Because two-parameter models can become very unstable, the stochastic simulation was limited to the temperature changes only. By using finite differences, (4.25) can be rewritten as

$$T_{n+1} = aS_{n+1} + \beta T_n + \delta\Delta T_{n,n+1}, \tag{4.26}$$

where $\Delta T_{n,n+1}$ is the randomly selected forward changes.

In a more recent paper, Dischel (1999) focuses on problems corresponding to the quality of the weather data and also to the quantity of the weather data that one must use in order to estimate the parameters of the model. Moreover, he discusses the problem of using constant volatility to a nonstationary time series. The solution that is proposed is to deconstruct and reconstruct the whole time series. A polynomial equation is proposed in order to capture the trend of the degree days. Then, the volatility of the history of the trend is calculated. A moving average scheme is used and then the time series is reconstructed using the new adjusted volatility.

McIntyre and Doherty (1999), in an attempt to model the DAT in Heathrow airport in the UK, concluded that a mean-reverting stochastic differential equation with constant volatility given by

$$dT(t) = \kappa(T(t) - S(t))dt + \sigma dB(t) \tag{4.27}$$

fits the data very well.

Dornier and Queruel (2000) use a more general ARMA model than the AR (1) model proposed by Dischel (1998a, b). Although they agree with the use of a mean-reverting model, they argue about the direct use of the Hull and White model. They prove that the models proposed by Dischel (1998b) and McIntyre and Doherty (1999) are mean-reverting models, but they revert to a value different from the historical mean. The only possibility for having a process with mean $S(t)$ using model (4.27) is to take $S(t)$ as a constant. However, the seasonality of daily temperature is clearly not constant.

Moreover, they show that this can be corrected by the addition of the term $dS(t)$ in (4.27), where $dS(t)$ is the changes in seasonal variations.

This means that today's temperature does not depend only on the previous day but depends also on the days before yesterday. Also, in their model, they allow the volatility to change among seasons, but in their analysis, they assume it as a constant.

Alaton et al. (2002) improves Dischel (1998a, b) model using the model (4.24) proposed by Dornier and Queruel (2000). Also, they incorporate seasonalities in the mean using a sinusoid function:

$$S(t) = A + Bt + C\sin(\omega t + \varphi), \tag{4.28}$$

where φ is the phase parameter that defines the day of the yearly minimum and maximum temperature. Since it is known that the DAT has a strong seasonality of a 1-year period, the parameter ω was set to $\omega = 2\pi/365$. This tactic was discouraged by Moreno (2000), since sinusoids do not fit well the asymmetric evolution of temperature, and as a result, a biased is induced in the out-of-sample forecasts despite the goodness of fit. The linear trend caused by urbanization or climate changes is represented by $A + Bt$. The time, measured in days, is denoted by t. The parameter C defines the amplitude of the difference between the yearly minimum and maximum DAT. Using the Itô formula, a solution of (4.24) is given by

$$T(t) = S(t) + (T(s) - S(s))e^{-\kappa(T-s)} + \int_s^t e^{-\kappa(t-\tau)}\sigma(\tau)dB(\tau). \tag{4.29}$$

In Alaton et al. (2002), in order to find numerical values of the constants, the function

$$Y_t = a_1 + a_2 t + a_3\sin(\omega t) + a_4\cos(\omega t) \tag{4.30}$$

is fitted to the temperature data using the method of least squares. The constants between the model (4.28) and (4.30) are connected with the following relations:

$$A = a_1, \tag{4.31}$$

$$B = a_2, \tag{4.32}$$

$$C = \sqrt{a_3^2 + a_4^2}, \tag{4.33}$$

$$\varphi = \arctan\left(\frac{a_4}{a_3}\right) - \pi. \tag{4.34}$$

Another innovative characteristic of Alaton et al. (2002) framework is the introduction of seasonalities in the standard deviation. In Alaton et al. (2002),

two estimators for the standard deviation of the temperature, σ, are obtained. The first estimator $\sigma_t = \{\sigma_i\}_{i=1}^{12}$ is a piecewise constant function, with a positive constant value each month. Following the notation of Alaton et al. (2002), the estimator is given by

$$\hat{\sigma}_\mu^2 = \frac{1}{N_\mu} \sum_{j=0}^{N_\mu - 1} (T(j+1) - T(j))^2, \tag{4.35}$$

where $j = 0, \ldots, N_\mu$ and N_μ is the DAT at day N at month $\mu = 1, \ldots, 12$.

By discretizing the mean-reverting O–U given by (4.24), a second estimator is obtained. Following again the notation of Alaton et al. (2002), the discretized equation is

$$T(j) = S(j) - S(j-1) + \kappa S(j-1) + (1-\kappa)T(j-1) + \sigma_\mu \varepsilon(j-1) \;\; j = 1, \ldots, N_\mu, \tag{4.36}$$

where $\varepsilon(j)$ are *i.i.d.* standard normally distributed.

By denoting

$$\tilde{T}(j) = T(j) - S(j), \tag{4.37}$$

Equation (4.36) can be rewritten as

$$\tilde{T}(j) = (1-\kappa)\tilde{T}(j-1) + \sigma_\mu \varepsilon(j-1). \tag{4.38}$$

Hence, an efficient estimator can be derived as

$$\hat{\sigma}_\mu^2 = \frac{1}{N_\mu - 2} \sum_{j=1}^{N_\mu} \left(\tilde{T}(j) - \hat{\kappa} S(j-1) - (1-\hat{\kappa})T(j-1)\right)^2 \tag{4.39}$$

which reduces to

$$\hat{\sigma}_\mu^2 = \frac{1}{N_\mu - 2} \sum_{j=1}^{N_\mu} \left(\tilde{T}(j) - T(j-1) + \hat{\kappa}\tilde{T}(j-1)\right)^2. \tag{4.40}$$

In Alaton et al. (2002), the mean between the two estimators is used for each month μ. Finally, in order to estimate (4.40), an estimator for parameter κ is needed. Alaton et al. (2002) uses a martingale estimation function method suggested by Bibby and Sorensen (1995):

$$\hat{\kappa}_n = -\log\left(\frac{\sum_{i=i}^{n} \tilde{T}(i) \cdot \tilde{T}(i-1)/\sigma_{i-1}^2}{\sum_{i=i}^{n} \tilde{T}^2(i)/\sigma_{i-1}^2}\right) \quad i = 1, 2, \ldots, n. \tag{4.41}$$

Alaton et al. (2002) use data from Bromma Airport at Sweden for a period of 40 years in order to estimate the parameters. Their results indicate that the model fits the temperature data well since it incorporates the main characteristics of the DAT. However, it is a simplification of the real world. The piecewise constant volatility results to underestimation of the real volatility and hence to underestimation of the prices of weather derivatives (Benth and Saltyte-Benth 2005). Finally, the proposed model assumes that the residuals ε_t are uncorrelated and normally distributed. Alaton et al. (2002) do not provide any statistical tests for the correlation or the normality of the residuals, while their results suggest that the small temperature differences have higher frequency than those predicted by the fitted normal distribution.

Torro et al. (2003) adapt interest rate models by expanding the frameworks of Bali (1999) and Chan et al. (1992). Based on the DAT in Spanish temperature index, a general single factor model that captures seasonality, mean reversion, GARCH structures in volatility, and relationships between volatility and temperature levels for modeling was developed:

$$dT(t) = (S(t) + a_2 T(t))dt + \sigma(t)dz, \tag{4.42}$$

where

$$\sigma(t) = \psi(t)T(t)^{\gamma} \tag{4.43}$$

and seasonality was captured as in Alaton et al. (2002) by a sinusoid function – $S(t) = a_0 + a_1 \cos(\omega t + \phi)$ – and ψ_t represents the structural changes in volatility captured by a GARCH model:

$$\psi_{t+1}^2 = \beta_0 + \beta_1 \varepsilon_t^2 + \beta_2 \psi_t^2. \tag{4.44}$$

Using different constraints, Torro et al. (2003) obtained a set of nested models. The results from the Spanish temperature index indicate that an appropriate model for the DAT should contain a mean reversion term. The volatility shows autoregressive behavior. Finally, the volatility and the temperature are negatively related while the sensitivity, between these too variables, is really low.

It is clear that Torro et al. (2003) have not included a trend that represents global or urban warming. Moreover the model (4.42), and the nested ones, does not revert to the appropriate value, according to Dornier and Queruel (2000), since the term $dS(t)$ was not included.

Brody et al. (2002) argue about the use of the standard BM. Although the temperature fluctuations are normally distributed, a slow decay in the ACF of the temperature can be observed, which BM fails to capture. To overcome this problem, a fractional Brownian motion (FBM) is proposed for the driving noise. FBM is a stochastic process that exhibits long-range dependence without significantly implicating the pricing equations of the future and option weather derivatives.

FBM models are the continuous analogous of the ARFIMA models proposed by Caballero et al. (2002).

As the name suggests, FBM is a modified general version of a standard BM that depends on the Hurst exponent H. The Hurst exponent, H, determines the level of the correlations of the increments. When $H > \frac{1}{2}$, the correlation is positive, while when $H < \frac{1}{2}$, the correlation is negative. If $H = 0$, then the correlation is zero and the standard BM is recovered.

Replacing the driving noise process with a FBM, the stochastic equation for the DAT can be written as

$$dT(t) = \kappa(t)(S(t) - T(t))dt + \sigma(t)dB^H(t), \tag{4.45}$$

where $0 < H < 1$.

Seasonality in the mean and volatility is captured by a sinusoid function similar to Alaton et al. (2002) and Torro et al. (2003):

$$S(t) = a_0 + a_1 \sin(\frac{2\pi t}{365} + \phi_1), \tag{4.46}$$

$$\sigma(t) = \beta_0 + \beta_1 \sin(\frac{2\pi t}{365} + \phi_2). \tag{4.47}$$

Moreover, $\kappa(t)$ is not limited to a constant value as in previous studies, but it is rather represented by a bounded deterministic function. Allowing $\kappa(t)$ to vary with time implies seasonalities in the speed of mean reversion. However, Brody et al. (2002) assume it constant, and they do not proceed in examining the dynamics of $\kappa(t)$. Finally, the factor $dS(t)$ should have been added for the temperature to revert to the seasonal mean.

In Brody et al. (2002), data from the daily central England temperature from the period 1772–1999 was taken. In contrast to previous studies that usually make use of 5–40 years of data, Brody et al. (2002) use a significantly larger sample. It is expected that measurements made during the last 200 years probably will not reflect the dynamics of the temperature of the next few years (Dischel 1999). Using a very large sample of historical data of DAT runs the danger for the estimated parameters to be affected by dynamics of the temperature that do not represent the future behavior of temperature anymore. Brody et al. (2002) found evidence of fractionality in the temperature since the estimated Hurst exponent was $H = 0.61$. However, the analysis should have been performed after all seasonalities have been removed from the data (Benth and Saltyte-Benth 2005; Bellini 2005).

Using a FBM process B^H, with H different than 0.5, has an impact in the stochastic calculus. The process is not a semimartingale or a Markov process which means that standard stochastic methods cannot be used. Recent studies have developed stochastic methods for FBM analogue to Itô calculus (Aldabe et al. 1998; Hu and Oksendal 2003; Lin 1995). Although these methods lead to arbitrage opportunities in finance, they can be used in temperature modeling since temperature cannot be traded or stored (Bellini 2005).

In Benth (2003), the findings of Brody et al. (2002) comprised the starting point to derive arbitrage-free pricing formulas for temperature derivatives. A FBM with a Hurst exponent between 0.5 and 1 is used to model the DAT. Benth (2003) first proved that the price of temperature derivatives is arbitrage-free using the quasi-conditional expectation. Next, Benth (2003) calculated the prices for European and Asian claims. Their results indicate that the derived theoretical pricing formulas using a FBM are not a function of $T - t$ as in the case of a BM. More precisely, the price of the temperature derivative no longer depends on the time to exercise $T - t$ but on the current time t and the exercise time T separately.

Bhowan (2003) expands the model proposed by Alaton et al. (2002) incorporating stochastic volatility. The stochastic volatility has the form of a mean-reverting process that reverts to a long-term trend:

$$d\sigma(t) = a(\sigma_{trend} - \sigma(t))dt + \gamma dB(t), \tag{4.48}$$

where σ_{trend} is the long-term trend and it is constant. The parameter γ is given by

$$\gamma^2 = \frac{1}{n}\sum_{j=0}^{n-1}(\sigma(j-1) - \sigma(j))^2 \tag{4.49}$$

and

$$\hat{a} = -\log\left(\frac{\sum_{i=i}^{n}\left(\frac{\sigma_{trend}-\sigma(t-1)}{\gamma^2}\right)(\sigma(t) - \sigma_{trend})}{\sum_{i=i}^{n}\left(\frac{\sigma_{trend}-\sigma(t-1)}{\gamma^2}\right)(\sigma(t-1) - \sigma_{trend})}\right) \quad i = 1, 2, \ldots, n. \tag{4.50}$$

Bhowan (2003) found that their model fits 20 years of data obtained from Pretoria very well, while Mraoua and Bari (2007) implement the same framework in 44 years of Moroccan data.

In a more recent paper, Benth and Saltyte-Benth (2005) fit Norwegian data by modeling the DAT variations with a mean-reverting O–U process where the noise process is modeled by a generalized hyperbolic Lévy process. Instead of the FBM used in their previous work, they expand the work of Dornier and Queruel (2000). Moreover, they argue with Brody et al. (2002) for not performing fractional analysis at the residuals in their regression model. As in previous works, a sine function captures the seasonal mean as in (4.46). Also the idea of the seasonal volatility proposed by Alaton et al. (2002) is expanded. The seasonal volatility has a continuous form, $\sigma(t) = \sigma(t + k \cdot 365)$ for $t = 1, \ldots, 365$ and $k = 1, 2, 3 \ldots$ and is repeated every year. More precisely, the model they used is given by

$$dT(t) = dS(t) + \kappa(T(t) - S(t))dt + \sigma(t)dL(t), \tag{4.51}$$

where $L(t)$ is a Lévy noise process.

Discretizing (4.51) with $\Delta t = 1$, the DAT can be written as

$$\begin{aligned} \Delta T(t) &= \Delta S(t) + \kappa(T(t) - S(t)) + \sigma(t)\Delta L(t) \\ T(t) - T(t-1) &= S(t) - S(t-1) + \kappa(T(t-1) - S(t-1)) + \sigma(t)\Delta L(t). \end{aligned} \tag{4.52}$$

By rearranging (4.52), we have that

$$T(t) = S(t) + (1+\kappa)(T(t-1) - S(t-1)) + \sigma(t)\varepsilon(t) \tag{4.53}$$

since

$$\Delta L(t) = \varepsilon(t)\sqrt{\Delta t}. \tag{4.54}$$

Hence, (4.51) can be written as an additive time series:

$$T(t) = S(t) + c(t) + \tilde{\varepsilon}(t), \tag{4.55}$$

where $S(t)$ is given by (4.46), $c(t) = a(T(t-1) - S(t-1))$, $a = 1 + \kappa$, and $\tilde{\varepsilon}(t) = \varepsilon(t)\sigma(t)$. The seasonal variance can be extracted from the residuals as follows. First, the residuals are grouped into 365 groups, where each group corresponds to a single day of the year. Then, by taking the average of the squared values of each group, the variance for that day is obtained. The result yields because $\varepsilon(t)$ has average value of 0 and variance of 1.

Lévy family-based distributions are flexible processes that allow heavy tails and skewness that often are observed in temperature time series. On the other hand, because of the nature of the distribution, no closed-form solution can be found. Benth and Saltyte-Benth (2005) confirm the existence of heavy tails and skewness in Norwegian data. Also they did not found any significant linear trend. This is probably a result of the use of a small dataset since Benth and Saltyte-Benth (2005) use only 13 years of data. Also, Benth and Saltyte-Benth (2005) examine if the parameter κ is constant or a time-varying function. More precisely κ was assumed to be a piecewise function with constant value during a month or a year. In contrast to Brody et al. (2002), they did not find any significant time dependency or variation in monthly or yearly basis. Again, this is probably a result of the use of a small dataset or their averaging method in monthly and yearly basis.

Benth and Saltyte-Benth (2005) removed the seasonality from the volatility; however, they found that the first few lags of the remaining residuals are still significantly correlated. They suggest that a moving average time series or a GARCH model will remove this effect but they did not proceed on estimating one.

In Benth and Saltyte-Benth (2007), 40 years of data were used to model the DAT in Stockholm, Sweden. In order to focus on pricing and to provide closed-form solution for the pricing of weather derivatives, Benth and Saltyte-Benth (2007) use a BM as the driving noise process. More precisely, a mean-reverting O–U process where the noise process is modeled by a simple BM as in (4.24) was suggested.

In this study, the speed of mean reversion parameter, κ, was considered constant. Both seasonal mean and (square of) daily volatility of temperature variations are modeled by truncated Fourier series:

$$S(t) = a + bt + \sum_{i=1}^{I_1} a_i \sin(2\pi i(t - f_i)/365) + \sum_{j=1}^{J_1} b_j \cos\big(2\pi j\big(t - g_j\big)/365\big), \tag{4.56}$$

$$\sigma^2(t) = c + \sum_{i=1}^{I_2} c_i \sin(2\pi it/365) + \sum_{j=1}^{J_2} d_j \cos(2\pi jt/365). \tag{4.57}$$

Using truncated Fourier series, a good fit for both the seasonality and the variance component can be obtained while keeping the number of parameters relatively low. The above representation simplifies the needed calculations for the estimation of the parameters and for the derivation of the pricing formulas. Equations (4.56) and (4.57) allow both larger and smaller periodicities than the classical 1-year temperature cycle. In Benth and Saltyte-Benth (2007), the order of both series was chosen arbitrarily, and no statistical tests were presented for the significance of each parameter.

The discrete form of (4.24) is given by

$$\Delta T(t) = \Delta S(t) - (1 - e^{-\kappa})(T(t) - S(t)) + e^{-\kappa} \int_t^{t+1} \sigma(u) e^{-(t-u)} dB(u), \tag{4.58}$$

where

$$\Delta T(t) = T(t+1) - T(t). \tag{4.59}$$

This can be easily shown as follows when $T(t)$ is given by (4.29)

$$\begin{aligned}
\Delta T(t) &= T(t+1) - T(t) \\
&= S(t+1) + (T(0) - S(0))e^{-\kappa(t+1)} + \int_0^{t+1} \sigma(u) e^{-\kappa(t+1-u)}\, dB(u) - S(t) \\
&\quad - (T(0) - S(0))e^{-\kappa t} - \int_0^{t} \sigma(u) e^{-\kappa(t-u)}\, dB(u) \\
&= S(t+1) - S(t) + (T(0) - S(0))e^{-\kappa t}\,(e^{-\kappa} - 1) \\
&\quad + e^{-\kappa} \int_0^{t+1} \sigma(u)\, e^{-\kappa(t-u)}\, dB(u) - \int_0^{t} \sigma(u) e^{-\kappa(t-u)}\, dB(u)
\end{aligned}$$

$$
\begin{aligned}
&= \Delta S(t) + \left(T(t) - S(t) - \int_0^t \sigma(u) e^{-\kappa(t-u)}\, dB(u) \right) (e^{-\kappa} - 1) \\
&\quad + e^{-\kappa} \int_0^{t+1} \sigma(u) e^{-\kappa(t-u)} dB(u) - \int_0^t \sigma(u) e^{-\kappa(t-u)}\, dB(u) \\
&= \Delta S(t) + (T(t) - S(t)) (e^{-\kappa} - 1) - e^{-\kappa} \int_0^t \sigma(u) e^{-\kappa(t-u)}\, dB(u) \\
&\quad + \int_0^t \sigma(u) e^{-\kappa(t-u)} dB(u) + e^{-\kappa} \int_0^{t+1} \sigma(u) e^{-\kappa(t-u)} dB(u) - \int_0^t \sigma(u) e^{-\kappa(t-u)}\, dB(u) \\
&= \Delta S(t) + (T(t) - S(t)) (e^{-\kappa} - 1) + e^{-\kappa} \int_t^{t+1} \sigma(u) e^{-\kappa(t-u)}\, dB(u)
\end{aligned}
$$

Approximating (4.58), we have that

$$\Delta T(t) \approx \Delta S(t) - (1 - e^{-\kappa})(T(t) - S(t)) + e^{-\kappa} \sigma(t) \Delta B(t). \tag{4.60}$$

By rearranging (4.60) and substituting $\Delta B(t) = B(t+1) - B(t) = \varepsilon(t)$, we have that the deseasonalized and detrended temperature is given by an AR(1) model:

$$\tilde{T}(t+1) = a\tilde{T}(t) + \tilde{\sigma}(t)\varepsilon(t), \tag{4.61}$$

where $\widetilde{T}(t)$ is given by (4.37) and

$$a = e^{-\kappa} \tag{4.62}$$

and

$$\tilde{\sigma}(t) = a\sigma(t). \tag{4.63}$$

In Benth and Saltyte-Benth (2007), the DAT is first detrended and deseasonalized by fitting $S(t)$ to the data. Next, the parameter a of the AR(1) model (4.61) is estimated. Then the ACF of the residuals is examined. More precisely, the ACF of the residuals and the squared residuals is examined for seasonalities. If seasonality in the residuals is found, then the seasonal variance of the DAT is removed. The seasonal variance can be extracted from the residuals as follows. First, the residuals are grouped into 365 groups, where each group corresponds to a single day of the year. Then by taking the average of the squared values of each group, the variance for that day is obtained. Finally, (4.57) is fitted to the estimated seasonal variance. In Benth and Saltyte-Benth (2007), the variance is considered a function of time, and it is repeated every year,

$$\text{that is,} \quad \sigma(t) = \sigma(t + k \cdot 365) \text{ for } t = 1, \ldots, 365 \text{ and } k = 1, 2, 3 \ldots$$

Their results indicate that their model is good enough to describe the main dynamics of temperature data. Moreover, the proposed model allows for closed-form solutions of the pricing formulas of the HDD and the CAT future and options contracts. Also, in Benth and Saltyte-Benth (2007), it is shown that the HDD future curves give higher prices when a seasonal volatility is considered compared to a constant volatility.

Their results indicate a clear linear trend, while the ACF of the squared residuals reveals time dependency and seasonality in the variance of the residuals (Benth and Saltyte-Benth 2007). As in previous studies, their results show higher level of volatility in the winter period. The seasonal variance is modeled and removed using (4.57) with $I_2 = J_2 = 4$. However, no statistical results were given for the choice of the length of the truncated Fourier series of the variance or of the significance of each parameter. Moreover, the autocorrelation for the first lags is still present. They suggest that a GARCH model will remove this effect but they did not proceed in estimating one. In addition, in contrast to the initial hypothesis of the BM, the normality test was rejected at the 1% significance level.

Zapranis and Alexandridis (2006) expand the model proposed by Benth and Saltyte-Benth (2007). They use 101 years of temperature in Paris in order to price European CAT derivatives. They model the seasonal cycle using an extension and a combination of discrete FT approach and the regression method. The seasonality and the seasonal variance were modeled as in Benth and Saltyte-Benth (2007); however, they propose a novel approach to correctly model (4.56) and (4.57) using WA. More specifically, they use WA in order to decompose the temperature series into a series of (orthogonal) basis functions (wavelets) with different time and frequency locations. As a result, the wavelet decomposition brings out the structure of the underlying dynamics of the temperature series as well as trends, periodicities, singularities, or jumps that could not be observed originally. Hence, seasonal mean and (square of) daily volatility of temperature variations can be modeled efficiently and accurately. However, the distribution of the residuals of AR(1) model, before and after dividing out the seasonal variance, differs significantly from the normal distribution. The next thing that was tried was to assess the impact of outliers to the original AR(1) model. The differences of today's average temperature from yesterday's average temperature were formed, and then the dates corresponding to the differences with a value greater than plus or minus 3.5 standard deviations were identified. In total, 40 outlier temperature observations were identified out of the 36,865 temperature values, which were then set equal to the average value of the temperature for that particular day, calculated from 101 years of data. The smoothness of the above outliers improved the distributional statistics of the residuals significantly. More precisely the skewness was only -0.005, the kurtosis was 3.04, and the Jarque–Bera (JB) statistic has fallen to only 3.77 which lead to the acceptance of the normal distribution.

In order to rectify the rejection of the normality hypothesis, in more recent chapters, Zapranis and Alexandridis (2007, 2009b) replaced the simple AR(1) model by more complex ones. They used ARMA(3,1), ARFIMA, and AFRIMA–FIGARCH models. Their results from the DAT in Paris indicate that,

as the model gets more complex, the noise part draws away from the normal distribution. They conclude that, although the AR(1) model probably is not the best model for describing temperature anomalies, increasing the model complexity and thus the complexity of theoretical derivations in the context of weather derivative pricing does not seem to be justified. Next, Zapranis and Alexandridis (2009b) model nonparametrically the seasonal residual variance with NNs. The improvement regarding the distributional properties of the original model is significant. The examination of the corresponding Q–Q plot reveals that the distribution is quite close to Gaussian, while the JB statistic of the original model is almost halved. The NN approach gives a good fit for the ACF and an improved and reasonable fit for the residuals.

In Zapranis and Alexandridis (2008), three different decades of DATs in Paris are examined using the mean-reverting O–U process proposed by Benth and Saltyte-Benth (2007). The seasonality and the seasonal variance were modeled using WA. Previous studies assume that the parameter of the speed of mean reversion, κ, is constant. However, the findings of Zapranis and Alexandridis (2008) indicate some degree of time dependency in $\kappa(t)$. Since $\kappa(t)$ is important for the correct and accurate pricing of temperature derivatives, a significant degree of time dependency in $\kappa(t)$ can be quite important (Alaton et al. 2002). A novel approach to estimate nonparametrically a nonlinear time-dependent $\kappa(t)$ with a NN was presented. Daily values of the speed of the mean reversion were computed. In contrast to averaging techniques, in a yearly or monthly basis, which run the danger of filtering out too much variation, it is expected that daily values will provide more information about the driving dynamics of the temperature process. Results from Zapranis and Alexandridis (2008) indicate that the daily variation of the value of the speed of mean reversion is quite high. Intuitively, it is expected of $\kappa(t)$ not to be constant. If the temperature today is far from the seasonal average (a cold day in summer), then it is expected that the mean reversion speed will be high, that is, the difference between today's temperature and tomorrow's temperature is expected to be high. In contrast, if the temperature today is close to the seasonal variance, we expect the temperature to revert to its seasonal average slowly. In Zapranis and Alexandridis 2008), $\kappa(t)$ is studied. Their data from Paris indicate that $\kappa(t)$ has a bimodal distribution with an upper threshold which is rarely exceeded. Also it was examined if $\kappa(t)$ is a stochastic process itself. Both an Augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests were used. Both tests conclude that $\kappa(t)$ is stationary. Finally, using a constant speed of mean reversion parameter, the normality hypothesis was rejected in all three cases, while in the case of the NN, the normality hypothesis was accepted in all three different samples.

Bellini (2005), motivated by the papers of Benth and Saltyte-Benth (2005, 2007), uses a Gaussian O–U model with time-dependent mean and volatility to describe the stochastic dynamics of DATs on four US cities: Chicago, Portland, Philadelphia, and Tucson. Seasonality in the mean and volatility was also modeled by a truncated Fourier series. For Chicago, Portland, and Philadelphia, the normal distribution provides a good fit; however, the estimated JB statistic is over 40 for all cities.

In the city of Tucson where the normality hypothesis provided the worst fit, a Lévy process was used. Finally, the study of Bellini (2005) indicates the absence of fractional characteristics in the standardized residuals after all seasonal cycles have been removed from the data.

In Benth et al. (2007) a continuous-time autoregressive process with lag p (CAR(p)-process) and seasonal variation is introduced. Using 40 years of data in Stockholm, their results indicate that a value of $p = 3$ is sufficient to explain the autoregressive temperature dynamics. The detrended and deseasonalized DATs follow an AR(3) process with seasonal residuals modeled as in (4.57). The overall fit is very good with $R^2 = 94.1\%$; although the distribution of the residuals is close to normal, the normality hypothesis is rejected.

In Geman and Leonardi (2005), three different modeling methods were compared: HDD index modeling, AccHDDs index modeling, and daily modeling. Their results from 50 years of DAT in Paris indicate that HDDs distribution differs significantly from the normal distribution, while in the case of the AccHDDs distribution, the hypothesis of normal distribution was accepted. Following McIntyre and Doherty (1999), Geman and Leonardi (2005) used (4.27) to model the daily temperature index. The discrete version of (4.27) leads to an AR(1) model. In order to remove the correlation in the residuals, Geman and Leonardi (2005) use a general AR(q) model with $q \geq 1$:

$$T(t) = S(t) + \mathrm{AR}(q) + \varepsilon_t. \tag{4.64}$$

Their results indicate that $q = 3$ and that the residuals ε_t are uncorrelated and follow the normal distribution. However, the proposed model does not revert to the appropriate seasonal value since the term $dS(t)$ was not included.

In Schiller et al. (2008), a spline model was proposed. Splines used to separate DAT into a trend and a seasonality component in the mean $\mu(t)$ and a trend and seasonality component in the standard deviation $\sigma(t)$:

$$T(t) = \mu(t) + \sigma(t)R(t). \tag{4.65}$$

The residuals $R(t)$ are modeled separately with an AR process. Their results indicate that the ACF of an AR(1) model decreases exponentially with time; hence, it is not able to capture the slow decay of the ACF. In order to rectify this problem, Schiller et al. (2008) model the residuals with an AROMA process previously suggested by Caballero et al. (2002) and Jewson and Caballero (2003a). As in Dubrovsky et al. (2004) and Oetomo and Stevenson (2005), their findings suggest that daily models tend to underestimate the variance of the error.

Yoo (2003) expands the work proposed by Alaton et al. (2002) by incorporating seasonal forecasts. The seasonal forecasts were incorporated using a linear combination of the warm, normal, and cool mean temperature processes:

$$S(t) = p_w S_t^w + p_n S_t^n + p_c S_t^c, \tag{4.66}$$

$$S_t^i = \beta_0^i + \beta_1^i \sin(\omega t) + \beta_2^i \cos(\omega t) \;\; i = w, n, c, \tag{4.67}$$

where p_w, p_n, p_c are the probabilities of each scenario (warm, normal, cool). Their model was tested in five cities in the USA. The volatility was assumed as a piecewise constant function, with a constant value during each season. Their results indicate that option values obtained by the MC simulation are very sensitive to the seasonal forecast probabilities.

In Richards et al. (2004), the DAT in Frenso, CA, was modeled by a general mean-reverting Brownian motion (MRBM) with lognormal jumps and time-varying volatility:

$$dT(t) = \kappa(S(t) - \lambda\varphi - T(t))dt + \sigma(t)dz + \varphi dq, \tag{4.68}$$

where $S(t)$ is the seasonal average daily temperature given similar to Alaton et al. 2002; Yoo (2003) and West (2002) slightly modified to incorporate correlated lags of the temperature process:

$$S(t) = \gamma_0 + \gamma_1 \sin(2\pi t/365) + \gamma_2 \cos(2\pi t/365) + \gamma_3 t + \sum_{j=1}^{p} \rho_j T(t-j). \tag{4.69}$$

Jumps occur according to a Poisson process q with average rate λ, and dz is a Wiener process. Finally, a time-varying volatility is incorporated by an autoregressive conditional heteroskedasticity (ARCH) process:

$$\sigma(t) = \gamma_0 + \gamma_1 (T(t-1) - S(t))^2. \tag{4.70}$$

Richards et al. (2004) estimate a series of nested models: a BM, a MRBM, a MRBM with lognormal jumps, and a MRBM with lognormal jumps and ARCH. Their results indicate that the latter outperforms the others. However, the term $dS(t)$ should have been added for a proper mean reversion. The results from Hamisultane (2006a, b, 2007, 2008) indicate that the above models produce very volatile prices when using MC simulations.

The presence of long memory in DAT was tested in New York data by Hamisultane (2006b). Following Brody et al. (2002), the Hurst exponent was calculated around 0.66, indicating the presence of long memory. However, the term dS should have been included for a proper mean reversion of the temperature (Dornier and Queruel 2000), and the Hurst exponent should have been estimated after all seasonal cycles have been removed from the data (Bellini 2005; Benth and Saltyte-Benth 2005). Hamisultane (2006b) compares long- and short-memory processes in both continuous and discrete process to calculate prices for New York CDD future prices. Both financial and actuarial approaches were used for pricing. Their results indicate that the financial approach produced better results while MC predictions have appeared to be very volatile. Finally, the results from

discrete time long-memory and short-memory processes used in the actuarial method were similar.

Oetomo and Stevenson (2005) compared various methods previously proposed in literature. More precisely, the following model were compared: a naïve temperature forecasting model which relies on the historical average temperature, the model proposed by Dischel (1998a, b), the model proposed by Alaton et al. (2002), an ARMA model, the model proposed by Cao and Wei (2004), and the model proposed by Campbell and Diebold (2005). The above models were tested in-sample and out-of-sample in various locations. Their results indicate that none model was able to constantly outperform the others. Also, their findings suggest that both stochastic models and time-series models tend to under forecast the DAT, while forecasts beyond 30 days are unreliable. Finally, their results indicate that forecasting AccCDDs index is more difficult than forecasting the AccHDDs index.

Finally, in Zapranis and Alexandridis (2009a), WNs were used in order to model a mean-reverting O–U temperature process, with seasonality in the level and volatility. The seasonality in the mean and volatility modeled as in Benth and Saltyte-Benth (2007). They forecast up to 2 months ahead out-of-sample daily temperatures, and the corresponding CAT and HDD indices were simulated. The proposed model is validated in eight European and five US cities, all traded in CME. Their results suggest that the proposed method outperforms alternative pricing methods proposed in prior studies in most cases. Their findings suggest that WNs can model the temperature process very well, and consequently, they constitute a very accurate and efficient tool for weather derivatives pricing.

4.6 Alternative Methods

Since the weather market is a classical form of an incomplete market, standard hedging-based pricing methods cannot be applied. As a result, many alternative methods have been proposed in literature.

Zeng (2000a) discusses the limitations of the actuarial methods that originate from the statistical properties of the weather indices. As an alternative, a modified MC method, named biased sampling MC, is proposed. The idea of Zeng (2000a) takes advantages of the seasonal forecasts for a particular underlying weather variable (temperature, precipitation). By denoting $p_A, p_N,$ and $p_B,$ the probabilities that the underlying weather variable will be above, near, or below the climate norm, it is assumed to approximate the probabilities that the corresponding index (e.g., CDD or HDD) will be above, near, or below the climate norm, respectively. Results from over 250 weather stations indicate that this assumption is justified (Zeng 2000b). In the proposed method, first a normal distribution is fitted on the historical index. Then the historical record is sorted and divided into three groups: the highest 33%, the middle 34%, and the lowest 33%. Then samples are taken from the three groups with replacement proportional to the probabilities $p_A, p_N,$ and p_B. However, the hypothesis that the CDD or the HDD indices follow the normal distribution contradicts the results of Geman and Leonardi (2005).

Platen and West (2005) use a fair pricing framework under the benchmark approach to price weather derivatives. More precisely, the growth optimal portfolio which is interpreted as a world stock index is used as a benchmark. The framework of Platen and West (2005) is based on the assumption that the weather market is liquid; however, the weather market is still emerging and it is illiquid (Platen and West 2005; Brockett et al. 2006).

Due to the absence of liquid secondary weather market, Brockett et al. (2006) use the indifference pricing approach to value weather derivatives taking into account portfolio effects. In this framework, an upper limit of the price of a weather derivative is the price at which a buyer is indifferent, in terms of expected utility, between buying and not buying the contract. Brockett et al. (2006) use a mean–variance utility function:

$$u(x) = E(x) - \lambda\sigma^2(x), \tag{4.71}$$

where λ is the risk aversion parameter and it is positive. It is clear that the choice of the utility function $u(x)$ as well as the value of the risk aversion parameter greatly affects pricing.

Xu et al. (2008) expand the framework proposed by Brockett et al. (2006). They argue about the use of HBA since it lacks of as sound theoretical basis and about equilibrium models since they have to resort to simplifying assumptions in order to become tractable (Xu et al. 2008). In the mean–variance framework, an exponential utility function was used:

$$u(x) = -e^{-\lambda x}, \tag{4.72}$$

where λ is the risk aversion parameter and it is positive. However, the simplifying assumptions made by Xu et al. (2008) reduce the model to a straightforward actuarial interpretation (Xu et al. 2008).

It is clear that the maximization of the expected utility framework is often proposed in the literature. However, utility functions are too much preference dependent and sensitive to the selection of the risk aversion parameter (Carr et al. 2001). Moreover, using temperature forecasts with a utility function to estimate the demand curve for the derivate reduces the proposed methodology to simply using the forecasts (Campbell and Diebold 2005) (Oetomo and Stevenson 2005).

4.7 Conclusions

In this chapter, the main methodologies proposed in literature for modeling temperature and pricing weather derivatives were presented and reviewed. Studying and understanding the advantages and disadvantages of prior studies, a new and effective model can be built.

The weather market is at its infancy and still developing. Similarly, the literature is evolving. However, a general accepted model still does not exist. In addition, practitioners and risk management companies keep weather market data private and do not publish their models.

The underlying variables of weather derivatives follow predictable recurrent patterns; hence, the actuarial method is not an appropriate pricing approach. The HBA is considered as a good first approximation of the price of a weather derivative. HBA is the simplest pricing method in terms of implementation and the most prone to large pricing errors.

Alternatively index or daily modeling can be used. In index modeling, the various temperature indices can be directly modeled. On the other hand, on daily modeling, the DAT is modeled and then any temperature index can be derived. Developing a daily model can be done by using either a discrete or a continuous process.

In this book, we focus on daily modeling. Modeling directly the daily temperature can, in principle, lead to more accurate pricing than modeling temperature indices. In the calculation of most indices, a lot of information concerning the temperature dynamics is lost. The risk with daily modeling is that small misspecifications in the models can lead to large mispricing of the temperature contracts.

References

Alaton P, Djehince B, Stillberg D (2002) On modelling and pricing weather derivatives. Appl Math Finance 9:1–20

Bali TG (1999) An empirical comparison of continuous time models of the short term interest rate. J Future Markets 7:777–797

Bellini F (2005) The weather derivatives market: modelling and pricing temperature. Ph.D. thesis, University of Lugano, Lugano

Benth FE (2003) On arbitrage-free pricing of weather derivatives based on fractional Brownian motion. Appl Math Finance 10:303–324

Benth FE, Saltyte-Benth J (2005) Stochastic modelling of temperature variations with a view towards weather derivatives. Appl Math Finance 12(1):53–85

Benth FE, Saltyte-Benth J (2007) The volatility of temperature and pricing of weather derivatives. Quant Finance 7(5):553–561

Benth FE, Saltyte-Benth J, Koekebakker S (2007) Putting a price on temperature. Scand J Stat 34:746–767

Benth FE, Saltyte-Benth J, Koekebakker S (2008) Stochastic modelling of electricity and related markets, vol 11, Advance Series on Statistical Science and Applied Probability. World Scientific, Singapore

Bhowan A (2003) Temperature derivatives. University of Wiwatersand, Johannesburg

Bibby BM, Sorensen M (1995) Martingale estimation functions for discretely observed diffusion processes. Bernoulli 1 ((I/II) March/June) 17–39

Box GEP, Jenkins GM (1970) Time series analysis, forecasting and control. Holden-Day, San Francisco

Brockett PL, Wang M, Yang C, Zou H (2006) Portfolio effects and valuation of weather derivatives. Financial Rev 41:55–76

Brody CD, Syroka J, Zervos M (2002) Dynamical pricing of weather derivatives. Quant Finance 2:189–198
Caballero R, Jewson S (2002) Multivariate long-memory modeling of daily surface air temperatures and the valuation of weather derivative portfolios. http://ssrn.com/abstract=405800. Accessed July 2002
Caballero R, Jewson S, Brix A (2002) Long memory in surface air temperature: detection modelling and application to weather derivative valuation. Clim Res 21:127–140
Campbell SD, Diebold FX (2005) Weather forecasting for weather derivatives. J Am Stat Assoc 100:6–16
Cao M, Wei J (1999) Pricing weather derivatives: an equilibrium approach. The Rotman Graduate School of Management, The University of Toronto. http://www.fields.utoronto.ca/programs/cim/financial_math/finance_seminar/99-00/cao_wei.pdf. Accessed 1 Oct 1999
Cao M, Wei J (2000) Pricing the weather. Risk weather risk special report. Energy Power Risk Manage 67–70 (May)
Cao M, Wei J (2003) Weather derivatives: a new class of financial instruments. University of Toronto. http://www.rotman.utoronto.ca/~wei/research/JAI.pdf. Accessed 1 Apr 2003
Cao M, Wei J (2004) Weather derivatives valuation and market price of weather risk. J Future Markets 24(11):1065–1089
Cao M, Li A, Wei J (2004) Watching the weather report. Can Invest Rev Summer 27–33
Carmona R (1999) Calibrating degree day options. In: 3rd seminar on stochastic analysis, random field and applications. Ecole Polytechnique de Lausanne, Ascona, Switzerland (23 September)
Carr P, Geman H, Madan DB (2001) Pricing and hedging in incomplete markets. J Financ Econ 62:131–167
Chan K, Karoly A, Longstaff F, Sanders A (1992) An empirical comparison of alternative models for the short term interest rate. J Finance 3:1209–1226
Davis M (2001) Pricing weather derivatives by marginal value. Quant Finance 1:1–4
Dischel B (1998a) At least: a model for weather risk. Weather risk special report, Energy and Power Risk Management, pp 30–32 (March)
Dischel B (1998b) Black-Scholes won't do. Weather risk special report, Energy and Power Risk Management, pp 8–9 (October)
Dischel B (1999) Shaping history for weather risk management. Energy Power Risk 12(8):13–15
Dorfleitner G, Wimmer M (2010) The pricing of temperature futures at the Chicago Mercantile Exchange. J Bank Finance. doi:10.1016/j.bankfin.2009.12.004
Dornier F, Queruel M (2000) Caution to the wind. Weather risk special report, Energy Power Risk Management,pp 30–32 (August)
Dubrovsky M, Buchtele J, Zalud Z (2004) High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modelling. Clim Change 63:145–179
Franses PH, Neele J, van Dijk D (2001) Modeling asymetric volatility in weekly Dutch temperature data. Environ Model Softw 16:37–46
Geman H, Leonardi M-P (2005) Alternative approaches to weather derivatives pricing. Manag Finance 31(6):46–72
Hamisultane H (2006a) Extracting information from the market to price the weather derivatives. http://halshs.archives-ouvertes.fr/docs/00/17/91/89/PDF/weathderiv_extraction.pdf. Accessed Nov 2006
Hamisultane H (2006b) Pricing the weather derivatives in the presence of long memory in temperatures. http://halshs.archives-ouvertes.fr/docs/00/08/87/00/PDF/weathderiv_longmemory.pdf. Accessed May 2006
Hamisultane H (2007) Utility-based pricing of the weather derivatives. http://halshs.archives-ouvertes.fr/docs/00/17/91/88/PDF/weathderiv_utility.pdf. Accessed Sept 2006
Hamisultane H (2008) Which method for pricing weather derivatives ? http://halshs.archives-ouvertes.fr/docs/00/35/58/56/PDF/wpaper0801.pdf. Accessed July 2008
Hull CJ (2003) Option, futures and other derivatives, 5th edn. Prentice Hall, Upper Saddle River

Hull CJ (2005) Fundamentals of futures and options markets, 5th edn. Prentice Hall, Upper Saddle River
Jewson S (2004) Introduction to weather derivative pricing. http://ssrn.com/abstract=557831. Accessed 21 June 2004
Jewson S, Caballero R (2003a) Seasonality in the dynamics of surface air temperature and the pricing of weather derivatives. Meteorol Appl 10(4):377–389
Jewson S, Caballero R (2003b) Seasonality in the statistics of surface air temperature and the pricing of weather derivatives. Meteorol Appl 10(4):367–376
Jewson S, Brix A, Ziehmann C (2005) Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations. Cambridge University Press, Cambridge, UK
Lucas R (1978) Asset Prices in an Exchange Economy. Econometrica 46:1429–1445
McIntyre R, Doherty S (1999) Weather Risk-An example from the UK. Energy and Power Risk Management (June)
Moreno M (2000) Riding the temp. Weather Derivatives, FOW Special Supplement (December)
Mraoua M, Bari D (2007) Temperature stochastic modeling and weather derivatives pricing: empirical study with Moroccan data. Afrika Statistika 2(1):22–43
Oetomo T, Stevenson M (2005) Hot or cold? A comparison of different approaches to the pricing of weather derivatives. J Emerg Market Finance 4(2):101–133
Platen E, West J (2005) A fair pricing approach to weather derivatives. Asia-Pacific Financ Markets 11:23–53
Richards TJ, Manfredo MR, Sanders DR (2004) Pricing weather derivatives. Am J Agric Econ 4 (86):1005–1017
Roustant O, Laurent J-P, Bay X, Carraro L (2003a) A bootstrap approach to price uncertainty of weather derivatives. Paper presented at the Ecole des Mines, Sent-Etienner and Ecole ISFA, Villeurbanne, Dec
Roustant O, Laurent J-P, Bay X, Carraro L (2003b) Model risk in the pricing of weather derivatives. Paper presented at the Ecole des Mines, Saint- Etienne and Ecole ISFA, Villeurbanne, Sept
Schiller F, Seidler G, Wimmer M (2008) Temperature models for pricing weather derivatives. http://ssrn.com/abstract=1280826. Accessed 5 Nov 2008
Svec J, Stevenson M (2007) Modelling and forecasting temperature based weather derivatives. Global Finance J 18(2):185–204
Taylor JW, Buizza R (2002) Neural network load forecasting with weather ensemble predictions. IEEE Trans Power Syst 17(3):626–632
Taylor JW, Buizza R (2004) A comparison of temperature density forecasts from GARCH and atmospheric models. J Forecast 23:337–355
Taylor JW, Buizza R (2006) Density forecasting for weather derivative pricing. Int J Forecast 22:29–42
Tol RSJ (1996) Autoregressive conditional heteroscedasticity in daily temperature measurements. Environmetrics 7:67–75
Torro H, Meneu V, Valor E (2003) Single factor stochastic models with seasonality applied to underlying weather derivatives variables. J Risk Finance 4(4):6–17
Turvey CG (2001) The pricing of degree-day weather options. http://ageconsearch.umn.edu/bitstream/34109/1/wp0205.pdf. Accessed Nov 2001
West J (2002) Benchmark pricing of weather derivatives. School of Finance and Economics, University of Technology. http://www.business.uts.edu.au/qfrc/conferences/qmf2002/West_J.pdf. Accessed Dec 2002
Xu W, Odening M, Musshof O (2008) Indifference pricing of weather derivatives. Am J Agric Econ 90(4):979–993
Yoo S (2003) Weather derivatives and seasonal forecast. Department of Applied Economics and Management, Cornell University. http://www.card.iastate.edu/faculty/profiles/bruce_babcock/Shiyong_Yoo_papers/WD_2.pdf. Accessed Jan 2003

Zapranis A, Alexandridis A (2006) Wavelet analysis and weather derivatives pricing. Paper presented at the 5th Hellenic Finance and Accounting Association (HFAA), Thessaloniki, 15–16 Dec

Zapranis A, Alexandridis A (2007) Weather derivatives pricing: modelling the seasonal residuals variance of an Ornstein-Uhlenbeck temperature process with neural networks. In: EANN 2007, Thessaloniki, 29–31 Aug 2007

Zapranis A, Alexandridis A (2008) Modelling temperature time dependent speed of mean reversion in the context of weather derivetive pricing. Appl Math Finance 15(4):355–386

Zapranis A, Alexandridis A (2009a) Modeling and forecasting CAT and HDD indices for weather derivative pricing. In: EANN 2009, London, 27–29 Aug 2009

Zapranis A, Alexandridis A (2009b) Weather derivatives pricing: modelling the seasonal residuals variance of an Ornstein-Uhlenbeck temperature process with neural networks. Neurocomputing 73:37–48

Zeng L (2000a) Pricing weather derivatives. J Risk Finance 1(3):72–78

Zeng L (2000b) Weather derivatives and weather insurance: concept, application and analysis. Bull Am Meteorol Soc 81(9):2075–2082

Chapter 5
Modeling the Daily Average Temperature

5.1 Introduction

The purpose of this chapter is to develop a model that accurately describes the dynamics of the DAT. The statistical properties of the DATs will be examined in order to propose a process that exhibits the same behavior.

Daily modeling can in principle lead to more accurate pricing than modeling temperature indices (Jewson et al. 2005), as a lot of information is lost due to existing boundaries in the calculation of temperature indices by a normal or lognormal process, such as HDDs being bounded by zero. On the other hand, deriving an accurate model for the daily temperature is not a straightforward process. Observed temperatures show seasonality in all of the mean, variance, distribution, and autocorrelation, and there is evidence of long memory in the autocorrelation. The risk with daily modeling is that small misspecifications in the models can lead to large mispricing of the temperature contracts (Jewson et al. 2005).

It is clear that when index modeling is used, a different model must be estimated for each index. On the other hand, when daily modeling is used, only one model is fitted to the data, and it can be used for all available contracts in the market on the same location. Using a daily model, an accurate representation of all indices and indices distribution can be obtained. Finally, in contrast to index modeling and HBA, it is easy to incorporate meteorological forecasts.

In this chapter, the DAT time series of seven different European cities will be examined. The seven European cities are Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, and Stockholm. Weather derivatives of these cities are traded in CME. Studying the past behavior of these time series will help us build a model that can predict the future behavior of the DATs because changes in temperature follow a cyclical pattern despite the large variability (Bellini 2005).

Following and expanding the previous studies (such as Bellini 2005; Benth and Saltyte-Benth 2005, 2007; Benth et al. 2007; Zapranis and Alexandridis 2008, 2009b, 2011), a stochastic process is selected for describing the temperature process. The stochastic process will be build upon the statistical properties found

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_5,

on the seven DAT time series. In this book, WA will be applied in order to correctly identify the seasonal mean of the temperature and the seasonal variance in the residuals. In addition, the speed of mean reversion parameter is not considered constant but rather a time-varying function. A wavelet network (WN) is used to estimate nonparametrically daily values of the speed of mean reversion. In our knowledge, we are the first to do so. Estimating daily values of the speed of mean reversion gives us a better insight of the temperature dynamics. Moreover, the impact of the false specification of the speed of mean reversion on the accuracy of the pricing of temperature derivatives is significant (Alaton et al. 2002). Then, our proposed model will be evaluated and compared against other models previously proposed in literature in-sample and out-of-sample. The in-sample comparison will be based upon the distributional statistics of the residuals and fitting criteria, while the out-of-sample will be based upon the accuracy of predicting the DAT. Finally, the inclusion of a Lévy process instead of a standard BM is investigated.

The rest of the chapter is organized as follows: in Sect. 5.2, the data is described and examined. In Sect. 5.3, a model for the DATs is proposed based on the results of the data examination. Next, in Sect. 5.4 the seasonal mean of the DATs is estimated using a linear approach in Sect. 5.4.1 and WA in Sect. 5.4.2. In Sect. 5.5 the speed of mean reversion is estimated. More precisely, a linear approach is applied in Sect. 5.5.1 while a nonlinear nonparametric WN is applied in Sect. 5.5.2. Next, in Sect. 5.6, WA is used in order to identify and model the seasonal variance that exists in the residuals. In Sect. 5.7, the distributional statistics of the residuals after removing the seasonal variance are examined. Moreover, our proposed model is compared in-sample against two popular model previously proposed in literature. Next, the residuals are tested under the assumption of a Lévy motion driving noise process in Sect. 5.7.3. In Sect. 5.8, an evaluation of our model out-of-sample is performed. Finally, in Sect. 5.9, we conclude.

5.2 Data Description

Since the underlying index of weather derivatives is a weather variable like temperature, rainfall, precipitation, or snowfall, weather data is important for pricing these derivatives. Not only an adequate amount of data is needed but also it has to be of high quality for an appropriate pricing and risk management of the weather risk (Dunis and Karalis 2003).

Easy access to high-quality weather data for long periods and for various stations would help the market evolve and would offer liquidity. Unfortunately, it is still very hard and costly to obtain this type of data. Moreover, the available to researchers' datasets have many flaws, like missing data, gaps, and errors (Nelken 2000).

Some stations had to be moved during the years or to be replaced by more modern equipment; as a result, jumps will occur on the data. Another aspect is the range of the data. Previous studies use datasets containing historical data from 5 to

230 years to fir various models. However, if a very long period is considered, then the datasets will be affected by trends like urban effects. On the other hand, when studying very small datasets, there is a possibility that important dynamics of the temperature process will not be revealed which will result to an incorrect model and to mispricing of the corresponding weather contracts. Finally, there are effects like urban heating or extreme weather patterns like the El Niño and La Niña that must be accounted when pricing a weather derivative.

For this book, we obtained data for the cities that are traded in CME. At the end of 2009, the CME trades weather products written on the following ten European cities: Amsterdam, Barcelona, Berlin, Essen, London, Madrid, Oslo, Paris, Rome, and Stockholm. In the USA, there are contracts for the following 24 cities: Atlanta, Baltimore, Boston, Chicago, Cincinnati, Colorado Springs, Dallas, Des Moines, Detroit, Huston, Jacksonville, Kansas City, Las Vegas, Little Rock, Los Angeles, Minneapolis – St. Paul, New York, Philadelphia, Portland, Raleigh, Sacramento, Salt Lake City, Tucson, and Washington D.C. Also, there are 6 Canadian cities (Calgary, Edmonton, Montreal, Toronto, Vancouver, and Winnipeg), 3 Australian cities (Brisbane, Melbourne, and Sydney), and finally 3 Japanese cities (Hiroshima, Tokyo, and Osaka). Unfortunately, quality data only for the European cities were obtained. The data corresponding to the European cities were provided by the ECAD.[1] The weather variable we are interested in is the DAT. In ECAD, the DAT is measured as the average of the daily maximum and minimum temperature and is measured in Celsius degrees (°C). European weather contracts traded on CME use the same measurement for the temperature. Precision with which temperature in ECAD is measured is 0.1 °C. Unfortunately, data from Essen were not available, while the missing values from Barcelona and London were more than 50% of the data; hence, these three cities are not included in our analysis.

The dataset consists of 18.615 values, corresponding to the DAT of 51 years (1951–2001) in cities that derivatives are actively traded in CME. In order for each year to have equal observations, the 29th of February was removed from the data.

One of the major problems of the data is the missing values. In Dunis and Karalis (2003), different methods for filling the missing data were described. In the naïve approach, the missing value is replaced be the temperature at the same day the previous year. This method is highly likely to produce large jumps in the temperature time series. Another approach is to fill the missing data using nearby weather stations to the one in interest. Dunis and Karalis (2003) propose and test more complex methods like the expectation maximization algorithm or the data augmentation algorithm, state space models, Kalman filter, NNs, and PCA with the later to outperform all other methods. However, PCA requires additional correlated cleaned temperature data (Dunis and Karalis 2003).

In this book, the procedure described below is followed in order to fill the missing values. Let T_t be the temperature at day t which value is missing. First, the average temperature of that particular day across the years is calculated denoted

[1] http://eca.knmi.nl/

by Avy. Next, the average temperature of 7 days ago and 7 days after the missing value is calculated denoted by Avd. Then, the missing value is replaced by the average of these two parameters:

$$T_{t,miss} = \frac{\left(T_{Avy,t} + T_{Avd,t}\right)}{2} \tag{5.1}$$

$$T_{Avy,t} = \frac{1}{N}\sum_{yr=1}^{N} T_{t,yr} \tag{5.2}$$

$$T_{Avd,t} = \frac{\sum_{i=1}^{7} T_{t-i} + \sum_{i=1}^{7} T_{t+i}}{14}. \tag{5.3}$$

The above procedure is very easy in implementation. More precisely, a normal average is obtained by Equation (5.2) which is balanced by the temporal temperature conditions around the missing values by (5.3).

In Fig. 5.1, the DATs for the seven European cities for the period 1/1/1991–31/12/2000 are presented. A closer inspection of Fig. 5.1 reveals a seasonal cycle of 1 year as it was expected. Moreover, extreme values in summer and winter are evident in all cities. In order to obtain a better insight of the temperature dynamics, the descriptive statistics of the DATs are examined. In Table 5.1, the descriptive statistics of the DATs in the seven cities is presented. The mean temperature ranges from 6.49 in Oslo to 15.57 in Rome. As it is shown on Table 5.1, the variation of the DAT is quite large in every city. The standard deviation ranges from 6.08 for Amsterdam, while it is 7.91 for Berlin. It is clear from Table 5.1 that cities with warmer climate like Amsterdam, Rome, and Paris have smaller standard deviation, while cities with colder climate with large periods of winter like Oslo, Berlin, and Stockholm have the largest standard deviation values. The difference between the maximum and minimum temperatures is around 30° for Rome and Madrid, while it is over 40° for Berlin, Oslo, and Stockholm. The maximum and minimum temperatures vary from city to city, but it is explained from their location. The above results indicate that temperature is very volatile, and it is expected to be difficult to accurately model and predict it.

Negative skewness is evident in all cities with the exception of Madrid, Stockholm, and Rome. Moreover, all cities exhibit excess negative kurtosis. The kurtosis is 2 for Madrid and Rome, while the largest kurtosis value is 2.6 for Amsterdam. The above results indicate that the distribution of the DAT in Europe is platykurtic with lower and wider peak where the mass of the distribution is concentrate on the left tail (on the right tail for Madrid and Rome). Finally, a normality test is reported on Table 5.1. In all cities, the normality is strongly rejected by a (Jarque–Bera) JB test. JB tests of the null hypothesis that the sample in vector sample data comes from a normal distribution with unknown mean and variance, against the alternative that it does not come from a normal distribution:

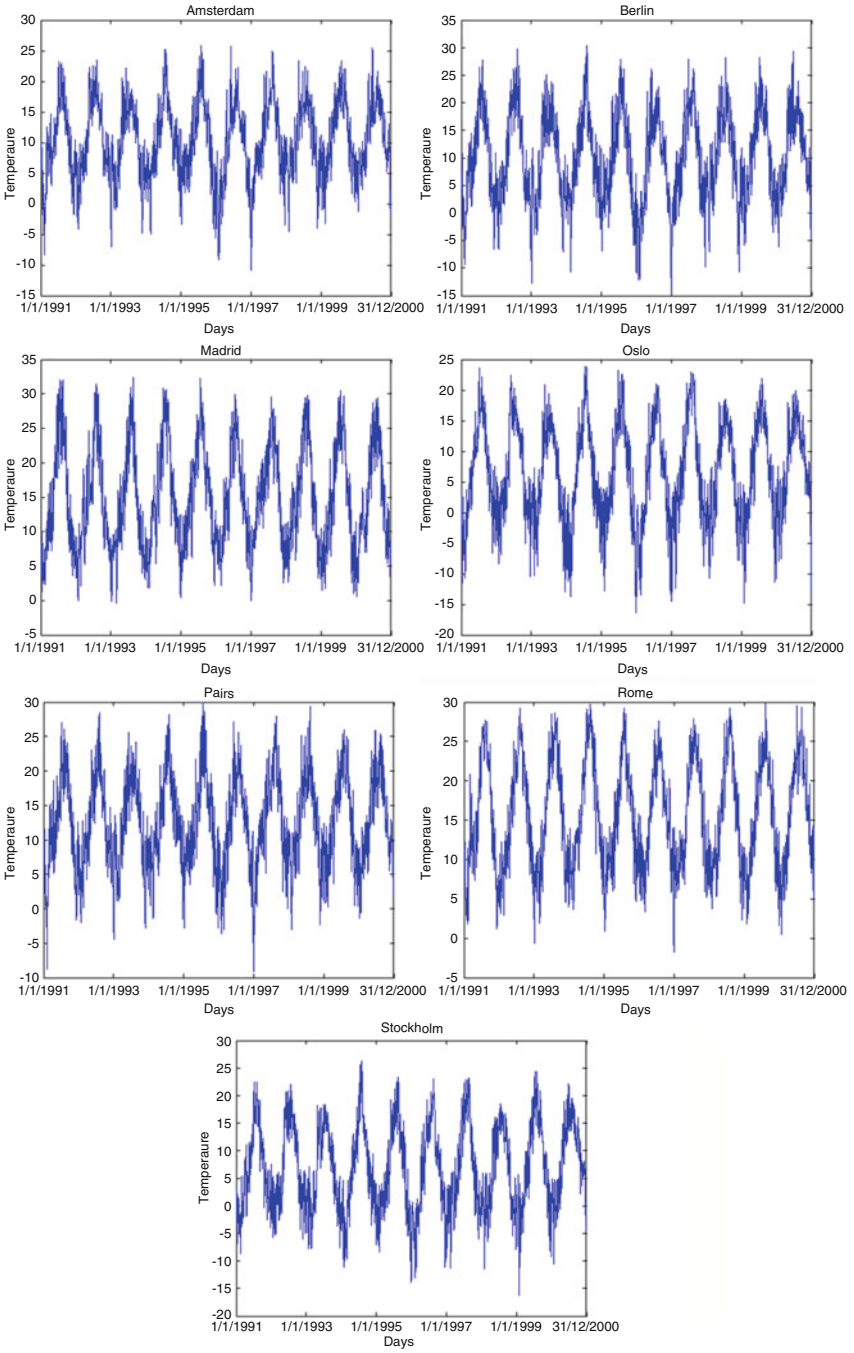

Fig. 5.1 Daily average temperature of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

Table 5.1 Descriptive statistics of the daily temperature for the period of 1991–2000

	Mean	St. dev	Max	Median	Min	Skewness	Kurtosis	JB-Stat	*P* value
Amsterdam	10.23	6.08	25.80	10.10	−10.90	−0.18	2.67	36.27	0.0000
Berlin	10.01	7.91	30.40	10.00	−14.70	−0.08	2.38	62.23	0.0000
Madrid	15.06	7.35	32.40	14.15	−0.40	0.30	2.06	189.47	0.0000
Oslo	6.49	7.88	23.90	6.10	−16.40	−0.10	2.29	83.28	0.0000
Paris	12.51	6.44	29.90	12.40	−9.10	−0.04	2.48	41.41	0.0000
Rome	15.57	6.73	29.80	15.10	−1.80	0.07	1.93	178.68	0.0000
Stockholm	6.84	7.68	26.30	6.30	−16.20	0.03	2.20	98.61	0.0000

St. dev standard deviation, *JB-Stat* Jarque–Bera statistic, *P value p* value of the JB test

$$\begin{aligned} H_0 &= \text{The sample comes from a normal distribution} \\ H_1 &= \text{The sample does not come from a normal distribution} \end{aligned} \tag{5.4}$$

The JB test is a two-sided goodness-of-fit test suitable when a fully specified null distribution is unknown and its parameters must be estimated. The test statistic is

$$JB = \frac{n}{6}\left(s^2 + \frac{(k-3)^2}{4}\right), \tag{5.5}$$

where n is the sample size, s is the sample skewness, and k is the sample kurtosis. For large sample sizes, the test statistic has a chi-square distribution with two degrees of freedom. The critical value at 5% significance level is 5.93. From (5.5), it is clear that the JB statistic is very sensitive to large values of kurtosis and skewness. The JB statistic is over 36 in all cases, and the p values are zero indicating the rejection of the null hypothesis that the temperature at the seven European cities follows a normal distribution. Figure 5.2 presents the empirical distributions of the DAT in the seven European cities. It is clear that all cities exhibit a bimodal distribution where the two peaks correspond to summer and winter temperatures.

In order to obtain better understanding of the temperature dynamics, the mean, the standard deviation, the skewness, and kurtosis of DAT were estimated. The mean of the DAT, $T_{avy,t}$, was estimated by (5.2) using only observations for each particular day t. In Fig. 5.3, the seasonal pattern is clear. For all cities, the temperature has its highest values during the end of July and the beginning of August, while the lowest values are observed during the end of December and until the beginning of February. A closer inspection of Fig. 5.3 reveals that the mean DAT in Amsterdam fluctuates from 1.9^oC to 19.7^oC. Similarly, in Berlin, the mean temperature fluctuates from -1.2^oC to 22.4^oC, from 4.9^oC to 28.3^oC in Madrid, from -5.7^oC to 18.8^oC in Oslo, from 3.1^oC to 23^oC in Paris, from 6^oC to 27^oC in Rome, and from -4.2^oC to 19.5^oC in Stockholm.

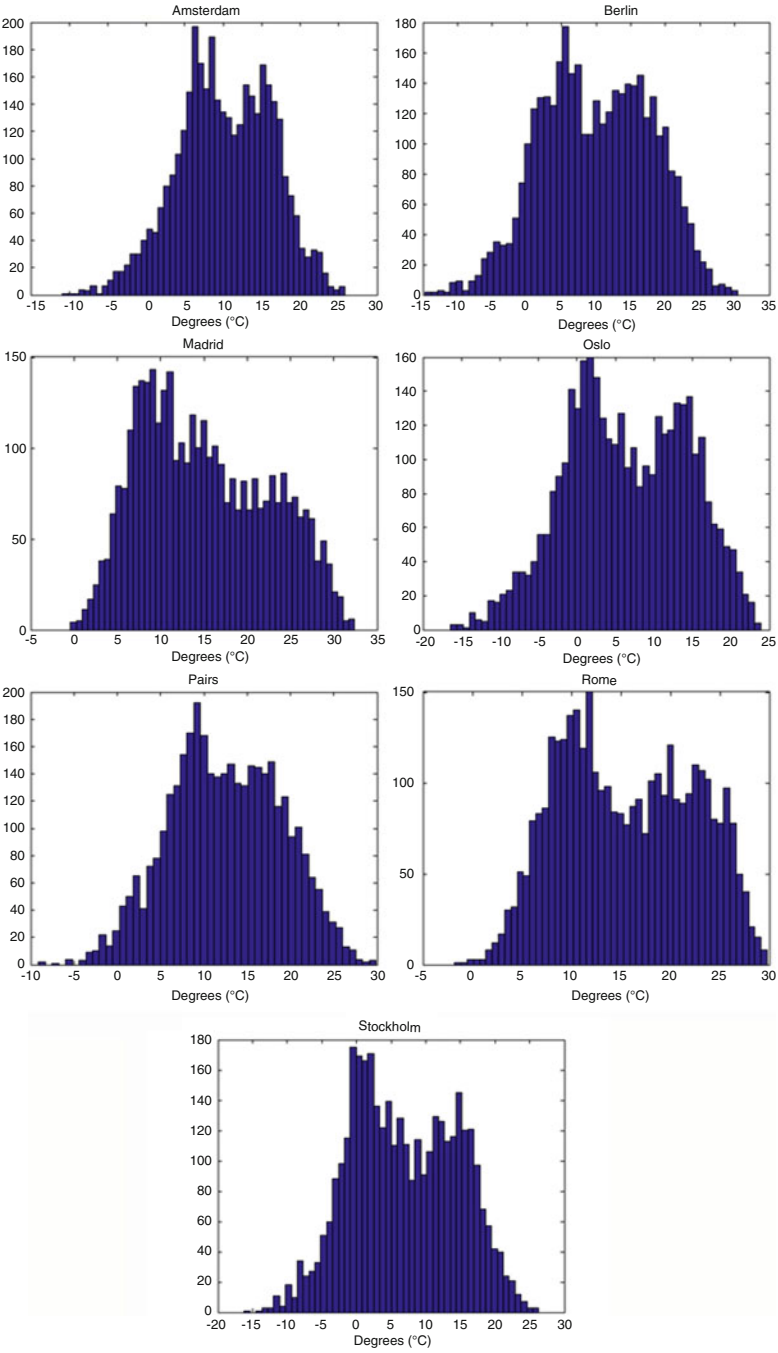

Fig. 5.2 Empirical distributions of the DAT of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

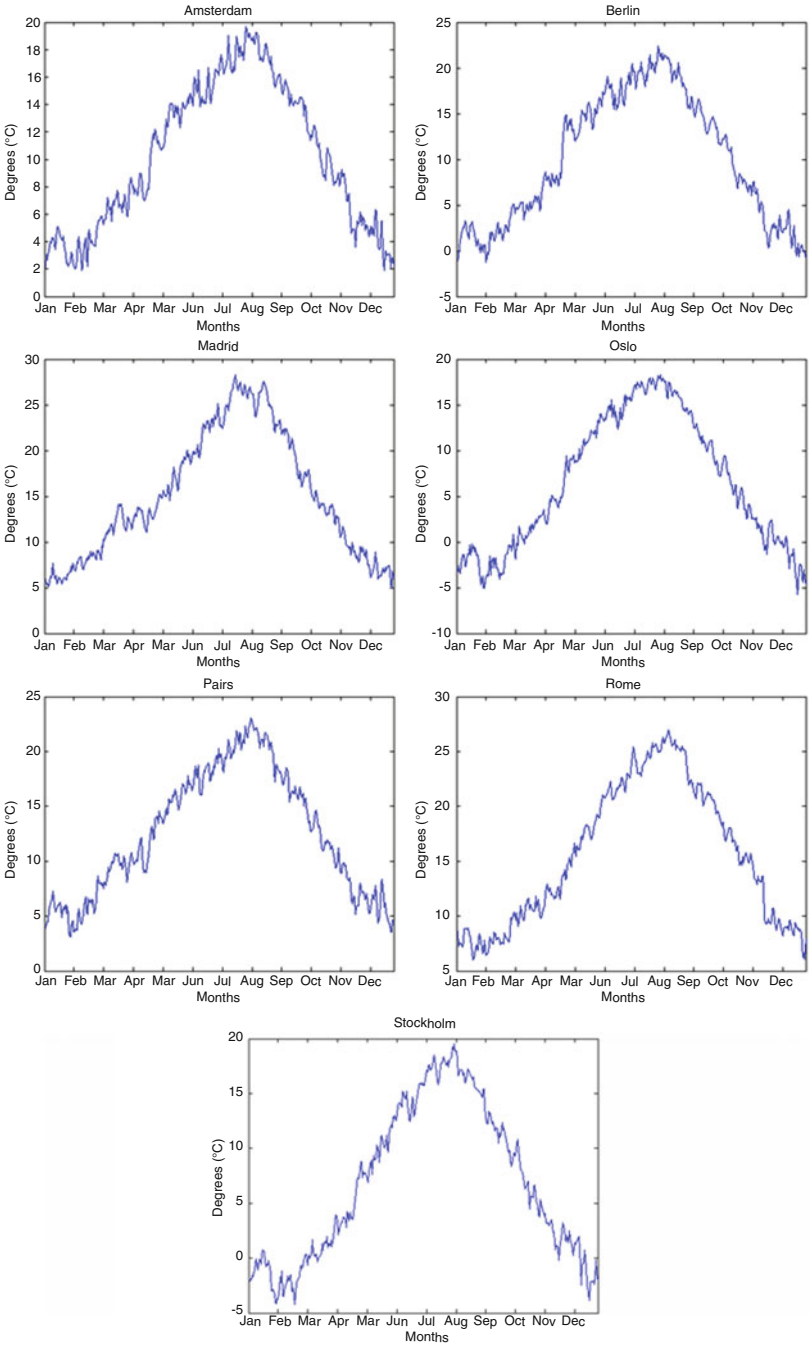

Fig. 5.3 Mean of the daily average temperature of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

Next, the standard deviation of the DAT is estimated. The standard deviation is given by

$$s_t = \sqrt{\frac{1}{N}\sum_{yr=1}^{N}\left(T_{t,yr} - T_{avy,t}\right)^2}. \tag{5.6}$$

In Fig. 5.4, the standard deviation for the seven cities is presented. Observing Fig. 5.4, it is clear that the standard deviation is higher in the winter period, while it is smaller in summer for all cities with exception of Madrid. Our results confirm the studies of Bellini (2005), Benth and Saltyte-Benth (2005, 2007), Benth et al. (2007), Zapranis and Alexandridis (2008, 2009b, 2011).

Figure 5.5 presents the estimated skewness for each day t for the seven cities. The skewness is given by

$$sk_t = \frac{1}{N}\sum_{yr=1}^{N}\left(\frac{T_{t,yr} - T_{avy,t}}{s_t}\right)^3. \tag{5.7}$$

Figure 5.5 reveals that the skewness tends to increases during the summer months, while it decreases during the winter months with exception of Rome and Madrid. In general, the skewness is negative at winter months and positive at summer months. This means that it is more likely to have warmer days than average in summer and colder days than average in winter (Bellini 2005).

Finally, the kurtosis on each day t is estimated by

$$k_t = \frac{1}{N}\sum_{yr=1}^{N}\left(\frac{T_{t,yr} - T_{avy,t}}{s_t}\right)^4 \tag{5.8}$$

and can be found in Fig. 5.6. Observing Fig. 5.6 does not reveal any seasonal pattern of the kurtosis. On the other hand, it is clear that for all cities the kurtosis has small deviations around two with many upward large spikes.

Next, the correlation of the temperature between different cities is examined. If strong correlation is present, then weather derivatives of correlated cities can be used for risk management and reduction of the basis risk. In Table 5.2, the correlation of the temperature between the different sites in Europe is shown. As it was expect, the correlation in general is very high and over 0.81, while it is 0.954 between Oslo and Stockholm. The correlation values are explained by the geographical location of each site. As it was expected, there is large correlation between Oslo and Stockholm and between Amsterdam and Berlin, while the correlation is smaller between distant cities like Madrid and Oslo or Stockholm. However, the correlation should be estimated after removing all seasonal components.

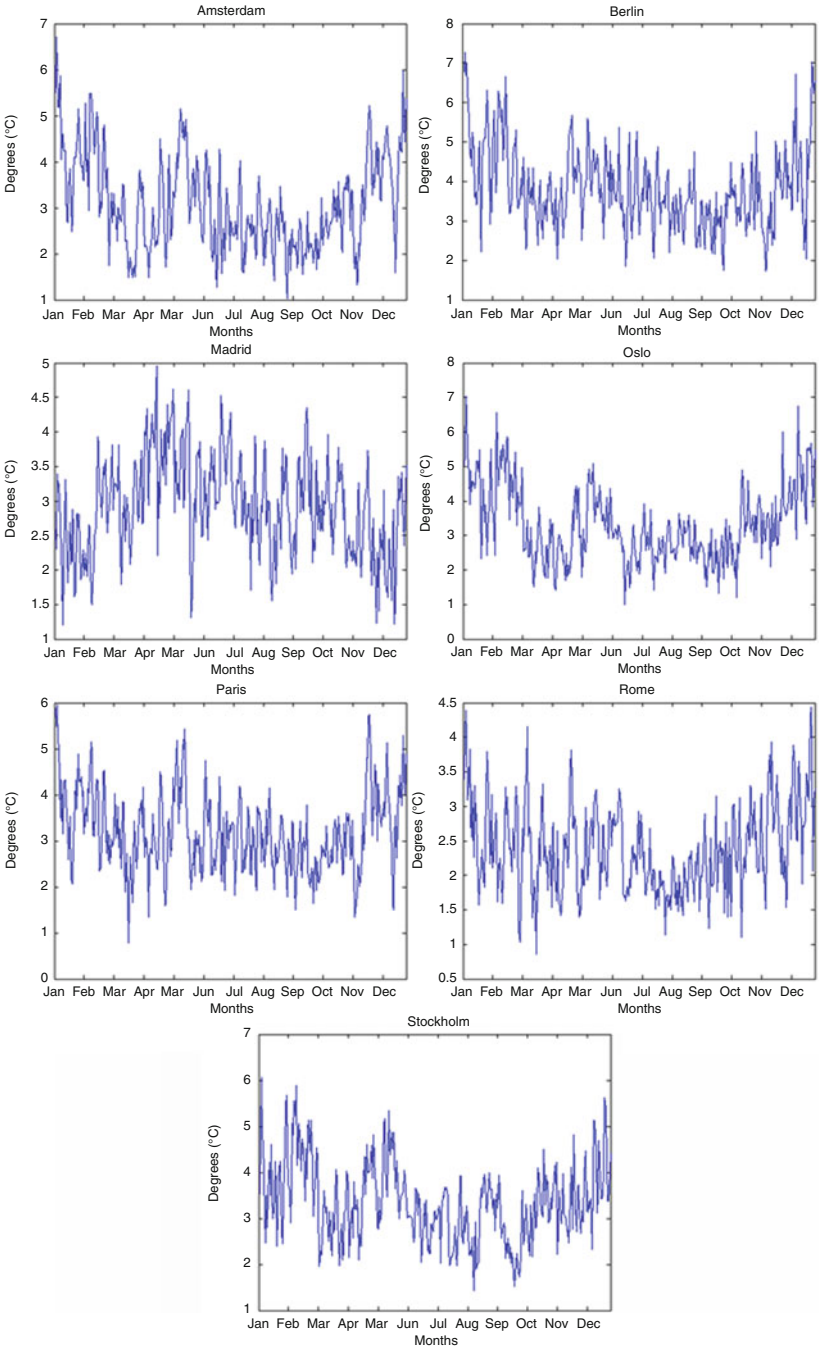

Fig. 5.4 Standard deviation of the daily average temperature of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

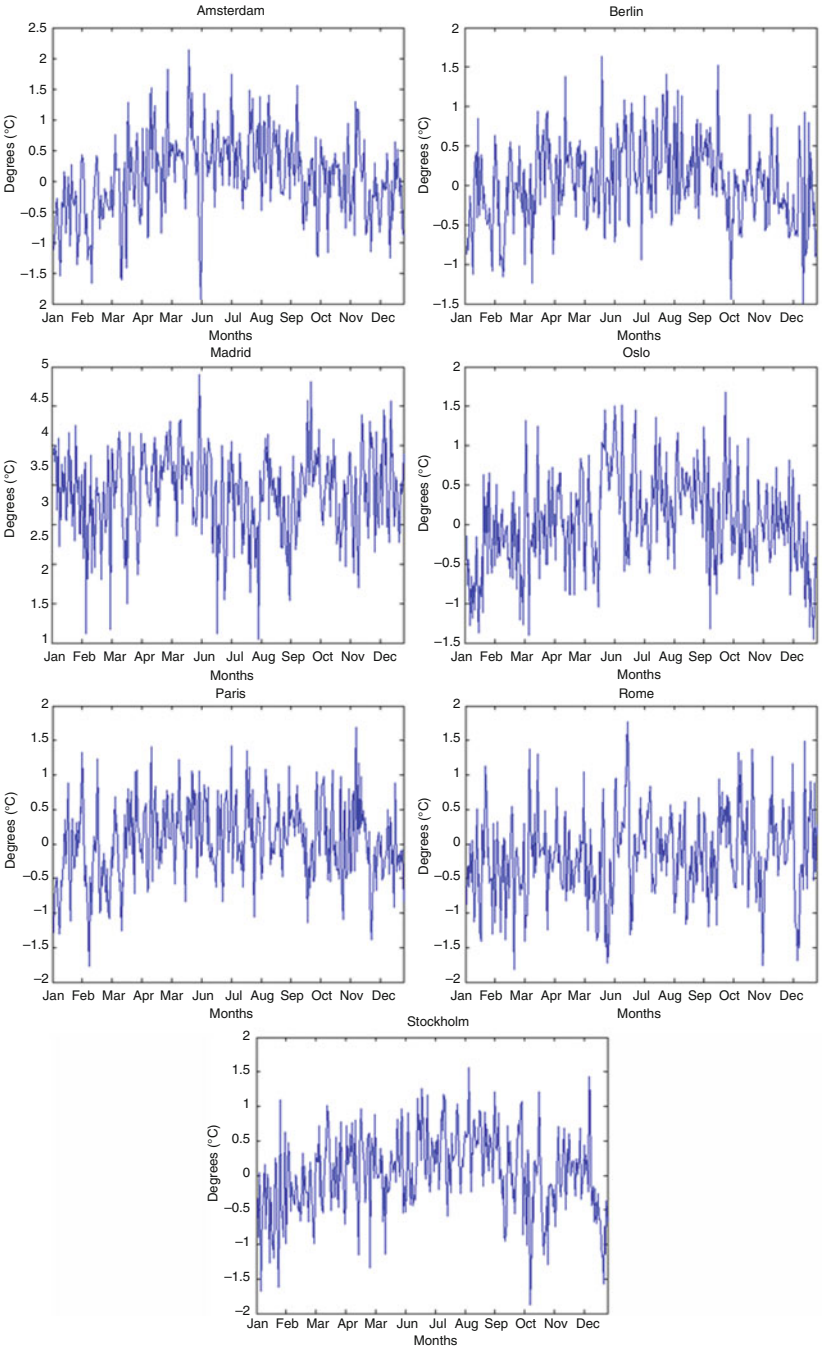

Fig. 5.5 Skewness of the daily average temperature of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

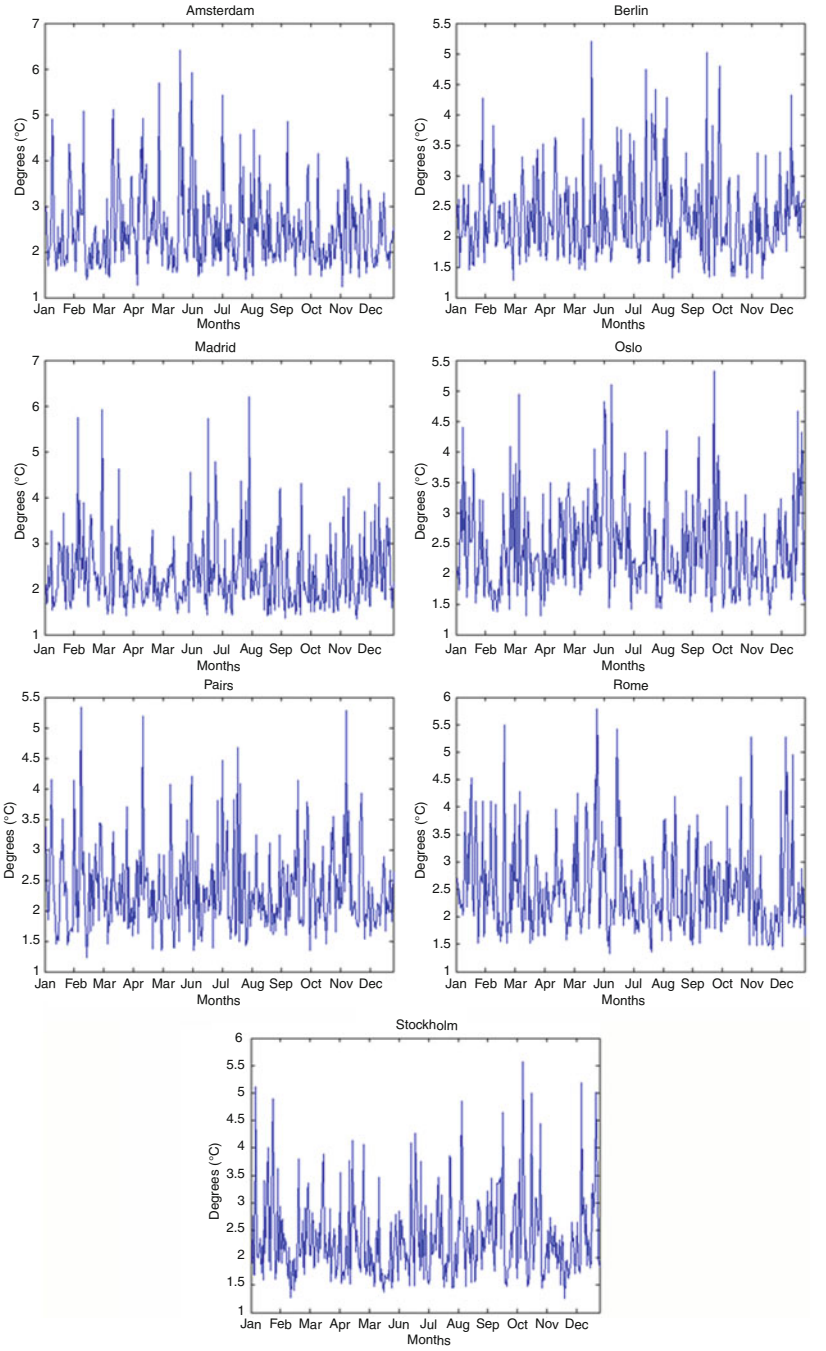

Fig. 5.6 Kurtosis of the daily average temperature of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

Table 5.2 Correlation matrix of the temperature before removing the seasonal components

	Amsterdam	Berlin	Madrid	Oslo	Paris	Rome	Stockholm
Amsterdam	1.000	0.936	0.819	0.875	0.943	0.833	0.874
Berlin	0.936	1,000	0.811	0.890	0.901	0.851	0.901
Madrid	0.819	0.811	1.000	0.812	0.855	0.858	0.812
Oslo	0.875	0.890	0.812	1.000	0.836	0.821	0.954
Paris	0.943	0.901	0.855	0.836	1.000	0.847	0.834
Rome	0.833	0.851	0.858	0.821	0.847	1.000	0.835
Stockholm	0.874	0.901	0.812	0.954	0.834	0.835	1.000

Table 5.3 Hurst exponent of the temperature before removing the seasonal components

	Amsterdam	Berlin	Madrid	Oslo	Paris	Rome	Stockholm
Hurst	0.7514	0.6414	0.7592	0.5222	0.6770	0.6161	0.6256

Table 5.4 Unit root tests of the temperature time series

	ADF	*P* value	Lags	KPSS	Bandwidth
Amsterdam	−6.0820	0.0000	6	0.0844	44
Berlin	−5.1116	0.0000	8	0.0505	44
Madrid	−4.6592	0.0001	7	0.0341	45
Oslo	−3.6377	0.0051	12	0.0531	45
Paris	−5.2100	0.0000	9	0.0543	44
Rome	−3.7785	0.0032	9	0.0380	45
Stockholm	−4.0654	0.0011	10	0.0753	45
Critical values 5%:	−2.8621			0.4630	

ADF augmented Dickey–Fuller, *KPSS* Kwiatkowski–Phillips–Schmidt–Shin

In Table 5.3, the Hurst exponent is estimated. In Brody et al. (2002), Benth (2003), and Caballero et al. (2002), fractional models were proposed with evidence that the Hurst exponent is greater than 0.5. However, as it was shown in Bellini (2005), not all seasonal effects were removed from the data before the estimation of the Hurst exponent which lead to an unsubstantial value. The Hurst exponent was estimated using iterative methods described in Koutsoyiannis (2003). Our results indicate that the Hurst exponent is significantly different than 0.5, with an exception of Oslo. More precisely, the Hurst exponent is 0.5222 for Oslo, and for the remaining cities, it varies from 0.6161 in Rome to 0.7592 in Madrid. However, the Hurst exponent must be calculated after all seasonal effects were removed (Bellini 2005).

In Table 5.4, two unit root tests were performed in the DAT for the seven cities. Each time series is tested for unit root using an ADF test. The ADF performed using the Schwartz information criterion in order to select the optimal number of lagged values. The null hypothesis of the ADF test is that the time series has a unit root versus the one-sided alternative that the root is less than one:

$$H_0 : \rho = 1$$
$$H_1 : \rho < 1 \quad . \tag{5.9}$$

In Table 5.4, the ADF value as well as the optimal number of lags used and the p values are reported. The lag length used to perform the test is selected by minimizing the Schwarz criterion. A close inspection of Table 5.4 reveals that the null hypothesis of a unit root is rejected since the ADF statistic is always smaller than the critical value at 5% significance level. Moreover, the p values are almost zero for the seven cities.

In order to obtain a more powerful test, the KPSS unit root test is also performed. In contrast to the ADF test, the KPSS tests the null hypothesis that the time series is stationary versus the alternative that the time series is nonstationary (a unit root exists):

$$H_0 : \rho < 1$$
$$H_1 : \rho = 1 \, . \tag{5.10}$$

The optimal bandwidth number was estimated using the Newey–West method. Table 5.4 reports both the KPSS values and the optimal bandwidth number for each city. The KPSS statistic has a value smaller than the critical value 0.463 for all cities; hence, the null hypothesis, which the time series is stationary, cannot be rejected for all cities.

5.3 Statistical Modeling of Daily Average Temperature

Many different models have been proposed in order to describe the dynamics of a temperature process. In this section, a model for the seven cities studied in the previous section will be derived. Studying temperature data, Cao and Wei (1999, 2000, 2003) and Cao et al. (2004) build their framework on the following five assumptions about DAT:

- It follows a predicted cycle.
- It moves around a seasonal mean.
- It is affected by global warming and urban effects.
- It appears to have autoregressive changes.
- Its volatility is higher in the winter than in summer.

As it will be shown in the rest of the chapter, our results confirm the above assumptions.

It is known that temperature follows a predicted cycle. As it was expected and it is shown on Fig. 5.1, a strong cycle of 1 year is evident in all cities. It is also known that temperature has a mean-reverting form. Temperature moves around a seasonal mean and cannot deviate from that seasonal mean for long periods. This can be verified by Figs. 5.1, 5.3 and 5.4. In other words, it is not possible to observe temperatures of 20^oC in winter in Oslo. Additionally, temperature is affected by

global warming and urban effects. In areas under development, the surface temperature rise as more people and buildings concentrate. This is due to the sun's energy absorbed by the urban buildings and the emissions of vehicles, industrial buildings, and cooling units. Hence, urbanization around a weather station results to an increment in the observed measurements of temperature. Finally, observing Fig. 5.4, it is clear that the temperature volatility is higher in winter than in summer. Following Benth and Saltyte-Benth (2007), a model that describes the temperature dynamics is given by a Gaussian mean-reverting O–U process defined as follows:

$$dT(t) = dS(t) + \kappa(T(t) - S(t))dt + \sigma(t)dB(t), \tag{5.11}$$

where $T(t)$ is the average daily temperature, κ is the speed of mean reversion, $S(t)$ is a deterministic function modeling the trend and seasonality, $\sigma(t)$ is the daily volatility of temperature variations, and $B(t)$ is the driving noise process. As it was shown in Dornier and Queruel (2000), the term $dS(t)$ should be added for a proper mean reversion toward the historical mean, $S(t)$.

In Benth and Saltyte-Benth (2007), both $S(t)$ and $\sigma^2(t)$ are being modeled as truncated Fourier series, that is,

$$S(t) = Trend_t + \sum_{i=1}^{I_1} a_i \sin(2i\pi(t - f_i)/365) + \sum_{j=1}^{J_1} b_j \cos(2j\pi(t - g_i)/365) \tag{5.12}$$

$$\sigma^2(t) = c + \sum_{i=1}^{I_2} c_i \sin(2i\pi t/365) + \sum_{j=1}^{J_2} d_j \cos(2j\pi t/365) \tag{5.13}$$

$$Trend_t = a + bt. \tag{5.14}$$

Intuitively, it is expected that the speed of mean reversion is not constant. If the temperature today is away from the seasonal average (a cold day in summer), then it is expected that the speed of mean reversion is high, that is, the difference of today and tomorrow's temperature is expected to be high. In contrast, if the temperature today is close to the seasonal variance, we expect the temperature to revert to its seasonal average slowly. To capture this feature, the speed of mean reversion is modeled by a time-varying function $\kappa(t)$. Hence, the structure to model the dynamics of the temperature evolution becomes

$$dT(t) = dS(t) + \kappa(t)(T(t) - S(t))dt + \sigma(t)dB(t). \tag{5.15}$$

Moreover, in Benth and Saltyte-Benth (2007), the historical mean is captured by a simple sinusoid, that is, $I_1 = 1$ and $J_1 = 0$ in (5.12). In addition, the length of the series of the $\sigma^2(t)$ is arbitrary set to $I_2 = 4$ and $J_2 = 4$. In this book, the true structure of the seasonal mean $S(t)$ and the seasonal variance $\sigma^2(t)$ are extracted using WA. Hence, we model them as follows:

$$S(t) = Trend_t + \sum_{i=1}^{I_1} a_i \sin(2i\pi(t - f_i)/(p_i \cdot 365)) + a_{I_1+1}\left(1 + \sin\left(2\pi\left(t - f_{I_1+1}\right)/(p_{I_1+1} \cdot 365)\right)\right)\sin(2\pi t/365) \tag{5.16}$$

$$\sigma^2(t) = c + \sum_{i=1}^{I_2} c_i \sin\left(2p_i'\pi t/365\right) + \sum_{j=1}^{J_2} d_j \sin\left(2p_j'\pi t/365\right), \tag{5.17}$$

and the trend is given by (5.14). In order to identify the terms I_1, p_i in (5.16) and I_2, J_2, p_i' in (5.17), we decompose the temperature series using a wavelet transform.

Finally, the driving noise process $B(t)$ is modeled by a standard BM. In Fig. 5.7, the histogram of the first difference of the DAT and the normal distribution (solid line) is presented. A closer inspection of Fig. 5.7 reveals that the empirical distribution of the first difference of the DAT is similar to the normal distribution. Hence, selecting the BM as the driving noise process seems logical. This hypothesis will be further tested later in this chapter.

5.4 The Seasonal Mean

In this section, a method for estimating and removing the trend and the seasonal component of the temperature series is described.

5.4.1 The Linear Approach

The simplest method is the linear approach, where the trend is modeled by linear trend. Similarly, the seasonal mean is described by a sinusoid. Thus, method very simple is accurate also. In Alaton et al. (2002), this method was followed. More precisely, the seasonal component and the trend were modeled by

$$S(t) = A + Bt + C\sin(2\pi t/365 + \varphi), \tag{5.18}$$

where A denotes the amplitude of the sinusoid which is the peak deviation of the function from its center position. In other words, it denotes the difference between the minimum and maximum temperature from the mean. Since the maximum or the minimum temperature is not observed in the 1st of January, the parameter φ is inserted to the model to calibrate the sinusoid. Equation (5.18) can be written equivalently

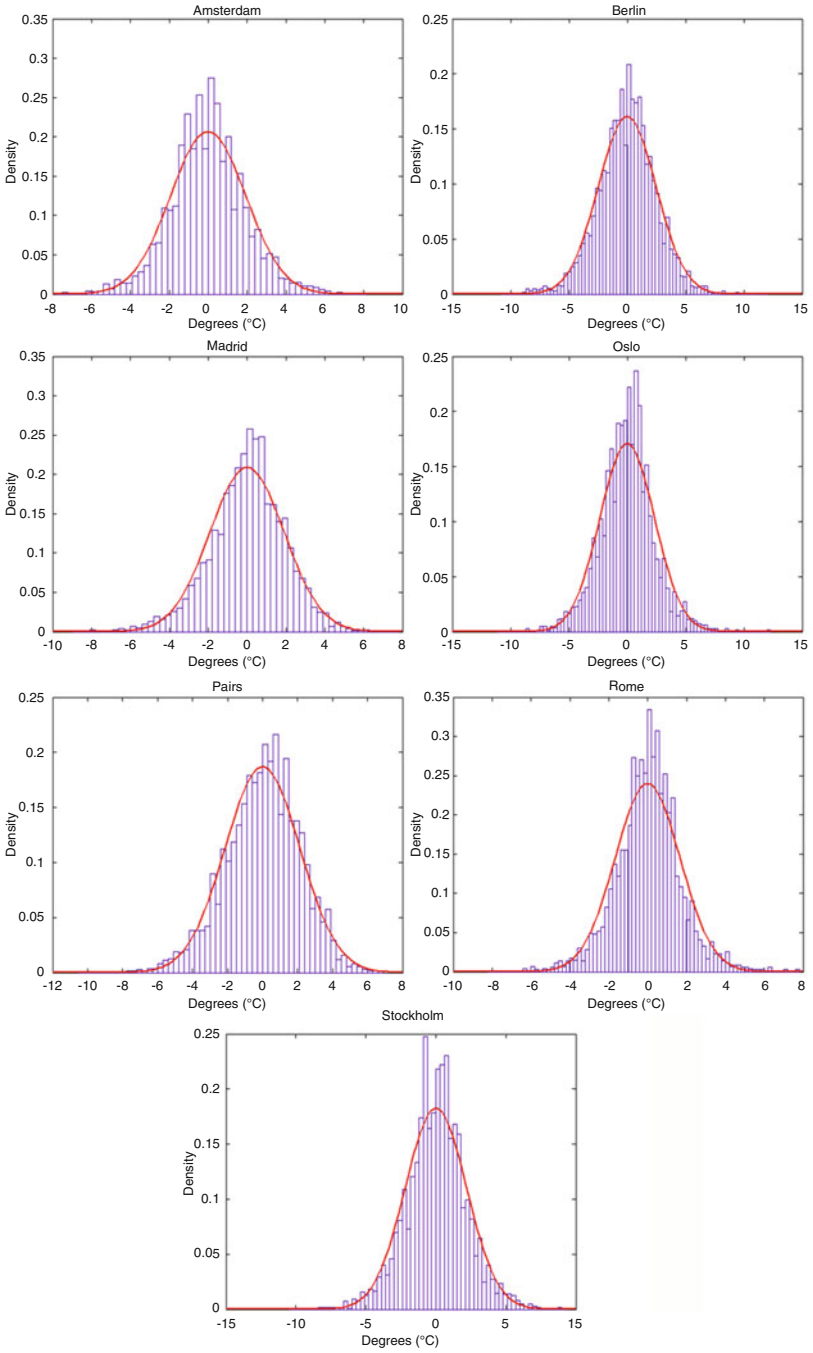

Fig. 5.7 Empirical and normal distribution (*solid line*) of the daily average temperature differences of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

$$S(t) = a + bt + c\sin(2\pi t/365) + d\cos(2\pi t/365). \tag{5.19}$$

The relation of the parameters between (5.18) and (5.19) is given by the following equations:

$$A = a \tag{5.20}$$

$$B = b \tag{5.21}$$

$$C = \sqrt{c^2 + d^2} \tag{5.22}$$

$$\varphi = \arctan\left(\frac{d}{c}\right) - \pi. \tag{5.23}$$

The estimated parameters of this model for the seven European cities are presented in Table 5.20. A closer inspection of Table 5.20 reveals a positive trend in all cities except Madrid. Hence, an upward trend in temperature for the last 10 years is evident in the remaining six cities. Also, the difference between the minimum and maximum ranges from 14.5^oC in Amsterdam to 19.8^oC in Oslo. In part (a) of Fig. 5.8, the DATs in Berlin and the seasonal mean are presented. A closer inspection of part (a) of Fig. 5.8 indicates that this simple method cannot capture the deviation of the temperature when there is a cold winter or a hot summer. In other words using only one sinusoid, the amplitude of the seasonal temperature is constant across the years. Clearly, this is not the case as it is shown in part (a) of Fig. 5.8.

5.4.2 A More Advanced Approach: Wavelet Analysis

Our results in the previous section indicate that a simple sinusoid is not sufficient to capture the changes in the dynamics of the seasonal mean. Alternatively, a truncated Fourier series can be used. Moreover, instead of only a summation of sinusoid, also multiplication between sinusoids is used. Hence, as it was described in the previous section, the historical mean was modeled by (5.16)

$$\begin{aligned} S(t) = Trend_t &+ \sum_{i=1}^{I_1} a_i \sin(2i\pi(t - f_i)/(p_i \cdot 365)) \\ &+ a_{I_1+1}\left(1 + \sin\left(2\pi\left(t - f_{I_1+1}\right)/(p_{I_1+1} \cdot 365)\right)\right)\sin(2\pi t/365), \end{aligned}$$

where the trend is given by (5.14).

In order to justify the structure of the seasonal part of the temperature and to identify the terms I_1, p_i in (5.16), the temperature series is decomposed using a WT.

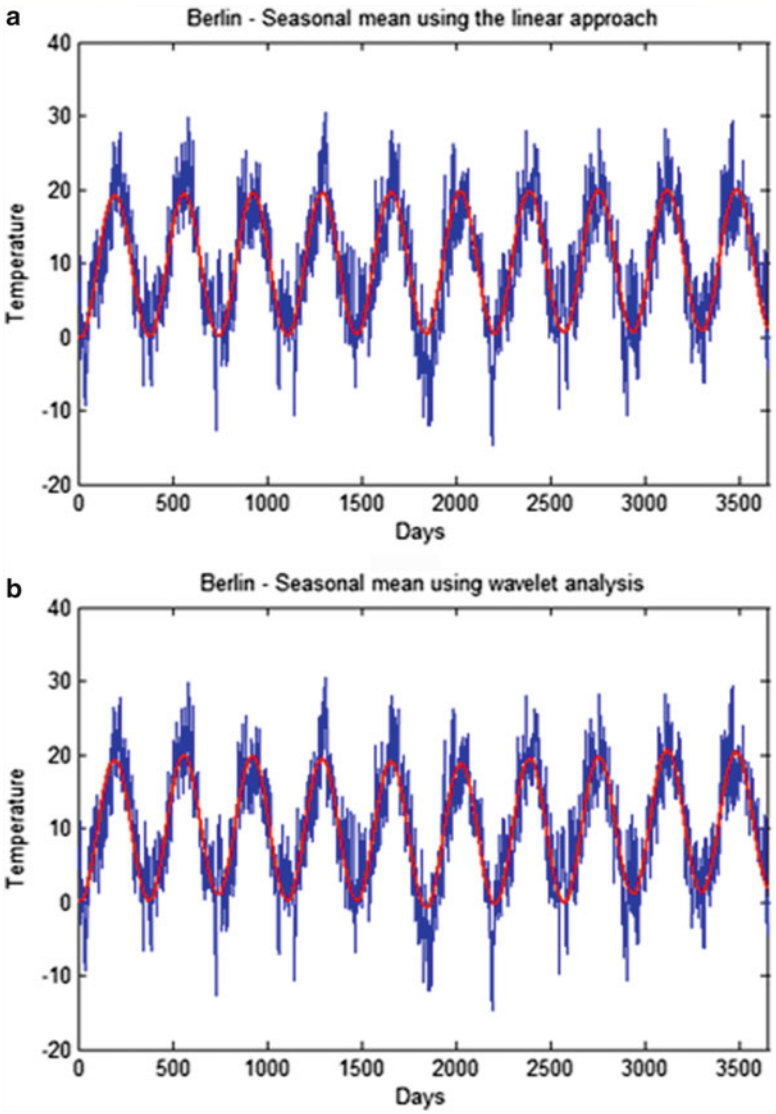

Fig. 5.8 Daily average temperature and the seasonal mean in Berlin using (**a**) the linear approach and (**b**) wavelet analysis for the years 1991–2000

Lau and Weng (1995) confirmed seasonalities in the temperature series with a period significantly greater than 1 year. Lau and Weng (1995) examined the monthly Northern Hemisphere surface temperature for the period January 1854–July 1993 using WA. They reported that the temperature has three main frequency branches: interannual (2–5 years), inter-decadal (10–12 years, 20–25 years, and 40–60 years), and century (~180 years) scales.

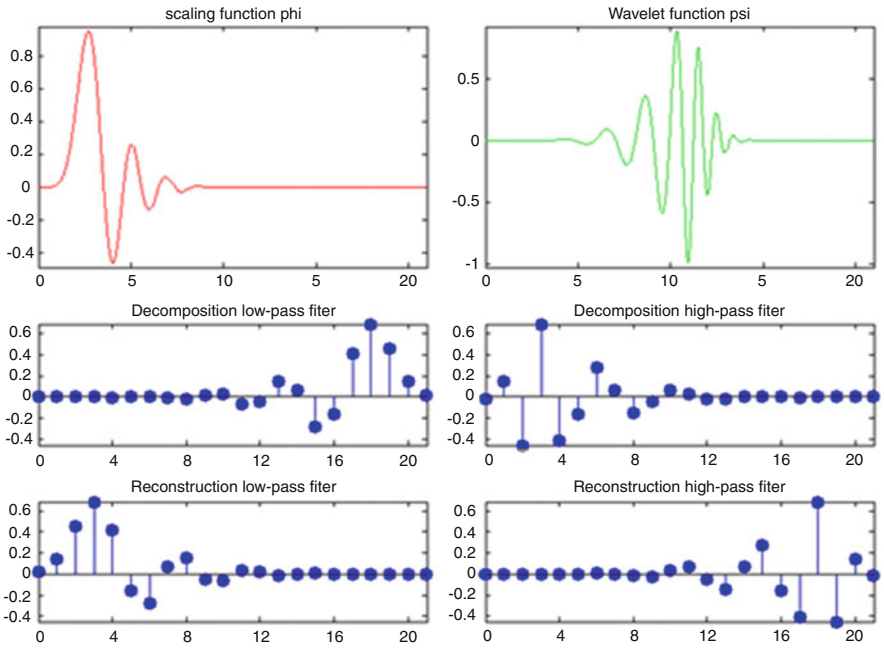

Fig. 5.9 The Daubechies 11 wavelet (*top right*) together with the scaling function (*top left*). The decomposition (*middle*) and reconstruction (*bottom*) filters are also presented with the low-pass (*left*) and high-pass (*right*) filters

This conclusion was also reached in Zapranis and Alexandridis (2006, 2008, 2009b), and Zapranis and Alexandridis (2009a). More precisely, Zapranis and Alexandridis (2006) used the Daubechies 11 wavelet at level 11 to decompose 100 years of the average daily temperature time series of Paris. Specifically, in this chapter, WA captured dynamics of temperature such as an upward trend and periodicities expanding up to 13 years.

In this book, the Daubechies wavelet family was chosen which has proved to outperform other wavelet families in various applications (Daubechies 1992). More precisely, the Daubechies 11 wavelet at level 11 was selected and applied in 50 years of DATs in each city. In Fig. 5.9, the Daubechies 11 wavelet can be found.

Here, for simplicity, we will refer analytically only to the results of the wavelet decomposition from Berlin. The results of the remaining cities are similar and can be found in Tables 5.5 and 5.6. Figure 5.10 refers to selected parts of the wavelet decomposition from Berlin; the results from the remaining cities are similar.

First, an upward trend exists in the DATs, reflecting urban and global warming. This is clearly shown in Fig. 5.10, in all approximations, a_j. Moreover, a series of cycles affects the dynamics of temperature. As expected, a 1-year cycle exists in the first seven approximations ($p_1 = 1$). Additionally, cycles of $p_2 = 2.12$, $p_3 = 6.88$, and $p_4 = 13.75$ years are evident and affect the temperature dynamics (details d_9, d_{11}

Table 5.5 Estimated parameters of the linear trend for the period 1991–2000

	a	P value	B	P value
Amsterdam	9.42	0.0000	0.000440	0.0000
Berlin	9.37	0.0000	0.000349	0.0038
Madrid	14.62	0.0000	0.000238	0.0335
Oslo	5.79	0.0000	0.000385	0.0011
Paris	11.86	0.0000	0.000353	0.0003
Rome	15.04	0.0000	0.000287	0.0056
Stockholm	5.91	0.0000	0.000509	0.0000

The coefficients of the linear trend and the corresponding p values. The parameter a is the intercept, and b is the slope

Table 5.6 Estimated parameters of the seasonal part using wavelet analysis

	Amsterdam	Berlin	Madrid	Oslo	Paris	Rome	Stockholm
Panel A							
p_1	1.00	1.00	1.00	1.00	1.00	1.00	1.00
p_2	2.12	2.12	3.93	2.12	2.20	1.96	2.29
p_3	5.50	6.88	9.17	4.58	6.88	4.23	3.93
p_4	4.23	13.75	11.00	6.88	13.75	6.11	4.23
p_5	7.86	–	–	7.86	–	13.75	7.86
p_6	13.75	–	–	9.17	–	–	13.75
p_7	–	–	–	13.75	–	–	–
p_{i+1}	7	8	–	8	7	7	6
Panel B							
a_1	−7.56	−9.79	9.27	9.72	−7.99	−8.89	9.39
a_2	−0.58	−0.27	−0.25	−0.87	−0.37	−0.29	0.97
a_3	4.95	0.56	−0.61	9.26	−0.23	−0.33	−3.00
a_4	−2.44	−0.37	0.67	272.14	0.26	0.16	3.23
a_5	5.85	–	–	650.51	–	−0.36	1.04
a_6	3.11	–	–	480.94	–	–	0.32
a_7	–	–	–	−101.47	–	–	–
a_{i+1}	0.73	0.43	–	−0.79	0.52	0.24	−0.95
f_1	−65.11	−73.79	−254.69	103.74	296.15	−1158.35	109.33
f_2	217.60	149.28	484.40	−588.68	111.07	−739.94	−411.27
f_3	−168.97	148.27	–	2508.23	43.38	3001.32	1626.54
f_4	279.21	981.76	–	1629.79	−935.66	951.38	73.52
f_5	370.59		–	184.04	–	1823.78	2938.66
f_6	1855.94	–	–	4952.47	–		−2508.79
f_7	–	–	–	1583.41	–	–	
f_{i+1}	–	–	–	1381.93	2647.40	−923.93	1359.63

In Panel A, the length of each cycle in years is presented. In Panel B, the estimated parameters of the seasonal mean are reported. Only the statistical significant parameters with p value < 0.05 are presented

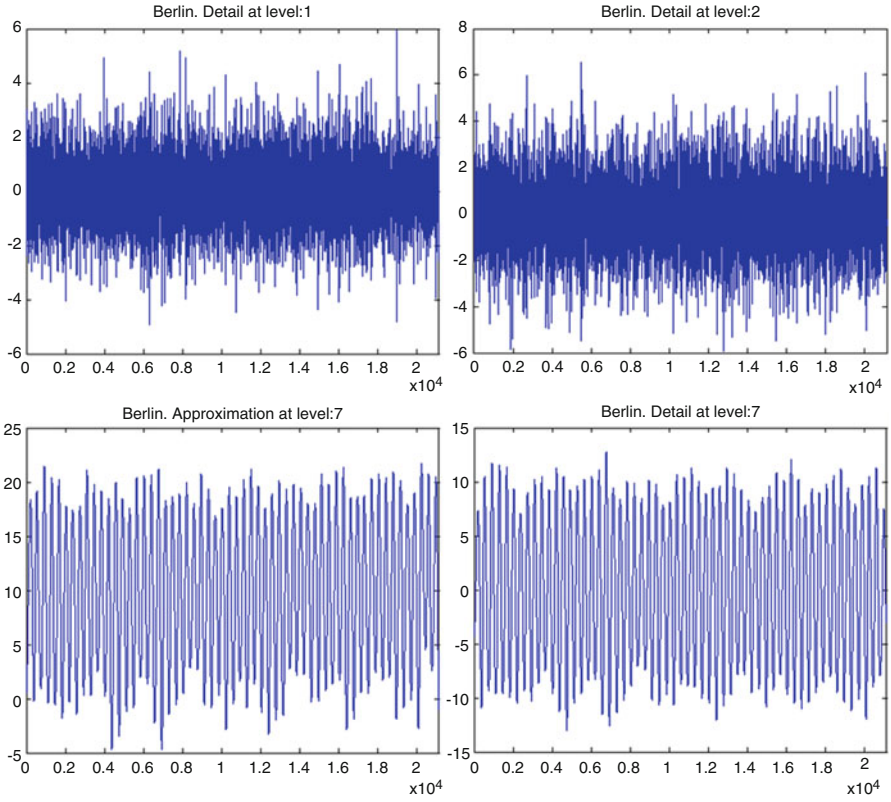

Fig. 5.10 Selected parts of the discrete wavelet decomposition in Berlin: approximations (a_j) and details (d_j). The Daubechies 11 at level 11 wavelet was applied

or a_{10}, and a_{11}, respectively). The above results indicate the periodicities in which the temperature is expected to be above or below the historical average.

Also, a product of two sinusoids was captured by WA, with period of 1 and p_{i+1} = 8 years, respectively, (d_8 and a_7). The above results indicate that every 8 years it is expected to have warmer than the usual summer and colder than the usual winter or colder than the usual summer and warmer than the usual winter.

Finally, the lower details (details d_1 and d_2) reflect the noise part of the time series. A closer inspection of the noise part reveals seasonalities, which will be extracted later. Hence, modeling the historical seasonal mean by (5.16) is justified by the results of the previous analysis.

Panel *A* of Table 5.6 reports all the cycles than can be found on the temperature dynamics using WA for the seven cities. In Table 5.6, only the statistical significant parameters with p values <0.05 are reported. Parameters with p values >0.05 are omitted and removed from our model.

First, the upward trend indicated by the results of the WA is quantified by fitting a linear regression to the temperature data. Our analysis will be focused on the last 10

years (1991–2000) since we want to emphasize on the dynamics that currently affect the temperature. Using a very large sample of historical data of DAT runs the danger for the estimated parameters to be affected by dynamics of the temperature that do not represent the future behavior of temperature anymore. Table 5.5 shows the estimated parameters a and b of the linear trend represented by (5.14). All p values are smaller than 0.05 suggesting that a trend exists, and it is statistically significant. Parameter b represents the slope of the trend. It is clear that a positive trend is present in all seven cities. The parameter b ranges from 0.000238 in Madrid to 0.000509 in Stockholm indicating an upward trend. The value of b indicates an increase in temperature from 0.9^oC in Madrid to 1.9^oC in Stockholm the last 10 years. Subtracting the trend form the original data, we obtain the detrended DAT series.

After removing the linear trend from the data, the seasonal part identified by the WA can be fitted. The results of the WA indicate that the seasonal part of the temperature takes the form of equation (5.16). Since parameters p_i were already identified by WA, next, least squares method can be applied in order to fit the parameters a_i and f_i.

The estimated parameters of the seasonal part in Berlin are as follows: $a_1 = -9.79$, $a_2 = -0.27$, $a_3 = 0.56$, $a_4 = -0.37$, $a_{i+1} = 0.43$, $f_1 = -73.79$, $f_2 = 149.28$, $f_3 = 148.27$, and $f_4 = -981.76$. On the other hand, f_{I_1+1} is not statistically significant different from zero. It is clear that the 1-year cycle has the biggest impact on the temperature dynamics since its coefficient has the largest absolute value. The estimated parameters of the seasonal part of $S(t)$ of the remaining cities are reported in panel B of Table 5.6. In Table 5.6, only the statistical significant parameters are reported. Hence, parameters with p value greater than 0.05 were considered not significant and were omitted (Aczel 1993). Next, the temperature series were deseasonalized by removing $S(t)$ from the detrended data.

In the case of Berlin, the DATs and the fitted seasonal mean together with the linear trend are presented in part (b) of Fig. 5.8. In contrast to the simple method, the seasonal mean has a better adjustment. A closer inspection of Fig. 5.8 reveals that the seasonal mean takes higher values at the beginning and at the end of the 10-year period where the temperature has higher values, while it takes lower values in the middle period where the minimum temperature has lower values.

5.5 The Speed of Mean Reversion

In the previous section, the temperature series were detrended and deseasonalized. In this section, a model for the detrended and deseasonalized DATs will be developed. Next, the derived model will be estimated by a nonparametric nonlinear WN. The variable significance testing framework described in Appendix A will be applied in order to construct an appropriate training set for the WN. Then, the model selection algorithm will be applied in order to construct a WN with the best generalization ability. Finally, the WN will be initialized by applying

the BE method and will be trained using the BP method. The WNs framework is analytically presented in Appendix A and Appendix B.

5.5.1 The Linear Approach

Following Benth and Saltyte-Benth (2007) and Zapranis and Alexandridis (2008), a discrete approximation of (5.11) is obtained and is given by

$$\Delta T(t) = \Delta S(t) + \kappa(T(t-1) - S(t-1)) + \sigma(t)\Delta B(t). \tag{5.24}$$

Expanding (5.24) with $\Delta t = 1$, we have that

$$T(t) - T(t-1) = S(t) - S(t-1) + \kappa(T(t-1) - S(t-1)) + \sigma(t)\varepsilon(t), \tag{5.25}$$

and by rearranging, we have that

$$T(t) - S(t) = T(t-1) - S(t-1) + \kappa(T(t-1) - S(t-1)) + \sigma(t)\varepsilon(t) \tag{5.26}$$

since

$$\Delta B(t) = \varepsilon(t)\sqrt{\Delta t} = \varepsilon(t). \tag{5.27}$$

Next, by setting

$$\tilde{T}(t) = T(t) - S(t), \tag{5.28}$$

we have that

$$\tilde{T}(t) = \tilde{T}(t-1) + \kappa\tilde{T}(t-1) + \sigma(t)\varepsilon(t) \tag{5.29}$$

or equivalently

$$\tilde{T}(t) = (1+\kappa)\tilde{T}(t-1) + \sigma(t)\varepsilon(t). \tag{5.30}$$

Substituting with

$$a = 1 + \kappa, \tag{5.31}$$

our model is reduced to

$$\tilde{T}(t) = a\tilde{T}(t-1) + \sigma(t)\varepsilon(t). \tag{5.32}$$

Model (5.32) is simple AR(1) model of the detrended and deseasonalized temperatures with a zero constant. Hence, the parameter a can be easily estimated

using the least squares method. We continue our analysis from Sect. 5.4.1 where the trend and the seasonal mean were removed by a simple method previously proposed by Alaton et al. (2002). On the detrended and deseasonalized data, we estimate the speed of mean reversion parameter a. The results are presented in the first row of Table 5.20. Note that in Table 5.20, the parameter κ is presented. The relation between κ and a is given by (5.31). The speed of mean reversion ranges from 0.178 in Madrid to 0.251 in Oslo.

5.5.2 *A More Advanced Approach: The Nonlinear Nonparametric Approach*

Model (5.32) is a lineal AR(1) model with a zero constant. Since in our analysis the speed of mean reversion is not considered constant but a time-varying function, Equation (5.32) can be written as follows:

$$\tilde{T}(t) = a(t-1)\tilde{T}(t-1) + \sigma(t)\varepsilon(t), \tag{5.33}$$

where

$$a(t) = 1 + \kappa(t). \tag{5.34}$$

The detrended and deseasonalized temperature series, $\tilde{T}(t)$, can be modeled with an AR(1) process with a zero constant term, as shown in (5.32). In the context of such a model, the mean reversion parameter a is typically assumed to be constant over time. In Brody et al. (2002), it was mentioned that in general a should be a function of time, but no evidence was presented. On the other hand, Benth and Saltyte-Benth (2005), using a dataset comprising of 10 years of Norwegian temperature data, calculated mean annual values of a. They reported that the variation of the values of a from year to year was not significant. They also investigated the seasonal structures in monthly averages of a, and they reported that none was found. However, since to date, no one has computed daily values of the mean reversion parameter, since there is no obvious way to do this in the context of model (5.33). On the other hand, averaging techniques, in a yearly or monthly basis, run the danger of filtering out too much variation and consequently presenting a distorted picture regarding the true nature of a. The impact of a false specification of a on the accuracy of the pricing of temperature derivatives is significant (Alaton et al. 2002).

In this section, we address that issue, by using a WN to estimate nonparametrical relationship (5.33) and then estimate a as a function of time. By computing the derivative of the network output with respect to (w.r.t.) the network input, we obtain a series of daily values for a. This is done for the first time, and it gives us a much better insight in temperature dynamics and in temperature derivative pricing. As we will see, the daily variation of a is quite significant after all.

Moreover, previous studies (Alaton et al. 2002; Bellini 2005; Benth and Saltyte-Benth 2005, 2007; Zapranis and Alexandridis 2008, 2009b) show that an AR(1) model is not complex enough to completely remove the autocorrelation in the residuals. Alternatively, more complex models were suggested (Carmona 1999; Geman and Leonardi 2005).

Using WNs the generalized version of (5.33) is estimated nonlinearly and nonparametrically, that is,

$$\tilde{T}(t+1) = \phi\left(\tilde{T}(t), \tilde{T}(t-1), \ldots\right) + e(t). \tag{5.35}$$

Model (5.35) uses past temperatures (detrended and deseasonalized) over one period. Using more lags, we expect to overcome the strong correlation found in the residuals in models such as in Alaton et al. (2002), Benth and Saltyte-Benth (2007), and Zapranis and Alexandridis (2008). However, the length of the lag series must be selected. Since the WN is a nonparametric nonlinear estimator results from the ACF or the Partial ACF (PACF) cannot be used. Similarly, criteria used in linear models like the Schwarz criterion cannot be applied. Hence, the variable significance algorithm presented in the previous section is applied in order to determine the number of significant lags in each city.

5.5.2.1 Variable Selection: Selecting the Significant Lags

In this section, our proposed variable selection framework will be applied on the detrended and deseasonalized DATs of the seven European cities in order to select the length of the lag series.

The target values of the WN are the DATs. The explanatory variables are lagged versions of the target variable. Choosing the length of a lag distribution in linear models can be done by minimizing an information criterion like Akaike or Schwarz criteria. Alternatively the ACF and the PACF can be studied. Figures 5.11 and 5.12 present the ACF and PACF of the detrended and deseasonalized DATs in Berlin. The ACF suggests that the first 35 lags are significant. On the other hand, the PACF suggests that the 6 first lags as well as the 8^{th} and the 11^{th} lag must be included on the model. However, results from these methods are not necessarily true in nonlinear nonparametric models. The results of the remaining cities are also inconclusive.

Alternatively, in order to select only the significant lags, the variable selection algorithm presented in Sect. 4.4 will be applied. Initially, the training set contains the dependent variable and 7 lags. Hence, the training set consists of 7 inputs, 1 output, and 3,643 training pairs.

In this study, the relevance of a variable to the model is quantified by the sensitivity-based pruning (SBP) criterion which was introduced in Appendix A.4. Our analysis in Appendix A.4 indicates that the SBP fitness criterion was found to significantly outperform alternative criteria in the variable selection algorithm.

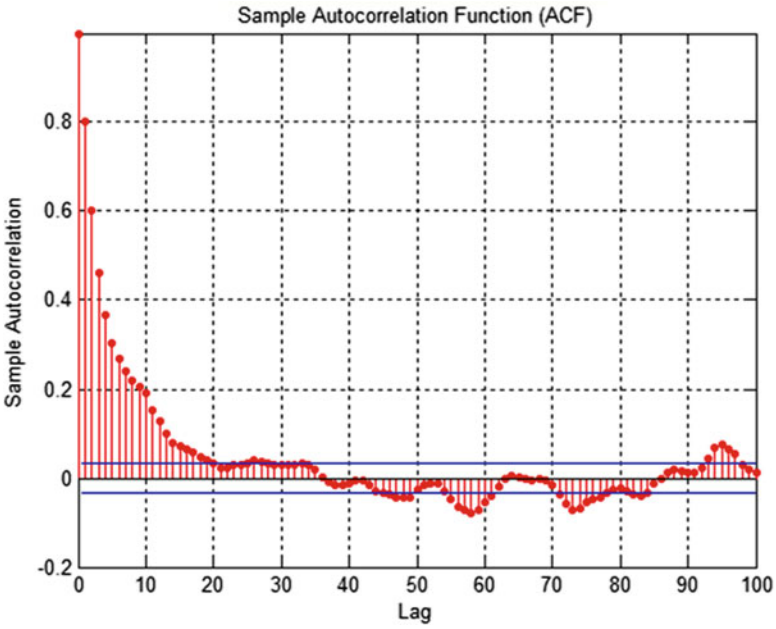

Fig. 5.11 Autocorrelation function of the detrended and deseasonalized daily average temperature in Berlin

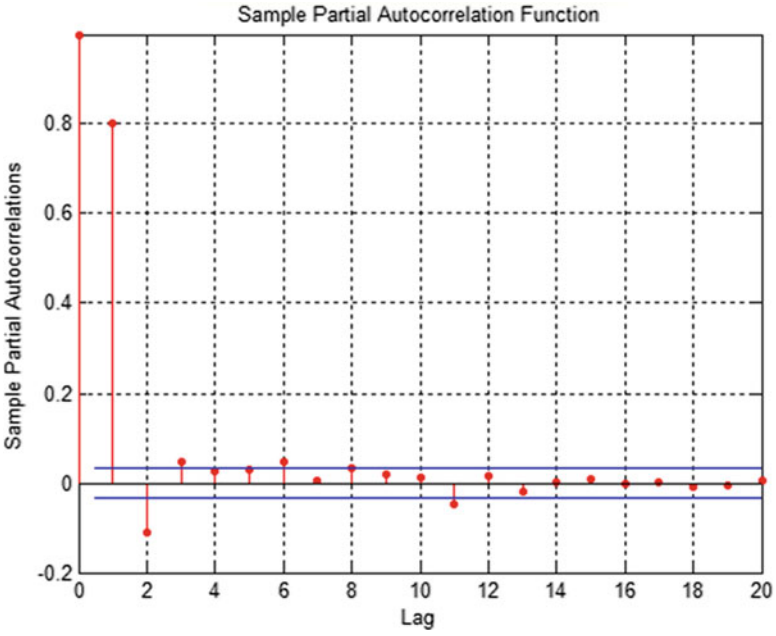

Fig. 5.12 Partial autocorrelation function of the detrended and deseasonalized daily average temperature in Berlin

The SBP quantifies the effect on the empirical loss of replacing a variable by its mean. The SBP is given by

$$SBP(x_j) = L_n(\mathbf{x}; \hat{\mathbf{w}}_n) - L_n(\bar{\mathbf{x}}^{(j)}; \hat{\mathbf{w}}_n), \tag{5.36}$$

where

$$\bar{\mathbf{x}}^{(j)} = \left(x_{1,t}, x_{2,t}, \ldots, \bar{x}_j, \ldots, x_{m,t}\right) \tag{5.37}$$

and

$$\bar{x}_j = \frac{1}{n} \sum_{t=1}^{n} x_{j,t}. \tag{5.38}$$

Here, bootstrap (BS) sampling is used in order to approximate the empirical distribution of the SBP criterion, since our previous results in Alexandridis (2010) indicate that BS performs better than the cross-validation (CV) in this part of the model identification algorithm. Using the empirical distribution of the SBP, hypothesis tests can be constructed. Hence, the removal of a variable is based on testing the following null hypothesis:

$$\begin{aligned} H_0 &: SBP = 0 \\ H_1 &: SBP \neq 0 \end{aligned}. \tag{5.36}$$

The p values of the hypothesis tests were used to identify insignificant variables. In statistical hypothesis testing, the p value is the probability of obtaining a value of the test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. Equivalently, p value is the smallest level of significance, a, at which a null hypothesis may be rejected using the observed value of the test statistic. The smaller the p values, the more convinced we are that that the null hypothesis should be rejected (Aczel 1993). The lower the p value, the more "significant" the result is in the sense of statistical significance. One often rejects a null hypothesis if the p value is less than 0.05 or 0.01, corresponding to a 5% or 1% significance level.

Following Aczel (1993), in this study, when the p value is less than 0.01, the variable is considered "very significant." When the p value is between 0.01 and 0.05, the variable is "significant." When the p value is between 0.05 and 0.1, the variable is "marginally significant." Finally, when the p value is larger than 0.1, the variable is "not significant." Hence, in each step of our algorithm, the variable with the larger p value greater than 0.1 will be removed from the training set of our model. After each variable removal, a new architecture of the WN will be selected, and a new WN will be trained. However, the correctness of the decision of removing a variable must be examined. This can be done either by examining the prediction risk or the $\bar{R}^2$.

Table 5.7 Variable selection with backward elimination in Berlin

Step	Variable to remove (lag)	Variable to enter (lag)	Variables in model	Hidden units (parameters)	n/p ratio	Empirical loss	Prediction risk
	–	–	7	5 (83)	43.9	1.5928	3.2004
1	X_6	–	6	2 (33)	110.4	1.5922	3.1812
2	X_7	–	5	1 (17)	214.3	1.5927	3.1902
3	X_5	–	4	1 (14)	260.2	1.6004	3.2056
4	X_4	–	3	1 (11)	331.2	1.5969	3.1914

The algorithm concluded in four steps. In each step, the following are presented: which variable is removed, the number of hidden units for the particular set of input variables, and the parameters used in the wavelet network, the empirical loss, and the prediction risk

If the new prediction risk is smaller than the new prediction risk multiplied by a threshold, then the decision of removing the variable was correct. If the prediction risk increased more than the allowed threshold, then the variable was reintroduced back to the model. We set this threshold at 5%.

Table 5.7 summarizes the results of the model identification algorithm for Berlin. Both the model selection and variable selection algorithms are included in Table 5.7. The algorithm concluded in four steps, and the final model contains only three variables. The prediction risk for the reduced model is 3.1914, while for the original model was 3.2004. On the other hand, the empirical loss slightly increased from 1.5928 for the initial model to 1.5969 for the reduced model indicating that the explained variability (unadjusted) slightly decreased. However, the explained variability (adjusted for degrees of freedom) was increased for the reduced model to 64.61%, while it was 63.98% initially. Finally, the number of parameters was significantly reduced in the final model. The initial model needed five hidden units (HUs) and seven inputs. Hence, 83 parameters were adjusted during the training phase. Hence, the ratio of the number of training pairs n to the number of parameters p was 43.9. In the final model, only 1 HU and 3 inputs were used. Hence, only 11 parameters were adjusted during the training phase, and the ratio of the number of training pairs n to the number of parameters p was 331.2.

In Table 5.8, the statistics for the full WN model can be found. More precisely, the first part of Table 5.8 reports the value of the SBP, its standard deviation, and its p value. It is clear that the value of the SBP for the last three variables is very small in contrast to the first two variables. Observing the p values, we conclude that the last four variables have p value greater than 0.1, while the 6^{th} lag has a p value of 0.8826 strongly indicating a "not significant" variable. In the second part of Table 5.8, various fitting criteria are reported, more precisely, the mean absolute error, the maximum absolute error (Max AE), the normalized mean-square error (NMSE), the mean absolute percentage error (MAPE), the $\bar{R}^2$, the empirical loss, and the prediction risk.

The Max AE, MAPE, MAPE, and NMSE are given by

$$MaxAE = \max(|y_i - \hat{y}_i|) \tag{5.37}$$

Table 5.8 Statistics for the full wavelet neural network model for Berlin (seven inputs, five hidden units)

Variable	SBP	St. dev	*P* value
X_7	0.0026	0.0129	0.7796
X_6	0.0032	0.0129	0.8826
X_5	0.0053	0.0163	0.6757
X_4	0.0161	0.0241	0.3500
X_3	0.2094	0.0944	0.0000
X_2	1.1123	0.1935	0.0000
X_1	9.8862	0.5112	0.0000
MAE	1.8080		
Max AE	11.1823		
NMSE	0.3521		
MAPE	3.7336		
$\bar{R}^2$	63.98%		
Empirical loss	1.5928		
Prediction risk	3.2004		
Iterations	43		

The average SBP for each variable of 50 bootstrapped samples, the standard deviation and the *p* value

SBP sensitivity-based pruning, *MAE* mean absolute error, *Max AE* maximum absolute error, *NMSE* normalized mean square error, *MSE* mean square error, *MAPE* mean absolute percentage error

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{5.38}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{5.39}$$

$$NMSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{5.40}$$

where y_i are the target values, $\hat{y}_i$ is the network output, and $\bar{y}$ is the average of the target values.

The WN was converged after 43 iterations. In general, a very good fit was obtained. The empirical loss is 1.5928, and the prediction risk is 3.2004. The Max AE is 11.1823, while the Mean Absolute Error (MAE) is 1.8080, and the NMSE is 0.3521. The MAPE is 3.7336. Finally, the $\bar{R}^2 = 63.98\%$.

Table 5.9 shows the statistics for the WN at step 1. The network had 6 inputs, 2 wavelets were used to construct the WN, and 33 weights adjusted during the training phase. The WN converged after 17 iterations. By removing X_6 from the model, we observe from Table 5.9 that the p value of X_5 became 0, while for

Table 5.9 Statistics for the wavelet neural network model at step 1 for Berlin (six inputs, two hidden units)

Variable	SBP	St. dev	*P* value
X_7	0.0031	0.0061	0.5700
X_5	0.0131	0.0156	0.0000
X_4	0.0149	0.0208	0.1403
X_3	0.2368	0.0789	0.0000
X_2	1.0318	0.1747	0.0000
X_1	10.0160	0.4584	0.0000
MAE	1.8085		
Max AE	11.1446		
NMSE	0.3529		
MAPE	3.7127		
$\bar{R}^2$	64.40%		
Empirical loss	1.5922		
Prediction risk	3.1812		
Iterations	17		

The average SBP for each variable of 50 bootstrapped samples, the standard deviation and the *p* value

SBP sensitivity-based pruning, *MAE* mean absolute error, *Max AE* maximum absolute error, *NMSE* normalized mean square error, *MSE* mean-square error, *MAPE* mean absolute percentage error

X_7 and X_4, the p values became 0.5700 and 0.1403, respectively. The empirical loss was slightly decreased to 1.5922. However, the MAE and NMSE were slightly increased to 1.8085 and 0.3529, respectively. On the other hand, the Max AE and the MAPE were decreased to 11.1446 and 3.7127, respectively. Next, the decision of removing X_6 is tested. The new prediction risk was reduced to 3.1812, while the explained variability adjusted for degrees of freedom increased to 64.40%. Hence, the removal of X_6 reduced the complexity of the model, while its predictive power was increased.

At step 2, X_7, which had the largest p value $= 0.5700$ at the previous step, was removed from the model. Table 5.10 shows the statistics for the WN at step 2. The new WN had 5 inputs, 1 HU was used, and 17 weights adjusted during the training phase. The WN converged after 19 iterations. A closer inspection of Table 5.10 reveals that the removal of X_7 resulted to an increase in the error measures, and a worse fit were obtained. The new $\bar{R}^2$ is 64.59%. The new prediction risk increased to 3.1902 which is smaller than the threshold. In other words, by removing X_7 the total predictive power of our model was slightly decreased; however, by adding the variable X_7 on the model, only 0.28% additional variability of our model was explained, while the computational burden was significantly increased.

Examining the values of the SBP on Table 5.10, it is observed that the first two variables still have significantly larger values than the remaining variables. The p values reveal that at in the third step the X_5 must be removed from the model since its p value is 0.1907.

Table 5.10 Statistics for the wavelet neural network model at step 2 for Berlin (five inputs, one hidden unit)

Variable	SBP	St. dev	*P* value
X_5	0.0206	0.0174	0.1907
X_4	0.0216	0.0250	0.1493
X_3	0.2285	0.0822	0.0000
X_2	1.0619	0.1568	0.0000
X_1	9.9858	0.4462	0.0000
MAE	1.8083		
Max AE	11.1949		
NMSE	0.3525		
MAPE	3.7755		
$\bar{R}^2$	64.59%		
Empirical loss	1.5927		
Prediction risk	3.1902		
Iterations	19		

The average SBP for each variable of 50 bootstrapped samples, the standard deviation and the *p* value
SBP sensitivity-based pruning, *MAE* mean absolute error, *Max AE* maximum absolute error, *NMSE* normalized mean square error, *MSE* mean-square error, *MAPE* mean absolute percentage error

Table 5.11 Statistics for the wavelet neural network model at step 3 for Berlin (four inputs, one hidden unit)

Variable	SBP	St. dev	*P* value
X_4	−0.0052	0.0064	0.4701
X_3	0.1991	0.0628	0.0000
X_2	0.9961	0.1502	0.0000
X_1	10.0537	0.4011	0.0000
MAE	1.8093		
Max AE	11.0800		
NMSE	0.3526		
MAPE	3.7348		
$\bar{R}^2$	64.61%		
Empirical loss	1.6004		
Prediction risk	3.2056		
Iterations	4		

The average SBP for each variable of 50 bootstrapped samples, the standard deviation and the *p* value
SBP sensitivity-based pruning, *MAE* mean absolute error, *Max AE* maximum absolute error, *NMSE* normalized mean square error, *MSE* mean-square error, *MAPE* mean absolute percentage error

Table 5.11 shows the statistics for the WN at step 3. The network had 4 inputs, 1 HU was used, and 14 weights were adjusted during the training phase. The WN converged after 4 iterations. When removing X_5 from the model, we observe from Table 5.11 that only X_4 has a p value greater than 0.1. Again, the empirical loss and the prediction risk were increased. More precisely, the empirical loss is 1.6004,

Table 5.12 Statistics for the wavelet neural network model at step 4 for Berlin (three inputs, one hidden unit)

Variable	SBP	St. dev	*P* value
X_3	0.2244	0.0573	0.0000
X_2	0.9363	0.1581	0.0000
X_1	10.1933	0.4442	0.0000
MAE	1.8095		
Max AE	11.0925		
NMSE	0.3530		
MAPE	3.7171		
$\bar{R}^2$	64.61%		
Empirical loss	1.5969		
Prediction risk	3.1914		
Iterations	19		

The average SBP for each variable of 50 bootstrapped samples, the standard deviation and the *p* value

SBP sensitivity-based pruning, *MAE* mean absolute error, *Max AE* maximum absolute error, *NMSE* normalized mean square error, *MSE* mean-square error, *MAPE* mean absolute percentage error

and the prediction risk increased 0.48% to 3.2056. The new prediction risk is greater than the estimated prediction risk of the initial model about 0.16%. Again, the increase in the prediction risk was significantly smaller than the threshold. On \the other hand, the $\bar{R}^2$ was increased to 64.61% indicating an improved fit. Hence, the decision of removing X_5 was accepted.

In the final step, the variable X_4 had p value $= 0.4701$, and it was removed from the model. Table 5.12 shows the statistics for the WN at step 4. The network had 3 inputs, 1 wavelet was used for the construction of the WN, and only 11 weights were adjusted during the training phase. The WN converged after 19 iterations. After the removal of X_4, a new WN was trained with only one wavelet. The new empirical loss was decreased to 1.5969. The MAE and NMSE are 1.8095 and 0.3530, respectively, while the Max AE and the MAPE are 11.0925 and 3.7171, respectively. Next, the decision of removing X_4 was tested. The new prediction risk was reduced to 3.1914, while the explained variability adjusted for degrees of freedom was 64.61%. Hence, the removal of X_4 reduced the complexity of the model, while its performance was increased. The p values of the remaining variables are zero indicating that the remaining variables are characterized as very significant variables. Hence, the algorithm stops. Our proposed algorithm indicates that only the 3 most recent lags should be used, while PACF suggested the first 6 lags as well as the 8^{th} and the 11^{th} lag.

Concluding, in the final model, only three of the seven variables were used. The complexity of the model was significantly reduced since from 83 parameters in the initial model only 11 parameters have to be trained in the final model. In addition in the reduced model, the prediction risk minimized when only one HU was used while five HUs were needed initially. Our results indicate that the in-sample fit was slightly decreased in the reduced model. However, when an adjustment for the

Table 5.13 Prediction risk at each step of the variable selection algorithm for the five first hidden units for Berlin

Step\HU	1	2	3	4	5
0	3.2009	3.2026	3.2023	3.2019	**3.2004**
1	3.1817	**3.1812**	3.1828	3.1861	3.1860
2	**3.1902**	3.1915	3.1927	3.1972	3.1974
3	**3.2056**	3.2077	3.2082	3.2168	3.2190
4	**3.1914**	3.2020	3.2182	3.2158	3.2169

degrees of freedom is made, we observe that the $\bar{R}^2$ was increased to 64.61% from 63.98% in the initial model. Finally, the prediction power of the final and less complex proposed model was improved since the prediction risk was reduced to 3.1914 from 3.2004.

On the first row of Table 5.14, the statistical significant lags for the seven cities are presented. The number of significant lags for each city is as follows: Oslo 2 lags; Berlin, Paris, and Stockholm 3 lags; Amsterdam 4 lags; Madrid 6 lags; and Rome 7 lags.

5.5.2.2 Model Selection: Selecting the Architecture of the Wavelet Network

In each step, the appropriate number of HUs is determined by applying the model selection algorithm. The model selection algorithm is presented in Appendix A.3. For simplicity, we refer only to results from Berlin. The results of the remaining cities are similar. Table 5.13 shows the prediction risk for the first 5 HUs at each step of the variable selection algorithm for Berlin. Ideally, the prediction risk will decrease (almost) monotonically until a minimum is reached, and then it will start to increase (almost) monotonically. The number of HUs that minimize the prediction risk is selected for the construction of the model.

In the initial model, where all seven inputs were used, the prediction risk with one HU is only 3.2009. When one additional HU is added to the model, the prediction risk increases. Then, as more HUs are added to the model, the prediction risk monotonically decreases. The minimum is reached when five HUs are used and is 3.2004. When additional HUs are added in the topology of the model, the prediction risk increases. Hence, the architecture of the WN contains five HUs. In other words, the five higher ranking wavelets should be selected form the wavelet basis in order to construct the WN. Observing Table 5.13, it is clear that the prediction risk at the initial model with only one HU is almost the same as in the model with five HUs. This due to the small number of parameters that were adjusted during the training phase when only 1 HU is used and not due to a better fit.

At the second step, when variable X_6 was removed, the prediction risk is minimized when two HUs are used. Similarly, at steps two, three, and four, the prediction risk was minimized when only one HU is used. Additional HUs does not improve the fitting or the predictive power of the model. Figure 5.13 presents the prediction risk for the first 5 HUs of the final model. It is clear that the prediction

Table 5.14 Model selection and fitness criteria of the wavelet network for the seven cities

	Amsterdam	Berlin	Madrid	Oslo	Paris	Rome	Stockholm
Lags k	4	3	6	2	3	7	3
HU	1	1	1	1	1	1	1
n/p ratio	260	332	182	456	332	158	332
MAE	1.3797	1.8090	1.3947	1.6717	1.5868	1.1709	1.5705
Max AE	8.3484	11.0931	8.3846	11.3632	8.2646	7.1735	9.1467
NMSE	0.3193	0.3523	0.2883	0.4202	0.3601	0.3702	0.3787
MSE	3.1829	5.4196	3.1842	4.7831	3.9800	2.4210	4.1678
MAPE	3.0692	3.7154	3.3918	2.9820	2.1585	2.0666	5.9942
POCID	61.62%	60.15%	60.86%	60.84%	59.90%	60.24%	60.12%
IPOCID	56.05%	52.30%	54.54%	47.87%	52.85%	51.13%	51.89%
POS	81.73%	81.49%	82.38%	78.18%	80.39%	80.87%	80.15%
$\overline{R}^2$	67.95%	64.61%	71.02%	57.9%	63.88%	62.75%	61.94%

The number of hidden units and lags used in each city to model the daily average temperature is presented. The fitting criteria using the wavelet network in each city are also presented
HU hidden units, *MAE* mean absolute error, *Max AE* maximum absolute error, *NMSE* normalized mean square error, *MSE* mean-square error, *MAPE* mean absolute percentage error, *POCID* position of change in direction, *IPOCID* independent position of change in direction, *POS* position of sign

risk is minimized when one HU is used, and then it increases almost monotonically. Table 5.14 presents the appropriate HUs for the construction of the final WN for each city. Our results indicate that a very simple model with only one HU is adequate to fit the DATs in the seven cities of our analysis.

5.5.2.3 Initializing and Training the Wavelet Network

After the training set and the correct topology of the WN are selected, the WN can be constructed and trained. The backward elimination (BE) method is used to initialize the WN. A wavelet basis is constructed by scanning the four first levels of the wavelet decomposition of the DAT of each city.

Focusing on Berlin again, the wavelet basis consists of 168 wavelets. However, not all wavelets in the wavelet basis contribute to the approximation of the original time series. Following Zhang (1997), the wavelets that contain less than five sample points of the training data in their support are removed. Seventy six wavelets that do not significantly contributed to the approximation of the original time series were indentified. The truncated basis contains 92 wavelet candidates. Applying the BE method, the wavelet are ranked in order of significance. The wavelets in the wavelet library are ranked as follows: the BE starts the regression by selecting all the available wavelets from the wavelet library. Then the wavelet that contributes the least in the fitting of the training data is repeatedly eliminated. Since only one HU is used on the architecture of the model, only the wavelet with the highest ranking is used to initialize the WN. Figure 5.14 presents the initialization of the final model using only one HU. The initialization is very good, and the WN

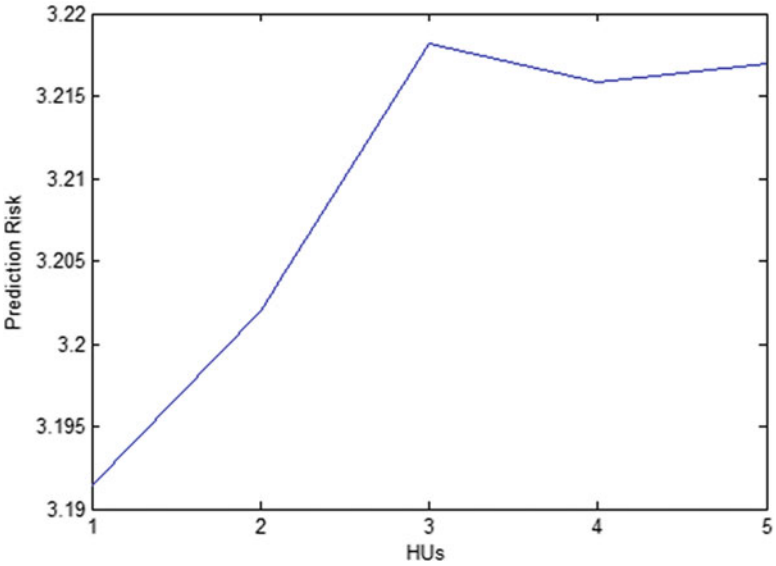

Fig. 5.13 The prediction risk for the first five hidden units of the final wavelet network model in Berlin

converged after only 19 iterations. The training stopped when the minimum velocity, 10^{-5}, of the training algorithm was reached. The final approximation of the WN to the daily detrended and deseasonalized temperature can be found in Figure 5.15. The results in the remaining cities are similar.

In Table 5.14, various fitness criteria of the seven WNs corresponding to the seven cities are presented. A closer inspection of Table 5.14 reveals that the WNs fit the DATs reasonably well. The overall fit for Oslo is $\bar{R}^2 = 57.9\%$, while for Madrid is $\bar{R}^2 = 71.02\%$. The smallest MSE is observed in Rome and is only 2.4210, while the largest one is observed in Berlin, and it is 5.4196. The MAE is only 1.1709 in Rome and 1.8090 in Berlin.

In Table 5.14, the prediction of sign (POS), the prediction of change in direction (POCID), and the independent prediction of change in direction (IPOCID) are also reported. These three criteria examine the ability of the network to predict changes, independently of the size of the change, and they are referred as percentages. The POS measures the ability of the network to predict the sign of the target values, positive or negative, and it is given by

$$POS = \frac{100}{n} \sum_{i-1}^{n} d_i, \tag{5.41}$$

where

$$d_i = \begin{cases} 1 & y_i \cdot \hat{y}_i > 0 \\ 0 & y_i \cdot \hat{y}_i \leq 0 \end{cases} \tag{5.42}$$

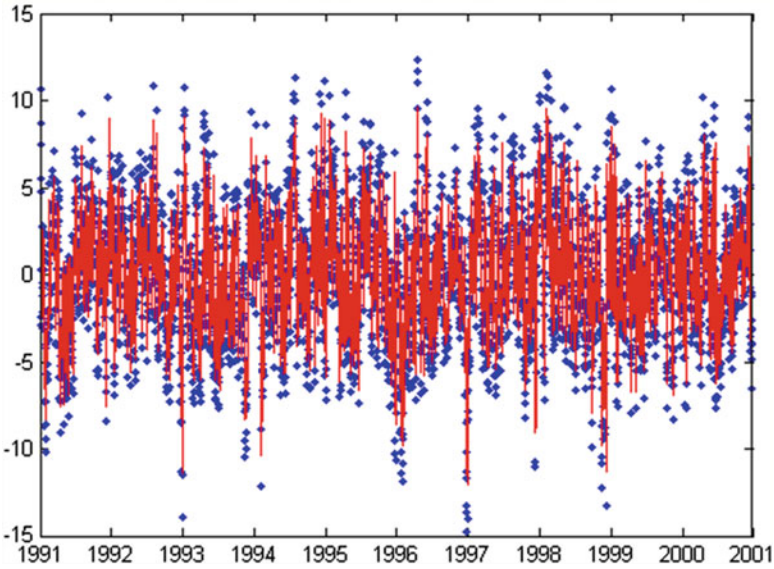

Fig. 5.14 Initialization of the final wavelet network model using the backward elimination method in Berlin. The daily average temperature (*dots*) and the wavelet network approximation (*line*) are presented

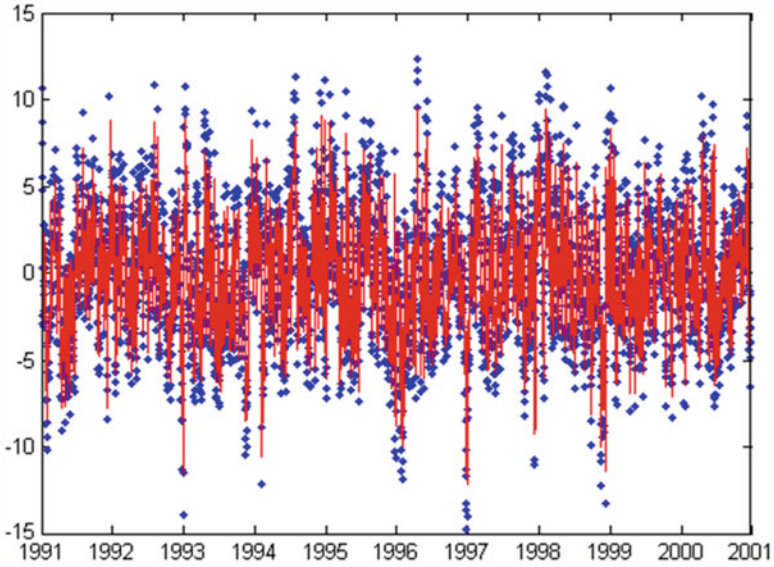

Fig. 5.15 Training the final wavelet network model with one hidden unit. The wavelet network converged after 19 iterations. The daily average temperature (*dots*) and the WN approximation (*line*) are presented

and n is the length of the forecasted patterns, y_i is the target value, and $\hat{y}_i$ is the WN approximation. The POCID is given by

$$POCID = \frac{100}{n} \sum_{i-1}^{n} d_i, \tag{5.43}$$

where

$$d_i = \begin{cases} 1 & (y_i - y_{i-1})(\hat{y}_i - y_{i-1}) > 0 \\ 0 & (y_i - y_{i-1})(\hat{y}_i - y_{i-1}) \leq 0 \end{cases}, \tag{5.44}$$

and finally, the IPOCID is given by

$$IPOCID = \frac{100}{n} \sum_{i-1}^{n} d_i, \tag{5.45}$$

where

$$d_i = \begin{cases} 1 & (y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) > 0 \\ 0 & (y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) \leq 0 \end{cases}. \tag{5.46}$$

The POS for the detrended and deseasonalized DATs is very high for all cities, and it ranges from 78.18% in Oslo to 81.73% in Amsterdam. The POCID ranges from 59.9% in Paris to 61.62% in Amsterdam. Similarly, the minimum IPOCID is 47.87%, and it is observed in Oslo, while the maximum is 56.05%, and it is observed in Amsterdam.

5.5.2.4 The Wavelet Neural Networks Approach: Time-Dependent Mean Reversion Variable

In this section, we focus on analyzing the speed of mean reversion, $\kappa(t)$. The DATs are modeled by a nonlinear AR model. By fitting the AR model nonlinearly and nonparametrically with a WN allows us to examine the time structure of the speed of the mean reversion of the temperature process. By computing the derivative of the WN output with respect to the network input, a series of the daily values for the mean reversion function are estimated. Since the relation between the "coefficient" of the nonlinear model and the speed of mean reversion function is linear, the "coefficient" of the nonlinear AR model is examined instead. The relation between the "coefficient" of the nonlinear AR model and the speed of mean reversion is given by (5.34).

Using WNs, the generalized version of (5.33) is estimated nonparametrically by (5.35)

$$\tilde{T}(t+1) = \phi\left(\tilde{T}(t), \tilde{T}(t-1), \ldots\right) + e(t).$$

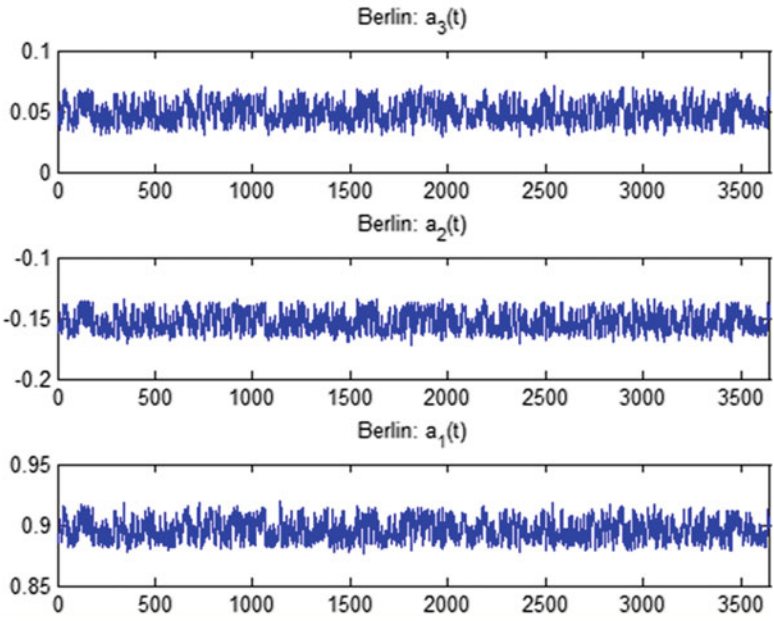

Fig. 5.16 Daily variation of the speed of mean reversion functions a_i in Berlin

Once we have the estimator of the underlying function ϕ, then the daily values of a can be computed as follows:

$$\alpha_1(t) = d\tilde{T}(t+1)/d\tilde{T}(t) = d\phi/d\tilde{T}. \tag{5.47}$$

The analytic expression for derivative of the WN w.r.t. the input variable $d\phi/d\tilde{T}$ can be found in appendix.

We estimate $\phi(\bullet)$ nonparametrically with a WN, $g(\bullet)$. Given an input vector $\mathbf{x}$ (the harmonics) and a set of weights $\mathbf{w}$ (parameters), the network response (output) $g_\lambda(\mathbf{x};\mathbf{w})$ is

$$g_\lambda(\mathbf{x};\mathbf{w}) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \Psi_j(\mathbf{x}) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i.$$

In that expression, $\Psi_j(\mathbf{x})$ is a multidimensional wavelet which is constructed by the product of m scalar wavelets, $\mathbf{x}$ is the input vector, m is the number of network inputs, λ is the number of HUs, and w stands for a network weight. The multidimensional wavelets are computed by (A.2). The mother wavelet is given by the Mexican hat function.

For Berlin, the daily values of $a(t)$ (3,647 values) are depicted in Fig. 5.16. Because in Berlin there are three significant lags, and there are three mean-reverting functions, $a_i(t)$. The corresponding frequency histograms are given in Fig. 5.17. The graphs for all cities are very similar. The relevant statistics of $a_i(t)$ for all cities

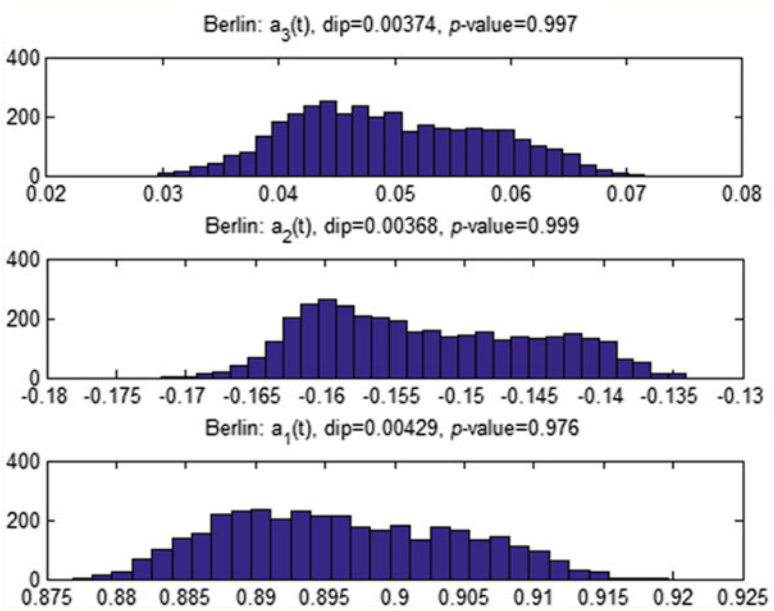

Fig. 5.17 Frequency distribution of the speed of mean reversion function a_i in Berlin

are presented in Table 5.15. Our results indicate that the mean reversion parameter is not constant. On the contrary, its daily variation is quite significant; this fact naturally has an impact on the accuracy of the pricing equations, and it has to be taken into account (Alaton et al. 2002). Intuitively, it was expected $a_i(t)$ not to be constant. If the temperature today is away from the seasonal average (a cold day in summer), then it is expected that the speed of mean reversion to be high, that is, the difference of today and tomorrows temperature, it is expected to be high. In contrast, if the temperature today is close to the seasonal average, then is expected the temperature to revert to its seasonal average slowly.

Referring now to Figs. 5.16 and 5.17, we observe that the spread between the maximum and minimum value is similar for the three mean-reverting parameters, 0.04. The standard deviation is 0.01, and the mean is 0.90, -0.15, and 0.05 for $a_1(t)$, $a_2(t)$ and $a_3(t)$, respectively. We also observe that there is an upper threshold in the values of $a_i(t)$ (0.915, -0.137, and 0.068) which is rarely exceeded. This can also be seen in the frequency distribution of $a_i(t)$ in Fig. 5.17. A closer inspection of Table 5.15 reveals that in every city, $a_1(t)$ has the largest value (over 0.79), and $a_2(t)$ is always negative. A closer inspection of Table 5.15 reveals that the absolute average value of $a_i(t)$ of higher order lags decreases when the lag order increases which was expected. The value of $a_1(t)$ ranges from 0.79 in Oslo to 0.99 in Amsterdam and Madrid. Finally, strong autocorrelation is present in the values of $a_i(t)$ in every city.

Next, the structure of $a_i(t)$ is examined. More precisely, it is examined if $a_i(t)$ are stochastic processes by themselves. Both an ADF and a KPSS tests are used. The ADF test statistic is -21.12, -25.65, and -21.38 for $a_1(t)$, $a_2(t)$ and $a_3(t)$,

Table 5.15 Descriptive statistics of the mean-reverting functions

Amsterdam	Mean	St. dev	Max	Median	Min	Skewness	Kurtosis	K–S	P value	LBQ	P value
$a_4(t)$	0.00	0.005	0.01	0.00	−0.01	0.08	2.31	29.90	0.0000	4150.98	0.0000
$a_3(t)$	0.14	0.004	0.15	0.14	0.13	0.13	2.07	33.26	0.0000	4264.02	0.0000
$a_2(t)$	−0.31	0.004	−0.30	−0.31	−0.32	0.14	2.06	37.35	0.0000	4327.80	0.0000
$a_1(t)$	0.99	0.005	1.00	0.99	0.98	0.08	2.32	50.46	0.0000	4289.41	0.0000
Berlin											
$a_3(t)$	0.05	0.010	0.07	0.05	0.03	0.19	2.21	30.91	0.0000	3979.24	0.0000
$a_2(t)$	−0.15	0.010	−0.13	−0.15	−0.17	0.27	2.00	33.42	0.0000	4180.54	0.0000
$a_1(t)$	0.90	0.010	0.92	0.90	0.88	0.22	2.160	48.90	0.0000	4099.78	0.0000
Madrid											
$a_6(t)$	0.05	0.004	0.08	0.05	0.02	0.07	10.80	31.09	0.0000	1041.98	0.0000
$a_5(t)$	−0.01	0.003	0.01	−0.01	−0.03	−0.69	12.30	30.07	0.0000	346.56	0.0000
$a_4(t)$	0.01	0.003	0.02	0.01	−0.01	−0.35	10.39	30.03	0.0000	392.80	0.0000
$a_3(t)$	0.05	0.003	0.07	0.05	0.02	−0.07	15.99	31.03	0.0000	479.11	0.0000
$a_2(t)$	−0.25	0.003	−0.23	−0.25	−0.27	0.73	13.19	35.60	0.0000	362.57	0.0000
$a_1(t)$	0.99	0.003	1.01	0.99	0.96	−0.82	11.95	50.26	0.0000	524.43	0.0000
Oslo											
$a_2(t)$	−0.04	0.010	−0.02	−0.05	−0.08	0.53	2.95	30.66	0.0000	1068.88	0.0000
$a_1(t)$	0.79	0.010	0.81	0.79	0.76	0.52	2.87	46.87	0.0000	1031.85	0.0000
Paris											
$a_3(t)$	0.07	0.020	0.12	0.07	0.03	0.45	2.72	30.91	0.0000	2966.06	0.0000
$a_2(t)$	−0.19	0.020	−0.14	−0.20	−0.23	0.67	2.66	33.65	0.0000	3279.65	0.0000
$a_1(t)$	0.91	0.020	0.97	0.91	0.88	0.48	2.64	48.91	0.0000	3074.21	0.0000
Rome											
$a_7(t)$	0.04	0.002	0.09	0.04	0.00	0.56	139.43	30.59	0.0000	188.80	0.0000
$a_6(t)$	−0.02	0.003	0.02	−0.02	−0.10	−6.35	335.22	30.08	0.0000	33.65	0.0286
$a_5(t)$	0.03	0.002	0.06	0.03	−0.01	−1.60	79.05	30.43	0.0000	29.45	0.0793
$a_4(t)$	−0.04	0.002	−0.01	−0.04	−0.09	0.33	91.04	30.76	0.0000	13.82	0.8393

(continued)

Table 5.15 (continued)

Amsterdam	Mean	St. dev	Max	Median	Min	Skewness	Kurtosis	K–S	P value	LBQ	P value
$a_3(t)$	0.05	0.003	0.09	0.05	−0.03	−7.00	256.78	30.80	0.0000	24.21	0.2333
$a_2(t)$	−0.14	0.003	−0.09	−0.14	−0.19	2.62	140.34	33.04	0.0000	20.23	0.4439
$a_1(t)$	0.88	0.002	0.91	0.88	0.85	−1.99	70.74	48.49	0.0000	21.10	0.3915
Stockholm											
$a_3(t)$	0.06	0.003	0.07	0.06	0.05	−0.38	2.43	31.52	0.0000	3696.99	0.0000
$a_2(t)$	−0.17	0.003	−0.16	−0.17	−0.18	−0.47	2.26	34.13	0.0000	3785.76	0.0000
$a_1(t)$	0.88	0.003	0.89	0.88	0.87	−0.39	2.39	48.83	0.0000	3752.39	0.0000

St. dev standard deviation, *K–S* Kolmogorov–Smirnov goodness-of-fit, *LBQ* Ljung–Box Q-statistic lack-of-fit

Table 5.16 Descriptive statistics of the residuals in each city using a wavelet network

City	Mean	St. dev	Max	Median	Min	Skewness	Kurtosis	K–S	*P* value	LBQ	*P* value
Amsterdam	0.00	1.78	6.97	−0.07	−8.35	0.14	3.74	7.57	0.0000	20.848	0.4062
Berlin	0.01	2.33	11.09	0.01	−9.78	−0.03	3.73	11.05	0.0000	31.469	0.0493
Madrid	0.01	1.78	7.26	0.14	−8.38	−0.40	3.73	8.54	0.0000	28.808	0.0916
Oslo	0.00	2.19	11.36	0.04	−9.64	0.06	4.19	10.08	0.0000	25.362	0.1879
Paris	0.00	1.99	5.50	0.04	−8.26	−0.19	3.03	10.20	0.0000	21.157	0.3880
Rome	0.01	1.55	7.17	0.04	−7.09	0.00	4.35	4.80	0.0000	26.177	0.1601
Stockholm	−0.01	2.04	8.29	0.05	−9.15	−0.12	3.87	8.90	0.0000	27.317	0.1266

St. dev standard deviation, *K–S* Kolmogorov–Smirnov goodness-of-fit, *LBQ* Ljung–Box Q-statistic lack-of-fit

respectively, for Berlin. The $p\ value = 0$ for the three mean reversion functions that leads to the rejection of the null hypothesis that $a_i(t)$ has a unit root. In order to have a more powerful test, the KPSS test is also applied. The KPSS test statistic is 0.043, 0.045, and 0.044 for $a_1(t)$, $a_2(t)$, and $a_3(t)$, respectively, and less than the critical values in 1%, 5%, and 10% confidence level. The previous results suggest the acceptance of the null hypothesis that $a_i(t)$ is stationary. The results of the remaining cities are similar. The null hypothesis of the ADF that $a_i(t)$ have a unit root is rejected for all cities. Similarly, the null hypothesis of the KPSS that $a_i(t)$ are stationary cannot be rejected for all cities.

The histogram in Table 5.16 may suggest that the distributions of $a_i(t)$ are bimodal. In order to test the hypothesis of bimodality, the Hartigan's DIP statistic is estimated. Hartigan's DIP statistic is a measure of departure from unimodality. If a distribution is unimodal, then the DIP converges to zero otherwise converges to a positive constant (Hartigan and Hartigan 1985). The null hypothesis test is that $a_i(t)$ follows a unimodal distribution versus the alternative that $a_i(t)$ follows a bimodal distribution:

$$\begin{array}{l} H_0 : a_i(t) \text{ follows a unimodal distribution, } DIP = 0 \\ H_1 : a_i(t) \text{ follows a bimodal distribution, } DIP \neq 0 \end{array}. \tag{5.48}$$

The estimated DIP statistics for Berlin are 0.0043, 0.0039, and 0.0037 for $a_1(t)$, $a_2(t)$ and $a_3(t)$, respectively, with $p\ values$ over 0.97. Hence, the null hypothesis that $a_i(t)$ follows a unimodal distribution cannot be rejected in Berlin. The results of the remaining cities are similar.

The results from Zapranis and Alexandridis (2008) indicate that $a_1(t)$ follows a bimodal distribution in Paris. However, in Zapranis and Alexandridis (2008), only one lag is used in order to estimate model (5.33) which may have a strong impact on the structure and values of $a_1(t)$.

Moreover, Fig. 5.16 may suggest seasonalities in the structure of $a_i(t)$. The ACF of $a_i(t)$ is shown in Fig. 5.18. A seasonality of a half year can be shown in the ACF. Also, the first 35 lags are statistically important and positively correlated, while the next 20 are negatively correlated.

In this section, the daily values of $a_i(t)$ were successfully estimated. Hence, the residuals $e(t)$ of model (5.35) can be obtained. In the next section, the residuals $e(t)$ will be examined.

5.6 The Seasonal Variance

In this section, the residuals of the WN will be examined. The initial hypothesis for the residuals $e(t)$ of model (5.35) was that they follow the normal distribution. However, a closer inspection of the noise part of the wavelet decomposition of Berlin's DAT (Fig. 5.10) reveals seasonalities.

The distributional statistics of the residuals of the WN for all cities can be found on Table 5.16. The mean value of the residuals is very close to zero for all cities;

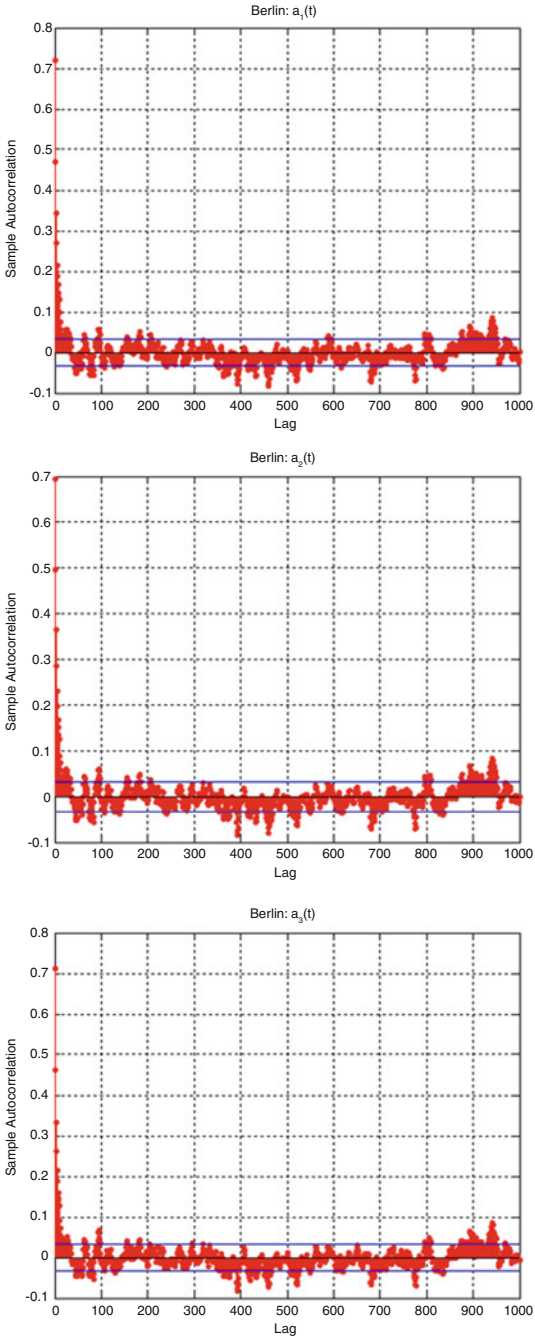

Fig. 5.18 Autocorrelation function of the speed of mean reversion functions of the nonlinear AR model, $a_i(t)$, in Berlin

however, the standard deviation is around 2. More precisely, the minimum standard deviation is observed in Rome and is 1.55, while the maximum is observed in Berlin, and it is 2.33. With an exception of Paris, all cities exhibit large positive kurtosis. On the other hand, the skewness is -0.40 for Rome, while it is 0.14 for Amsterdam.

Next, a normality test will be performed on the estimated residuals of the WN. More precisely, the distance of the empirical distribution of the residuals and the standard normal distribution will be estimated. The distance between the empirical distribution of the data and a benchmark distribution is estimated by the Kolmogorov–Smirnov test or the Kolmogorov distance. The Kolmogorov distance is defined as follows:

$$KS = \sqrt{n} \sup_x |F(x) - F_n(x)|, \tag{5.49}$$

where n is the sample size and $F_n(x)$ is the empirical cumulative density function (CDF), while $F(x)$ is the estimated CDF. The hypothesis test of the Kolmogorov–Smirnov is defined as

$$\begin{aligned} H_0 &= \text{The data have the hypothesized, continuous CDF} \\ H_1 &= \text{The data do not have the hypothesized, continuous CDF} \end{aligned}. \tag{5.50}$$

In this case, the hypothesized continuous CDF is the standard normal distributions with mean zero and variance 1, $N(0, 1)$. The normality hypothesis is rejected for all cities since the Kolmogorov–Smirnov statistics are larger than 4.5 for all cities. The critical values of the Kolmogorov–Smirnov test is 1.36 for confidence level of 5%. Moreover, the p values are 0 for all cities indicating the rejection of the null hypothesis that the residuals are drawn from the standard normal distribution.

Finally, a Ljung–Box lack-of-fit hypothesis test is performed. The Ljung–Box tests for model misspecification is based on the Q-statistic. The Ljung–Box test can be defined as follows:

$$\begin{aligned} H_0 &: \text{ The data are random} \\ H_1 &: \text{ The data are not random} \end{aligned} \tag{5.51}$$

and the test statistic is given by

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}, \tag{5.52}$$

where n is the sample size, $\hat{\rho}_k^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested. The corresponding statistics and p values can be found on Table 5.16. All p values are larger than 0.05 with an exception of Berlin where the p values is 0.0493, indicating the absence of autocorrelation in the residuals of the WN in 5% significance level.

The above results are confirmed by the ACF of the residuals. The ACF of the residuals is shown on Fig. 5.19. However, a closer inspection of the ACF reveals a seasonal component in the residuals in Madrid and Rome.

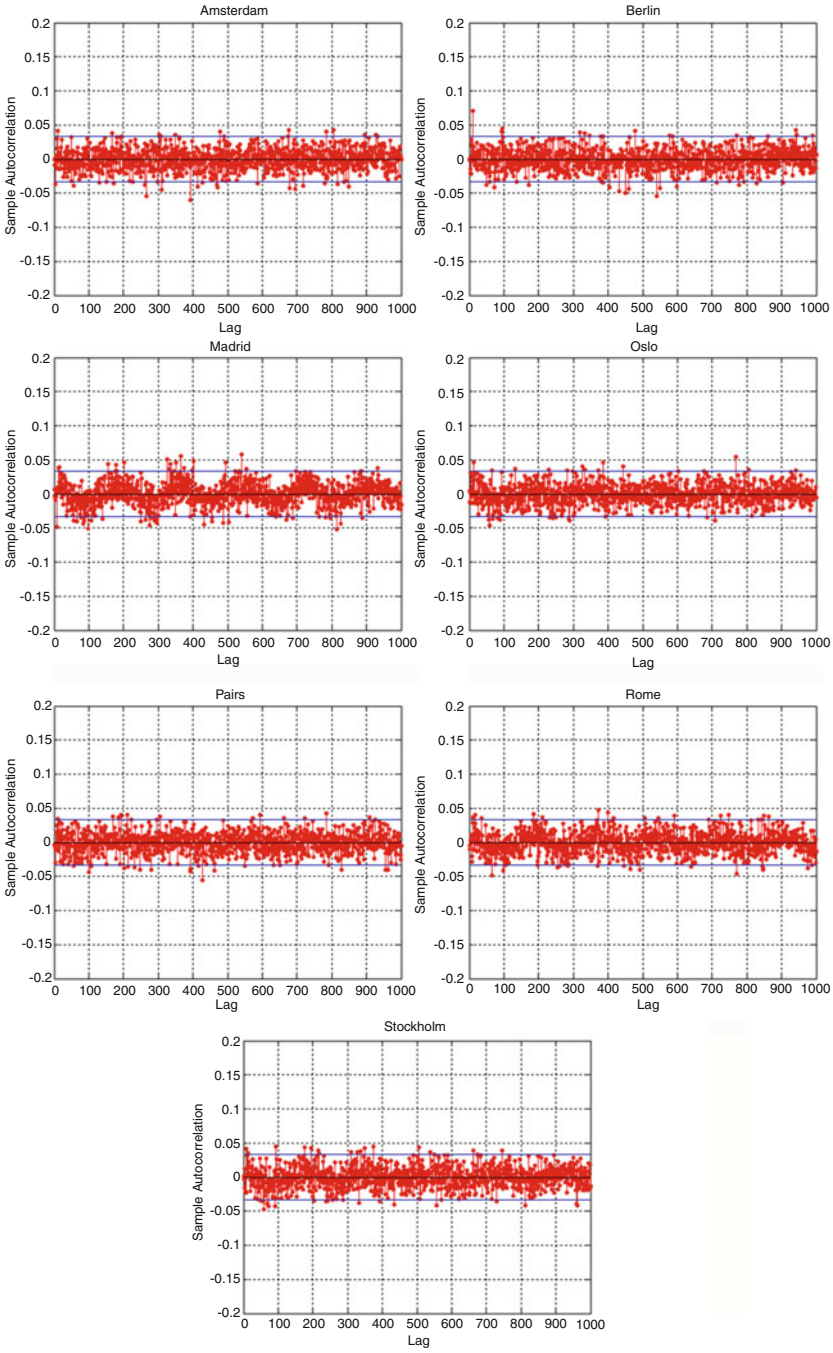

Fig. 5.19 Autocorrelation function of the residuals of the wavelet network of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

Previous studies identified the existence of seasonal variance in the residuals of either the linear or the nonlinear AR model (Benth and Saltyte-Benth 2005, 2007; Zapranis and Alexandridis 2008, 2009b). Hence, the residuals are further examined. More precisely, the ACF of the squared residuals are inspected. The ACF of the squared residuals can be found in Fig. 5.20. By squaring the residuals, the seasonal pattern in the ACF is clear in every city as it is shown in Fig. 5.20.

As it was mentioned earlier, the seasonal variance can be modeled by a series of truncated Fourier series. However, we follow a more advance method where the selection of the parts of the Fourier series is based on WA. This method is similar as the one presented in Sect. 5.4.2 for the seasonal mean. Hence, the seasonal variance is modeled by (5.17) as follows:

$$\sigma^2(t) = c + \sum_{i=1}^{I_2} c_i \sin\left(2p_i'\pi t/365\right) + \sum_{j=1}^{J_2} d_j \cos\left(2p_j'\pi t/365\right).$$

Since for the residuals $e(t)$ of the nonlinear AR model it is true that

$$e(t) = \sigma(t)\varepsilon(t), \tag{5.53}$$

where $\varepsilon(t)$ are i.i.d. $N(0, 1)$, the seasonal variance of the residuals can be extracted as follows: first, the residuals are grouped into 365 groups, comprising 10 observations each (each group corresponds to a single day of the year). Then, by taking the average of the 10 squared values, the variance of that day is obtained. That is, we assume that the seasonal variance is repeated every year:

$$\sigma^2(t + 365) = \sigma^2(t), \tag{5.54}$$

where $t = 1, \ldots, 3650$.

In Fig. 5.21, the empirical variance in Berlin is presented. To decide which terms of the truncated Fourier series to use in order to model the variance $\sigma^2(t)$, WA is performed again. The Daubechies 8 wavelet at level 8 was used. The Daubechies 8 wavelet can be found in Fig. 5.22.

In Fig. 5.23, selected parts of the wavelet decomposition of the squared residuals for Berlin are presented. It is clear that a cycle of 1 year exists (approximation at level 8, a_8) as it was assumed by (5.54). Moreover, a half-year cycle (a_6 and d_7) as well as a seasonal cycle exist (d_6). Hence, in (5.17), we set $I_2 = 3$ and $J_2 = 3$. Moreover, the results from WA indicate that $p_1' = 1, p_2' = 2$, and $p_3' = 3$. In panel A of Table 5.17, the results of the wavelet decomposition for the remaining cities are presented. Since parameters p_i' were identified by WA, least squares method were used to fit the parameters c_i and d_j of (5.17).

The estimated parameters of the seasonal variance in Berlin are as follows: $c_0 = 5.42, c_1 = 0.94, c_2 = -0.53, d_2 = 1.13$, and $d_3 = -0.31$. Note that parameters c_3 and d_1 are not statistically significant, and they are not reported. In panel B of Table 5.17, the estimated parameters of the remaining cities are reported. Note that only the statistically significant parameters (p value < 0.05) are reported. Parameters with p value > 0.05 are omitted and removed from the model. Hence, in

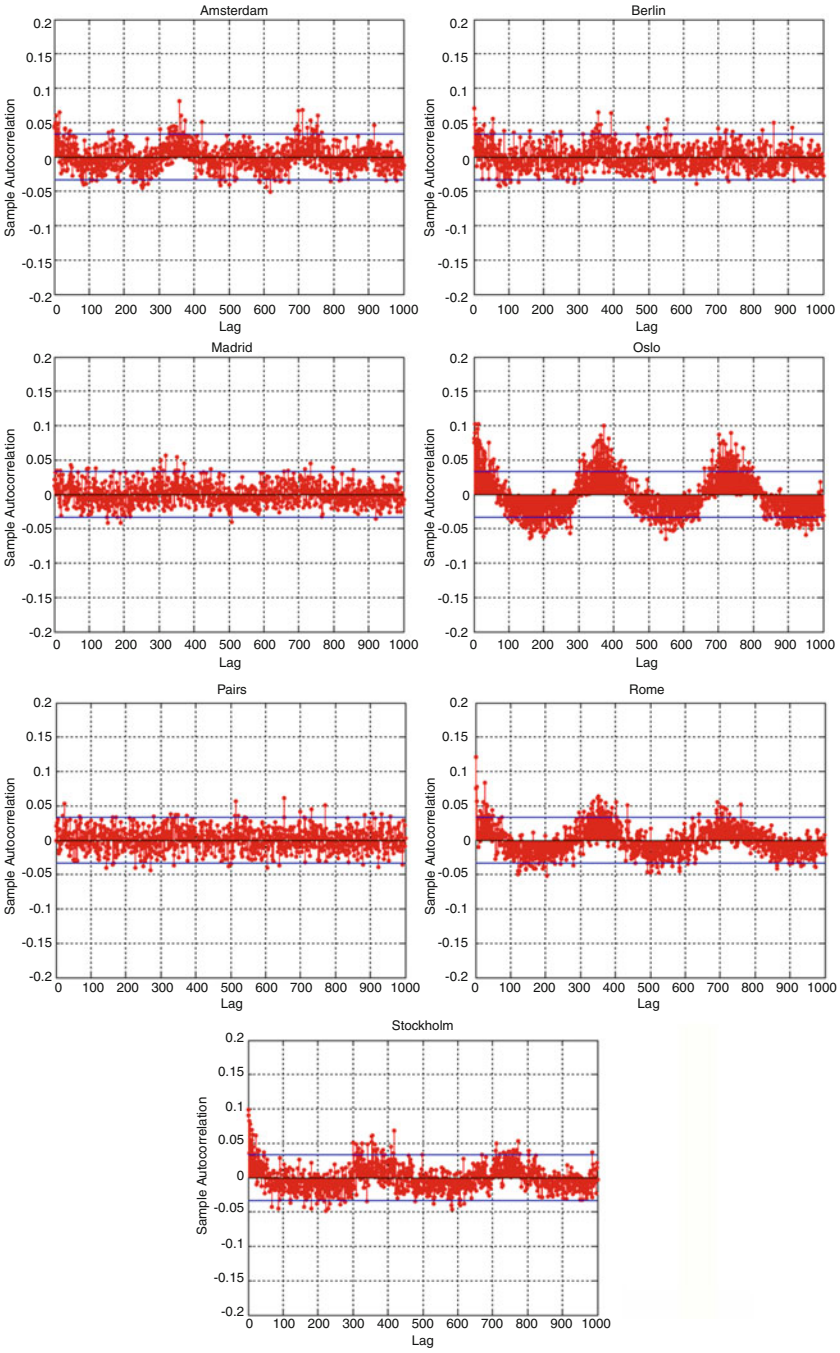

Fig. 5.20 Autocorrelation function of the squared residuals of the wavelet network of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

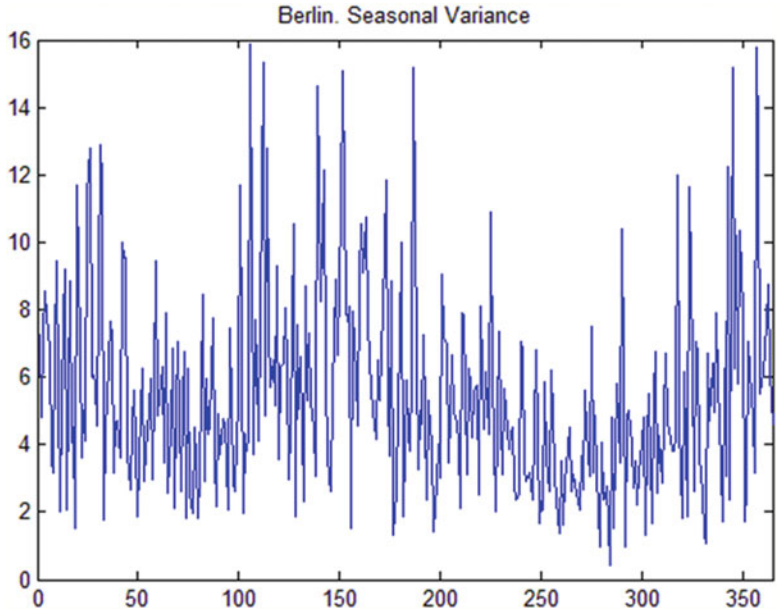

Fig. 5.21 Empirical seasonal variance in Berlin

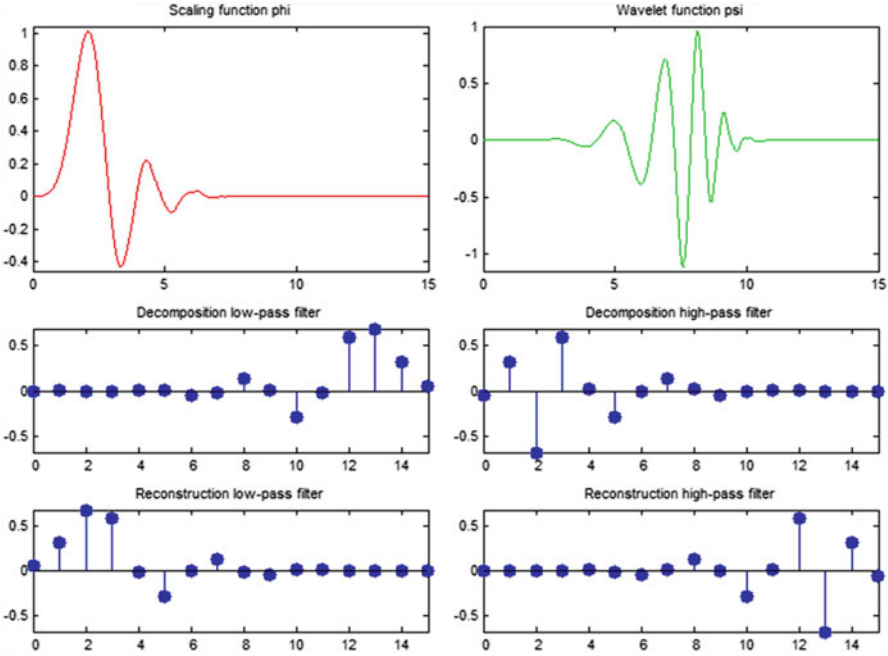

Fig. 5.22 The Daubechies 8 wavelet (*top right*) together with the scaling function (*top left*). The decomposition (*middle*) and reconstruction (*bottom*) filters are also presented with the low-pass (*left*) and high-pass (*right*) filters

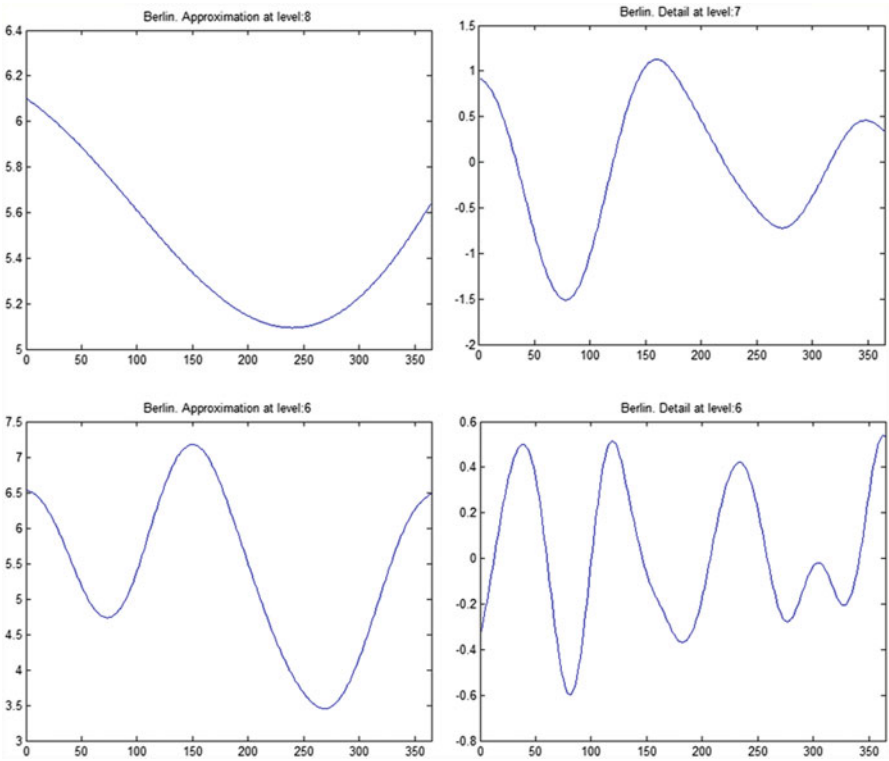

Fig. 5.23 Selected parts of the discrete wavelet decomposition of the seasonal variance in Berlin: approximations (a_j) and details (d_j). The Daubechies 8 wavelet at level 8 was used

Table 5.17 Estimated parameters of the seasonal variance using wavelet analysis

	Amsterdam	Berlin	Madrid	Oslo	Paris	Rome	Stockholm
Panel A							
p'_1	1	1	1	1	1	1	1
p'_2	2	2	–	2	1.5	2	2
p'_3	5	3	–	–	–	–	4
Panel B							
c_0	3.18	5.42	3.18	4.78	4.44	2.41	4.16
c_1	0.34	0.94	0.38	0.68	1.07	0.25	0.85
c_2	−0.42	−0.53	–	–	−1.28	−0.32	−0.40
c_3	–	–	–	–	–	–	0.43
d_1	0.69	–	−0.46	2.51	−0.73	–	1.10
d_2	0.72	1.13		1.27	–	1.02	0.75
d_3	−0.31	0.47	–	–	–	–	–

In Panel A, the length of each cycle in years is presented. In Panel B, the estimated parameters of the seasonal mean are reported. Only the statistical significant parameters with p value < 0.05 are presented

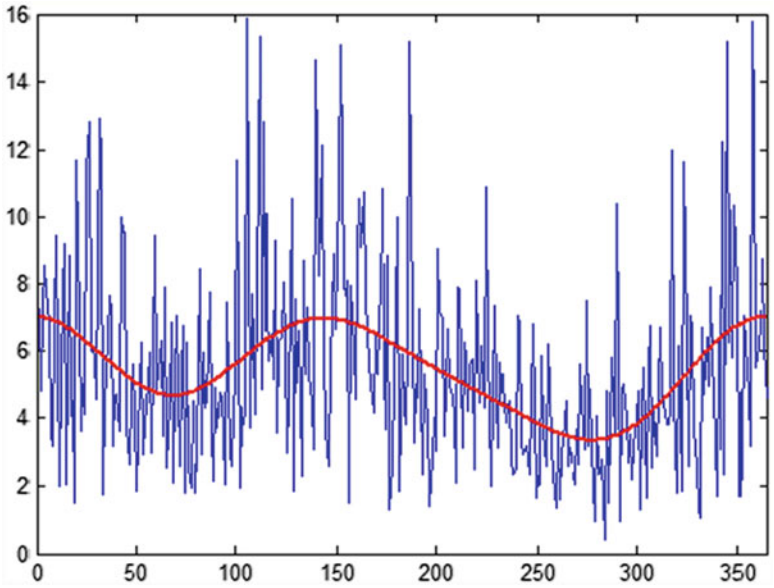

Fig. 5.24 Empirical and fitted variance in Berlin

Madrid, only three parameters were needed in order to fit and remove the seasonal variance, while in Amsterdam and Stockholm, six parameters were needed.

The empirical values of the variance of the residuals (365 values) in Berlin together with the fitted variance can be seen in Fig. 5.24. We observe that the variance takes its highest values during the winter months, while it takes its lowest values during the summer months. This is consistent with our initial hypothesis in Sect. 5.3. Moreover, an increase in variance is observed during May.

In this section, WA was applied in order to indentify the various periodicities that exist in the seasonal variance. The seasonal variance was successfully captured by a truncated Fourier series, and then it was divided out from our data. Since the seasonal variance was successfully fitted, it can be removed from the data to obtain the final residuals (noise $\varepsilon(t)$). Note that the residuals of the nonlinear AR model and the residuals after dividing out the seasonal variance are connected with (5.53). In the next section, the remaining residuals $\varepsilon(t)$ will be further tested.

5.7 Examination of the Residuals

In this section, the residuals $\varepsilon(t)$ after dividing out the seasonal variance will be examined. Various statistics of the remaining residuals will be presented as well as distributional tests will be performed. Finally, a comparison between the proposed model and previous studies will be presented. More precisely, our model will be compared against the models proposed by Alaton et al. (2002) and by Benth and Saltyte-Benth (2007).

5.7.1 Testing the Normality Hypothesis

First, the ACF of the residuals after dividing out the seasonal variance is examined. Figure 5.25 presents the ACF of the squared residuals after dividing out the seasonal variance for the seven cities. We observe that the seasonality has been successfully removed from all cities.

In Table 5.18, the descriptive statistics of the residuals after dividing out the seasonal variance are presented. The residuals for the seven cities have a mean of almost 0 and standard deviation of 1. In all cities, a negative skewness is present with an exception of Amsterdam where the skewness is positive. In addition, positive kurtosis is evident in all cities. Moreover, a Ljung–Box lack-of-fit hypothesis test is performed. The corresponding statistics and p values can be found on Table 5.18. All p values are larger than 0.05 indicating the absence of autocorrelation in the residuals in confidence level of 5%. Finally, a Kolmogorov–Smirnov is performed to test the normality hypothesis. In Table 5.18, the corresponding statistics and p values are presented. In Berlin, Oslo, Paris, and Stockholm, the null hypothesis that the residuals are drawn from the normal distribution cannot be rejected in 10% confidence level. Similarly, in Amsterdam, the null hypothesis cannot be rejected in 1% confidence level. Only in two cities, in Madrid and Rome, the normality hypothesis is rejected.

Next, the hypothesis of long range dependence in the estimated residuals should be tested. The Hurst exponent is related to the fractional differencing parameter d and is given by

$$H = d + \frac{1}{2}. \tag{5.55}$$

The Hurst exponent takes values in the interval (0,1). For $\frac{1}{2} < H < 1$, the process has long memory, and for $0 < H < \frac{1}{2}$, the process has short memory, while for $H = \frac{1}{2}$, the BM is retrieved (Bellini 2005). In Table 5.19, the Hurst exponent for the seven cities is presented. The Hurst exponent was estimated after all seasonal component were removed from the data. The iterative method described in Koutsoyiannis (2003) is followed in order to estimate the Hurst exponent.

Results from Table 5.19 indicate that the Hurst exponent does not differ significantly from 0.5. The smallest Hurst exponent was observed in Amsterdam with value of 0.4874 and standard deviation of 1.0018, while the largest Hurst exponent was observed in Oslo with value of 0.5201 and standard deviation of 0.99847. The standard deviation for all cities is very close to 1. The above results indicate the absence of fractionality characteristics in the dynamics of the temperature process. Therefore, the assumption of a BM instead of FBM is justified.

Our results are in contrast to those of Brody et al. (2002) and Benth (2003). In Brody et al. (2002), the Hurst exponent was calculated before the elimination of any seasonal components. Their results are in line with the results presented in Table 5.3. Note that results from Table 5.3, where the Hurst exponent was calculated

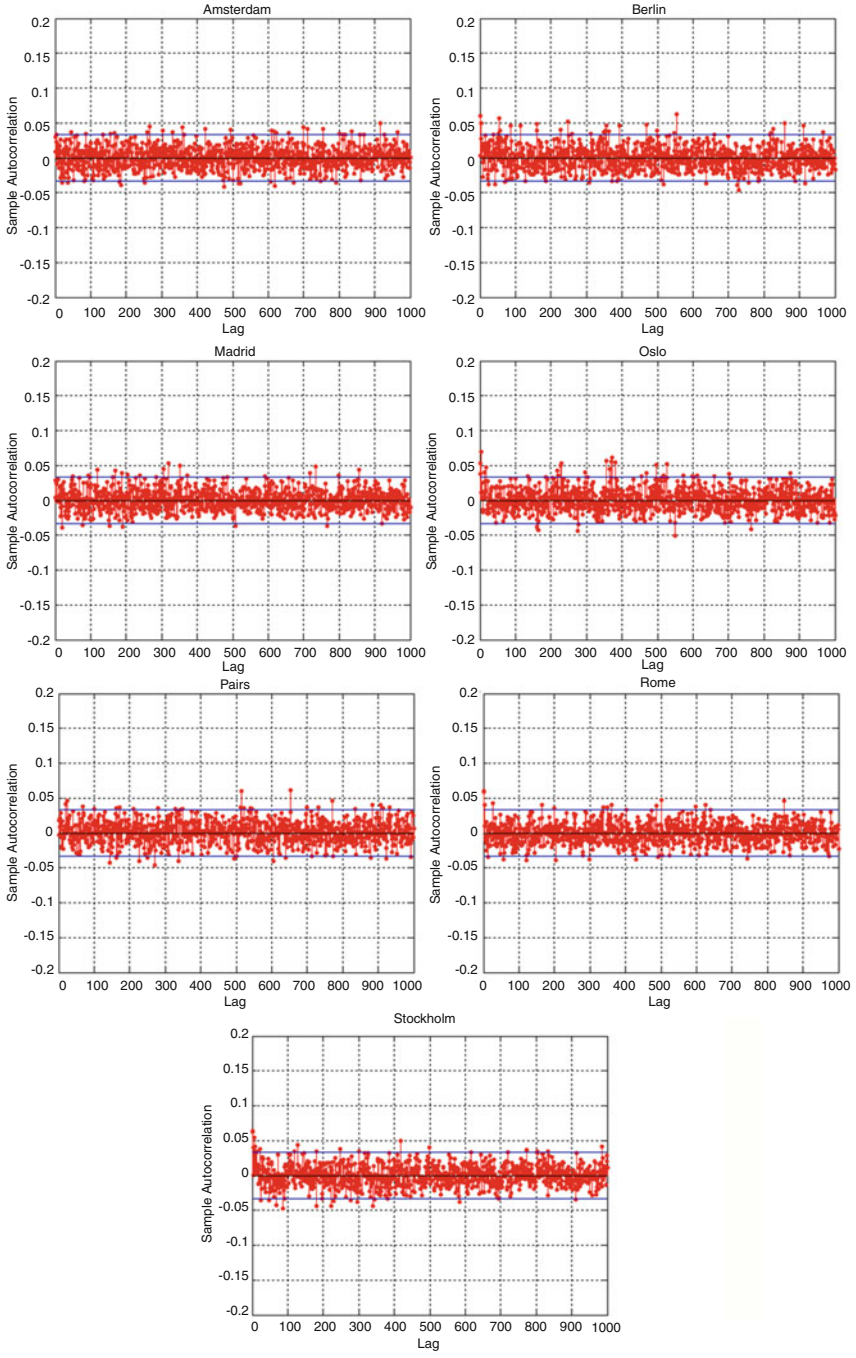

Fig. 5.25 Autocorrelation function of the squared residuals after dividing out the volatility function of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

Table 5.18 Descriptive statistics of the residuals of the proposed model after dividing out the seasonal variance

City	Mean	St. dev	Max	Median	Min	Skewness	Kurtosis	K–S	*P* value	LBQ	*P* value
Amsterdam	0.00	1.00	3.80	−0.04	−4.16	0.13	3.50	1.49	0.0237	23.068	0.2855
Berlin	0.00	1.00	4.45	0.00	−4.02	−0.02	3.53	0.96	0.3086	29.616	0.0763
Madrid	0.01	1.00	4.40	0.08	−4.37	−0.34	3.64	2.41	0.0010	27.937	0.1109
Oslo	0.00	1.00	3.95	0.02	−4.37	−0.08	3.67	1.06	0.2125	29.681	0.0750
Paris	0.00	1.00	2.89	0.02	−4.23	−0.17	3.01	0.90	0.3960	21.192	0.3859
Rome	0.01	1.00	3.94	0.02	−4.21	−0.10	3.90	1.80	0.0030	23.802	0.2512
Stockholm	−0.01	1.00	3.74	0.03	−4.49	−0.16	3.64	1.21	0.1084	28.340	0.1016

St. dev standard deviation, *K–S* Kolmogorov–Smirnov goodness-of-fit, *LBQ* Ljung–Box Q-statistic lack-of-fit

Table 5.19 Hurst exponent of the residuals after removing all seasonal components

	Amsterdam	Berlin	Madrid	Oslo	Paris	Rome	Stockholm
Hurst	0.4874	0.5078	0.4951	0.5201	0.4928	0.5138	0.5069

Table 5.20 Estimated parameters using the Alaton model for the seven cities

City	κ	A	B	C	φ	$\overline{R}^2$
Amsterdam	0.194	9.66	0.000312	7.251	−1.923	71.58
Berlin	0.216	9.59	0.000226	9.658	−1.825	74.79
Madrid	0.178	15.06	–	9.264	−1.898	79.37
Oslo	0.251	6.06	0.000239	9.966	−1.865	80.30
Paris	0.226	12.08	0.000233	7.766	−1.880	72.91
Rome	0.231	15.41	0.000087	8.788	−2.030	85.28
Stockholm	0.220	6.26	0.000319	9.663	−1.966	79.54

The parameters using the Alaton model. κ is the speed of mean reversion, A is the intercept, B is the slope of the linear trend, C is the amplitude of the seasonal component, and φ is the angle referring to the day of the maximum temperature. Only the statistical significant parameters with p value < 0.05 are presented

before removing the seasonal components, indicate the presence of long memory. In this study, WA was used in order to successfully remove all seasonal effects in temperature and in the seasonal variance. Hence, any possible fractionality was successfully removed. The same conclusion achieved in (Bellini 2005) using Fourier theory in order to indentify periodicities in the temperature data.

5.7.2 *In-Sample Comparison*

Next, the proposed model will be compared in-sample against two models previously proposed in the literature. The first model was proposed by Alaton et al. (2002), while the second model was proposed by Benth and Saltyte-Benth (2007). For simplicity, we name the two models as the Alaton and the Benth model, respectively.

In Table 5.20, the estimated parameters from Alaton model are presented, while in Table 5.21, the descriptive statistics of the residuals can be found. In Table 5.20, only the statistical significant parameters at significance level 5% are reported. A closer inspection of Table 5.21 indicates that the distributional statistics are similar to the statistics of the residuals of our proposed model. The mean is almost zero, and the standard deviation is almost 1 for all cities. With an exception of Paris, there is positive kurtosis. On the other hand, negative skewness is present in all cities with the exception of Amsterdam and Berlin. The results of the normality hypothesis test performed by the Kolmogorov–Smirnov test indicate that the normality hypothesis is rejected in Amsterdam, Madrid, and Rome, while there is not enough evidence to reject the normality hypothesis in Oslo, Berlin, Paris, and

Table 5.21 Descriptive statistics of the residuals of the Alaton model

City	Mean	St. dev	Max	Median	Min	Skewness	Kurtosis	K–S	*P* value	LBQ	*P* value
Amsterdam	0.00	0.99	3.42	−0.04	−4.05	0.16	3.40	1.89	0.0015	193.43	0.0000
Berlin	0.00	0.99	4.40	−0.02	−3.85	0.01	3.43	0.99	0.2799	87.82	0.0000
Madrid	0.00	1.00	3.91	0.08	−4.46	−0.31	3.44	2.17	0.0002	188.45	0.0000
Oslo	0.00	0.99	3.51	0.03	−4.84	−0.07	3.51	1.20	0.1126	60.96	0.0000
Paris	0.00	0.99	3.03	0.00	−3.61	−0.13	2.95	0.75	0.6156	100.63	0.0000
Rome	0.01	0.99	3.92	0.01	−4.20	−0.07	3.85	2.07	0.0004	99.13	0.0000
Stockholm	0.00	0.99	3.64	0.02	−4.32	−0.12	3.50	1.13	0.1567	100.15	0.0000

St. dev standard deviation, *K–S* Kolmogorov–Smirnov goodness-of-fit, *LBQ* Ljung–Box Q-statistic lack-of-fit

Table 5.22 Estimated parameters of the Benth model for the seven cities

Parameter	Amsterdam	Berlin	Madrid	Oslo	Paris	Rome	Stockholm
a	9.66	9.59	15.06	6.06	12.08	15.41	6.26
b	0.000312	0.000226	–	0.000239	0.000233	0.000087	0.000319
κ	0.194	0.216	0.178	0.253	0.226	0.231	0.220
b_1	−7.246	−9.654	−9.256	−9.961	−7.762	−8.777	−9.655
g_1	20.365	14.676	18.830	16.989	17.870	26.551	22.820
c_0	4.983	8.490	4.765	8.020	6.432	3.916	6.658
c_1	0.581	1.461	0.416	1.158	0.680	0.384	1.358
c_2	−0.675	−0.963	−0.237	−0.292	−0.233	−0.529	−0.618
c_3	0.230	0.288	−0.297	0.872	−0.356	−0.064	0.664
c_4	0.360	0.071	0.285	0.041	0.303	−0.271	0.739
d_1	1.078	−0.011	−0.913	4.292	0.056	1.652	1.803
d_2	1.164	1.777	0.231	2.160	0.851	0.316	1.258
d_3	0.235	0.726	0.245	0.912	0.011	0.125	0.507
d_4	−0.330	−0.251	0.410	0.062	−0.073	0.271	−0.321

The parameters using the Benth model. κ is the speed of mean reversion, a is the intercept, and b is the slope of the linear trend. b_1 and g_1 is the amplitude and the angle of the seasonal mean. c_1 and d_1 are the parameters of the seasonal variance. Only the statistical significant parameters with p value < 0.05 are presented

Stockholm in 10% confidence level. However, the Ljung–Box Q-statistic lack-of-fit reveals strong autocorrelation in the residuals. Hence, the results of the previous test for normality may not lead to substantial values of the Kolmogorov–Smirnov test.

In Table 5.22, the estimated parameters from the Benth model are presented, while in Table 5.23, the descriptive statistics of the residuals can be found. In Table 5.22, only the statistical significant parameters at significance level 5% are reported. A closer inspection of Table 5.23 indicates that the standard deviation is close to 0.8 in contrast to the initial hypothesis that the residuals follow a $N(0, 1)$ distribution. This results to an implication of the estimation of the seasonal variance. In addition, the normality hypothesis is rejected in all cities. More precisely, the Kolmogorov–Smirnov value is over 3.5 with p value of 0 for all cities. Finally, the Ljung–Box Q-statistic lack-of-fit reveals strong autocorrelation in the residuals.

The findings of Benth and Saltyte-Benth (2007) for the Stockholm temperature series are very similar. Although they did not use WA to calibrate their model, they had managed to remove seasonality from the residuals, but their distribution proved to be non-normal. They suggested that a more refined model would probably rectify this problem, but they did not proceed in estimating one. In an earlier paper regarding Norwegian temperature data, Benth and Saltyte-Benth (2005) suggested to model the residuals by a generalized hyperbolic distribution. However, as the same authors comment the inclusion of a non-normal model leads to a complicated Lévy process dynamics. Recently, Benth et al. (2007) proposed a continuous time autoregressive process with lag p (CAR(p)-process). Although they managed to correct the autocorrelation on the residuals, their distribution proved again to be non-normal.

Table 5.23 Descriptive statistics of the residuals of the Benth model for the seven cities

City	Mean	St. dev	Max	Median	Min	Skewness	Kurtosis	K–S	*P* value	LBQ	*P* value
Amsterdam	0.00	0.82	2.86	−0.03	−3.18	0.15	3.42	3.82	0.0000	197.39	0.0000
Berlin	0.00	0.81	3.60	−0.01	−3.25	0.00	3.49	3.91	0.0000	82.29	0.0000
Madrid	0.00	0.84	3.23	0.07	−3.45	−0.31	3.48	3.76	0.0000	181.94	0.0000
Oslo	0.00	0.78	2.91	0.02	−3.89	−0.06	3.58	4.60	0.0000	40.21	0.0047
Paris	0.00	0.80	2.53	0.00	−3.10	−0.15	2.99	3.57	0.0000	98.97	0.0000
Rome	0.01	0.79	3.31	0.00	−3.44	−0.08	3.90	4.89	0.0000	83.79	0.0000
Stockholm	0.00	0.80	2.90	0.02	−3.39	−0.14	3.58	4.11	0.0000	88.77	0.0000

St. dev standard deviation, *K–S* Kolmogorov–Smirnov goodness-of-fit, *LBQ* Ljung–Box Q-statistic lack-of-fit

Zapranis and Alexandridis (2006) estimated a number of alternatives to the original AR(1) model. In particular, they estimated an ARMA(3,1) model, a long-memory homoscedastic ARFIMA model, and a long-memory heteroscedastic ARFIMA–FIGARCH model. Their findings suggest that increasing the model complexity and thus the complexity of theoretical derivations in the context of weather derivative pricing does not seem to be justified.

Our model outperformed the two models in the sense of distributional statistics. First of all, in contrast to the models of Alaton and Benth, our tests indicate the absence of autocorrelation in the residuals. Next, only in two of the seven cities the normality hypothesis was rejected justifying our initial hypothesis of a BM as the driving noise process. Finally, WA successfully indentified all the seasonal cycles that affect the temperature dynamics.

5.7.3 *Testing Alternative Distributions for the Residuals*

In the previous section, the residuals of our proposed model were examined. We concluded that the use of a BM is justified since the normality hypothesis was rejected in only two cities. In order to obtain a better understanding of the distributions of the residuals, we expand our analysis by fitting additional distributions. More precisely, a Lévy family distribution is fitted to the residuals. The Lévy family contains many known distributions as subclasses. To our knowledge, only Benth and Saltyte-Benth (2005) and Bellini (2005) used a Lévy process as the driving noise process. In particular, Benth and Saltyte-Benth (2005) used a generalized hyperbolic distribution. In Bellini (2005), a hyperbolic distribution was used which is a limiting case of the generalized hyperbolic distribution. In this study, two limiting cases of the generalized hyperbolic, the hyperbolic and the NIG, and one limiting case of the Lévy distribution, the stable distribution, are examined.

The generalized hyperbolic probability density function (PDF) is given by

$$f_{GH}(x,\lambda,a,b,\mu,\delta) = \frac{\left(\sqrt{a^2-b^2}/\delta\right)^{\lambda} e^{\beta(x-\mu)}}{\sqrt{2\pi}K_{\lambda}\left(\delta\sqrt{a^2-b^2}\right)} \cdot \frac{K_{\lambda-\frac{1}{2}}\left(a\sqrt{\delta^2+(x-\mu)^2}\right)}{\left(\sqrt{\delta^2+(x-\mu)^2}/a\right)^{\frac{1}{2}-\lambda}}, \quad (5.56)$$

where μ is the location, a controls the steepness of the distribution, b is the asymmetry parameter, and δ is the scale parameter. The parameter λ is indentifying the subfamily within the generalized hyperbolic class, while K_{λ} is the modified Bessel function of the third kind.

The moment-generating function of the generalized hyperbolic distribution is given by

$$\mathrm{E}\left[e^{uX}\right] = e^{\mu u}\left(\frac{\sqrt{a^2-b^2}}{\sqrt{a^2-(b+u)^2}}\right)^{\lambda} \cdot \frac{K_\lambda\left(\delta\sqrt{a^2-(b+u)^2}\right)}{K_\lambda\left(\delta\sqrt{a^2-b^2}\right)}. \tag{5.57}$$

Hence, if $L(t)$ is a Lévy process with $L(1)$ being a generalized hyperbolic distributed random variable, then $L(t)$ is called a generalized hyperbolic Lévy process. Then, the Lévy measure if $\lambda \geq 0$ is given by

$$l_{GH} = |z|^{-1} e^{bz}\left\{\frac{1}{\pi^2}\int_0^\infty \frac{1}{y}\frac{\exp\left(-\sqrt{2y+a^2}|z|\right)}{J_\lambda^2(\delta\sqrt{2y})+Y_\lambda^2(\delta\sqrt{2y})}dy + \lambda e^{-a|z|}\right\}dz \tag{5.58}$$

and when $\lambda<0$ by

$$l_{GH}(dz) = |z|^{-1} e^{bz}\frac{1}{\pi^2}\int_0^\infty \frac{1}{y}\frac{\exp\left(-\sqrt{2y+a^2}|z|\right)}{J_{-\lambda}^2(\delta\sqrt{2y})+Y_{-\lambda}^2(\delta\sqrt{2y})}dydz, \tag{5.59}$$

where J_λ and Y_λ are the Bessel functions of the first and second kind, respectively, with index λ.

The hyperbolic PDF f_H is obtained from (5.56) with $\lambda = 1$ and the normal inverse Gaussian (NIG) PDF f_{NIG} with $\lambda = -\frac{1}{2}$. Hence, for the hyperbolic distribution, we have that

$$f_H(x,a,b,\mu,\delta) = \frac{\sqrt{a^2-b^2}}{\sqrt{2\pi}\delta K_1\left(\delta\sqrt{a^2-b^2}\right)} e^{-a\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)} \tag{5.60}$$

with Lévy measure given by

$$l_H(dz) = |z|^{-1} e^{bz}\left\{\frac{1}{\pi^2}\int_0^\infty \frac{1}{y}\frac{\exp\left(-\sqrt{2y+a^2}|z|\right)}{J_1^2(\delta\sqrt{2y})+Y_1^2(\delta\sqrt{2y})}dy + e^{-a|z|}\right\}dz. \tag{5.61}$$

Similarly, for the NIG distribution, we have that

$$f_{NIG}(x,a,b,\mu,\delta) = \frac{a\delta K_1\left(a\sqrt{\delta^2+(x-\mu)^2}\right)}{\pi\sqrt{\delta^2+(x-\mu)^2}} e^{\delta\sqrt{a^2-b^2}+b(x-\mu)} \tag{5.62}$$

with Lévy measure given by

$$l_{NIG}(dz) = \frac{a\delta}{\pi|z|}K_1(a|z|)e^{bz}dz. \tag{5.63}$$

The Lévy distribution PDF is given by

$$f_L = \sqrt{\frac{c}{2\pi}} \cdot \frac{e^{\frac{-c}{2(x-\mu)}}}{(x-\mu)^{\frac{3}{2}}}, \tag{5.64}$$

where c is a positive constant. Hence, if $X \sim Levy(\mu, c)$, then $X \sim Stable\left(\frac{1}{2}, 1, \mu, c\right)$.

The PDF for a general stable distribution cannot be written analytically; however, the general characteristic function can be. The PDF is determined by its characteristic function $\varphi(t)$ by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t) e^{-ixt} dt. \tag{5.65}$$

The random variable X is called stable if its characteristic function is given by

$$\varphi(t, \alpha, \beta, \mu, c) = \exp[it\mu - |ct|^a (1 - i\beta \text{sgn}(t)\Phi)], \tag{5.66}$$

where $\text{sgn}(t)$ is the sign of t and Φ is given by

$$\Phi = \tan\left(\frac{\pi\alpha}{2}\right) \tag{5.67}$$

for all α except $\alpha = 1$ in which case

$$\Phi = -\frac{2}{\pi} \log(t). \tag{5.68}$$

In expression (5.66), μ is a shift (location) parameter, $\beta \in [-1, 1]$, and is a measure of asymmetry, c is the scale parameter, and $\alpha \in (0, 2]$ is the exponent or index of the distribution and specifies the asymptotic behavior of the distribution. Both the normal and the Lévy distributions are special cases of the stable distribution.

The distance between the empirical distribution of the residuals and the four distributions is estimated by the Kolmogorov distance. In addition, the Anderson–Darling test which gives additional weight to the tails of the distribution is also performed (Bellini 2005). Kolmogorov distance is given by (5.49), while the Anderson–Darling statistics is given by

$$AD = \sup_x \frac{|F_n(x) - F(x)|}{\sqrt{F(x)(1 - F(x))}}. \tag{5.69}$$

The Kolmogorov–Smirnov test does not depend on the specific distribution on calculating critical values. On the other hand, critical values for the Anderson–Darling test exist only for the normal, lognormal, exponential, Weibull, extreme value type I, and logistic distributions. In Table 5.24, the estimated Kolmogorov distance and the

Table 5.24 Distributional tests

	Amsterdam		Berlin		Madrid		Oslo		Paris		Rome		Stockholm	
	A–D	K–S	A–D	K–S	A–D	K–S	A–D	K–S	A–D	K–S	A–D	K–S	A–D	K–S
Normal	3.96	1.49	1.39	0.96	8.68	2.41	1.96	1.06	1.26	0.90	7.40	1.80	2.82	1.21
Hyperbolic	0.47	0.53	0.17	0.48	0.48	0.58	0.29	0.54	0.44	0.80	0.20	0.50	0.34	0.57
NIG	0.52	0.55	0.17	0.48	0.53	0.61	0.29	0.54	inf	0.80	0.22	0.52	0.34	0.59
Stable	1.04	0.61	0.31	0.62	8.88	2.57	0.98	0.83	0.65	0.61	0.84	0.69	1.15	0.83

Kolmogorov distances (K–S) and the Anderson–Darling (A–D) statistics, performed to test if the residuals come from the specified distribution. The normal, hyperbolic, normal inverse Gaussian (NIG), and stable distributions are tested. The critical value of the Kolmogorov distribution is 1.36 at confidence level of 5% and 1.63 at 1%. For the A–D statistic, critical values are not available for the hyperbolic, NIG, and stable distributions

Anderson–Darling statistics are presented for four distributions: normal, hyperbolic, NIG, and stable. From Table 5.24, it is clear that both statistics have the smallest values when the hyperbolic distribution is used with an exception of Paris where the stable distribution provides the smallest Kolmogorov–Smirnov statistic.

Concluding, the hyperbolic distribution provides a slightly better fit than the normal distribution. However, introducing a Lévy process in the temperature dynamics does not allow to find closed form solutions for the temperature derivatives. The increased complexity of the pricing formulas of the weather derivatives makes the use of the normal distribution more favorable.

5.8 The Forecasting Ability of the Daily Models

In this section, our proposed model will be validated out-of-sample. Our method is validated and compared against two forecasting methods proposed in prior studies, the Alaton's and Benth's models. The three models will be used for forecasting out-of-sample DATs for different periods. Usually, temperature derivatives are written for a period of a month or a season and sometimes even for a year. Hence, DATs for 1, 2, 3, 6, and 12 months will be forecasted. The out-of-sample period corresponds to the period of 1st January–31st December 2001 and every time interval starts at 1st January of 2001. Note that the DATs from 2001 were not used for the estimation of the parameters of the three models. Next, the corresponding HDDs and CAT indices will be constructed.

The predictive power of the three models will be evaluated using two out-of-sample forecasting methods. First, we will estimate out-of-sample forecasts over a period and then 1-day-ahead forecasts over a period. The first case, in the out-of-sample forecasts, today (time step 0) temperature is known and is used to forecast the temperature tomorrow (time step 1). However, tomorrow's temperature is unknown and cannot be used to forecast the temperature 2 days ahead. Hence, we use the forecasted temperature at time step 1 to forecast the temperature at time step 2 and so on. We call this method the out-of-sample over a period forecast. The second case, the 1-day-ahead forecast, the procedure is as follows. Today (time step 0), temperature is known and is used to forecast the temperature tomorrow (time step 1). Then tomorrow's real temperature is used to forecast the temperature at time step 2 and so on. We will refer to this method as the 1-day-ahead over a period forecast. The first method can be used for out-of-period valuation of a temperature derivative, while the second one for in-period valuation. Naturally, it is expected the first method to cause larger errors.

In order to forecast the future DATs in the seven cities, the MC method was applied. So far, we have modeled temperature by a stochastic differential equation. The form of the temperature model reveals that the temperature is a path depended process. However, our temperature model depends on the driving noise process which is modeled by a BM. Hence, in order to estimate the expected temperature at some future point, the following procedure is pursued. First, we create a large

Table 5.25 Out-of-sample comparison for a period of 1 month using the HDD index and the relative percentage errors

HDD/t1	Real	Historical	Alaton	Benth	WN
Amsterdam	463.6	449.49	445.15	439.38	460.57
Berlin	522.4	517.9	531.05	528.1	532.26
Madrid	338.8	366.33	369.61	355.16	369.18
Oslo	646.7	630.45	652.8	679	676.08
Paris	378.6	394.68	395.71	395.43	398.13
Rome	260.1	322.8	335.39	311.02	299.52
Stockholm	555.7	602.13	622.15	599.27	590.74
Relative percentage errors					
Amsterdam			3.98%	5.22%	*0.65%*
Berlin			1.66%	*1.09%*	1.89%
Madrid			9.09%	*4.83%*	8.97%
Oslo			*0.94%*	4.99%	4.54%
Paris			4.52%	*4.45%*	5.16%
Rome			28.95%	19.58%	*15.16%*
Stockholm			11.96%	7.84%	*6.31%*

Real and historical HDDs for the period 1–31 January 2001 and estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

number of paths for the future evolution of temperature. Then, the average of all paths is estimated. The MC simulation method provides with an unbiased estimate of the temperatures values. As the number of sample paths, N, increases, the standard error of the estimate, which is given by $1/\sqrt{N}$, decreases. In this thesis, we create 10.000 sample paths for each model that represent the future evolution of temperature over a specified period.

Since we are studying 7 cities and 2 indices for 5 different time periods, the three models are compared in 70 cases for each method. Our results are very promising. In the out-of-sample forecasts our method outperformed alternative methods in 34 cases out of the 70. In the 1-day-ahead forecasts, our model performed even better outperforming the Alaton and Benth models in 47 times out of 70.

From Tables 5.25, 5.26, 5.27, 5.28, 5.29, 5.30, 5.31, 5.32, 5.33 and 5.34, the out-of-sample forecasts for the HDD and CAT indices over a period of 1, 2, 3, 6, and 12 months are presented. The first column corresponds to the real index over the specified index, while the second column to the average historical mean. The next three columns corresponds to the three models, Alaton, Benth, and the proposed one using WNs. The second panel of each table corresponds to the relative absolute percentage errors for each method given by

$$APE = \left| \frac{y - \hat{y}}{y} \right|, \tag{5.70}$$

where y is the corresponding index and $\hat{y}$ is the estimated index.

Table 5.26 Out-of-sample comparison for a period of 1 month using the CAT index and the relative percentage errors

CAT/t1	Real	Historical	Alaton	Benth	WN
Amsterdam	94.4	108.5	112.9	118.6	97.4
Berlin	35.6	40.1	27.0	29.9	25.7
Madrid	219.2	191.7	188.4	202.8	188.8
Oslo	−88.7	−72.5	−94.8	−121.0	−118.1
Paris	179.4	163.3	162.3	162.6	159.9
Rome	297.9	235.2	222.6	247.0	258.5
Stockholm	2.3	−44.1	−64.2	−41.3	−32.7
Relative percentage errors					
Amsterdam			19.54%	25.66%	*3.21%*
Berlin			24.29%	*16.02%*	27.71%
Madrid			14.06%	*7.46%*	13.86%
Oslo			*6.87%*	36.41%	33.12%
Paris			9.54%	*9.38%*	10.89%
Rome			25.27%	*17.09%*	13.23%
Stockholm			2889.30%	1894.43%	*1523.26%*

Real and historical CAT for the period 1–31 January 2001 and estimated CAT using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.27 Out-of-sample comparison for a period of 12 months using the HDD index and the relative percentage errors

HDD/t2	Real	Historical	Alaton	Benth	WN
Amsterdam	840.1	846.8	827.7	814.6	853.8
Berlin	968.3	962.6	971.7	956.6	973.56
Madrid	602.7	642.5	676.1	649.2	678.6
Oslo	1300.4	1195.7	1204.2	1217.3	1228.1
Paris	699.7	733.6	727.5	719.0	723.1
Rome	518.2	602.3	628.6	587.6	566.1
Stockholm	1147.7	1162.0	1163.6	1121.9	1109.9
Relative percentage errors					
Amsterdam			*1.47%*	3.04%	1.63%
Berlin			*0.35%*	1.21%	0.54%
Madrid			12.17%	*7.71%*	12.59%
Oslo			7.40%	6.39%	*5.56%*
Paris			3.97%	*2.76%*	3.35%
Rome			21.30%	13.39%	*9.24%*
Stockholm			*1.39%*	2.25%	3.29%

Real and historical HDDs for the period 1 January–28 February 2001 and estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.28 Out-of-sample comparison for a period of 2 months using the CAT index and the relative percentage errors

CAT/t2	Real	Historical	Alaton	Benth	WN
Amsterdam	221.9	215.3	234.3	247.4	208.2
Berlin	93.7	99.4	90.3	105.4	88.4
Madrid	459.3	419.5	386.0	412.8	383.4
Oslo	−238.4	−133.7	−142.2	−155.3	−166.1
Paris	362.3	328.4	334.5	343.0	338.9
Rome	543.8	459.7	433.4	474.4	495.9
Stockholm	−85.7	−100.0	−101.6	−59.9	−47.9
Relative percentage errors					
Amsterdam			*5.57%*	11.51%	6.18%
Berlin			*3.60%*	12.52%	5.61%
Madrid			15.97%	*10.12%*	16.52%
Oslo			40.36%	34.87%	*30.34%*
Paris			7.68%	*5.34%*	6.47%
Rome			20.29%	12.76%	*8.80%*
Stockholm			*18.59%*	30.15%	44.07%

Real and historical CAT for the period 1 January–28 February 2001 and estimated CAT using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.29 Out-of-sample comparison for a period of 3 months using the HDD index and the relative percentage errors

HDD/t3	Real	Historical	Alaton	Benth	WN
Amsterdam	1251.9	1199.7	1178.6	1155.2	1217.2
Berlin	1429.9	1364.1	1351.7	1324.7	1359.5
Madrid	792.5	832.4	919.6	878.0	927.3
Oslo	1907.3	1717.3	1708.3	1712.3	1735.6
Paris	966.0	1001.1	1012.3	994.1	1002.9
Rome	661.8	836.3	876.4	817.1	788.0
Stockholm	1730.1	1694.9	1671.9	1616.0	1595.6
Relative percentage errors					
Amsterdam			5.86%	7.72%	*2.77%*
Berlin			5.47%	7.36%	*4.92%*
Madrid			16.03%	*10.78%*	17.01%
Oslo			10.43%	10.22%	*9.00%*
Paris			4.79%	*2.91%*	3.82%
Rome			32.42%	23.46%	*19.07%*
Stockholm			*3.36%*	6.59%	7.77%

Real and historical HDDs for the period 1 January–31 March 2001 and estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.30 Out-of-sample comparison for a period of 3 months using the CAT index and the relative percentage errors

CAT/t3	Real	Historical	Alaton	Benth	WN
Amsterdam	368.1	420.3	441.4	464.8	402.9
Berlin	190.1	255.9	268.3	295.3	260.5
Madrid	827.5	787.7	700.4	742.0	692.7
Oslo	−287.3	−97.3	−88.3	−92.3	−115.6
Paris	654.0	618.9	607.7	625.9	617.1
Rome	958.5	784.1	743.7	803.0	832.0
Stockholm	−110.1	−74.9	−51.9	4.0	24.4
Relative percentage errors					
Amsterdam			19.91%	26.26%	*9.44%*
Berlin			41.15%	55.35%	*37.05%*
Madrid			15.35%	*10.33%*	16.29%
Oslo			69.27%	67.87%	*59.77%*
Paris			7.08%	*4.30%*	5.65%
Rome			22.42%	16.23%	*13.20%*
Stockholm			*52.83%*	103.64%	122.16%

Real and historical CAT for the period 1 January–31 March 2001 and estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.31 Out-of-sample comparison for a period of 6 months using the HDD index and the relative percentage errors

HDD/t6	Real	Historical	Alaton	Benth	WN
Amsterdam	1782.4	1727.9	1580.5	1534.0	1644.7
Berlin	1914.2	1794.8	1661.4	1622.9	1689.7
Madrid	1018.5	1064.0	1022.7	966.2	1038.0
Oslo	2618.6	2437.3	2325.8	2299.7	2362.8
Paris	1324.9	1339.8	1222.2	1195.0	1203.8
Rome	843.8	1035.1	1005.1	925.8	887.9
Stockholm	2379.2	2472.3	2327.9	2228.8	2194.0
Relative percentage errors					
Amsterdam			11.33%	13.94%	*7.73%*
Berlin			13.21%	15.22%	*11.73%*
Madrid			*0.41%*	5.13%	1.91%
Oslo			11.18%	12.18%	*9.77%*
Paris			*7.75%*	9.80%	9.14%
Rome			19.12%	9.72%	*5.23%*
Stockholm			*2.16%*	6.32%	7.78%

Real and historical HDDs for the period 1 January–30 June 2001 and estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.32 Out-of-sample comparison for a period of 6 months using the CAT index and the relative percentage errors

CAT/t6	Real	Historical	Alaton	Benth	WN
Amsterdam	1485.0	1549.0	1677.5	1724.0	1613.3
Berlin	1357.6	1523.0	1621.3	1667.8	1588.3
Madrid	2445.9	2350.1	2424.4	2499.5	2400.7
Oslo	648.9	829.2	932.2	958.3	895.2
Paris	2013.6	1973.7	2079.3	2116.0	2104.9
Rome	2568.3	2360.3	2408.5	2518.7	2557.9
Stockholm	890.0	790.6	930.1	1029.2	1064.0
Relative percentage errors					
Amsterdam			12.96%	16.09%	*8.64%*
Berlin			19.42%	22.85%	*16.99%*
Madrid			*0.88%*	2.19%	1.85%
Oslo			43.66%	47.68%	*37.96%*
Paris			*3.26%*	5.09%	4.53%
Rome			6.22%	1.93%	*0.40%*
Stockholm			*4.50%*	15.64%	19.55%

Real and historical CAT for the period 1 January – 30 June 2001 and estimated CAT using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.33 Out-of-sample comparison for a period of 12 months using the HDD index and the relative percentage errors

HDD/t12	Real	Historical	Alaton	Benth	WN
Amsterdam	2786.1	2914.0	2612.1	2490.4	2578.4
Berlin	3174.9	3131.1	2861.2	2744.3	2865.9
Madrid	1813.4	1788.7	1629.0	1515.5	1681.4
Oslo	4344.2	4249.3	4030.4	3899.6	4031.6
Paris	2239.3	2261.6	2016.5	1924.5	1949.6
Rome	1371.3	1563.9	1466.1	1323.8	1331.3
Stockholm	3820.7	4133.6	3838.3	3593.9	3620.8
Relative percentage errors					
Amsterdam			*6.25%*	10.61%	7.45%
Berlin			9.88%	13.56%	*9.73%*
Madrid			10.17%	16.43%	*7.28%*
Oslo			7.22%	10.23%	*7.20%*
Paris			*9.95%*	14.06%	12.94%
Rome			6.91%	3.46%	*2.92%*
Stockholm			*0.46%*	5.94%	5.23%

Real and historical HDDs for the period 1 January–31 December 2001 and estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.34 Out-of-sample comparison for a period of 12 months using the CAT index and the relative percentage errors

CAT/t12	Real	Historical	Alaton	Benth	WN
Amsterdam	3873.7	3732.4	3959.5	4095.5	3997.9
Berlin	3569.7	3652.0	3817.8	3958.1	3810.7
Madrid	5493.7	5495.4	5560.3	5749.5	5485.1
Oslo	2266.4	2370.1	2539.6	2670.4	2538.4
Paris	4615.8	4565.4	4735.1	4864.5	4839.9
Rome	5931.8	5682.7	5747.9	5998.2	5965.4
Stockholm	2867.1	2495.9	2731.7	2976.1	2949.2
Relative percentage errors					
Amsterdam			*2.21%*	5.73%	3.21%
Berlin			6.95%	10.88%	*6.75%*
Madrid			1.21%	4.66%	*0.16%*
Oslo			12.05%	17.83%	*12.00%*
Paris			*2.58%*	5.39%	4.86%
Rome			3.10%	1.12%	*0.57%*
Stockholm			4.72%	3.80%	*2.86%*

Real and historical CAT for the period 1 January–31 December 2001 and estimated CAT using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Over the 5 different periods, our method gives the best results in 17 times for the HDD index and 17 times for the CAT. On the other hand, the Alaton method causes the smallest errors in 11 cases for both indices, while the Benth model in 7 and 8 cases for the HDD and CAT indices, respectively. Observing Tables 5.25, 5.26, 5.27, 5.28, 5.29, 5.30, 5.31, 5.32, 5.33 and 5.34, we conclude that the results for the HDD and the CAT index are the same. Moreover, we observe that our proposed method gives almost always better results for the following cities: Berlin, Oslo, and Rome. On the other hand, Alaton method performs better in Stockholm. Finally, a closer inspection of Tables 5.25, 5.26, 5.27, 5.28, 5.29, 5.30, 5.31, 5.32, 5.33 and 5.34 reveals that the forecasts of the Benth model deteriorate as the forecast window increases.

Next, our model is validated using the 1-day-ahead forecasts over 5 different periods. From Tables 5.35, 5.36, 5.37, 5.38, 5.39, 5.40, 5.41, 5.42, 5.43 and 5.44, the 1-day-ahead out-of-sample forecasts for the HDD and CAT indices over a period of 1, 2, 3, 6, and 12 months are presented. The first column corresponds to the real index over the specified index, while the second column to the average historical mean. The next three columns corresponds to the three models, Alaton, Benth, and the proposed one using WNs. The second panel of each table corresponds to the relative absolute percentage errors for each method.

Over the 5 different periods, our method gives the best results in 23 times for the HDD index and 24 times for the CAT. On the other hand, the Alaton method causes the smallest errors only in 7 cases for both indices, while the Benth model only in 5 and 4 cases for the HDD and CAT indices, respectively. Observing Tables 5.35, 5.36, 5.37, 5.38, 5.39, 5.40, 5.41, 5.42, 5.43 and 5.44, we conclude that the results for the HDD and the CAT index are the same.

Table 5.35 Day-ahead comparison for a period of 1 month using the HDD index and the relative percentage errors

HDD/t1	Real	Historical	Alaton	Benth	WN
Amsterdam	463.6	449.5	460.4	458.3	463.8
Berlin	522.4	517.9	524.8	523.0	523.8
Madrid	338.8	366.3	341.9	340.8	338.5
Oslo	646.7	630.5	654.9	650.6	651.0
Paris	378.6	394.7	381.3	379.9	380.2
Rome	260.1	322.8	272.1	270.6	265.2
Stockholm	555.7	602.1	565.5	562.7	556.8
Relative percentage errors					
Amsterdam			0.69%	1.14%	*0.04%*
Berlin			0.46%	*0.11%*	0.27%
Madrid			0.90%	0.59%	*0.10%*
Oslo			1.27%	*0.60%*	0.67%
Paris			0.72%	*0.35%*	0.41%
Rome			4.60%	4.04%	*1.95%*
Stockholm			1.76%	1.26%	*0.20%*

Real and historical HDDs for the period 1–31 January 2001 and 1-day-ahead estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.36 Day-ahead comparison for a period of 1 month using the CAT index and the relative percentage errors

CAT/t1	Real	Historical	Alaton	Benth	WN
Amsterdam	94.4	108.5	97.6	99.7	94.2
Berlin	35.6	40.1	33.2	35.0	34.2
Madrid	219.2	191.7	216.2	217.2	219.5
Oslo	−88.7	−72.5	−96.9	−92.5	−93.0
Paris	179.4	163.3	176.7	178.1	177.9
Rome	297.9	235.2	285.9	287.4	292.8
Stockholm	2.3	−44.1	−7.5	−4.7	1.2
Relative percentage errors					
Amsterdam			3.40%	5.58%	*0.19%*
Berlin			6.72%	1.64%	*3.92%*
Madrid			1.39%	0.91%	*0.15%*
Oslo			9.29%	*4.33%*	4.86%
Paris			1.53%	*0.74%*	0.86%
Rome			4.02%	3.53%	*1.71%*
Stockholm			425.22%	304.50%	*48.43%*

Real and historical CAT for the period 1–31 January 2001 and 1-day-ahead estimated CAT using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.37 Day-ahead comparison for a period of 2 months using the HDD index and the relative percentage errors

HDD/t2	Real	Historical	Alaton	Benth	WN
Amsterdam	840.1	846.8	836.6	832.7	838.5
Berlin	968.3	962.6	966.2	961.4	964.46
Madrid	602.7	642.5	610.1	607.0	599.8
Oslo	1300.4	1195.7	1282.9	1272.7	1280.6
Paris	699.7	733.6	704.0	700.6	701.2
Rome	518.2	602.3	535.4	530.5	523.4
Stockholm	1147.7	1162.0	1142.4	1134.0	1133.0
Relative percentage errors					
Amsterdam			0.42%	0.88%	*0.19%*
Berlin			*0.22%*	0.72%	0.40%
Madrid			1.23%	0.72%	*0.49%*
Oslo			*1.35%*	2.13%	1.52%
Paris			0.61%	*0.12%*	0.21%
Rome			3.31%	2.36%	*1.00%*
Stockholm			*0.46%*	1.19%	1.28%

Real and historical HDDs for the period 1 January–28 February 2001 and 1-day-ahead estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.38 Day-ahead comparison for a period of 2 months using the CAT index and the relative percentage errors

CAT/t2	Real	Historical	Alaton	Benth	WN
Amsterdam	221.9	215.3	225.4	229.3	223.5
Berlin	93.7	99.4	95.8	100.7	97.545
Madrid	459.3	419.5	451.9	455.0	462.2
Oslo	−238.4	−133.7	−220.9	−210.7	−218.6
Paris	362.3	328.4	358.0	361.4	360.8
Rome	543.8	459.7	526.6	531.6	538.6
Stockholm	−85.7	−100.0	−80.4	−72.0	−71.0
Relative percentage errors					
Amsterdam			1.59%	3.34%	*0.72%*
Berlin			*2.28%*	7.42%	4.10%
Madrid			1.61%	0.94%	*0.64%*
Oslo			*7.35%*	11.63%	8.32%
Paris			1.19%	*0.24%*	0.41%
Rome			3.16%	2.25%	*0.96%*
Stockholm			*6.23%*	15.94%	17.12%

Real and historical CAT for the period 1 January–28 February 2001 and 1-day-ahead estimated CAT using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.39 Day-ahead comparison for a period of 3 months using the HDD index and the relative percentage errors

HDD/t3	Real	Historical	Alaton	Benth	WN
Amsterdam	1251.9	1199.7	1239.2	1232.8	1250.6
Berlin	1429.9	1364.1	1415.7	1408.7	1422.0
Madrid	792.5	832.4	812.2	808.3	795.7
Oslo	1907.3	1717.3	1875.1	1861.1	1876.1
Paris	966.0	1001.1	975.4	971.0	971.4
Rome	661.8	836.3	699.7	693.5	677.9
Stockholm	1730.1	1694.9	1716.1	1704.7	1710.4
Relative percentage errors					
Amsterdam			1.01%	1.53%	*0.10%*
Berlin			0.99%	1.48%	*0.55%*
Madrid			2.49%	1.99%	*0.41%*
Oslo			1.69%	2.42%	*1.64%*
Paris			0.97%	*0.52%*	0.56%
Rome			5.73%	4.79%	*2.44%*
Stockholm			*0.81%*	1.47%	1.14%

Real and historical HDDs for the period 1 January–31 March 2001 and 1-day-ahead estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.40 Day-ahead comparison for a period of 3 months using the CAT index and the relative percentage errors

CAT/t3	Real	Historical	Alaton	Benth	WN
Amsterdam	368.1	420.3	380.8	387.2	369.4
Berlin	190.1	255.9	204.3	211.3	198.0
Madrid	827.5	787.7	807.8	811.7	824.3
Oslo	−287.3	−97.3	−255.1	−241.1	−256.1
Paris	654.0	618.9	644.6	649.0	648.6
Rome	958.5	784.1	920.3	926.5	942.1
Stockholm	−110.1	−74.9	−96.1	−84.7	−90.4
Relative percentage errors					
Amsterdam			3.46%	5.20%	*0.36%*
Berlin			7.45%	11.17%	*4.16%*
Madrid			2.38%	1.91%	*0.39%*
Oslo			11.22%	16.07%	*10.88%*
Paris			1.43%	*0.76%*	0.82%
Rome			3.98%	3.34%	*1.71%*
Stockholm			*12.75%*	23.03%	17.92%

Real and historical CAT for the period 1 January–31 March 2001 and 1-day-ahead estimated CAT using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.41 Day-ahead comparison for a period of 6 months using the HDD index and the relative percentage errors

HDD/t6	Real	Historical	Alaton	Benth	WN
Amsterdam	1782.4	1727.9	1749.4	1736.2	1783.0
Berlin	1914.2	1794.8	1867.5	1850.3	1893.0
Madrid	1018.5	1064.0	1013.8	1004.3	1015.8
Oslo	2618.6	2437.3	2566.4	2543.2	2570.0
Paris	1324.9	1339.8	1304.3	1291.1	1320.7
Rome	843.8	1035.1	870.6	858.4	851.4
Stockholm	2379.2	2472.3	2368.5	2349.2	2355.6
Relative percentage errors					
Amsterdam			1.85%	2.59%	*0.03%*
Berlin			2.44%	3.34%	*1.11%*
Madrid			0.46%	1.39%	*0.27%*
Oslo			1.99%	2.88%	*1.86%*
Paris			1.55%	2.55%	*0.32%*
Rome			3.17%	1.73%	*0.91%*
Stockholm			*0.45%*	1.26%	0.99%

Real and historical HDDs for the period 1 January–30 June 2001 and 1-day-ahead estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.42 Day-ahead comparison for a period of 6 months using the CAT index and the relative percentage errors

CAT/t6	Real	Historical	Alaton	Benth	WN
Amsterdam	1485.0	1549.0	1513.8	1526.7	1482.9
Berlin	1357.6	1523.0	1402.5	1417.3	1377.0
Madrid	2445.9	2350.1	2443.1	2453.8	2444.8
Oslo	648.9	829.2	698.6	720.6	695.6
Paris	2013.6	1973.7	2021.9	2033.7	2011.4
Rome	2568.3	2360.3	2540.8	2557.5	2561.2
Stockholm	890.0	790.6	896.7	916.0	911.3
Relative percentage errors					
Amsterdam			1.94%	2.81%	*0.14%*
Berlin			3.31%	4.40%	*1.43%*
Madrid			0.11%	0.32%	*0.04%*
Oslo			7.65%	11.04%	*7.19%*
Paris			0.41%	1.00%	*0.11%*
Rome			1.07%	0.42%	*0.28%*
Stockholm			*0.75%*	2.92%	2.39%

Real and historical CAT for the period 1 January–30 June 2001 and 1-day-ahead estimated CAT using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.43 Day-ahead comparison for a period of 12 months using the HDD index and the relative percentage errors

HDD/t12	Real	Historical	Alaton	Benth	WN
Amsterdam	2786.1	2914.0	2754.7	2725.4	2781.5
Berlin	3174.9	3131.1	3114.6	3079.8	3142.8
Madrid	1813.4	1788.7	1776.1	1752.1	1805.8
Oslo	4344.2	4249.3	4279.6	4227.6	4274.6
Paris	2239.3	2261.6	2189.0	2158.3	2,219
Rome	1371.3	1563.9	1379.8	1352.7	1369.2
Stockholm	3820.7	4133.6	3809.5	3766.0	3785.5
Relative percentage errors					
Amsterdam			1.13%	2.18%	*0.17%*
Berlin			1.90%	3.00%	*1.01%*
Madrid			2.06%	3.38%	*0.42%*
Oslo			*1.49%*	2.68%	1.60%
Paris			2.25%	3.62%	*0.91%*
Rome			0.62%	1.36%	*0.15%*
Stockholm			*0.29%*	1.43%	0.92%

Real and historical HDDs for the period 1 January–31 December 2001 and 1-day-ahead estimated HDDs using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.44 Day-ahead comparison for a period of 12 months using the CAT index and the relative percentage errors

CAT/t12	Real	Historical	Alaton	Benth	WN
Amsterdam	3873.7	3732.4	3889.0	3917.0	3877.3
Berlin	3569.7	3652.0	3614.3	3646.9	3592.8
Madrid	5493.7	5495.4	5507.6	5537.7	5496.7
Oslo	2266.4	2370.1	2321.2	2369.0	2328.7
Paris	4615.8	4565.4	4638.5	4668.8	4624.2
Rome	5931.8	5682.7	5900.4	5944.3	5922.7
Stockholm	2867.1	2495.9	2849.0	2892.2	2888.9
Relative percentage errors					
Amsterdam			0.39%	1.12%	*0.09%*
Berlin			1.25%	2.16%	*0.65%*
Madrid			0.25%	0.80%	*0.05%*
Oslo			*2.42%*	4.53%	2.75%
Paris			0.49%	1.15%	*0.18%*
Rome			0.53%	0.21%	*0.15%*
Stockholm			*0.63%*	0.88%	0.76%

Real and historical CAT for the period 1 January–31 December 2001 and 1-day-ahead estimated CAT using the Alaton's, Benth's, and proposed (WN) methods. The second panel corresponds to the relative absolute percentage errors

Table 5.45 Out-of-sample performance of the proposed, the Alaton and Benth models

	1-day-ahead	Out-of-sample	Total
Proposed	46 (66%)	35 (50%)	81 (58%)
Alaton	14 (20%)	21 (30%)	35 (25%)
Benth	10 (14%)	14 (20%)	24 (17%)

As it was expected, the absolute percentage errors are very small. Modeling the DATs using WNs, a very good estimate of the real indices is obtained. The absolute percentage error is less than 2.5% in all cases for the HDD index. The worst predicted estimated level of HDD index produced when approximating the 3-month HDD in Rome. In general, the proposed method produces the worst results when forecasting the DAT in Rome, while the best 1-day-ahead out-of-sample forecasts are obtained in Amsterdam, Madrid, and Paris with absolute percentage errors less than 0.2%, 0.5%, and 0.9%, respectively.

Our results corresponding to the CAT index are similar. Finally, as in the case of the out-of-sample forecasts, a closer inspection of Tables 5.35, 5.36, 5.37, 5.38, 5.39, 5.40, 5.41, 5.42, 5.43 and 5.44 reveals that the forecasts of the Benth model deteriorate as the forecast window increases.

Table 5.45 summarizes the results of the performance of each method. The proposed model outperformed the other two methods in 81 cases out of 140 resulting to a success ratio of 58%. On the other hand, the Alaton model gave the best results in only 35 cases with a success ratio of 25% and the Benth model in only 24 cases with a success ratio of 17%. Our results suggest that the proposed method significantly outperforms other methods previously proposed in literature.

The previous extensive analysis indicates that our results are very promising. Modeling the DAT using WA and WNs enhanced the fitting and the predictive accuracy of the temperature process. Modeling the DAT assuming a time-varying speed of mean reversion resulted to a better out-of-sample predictive accuracy of our model. The additional accuracy of the proposed model will have an impact on the accurate pricing of temperature derivatives.

In the proposed model, weather forecasts can easily be implemented. It is expected that the use of weather forecasts would further improve the forecasting ability of the WN model and hence the accuracy of the pricing of weather derivatives.

5.9 Conclusions

In this chapter, several temperature time series were studied in order to develop a model that describes the temperature evolution. A mean-reverting O–U with seasonal mean and variance and time-varying speed of mean reversion was proposed.

In the context of an O–U temperature process, the time dependence of the speed of the mean reversion $\kappa(t)$ was examined using a WN. By computing the derivative $d\tilde{T}(t+1)/d\tilde{T}(t)$ of the fitted WN model, daily values of $\kappa(t)$ were obtained. To our

knowledge, we are the first to do so. Our results indicate a strong time dependence in the daily variations of the values of $\kappa(t)$.

We compared the fit of the residuals with the normal distributions with two types of models. The first type was the proposed nonlinear nonparametric model where κ is a function of time. The second type of models were two linear models previously proposed and often cited in literature where κ is constant. It follows that by setting the speed of mean reversion to be a function of time, the accuracy of the pricing of temperature derivatives improves. Generally, in our model, a better fit was obtained. Only in two of the seven cities the normality hypothesis was rejected. Moreover, the framework presented for selecting the significant lags of the temperature completely removed the autocorrelation in the residuals. On the other hand, on both Alaton and Benth models, strong autocorrelation in the residuals was evident. Furthermore, the normality hypothesis was rejected in every city when the Benth model was applied.

Also, since small misspecifications in the dynamic models lead to large mispricing errors, an approach to estimate and calibrate the seasonal component in both the mean and variance using WA was presented. WA is an efficient and accurate tool that can be successfully used in the analysis of temperature data. WA was successfully applied in order to indentify and quantify all the statistical significant cycles in the seasonal mean and variance of DATs.

Finally, the proposed model was evaluated out-of-sample. The predictive power of the proposed model was evaluated using two out-of-sample forecasting methods. First, out-of-sample forecasts over a period and then 1-day-ahead forecasts over a period were estimated. Modeling the DAT using WA and WNs enhanced the fitting and the predictive accuracy of the temperature process. Modeling the DAT assuming a time-varying speed of mean reversion resulted to a model with better out-of-sample predictive accuracy. The additional accuracy of our model has an impact on the accurate pricing of temperature derivatives.

In order to obtain a better understanding of the distributions of the residuals, we expanded our analysis by fitting additional distributions. Of the four distributions (normal, hyperbolic, NIG, stable), the hyperbolic distribution provides a slightly better fit than the normal distribution. However, introducing a Lévy process in the temperature dynamics does not allow to find closed form solutions for the temperature derivatives. The increased complexity of the pricing formulas of the weather derivatives makes the use of the normal distribution more favorable.

References

Aczel AD (1993) Complete business statistics, 2nd edn. Irwin, Homewood

Alaton P, Djehince B, Stillberg D (2002) On modelling and pricing weather derivatives. Appl Math Finance 9:1–20

Alexandridis A (2010) Modelling and pricing temperature derivatives using wavelet networks and wavelet analysis. Ph.D. thesis, University of Macedonia, Thessaloniki

Bellini F (2005) The weather derivatives market: modelling and pricing temperature. Ph.D. thesis, University of Lugano, Lugano

Benth FE (2003) On arbitrage-free pricing of weather derivatives based on fractional Brownian motion. Appl Math Finance 10:303–324
Benth FE, Saltyte-Benth J (2005) Stochastic modelling of temperature variations with a view towards weather derivatives. Appl Math Finance 12(1):53–85
Benth FE, Saltyte-Benth J (2007) The volatility of temperature and pricing of weather derivatives. Quant Finance 7(5):553–561
Benth FE, Saltyte-Benth J, Koekebakker S (2007) Putting a price on temperature. Scand J Stat 34:746–767
Brody CD, Syroka J, Zervos M (2002) Dynamical pricing of weather derivatives. Quant Finance 2:189–198
Caballero R, Jewson S, Brix A (2002) Long memory in surface air temperature: detection modelling and application to eeather derivative valuation. Clim Res 21:127–140
Cao M, Wei J (1999) Pricing weather derivatives: an equilibrium approach. The Rotman Graduate School of Management, The University of Toronto. http://www.fields.utoronto.ca/programs/cim/financial_math/finance_seminar/99-00/cao_wei.pdf
Cao M, Wei J (2000) Pricing the weather. Risk weather risk special report, Energy Power Risk Manage 67–70
Cao M, Wei J (2003) Weather derivatives: a new class of financial instruments. University of Toronto. http://www.rotman.utoronto.ca/~wei/research/JAI.pdf
Cao M, Li A, Wei J (2004) Watching the weather report. Can Invest Rev (Summer) 27–33
Carmona R (1999) Calibrating degree day options. In: 3rd seminar on stochastic analysis, Random Field and Applications, Ecole Polytechnique de Lausanne
Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia
Dornier F, Queruel M (2000) Caution to the wind. Weather risk special report, Energy Power Risk Management 30–32
Dunis CL, Karalis V (2003) Weather derivative pricing and filling analysis for missing temperature data. Deriv Use Trad Regul 9(1):61–83
Geman H, Leonardi M-P (2005) Alternative approaches to weather derivatives pricing. Manag Finance 31(6):46–72
Hartigan JA, Hartigan PM (1985) The dip test of unimodality. Ann Stat 13:70–84
Jewson S, Brix A, Ziehmann C (2005) Weather derivative valuation: the Meteorological, Statistical, Financial and Mathematical Foundations. Cambridge University Press, Cambridge, UK
Koutsoyiannis D (2003) Climate change, the Hurst phenomenon, and hydrological statistics/Changement climatique, phénomène de Hurst et statistiques hydrologiques. Hydrol Sci J 48 (1):3–24
Lau KM, Weng HY (1995) Climate signal detecting using wavelet signal transform. Bull Am Meteorol Soc 76:2391–2401
Nelken I (2000) Weather derivatives – pricing and hedging. http://www.supercc.com. Accessed 3 Jan 2000
Zapranis A, Alexandridis A (2006) Wavelet analysis and weather derivatives pricing. Paper presented at the 5th Hellenic Finance and Accounting Association (HFAA), Thessaloniki, 15–16 Dec 2006
Zapranis A, Alexandridis A (2008) Modelling temperature time dependent speed of mean reversion in the context of weather derivative pricing. Appl Math Finance 15(4):355–386
Zapranis A, Alexandridis A (2009a) Modeling and forecasting CAT and HDD indices for weather derivative pricing. In: EANN 2009, London, 27–29 Aug 2009
Zapranis A, Alexandridis A (2009b) Weather derivatives pricing: modelling the seasonal residuals variance of an Ornstein-Uhlenbeck temperature process with neural networks. Neurocomputing 73:37–48
Zapranis A, Alexandridis A (2011) Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing. Neural Comput Appl 20(6):787–801. doi:10.1007/s00521-010-0494-1
Zhang Q (1997) Using wavelet network in nonparametric estimation. IEEE Trans Neural Netw 8(2):227–236

Chapter 6
Pricing Temperature Derivatives

6.1 Introduction

The analysis that was performed in the previous chapter indicates that assuming a normal distribution is justified. In general, the normal distributions fit the final residuals after dividing out the seasonal variance reasonably well, while only in two of the seven cities, the normality hypothesis was rejected. Expanding our research, three more distributions were tested: the hyperbolic, NIG, and stable distribution. Our results indicate that the hyperbolic distribution provides the best fit to the residuals. The Anderson–Darling statistic and the Kolmogorov distance had the smallest value in every city when a hyperbolic distribution was used. In this chapter, the pricing formulas of various temperature derivatives will be presented first under the assumption of normal distribution and then under the assumption of a Lévy motion noise. More precisely, the pricing formulas for the following indices will be derived: CAT, AccHDD, AccCDD, and Pacific Rim.

When the market is complete, a unique risk-neutral probability measure $Q \sim P$ can be obtained, where P is the real-world probability measure. This change of measure turns the stochastic process into a martingale. Hence, financial derivatives can be priced under the risk-neutral measure by the discounted expectation of the derivative payoff.

The weather market is an incomplete market in the sense that the underlying weather derivative cannot be stored or traded. Moreover, the market is relatively illiquid. In principle, (extended) risk-neutral valuation can be still carried out in incomplete markets (Xu et al. 2008). However, in incomplete markets, a unique price cannot be obtained using the no-arbitrage assumption. In other words, many equivalent martingales exist, and as a result, only bounds for prices on contingent claims can be provided (Xu et al. 2008).

The change of measure from the real world to the risk-neutral world under the dynamics of a BM can be performed using the Girsanov theorem (or the Esscher transform for a jump process). The Girsanov theorem tells us how a stochastic process changes under changes in the measure. Then, the discounted expected

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_6,

payoff of the various weather contracts can be estimated. However, in order to estimate the expected payoff of each derivative, the solution of the stochastic differential equation that describes the temperature dynamics must be solved. This can be done by applying the Itô Lemma when a BM is considered or the Itô formula for semimartingales when a Lévy motion is considered.

The rest of the chapter is organized as follows. In Sect. 6.2, the organized market is described, and the various temperature indices are presented. The solution of the stochastic differential equation that describes the dynamics of the temperature is presented in Sect. 6.3. In Sect. 6.4, the pricing formulas of the weather derivatives on various temperature indices under the assumption of the normal distributions are presented. More precisely, in Sect. 6.4.1, the pricing formulas for the CAT and Pacific Rim indices are derived, while the prices of HDDs and CDDs future contracts are presented in Sect. 6.4.2. In Sect. 6.5, the prices of the weather derivatives are derived under the assumption of a Lévy noise process. Next, the importance of the market price of risk is analyzed and discussed in Sect. 6.6. Finally, in Sect. 6.7, we conclude.

6.2 Temperature Derivatives Traded in the CME

The list of traded contracts on the weather derivatives market is extensive and constantly evolving. CME offers various weather future and option contracts. They are index-based products geared to average seasonal and monthly weather in 46 cities[1] around the world – 24 in the USA, 10 in Europe, 6 in Canada, 3 Australian, and 3 in Japan.

In Europe, CME weather contracts for the summer months are based on an index of CAT. The CAT index is the sum of the DATs over the contract period. The average temperature is measured as the simple average of the minimum and maximum temperature over 1 day. The value of a CAT index for the time interval $[\tau_1, \tau_2]$ is given by the following expression:

$$\text{CAT} = \int_{\tau_1}^{\tau_2} T(s)ds, \tag{6.1}$$

where the temperature is measured in degrees Celsius. In London, one CAT index future contract costs £20 per index point, while it costs €20 per index unit in all other European locations. CAT contracts have monthly or seasonal duration. CAT futures and options are traded on the following months: May, June, July, August, September, April, and October.

In the USA, Canada, and Australia, CME weather derivatives are based on the HDD or CDD index. A HDD is the number of degrees by which the daily

[1] The number of cities that the CME trades weather contracts at the end of 2009.

temperature is below a base temperature, and a CDD is the number of degrees by which the daily temperature is above the base temperature, that is,

$$\text{Daily HDD} = \max(0, \text{base temperature} - \text{daily average temperature}),$$
$$\text{Daily CDD} = \max(0, \text{daily average temperature} - \text{base temperature}).$$

The base temperature is usually 65 °F in the USA and 18 °C in Europe and Japan. HDDs and CDDs are usually accumulated over a month or over a season. CME also trades HDDs contracts for the European cities. In USA, one HDD or CDD index future contract costs $20 per index point. Contacts on the following months can be found: November, December, January, February, March, October, and April.

For the three Japanese cities, weather derivatives are based on the Pacific Rim index. The Pacific Rim index is simply the average of the CAT index over the specific time period:

$$\text{PAC} = \frac{1}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} T(s)ds. \tag{6.2}$$

However, in PAC the average temperature is calculated as the average of 24 hourly measurements. In Asia-Pacific area, one PAC index future contract costs ¥2,500 per index point. The contracts are traded for all 12 calendar months.

Finally, in Australia contracts are based on HDD and CDD indices, while on Canada, the temperature contracts are base on HDD, CDD, and CAT indices. In Table 6.1, the contract specifications, the months that are traded, and the contract size of the temperature contracts traded in the CME are presented.

6.3 Solving the Temperature Stochastic Differential Equation

In the previous chapter, a stochastic differential equation was proposed in order to model the dynamics of the daily average temperature. The proposed model can be used in order to derive the theoretical prices of the temperature contracts presented in the previous section. However, in order to proceed to the pricing of the temperature contracts, the solution of the stochastic differential equation must be derived.

The pricing of these contracts using daily models is not a straightforward process. In Alaton et al. (2002), a numerical approach was adapted in order to find the fair price of HDD option contract; however, strong simplifications were made that significantly reduced the complexity of the pricing formulas. In Brody et al. (2002) and later in Benth (2003), the price of various temperature options was estimated under the assumption that the driving noise process of the temperature is a FBM. In a more recent paper, Benth and Saltyte-Benth (2005) estimate the prices of a CAT future and option contracts under the assumption of a Lévy noise process. More precisely, Benth and Saltyte-Benth (2005) propose that the residuals follow the generalized hyperbolic distribution. Similarly, Bellini (2005) presents the

Table 6.1 Specifications of temperature contracts traded in the CME

Contracts	Contract months	Contract size
US cooling monthly	May, Jun, Jul, Aug, Sep plus Apr and Oct	$20.00
US cooling seasonal	Minimum of two and maximum of seven consecutive calendar months, Apr–Oct	$20.00
US heating monthly	Nov, Dec, Jan, Feb, Mar plus Oct and Apr	$20.00
US heating seasonal	Minimum of two and maximum of seven consecutive calendar months, Oct–Apr	$20.00
US weekly weather	All weeks (Monday–Friday)	$100.00
Canada CAT monthly	May, Jun, Jul, Aug, Sep plus Apr and Oct	$20.00
Canada CAT seasonal	Minimum of two and maximum of seven consecutive calendar months, Apr–Oct	$20.00
Canada cooling monthly	May, Jun, Jul, Aug, Sep plus Apr and Oct	$20.00
Canada cooling seasonal	Minimum of two and maximum of seven consecutive calendar months, Apr–Oct	$20.00
Canada heating monthly	Nov, Dec, Jan, Feb, Mar plus Oct and Apr	$20.00
Canada heating seasonal	Minimum of two and maximum of seven consecutive calendar months, Oct–Apr	$20.00
Europe CAT monthly	May, Jun, Jul, Aug, Sep plus Apr and Oct	€ 20.00
Europe CAT seasonal	Minimum of two and maximum of seven consecutive calendar months, Apr–Oct	€ 20.00
Europe heating monthly	Nov, Dec, Jan, Feb, Mar plus Oct and Apr	€ 20.00
Europe heating seasonal	Minimum of two and maximum of seven consecutive calendar months, Oct–Apr	€ 20.00
Asia-Pacific monthly	All 12 calendar months	¥2,500.00
Asia-Pacific seasonal	Minimum of two and maximum of seven consecutive calendar months	¥2,500.00
Australia cooling monthly	Nov, Dec, Jan, Feb, Mar plus Oct and Apr	$20.00
Australia cooling seasonal	Minimum of two and maximum of seven consecutive calendar months, Oct–Apr	$20.00
Australia heating monthly	May, Jun, Jul, Aug, Sep plus Apr and Oct	$20.00
Australia heating seasonal	Minimum of two and maximum of seven consecutive calendar months, Apr–Oct	$20.00

pricing of HDDs and CDDs contracts under the assumption of a Lévy noise process where the residuals follow the hyperbolic distributions. More recently, Benth and Saltyte-Benth (2007) presented the pricing formulas of derivatives on various temperature indices under the normality assumption. More precisely, prices of futures and options of the following indices were derived: the CAT, Pacific Rim, HDDs, and CDDs indices. In Benth et al. (2008), the temperature dynamics were modeled by a CAR(p) process first introduced by Brockwell and Marquardt (2005).

Under the normality assumption, pricing formulas for the CAT, HDDs, and CDDs indices were presented. In Zapranis and Alexandridis (2008), the price of CAT futures were derived when the speed of mean reversion is a time-varying function.

In Geman (1999) and Jewson et al. (2005), various pricing approaches were presented. These approaches were derived either from daily or index models or actuarial-based methods. Davis (2001) prices weather derivatives by marginal value using a modified Black–Scholes equation, while Platen and West (2005) suggest a fair pricing approach based on an equilibrium method. On the other hand, Garman et al. (2000) introduce MC to price weather derivatives, while Xu et al. (2008) apply an indifference pricing approach for weather derivatives that are traded OTC.

Thus far, we have modeled the temperature using an O–U process with time-varying speed of mean reversion function. We have also used WA to identify and filter out the seasonal component. Moreover, we have shown that the coefficient a in the nonlinear AR model (5.27) is characterized by significant daily variation. Recall that parameter a is connected to our initial model (5.15) as $a = 1 + \kappa$, where κ is the speed of mean reversion. It follows that the assumption of a constant mean reversion parameter introduces significant error in the pricing of weather derivatives. In this chapter, we give the pricing formulas for a future and an option contract written on the indices presented above that incorporates the time dependency of the speed of the mean reversion parameter. First, we rewrite our model that describes the temperature dynamics and solve the stochastic differential equation using the Itô Lemma.

Recall that if the stochastic process of a variable x is known, then Itô Lemma gives us the stochastic process that a variable $G(x, t)$ follows.

Itô Lemma. Let x be variable that follows an Itô process:

$$dx = a(x, t)dt + b(x, t)dz, \tag{6.3}$$

where dz is a Wiener process. The variable x has drift rate a and variance b^2. Then, the process $G(x, t)$ follows also an Itô process:

$$dG = \left(\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right)dt + \frac{\partial G}{\partial x}bdz \tag{6.4}$$

with drift rate

$$\left(\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right) \tag{6.5}$$

and variance

$$\left(\frac{\partial G}{\partial x}b\right)^2. \tag{6.6}$$

Hence, the following proposition follows.

Proposition 6.1 *If the* DAT *follows a mean-reverting* O–U *process with time-varying speed of mean reversion and seasonal mean and variance*:

$$dT(t) = dS(t) + \kappa(t)(T(t) - S(t))dt + \sigma(t)dB(t), \tag{6.7}$$

an explicit solution can be derived from the Itô formula:

$$T(t) = S(t) + e^{\int_0^t \kappa(u)du}(T(0) - S(0)) + e^{\int_0^t \kappa(u)du} \int_0^t \sigma(s)e^{-\int_0^s \kappa(u)du} dB(s). \tag{6.8}$$

Proof Let us rewrite (6.7) as

$$d\tilde{T}(t) = \kappa(t)\tilde{T}(t)dt + \sigma(t)dB(t),$$

where $\tilde{T}(t) = (T(t) - S(t))$. To solve the above stochastic equation, the following transformation is convenient:

$$G(\tilde{T}, t) = e^{-\int_0^t \kappa(u)du}\tilde{T}(t).$$

Note that both processes X and Y satisfy the same initial condition:

$$\tilde{T}(0) = G(0).$$

Applying the Itô Lemma with

$$G_t = -\kappa(t)e^{-\int_0^t \kappa(u)du}\tilde{T},$$

$$G_{\tilde{T}} = e^{-\int_0^t \kappa(u)du},$$

$$G_{\tilde{T}\tilde{T}} = 0,$$

and

$$a = \kappa(t)\tilde{T}(t),$$

$$b = \sigma(t),$$

we have that

$$dG(t) = \left(\kappa(t)e^{-\int_0^t \kappa(u)du}\tilde{T}(t) - \kappa(t)e^{-\int_0^t \kappa(u)du}\tilde{T}(t)\right)dt + \sigma(t)e^{-\int_0^t \kappa(u)du}dB(t)$$

which reduces to

$$dG(t) = \sigma(t)e^{-\int_0^t \kappa(u)du} dB(t).$$

Integrating the above equation in the interval $[0, t]$, we have that

$$G(t) - G(0) = \int_0^t \sigma(t)e^{-\int_0^s \kappa(u)du} dB(t),$$

and by replacing G, we have that

$$e^{-\int_0^t \kappa(u)du}\tilde{T}(t) - \tilde{T}(0) = \int_0^t \sigma(t)e^{-\int_0^s \kappa(u)du} dB(t).$$

By rearranging, we have that

$$\tilde{T}(t) = e^{\int_0^t \kappa(u)du}\tilde{T}(0) + e^{\int_0^t \kappa(u)du} \int_0^t \sigma(t)e^{-\int_0^s \kappa(u)du} dB(t).$$

Since $\tilde{T}(t) = T(t) - S(t)$,

$$T(t) - S(t) = e^{\int_0^t \kappa(u)du}(T(0) - S(0)) + e^{\int_0^t \kappa(u)du} \int_0^t \sigma(t)e^{-\int_0^s \kappa(u)du} dB(t).$$

Finally, by rearranging, we prove the proposition

$$T(t) = S(t) + e^{\int_0^t \kappa(u)du}(T(0) - S(0)) + e^{\int_0^t \kappa(u)du} \int_0^t \sigma(t)e^{-\int_0^s \kappa(u)du} dB(t).$$

□

Note that we assumed that driving noise process of the temperature model follows a Brownian motion $B(t)$. Later in the chapter, the BM will be replaced by a Lévy noise process. As we will see in the next section, the result of Proposition 6.1 will be used directly in the derivation of the theoretical prices of temperature derivatives written on CAT, HDD, CDD, and PAC indices.

6.4 Pricing Under the Normal Assumption

In this section, the pricing formulas of the weather derivatives on various temperature indices under the assumption of the normal distribution are presented. More precisely, the pricing formulas of futures and options on futures for the CAT, AccHDD, AccCDD, and Pacific Rim indices are derived.

6.4.1 CAT and Pacific Rim: Futures and Options

Our aim is to give a mathematical expression for the CAT future price. The weather derivatives market is a classical incomplete market. Moreover, the market is relatively illiquid. In principle, risk-neutral valuation can be still carried out in incomplete markets (Xu et al. 2008). However, in incomplete markets, a unique price cannot be obtained using the no-arbitrage assumption. Temperature contracts are written on a temperature index which is not a tradable or storable asset. In order to derive the pricing formula, first we must find a risk-neutral probability measure $Q \sim P$, where all assets are martingales after discounting. In the case of weather derivatives, any equivalent measure Q is a risk-neutral probability. If Q is the risk-neutral probability and r is the constant compounding interest rate, then the arbitrage-free future price of a CAT contract at time $t \leq \tau_1 < \tau_2$ is given by

$$e^{-r(\tau_2 - t)} \mathrm{E}_Q \left[\int_{\tau 1}^{\tau_2} T(\tau) d\tau - F_{\mathrm{CAT}}(t, \tau_1, \tau_2) | \mathbf{F}_t \right] = 0, \tag{6.9}$$

and since $\mathrm{F}_{\mathrm{CAT}}$ is $\mathbf{F}_t$ adapted, we derive the price of a CAT future to be

$$F_{\mathrm{CAT}}(t, \tau_1, \tau_2) = \mathrm{E}_Q \left[\int_{\tau 1}^{\tau_2} T(\tau) d\tau | \mathbf{F}_t \right]. \tag{6.10}$$

Using Girsanov's theorem, under the equivalent measure Q, we have

$$W(t) = B(t) - \int_0^t \theta(u) du \tag{6.11}$$

or equivalently

$$dW(t) = dB(t) - \theta(t) dt, \tag{6.12}$$

and note that $\sigma(t)$ is bounded away from zero. Hence, by combining (6.7) and (6.12), the stochastic process of the temperature in the risk-neutral probability Q^θ is

$$dT(t) = dS(t) + (\kappa(t)(T(t) - s(t)) + \sigma(t)\theta(t))dt + \sigma(t) dW(t), \tag{6.13}$$

where $\theta(t)$ is a real-valued measurable and bounded function denoting the market price of risk. The market price of risk can be calculated from historical data. More specifically, $\theta(t)$ can be calculated by looking at the market price of contracts. The value that makes the price of the model fit the market price is the market price of risk. Using the Itô formula, the solution of (6.13) under Q^θ is

$$\begin{aligned} T(t) = S(t) + e^{\int_0^t \kappa(u) du}(T(0) - S(0)) + e^{\int_0^t \kappa(u) du} \int_0^t \sigma(s)\theta(s) e^{-\int_0^s \kappa(u) du} ds \\ + e^{\int_0^t \kappa(u) du} \int_0^t \sigma(s) e^{-\int_0^s \kappa(u) du} dW(s). \end{aligned} \tag{6.14}$$

The proof of (6.14) is similar to the proof of Proposition 6.1. Note that Q is the risk-neutral probability measure, where $Q \sim P$, while Q^θ is a subclass of these probabilities defined by the Girsanov theorem. Since we restrict our attention in these probabilities, in order to simplify the notation in the remaining of the chapter, we will define this subclass of probabilities also with the same letter Q.

Replacing expression (6.14) in (6.10), we find the price of a future contract on the CAT index at time t, where $t \leq \tau_1 < \tau_2$.

Proposition 6.2 *The* CAT *future price for* $t \leq \tau_1 < \tau_2$ *is given by*

$$F_{\text{CAT}}(t, \tau_1, \tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds|\mathbf{F}_t\right] = \int_{\tau_1}^{\tau_2} S(s)ds + I_1 + I_2, \tag{6.15}$$

where

$$I_1 = \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}\widetilde{T}(t)ds, \tag{6.16}$$

$$I_2 = \int_t^{\tau_1}\int_{\tau_1}^{\tau_2} e^{\int_0^s \kappa(z)dz}\sigma(u)\theta(u)e^{\int_u^0 \kappa(z)dz}dsdu + \int_{\tau_1}^{\tau_2}\int_u^{\tau_2} e^{\int_0^s \kappa(z)dz}\sigma(u)\theta(u)e^{\int_u^0 \kappa(z)dz}dsdu. \tag{6.17}$$

Proof From (6.10) and (6.14), we have

$$\mathbf{F}_{\text{CAT}}(t, \tau_1, \tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds|\mathbf{F}_t\right] = \int_{\tau_1}^{\tau_2} S(s)ds + \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \tilde{T}(s)ds|\mathbf{F}_t\right],$$

and using Itô isometry, we can interchange the expectation and the integral

$$\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \tilde{T}(s)ds|\mathbf{F}_t\right] = \int_{\tau_1}^{\tau_2} \mathrm{E}_Q\left[\tilde{T}(s)|\mathbf{F}_t\right]ds$$

$$= \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}\tilde{T}(t)ds + \int_{\tau_1}^{\tau_2}\int_t^s \sigma(u)\theta(u)e^{\int_u^s \kappa(z)dz}duds$$

$$= I_1 + I_2.$$

Hence,

$$I_1 = \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}\tilde{T}(t)ds$$

and

$$I_2 = \int_{\tau_1}^{\tau_2} \int_t^s \sigma(u)\theta(u) e^{\int_u^s \kappa(z)dz} duds = \int_{\tau_1}^{\tau_2} \int_t^{\tau_2} 1_{[t,s]}(u)\sigma(u)\theta(u) e^{\int_u^s \kappa(z)dz} duds,$$

where $1_{[t,s]}$ is zero outside the interval $[t, s]$. Then, we can change the order of the integrals,

$$= \int_t^{\tau_2} \int_{\tau_1}^{\tau_2} 1_{[t,s]}(u)\sigma(u)\theta(u) e^{\int_u^s \kappa(z)dz} dsdu.$$

Next, we split the outer integral in two parts:

$$= \int_t^{\tau_1} \int_{\tau_1}^{\tau_2} 1_{[t,s]}(u)\sigma(u)\theta(u) e^{\int_u^s \kappa(z)dz} dsdu + \int_{\tau_1}^{\tau_2} \int_{\tau_1}^{\tau_2} 1_{[t,s]}(u)\sigma(u)\theta(u) e^{\int_u^s \kappa(z)dz} dsdu.$$

The second part is zero when $s>u$. Hence, we can change the limits of the inner integral

$$= \int_t^{\tau_1} \int_{\tau_1}^{\tau_2} \sigma(u)\theta(u) e^{\int_u^s \kappa(z)dz} dsdu + \int_{\tau_1}^{\tau_2} \int_u^{\tau_2} \sigma(u)\theta(u) e^{\int_u^s \kappa(z)dz} dsdu$$

or, equivalently,

$$= \int_t^{\tau_1} \int_{\tau_1}^{\tau_2} e^{\int_0^s \kappa(z)dz} \sigma(u)\theta(u) e^{\int_u^0 \kappa(z)dz} dsdu + \int_{\tau_1}^{\tau_2} \int_u^{\tau_2} e^{\int_0^s \kappa(z)dz} \sigma(u)\theta(u) e^{\int_u^0 \kappa(z)dz} dsdu.$$

□

Proposition 6.2 gives the price of a CAT future at time $t \leq \tau_1 < \tau_2$. In other words, the price of a CAT future before the contract period. Hence, (6.15) corresponds to out-of-period valuation. In order to evaluate the future price inside the contract period, the above formula can be easily modified.

Proposition 6.3 *The* CAT *future price for* $\tau_1 \leq t \leq \tau_2$ *is given by*

$$F_{\text{CAT}}(t, \tau_1, \tau_2) = \int_{\tau_1}^{t} T(s)ds + F_{\text{CAT}}(t, t, \tau_2). \tag{6.18}$$

Proof We have that the CAT future price is

$$\begin{aligned} F_{\text{CAT}}(t, \tau_1, \tau_2) &= \text{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds | \mathbf{F}_t\right] \\ &= \text{E}_Q\left[\int_{\tau_1}^{t} T(s)ds + \int_t^{\tau_2} T(s)ds | \mathbf{F}_t\right] \end{aligned}$$

$$= \int_{\tau_1}^{t} T(s)ds + \mathrm{E}_Q\left[\int_{t}^{\tau_2} T(s)ds|F_t\right]$$
$$= \int_{\tau_1}^{t} T(s)ds + F_{\mathrm{CAT}}(t,t,\tau_2).$$

Note that the first term is known at time t since it refers to past temperatures, while the second term is stochastic.

□

Similarly, the in-period pricing formulas of the remaining indices can be easily extracted from the pricing formulas of the out-of-period valuation. Hence, the dynamics of the CAT future price under Q is given in the following proposition.

Proposition 6.4 *The dynamics of* $F_{\mathrm{CAT}}(t,\tau_1,\tau_2)$ *under the risk-neutral measure* Q *is*

$$dF_{\mathrm{CAT}}(t,\tau_1,\tau_2) = \Sigma_{\mathrm{CAT}}(t,\tau_1,\tau_2)dW(t), \tag{6.19}$$

where

$$\Sigma_{\mathrm{CAT}}(t,\tau_1,\tau_2) = \sigma(t)\int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}ds. \tag{6.20}$$

Proof $F_{\mathrm{CAT}}(t,\tau_1,\tau_2)$ is Q martingale; hence, the proposition follows after a direct application of the Itô formula. We focus only on the part $dW(t)$ since the drift part is zero. We have that

$$\frac{dF_{\mathrm{CAT}}}{dT} = \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}ds;$$

hence,

$$dF_{\mathrm{CAT}}(t,\tau_1,\tau_2) = \sigma(t)\int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}dsdW(t).$$

□

Using Proposition 6.4, the price of call option written on CAT futures can be estimated.

Proposition 6.5 *The price at time* $t \le \tau$ *of a call option written on a* CAT *future with strike price* K *at exercise time* $\tau \le \tau_1$ *is*

$$C_{\mathrm{CAT}}(t,\tau,\tau_1,\tau_2) = e^{-r(\tau-t)}\{(F_{\mathrm{CAT}}(t,\tau_1,\tau_2)-K)\Phi(d(t,\tau,\tau_1,\tau_2))$$
$$+\Phi'(d(t,\tau,\tau_1,\tau_2))\int_t^{\tau}\Sigma^2_{\mathrm{CAT}}(t,\tau_1,\tau_2)ds\Big\}, \tag{6.21}$$

where

$$d(t,\tau,\tau_1,\tau_2) = \frac{F_{\mathrm{CAT}}(t,\tau_1,\tau_2) - K}{\sqrt{\Sigma^2_{t,\tau}}} \tag{6.22}$$

and

$$\Sigma^2_{t,\tau} = \int_t^\tau \Sigma^2_{\mathrm{CAT}}(t,\tau_1,\tau_2)ds \tag{6.23}$$

and Φ *is the cumulative standard normal distribution function.*

Proof The option price by definition is given by

$$C_{\mathrm{CAT}}(t,\tau,\tau_1,\tau_2) = e^{-r(\tau-t)}\mathrm{E}_Q[\max(F_{\mathrm{CAT}}(\tau,\tau_1,\tau_2) - K, 0)|\mathbf{F}_t].$$

From Proposition 6.4, we have that the Q dynamics of the future price can be written as

$$F_{\mathrm{CAT}}(\tau,\tau_1,\tau_2) = F_{\mathrm{CAT}}(t,\tau_1,\tau_2) + \int_t^\tau \Sigma_{\mathrm{CAT}}(s,\tau_1,\tau_2)dW(s).$$

From this, it follows that $F_{\mathrm{CAT}}(\tau,\tau_1,\tau_2)$ conditioned on $F_{\mathrm{CAT}}(t,\tau_1,\tau_2)$ follows the normal distribution with mean $F_{\mathrm{CAT}}(t,\tau_1,\tau_2)$ and variance given by

$$\int_t^\tau \Sigma^2_{\mathrm{CAT}}(s,\tau_1,\tau_2)ds.$$

Hence,

$$\begin{aligned} &E_Q[\max(F_{\mathrm{CAT}}(\tau,\tau_1,\tau_2) - K, 0)|\mathbf{F}_t] \\ &\quad = E_Q\left[\max\left(F_{\mathrm{CAT}}(t,\tau_1,\tau_2) + \int_t^\tau \Sigma_{\mathrm{CAT}}(s,\tau_1,\tau_2)dW(s) - K, 0\right)|\mathbf{F}_t\right]. \end{aligned}$$

The price C_{CAT} follows by a straightforward calculation using the properties of the normal distribution.

□

As it was mentioned earlier, the Pacific Rim index is simply the average of the CAT index over the specific time period. Then, the arbitrage-free future price of a Pacific Rim contract at time $t \leq \tau_1 < \tau_2$ is given by

$$e^{-r(\tau_2-t)}\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \frac{1}{\tau_2-\tau_1} T(\tau)d\tau - F_{\mathrm{PAC}}(t,\tau_1,\tau_2)|\mathbf{F}_t\right] = 0, \tag{6.24}$$

and since F_{PAC} is $\mathbf{F}_t$ adapted, we derive the price of a PAC future to be

$$F_{\mathrm{PAC}}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\frac{1}{\tau_2-\tau_1}\int_{\tau_1}^{\tau_2} T(s)ds|\mathbf{F}_t\right]. \tag{6.25}$$

Observing (6.10) and (6.25), we conclude that

$$F_{\mathrm{PAC}}(t,\tau_1,\tau_2) = \frac{1}{\tau_2-\tau_1}F_{\mathrm{CAT}}(t,\tau_1,\tau_2) \tag{6.26}$$

and, similarly, that the price of call option written on a PAC future is given by

$$C_{\mathrm{PAC}}(t,\tau_1,\tau_2) = \frac{1}{\tau_2-\tau_1}C_{\mathrm{CAT}}(t,\tau_1,\tau_2). \tag{6.27}$$

6.4.2 HDD and CDD Indices: Futures and Options

Next, the pricing formulas for the CDDs and HDDs are presented. The AccCDD and AccHDD indices over a period $[\tau_1,\tau_2]$ are given by

$$\mathrm{HDD} = \int_{\tau_1}^{\tau_2} \max(c - T(s), 0)ds, \tag{6.28}$$

$$\mathrm{CDD} = \int_{\tau_1}^{\tau_2} \max(T(s) - c, 0)ds. \tag{6.29}$$

Hence, the pricing equations are similar for both indices. Our aim is to give a mathematical expression for the HDD future price. If Q is the risk-neutral probability and r is the constant compounding interest rate, then the arbitrage-free future price of a HDD contract at time $t \leq \tau_1 < \tau_2$ is given by

$$e^{-r(\tau_2-t)}\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(0, c - T(\tau))d\tau - F_{\mathrm{HDD}}(t,\tau_1,\tau_2)|\mathbf{F}_t\right] = 0, \tag{6.30}$$

and since F_{HDD} is $\mathbf{F}_t$ adapted, we derive the price of a HDD future to be

$$F_{\mathrm{HDD}}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(c - T(\tau), 0)d\tau|\mathbf{F}_t\right]. \tag{6.31}$$

Similarly, we have that the arbitrage-free future price of a CDD contract at time $t \leq \tau_1 < \tau_2$ is given by

$$F_{\mathrm{CDD}}(t,\tau_1,\tau_2)=\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\max(T(\tau)-c,0)d\tau|\mathbf{F}_t\right]. \tag{6.32}$$

Observing (6.10), (6.31), and (6.32), we have the following proposition.

Proposition 6.6 *The* CDD, HDD, *and* CAT *prices are linked by the following relation:*

$$F_{\mathrm{HDD}}(t,\tau_1,\tau_2)=c(\tau_2-\tau_1)-F_{\mathrm{CAT}}(t,\tau_1,\tau_2)+F_{\mathrm{CDD}}(t,\tau_1,\tau_2). \tag{6.33}$$

Proof We have that

$$\max(c-T(\tau),0)=c-T(\tau)+\max(T(\tau)-c,0).$$

Hence, by replacing the above relation to (6.31), we have that

$$\begin{aligned}
F_{\mathrm{HDD}}(t,\tau_1,\tau_2)&=\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\max(c-T(\tau),0)d\tau|\mathbf{F}_t\right]\\
&=\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}(c-T(\tau)+\max(T(\tau)-c,0))d\tau|\mathbf{F}_t\right]\\
&=\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}cd\tau|\mathbf{F}_t\right]-\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}T(\tau)d\tau|\mathbf{F}_t\right]\\
&\quad+\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\max(T(\tau)-c,0)d\tau|\mathbf{F}_t\right]\\
&=c(\tau_2-\tau_1)-F_{\mathrm{CAT}}(t,\tau_1,\tau_2)+F_{\mathrm{CDD}}(t,\tau_1,\tau_2).
\end{aligned}$$

□

Proposition 6.6 indicates that the pricing formulas of futures on CDD and HDD indices are similar. Hence, we can focus only on the pricing formulas of the CDD indices.

Proposition 6.7 *The* CDD *future price for* $0\le t\le\tau_1<\tau_2$ *is given by*

$$\begin{aligned}
F_{\mathrm{CDD}}(t,\tau_1,\tau_2)&=\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\max(T(s)-c)ds|F_t\right]\\
&=\int_{\tau_1}^{\tau_2}v(t,s)\Psi\left(\frac{m\left(t,s,e^{\int_t^s\kappa(z)dz}\tilde{T}(t)\right)}{v(t,s)}\right)ds,
\end{aligned} \tag{6.34}$$

where

$$m(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)) = S(s) + e^{\int_t^s \kappa(z)dz}\tilde{T}(t) + e^{\int_t^s \kappa(z)dz}\int_t^s \sigma(u)\theta(u)e^{-\int_t^u \kappa(z)dz}du - c, \tag{6.35}$$

$$v^2(t,s) = e^{2\int_t^s \kappa(z)dz}\int_t^s \sigma^2(u)e^{-2\int_t^u \kappa(z)dz}du, \tag{6.36}$$

and $\Psi(x) = x\Phi(x) + \Phi'(x)$, *where* Φ *is the cumulative standard normal distribution function.*

Proof From (6.31) and (6.14), we have that

$$F_{\text{CDD}}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(T(s) - c)ds|\mathbf{F}_t\right]$$

and using Itô isometry, we can interchange the expectation and the integral

$$\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(T(s) - c)ds|\mathbf{F}_t\right] = \int_{\tau_1}^{\tau_2} \mathrm{E}_Q[\max(T(s) - c)|\mathbf{F}_t]ds.$$

$T(s)$ is normally distributed under the probability measure Q with mean and variance given by

$$E_Q[T(s)|\mathbf{F}_t] = S(s) + e^{\int_t^s \kappa(z)dz}\tilde{T}(t) + e^{\int_t^s \kappa(z)dz}\int_t^s \sigma(u)\theta(u)e^{-\int_t^u \kappa(z)dz}du,$$

$$\mathrm{Var}_Q[T(s)|\mathbf{F}_t] = e^{2\int_t^s \kappa(z)dz}\int_t^s \sigma^2(u)e^{-2\int_t^u \kappa(z)dz}du.$$

Hence, $T(s) - c$ is normally distributed with mean given by $m(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t))$ and variance given by $v^2(t,s)$, and the proposition follows by standard calculations using the properties of the normal distribution.

□

Proposition 6.7 gives the price of a future CDD at time $t \leq \tau_1 < \tau_2$. In other words, the price of a future CDD before the contract period. Hence, (6.34) corresponds to out-of-period valuation. In order to evaluate the future price inside the contract period, the above formula can be easily modified.

Proposition 6.8 *The* CDD *future price for* $\tau_1 \leq t < \tau_2$ *is given by*

$$F_{\text{CDD}}(t,\tau_1,\tau_2) = \int_{\tau_1}^{t} \max(T(s) - c)ds + F_{\text{CDD}}(t,t,\tau_2). \tag{6.37}$$

Proof We have that the future price of a CDD is given by

$$\begin{aligned} F_{\mathrm{CDD}}(t,\tau_1,\tau_2) &= \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(T(s)-c)ds|\mathbf{F}_t\right] \\ &= \mathrm{E}_Q\left[\int_{\tau_1}^{t} \max(T(s)-c)ds + \int_{t}^{\tau_2} \max(T(s)-c)ds|\mathbf{F}_t\right] \\ &= \int_{\tau_1}^{t} \max(T(s)-c)ds + \mathrm{E}_Q\left[\int_{t}^{\tau_2} \max(T(s)-c)ds|\mathbf{F}_t\right] \\ &= \int_{\tau_1}^{t} \max(T(s)-c)ds + F_{\mathrm{CDD}}(t,t,\tau_2). \end{aligned}$$

Note that the first term is known at time t since it refers to past temperatures, while the second term is stochastic.

□

Similarly to the case of the CAT contracts, the dynamics of the CDD future price under Q is given in the following proposition.

Proposition 6.9 *The dynamics of* $F_{\mathrm{CDD}}(t,\tau_1,\tau_2)$ *for* $0 \le t \le \tau_1$ *under* Q *is given by*

$$dF_{\mathrm{CDD}}(t,\tau_1,\tau_2) = \Sigma_{\mathrm{CDD}}(t,\tau_1,\tau_2)dW(t), \tag{6.38}$$

where

$$\Sigma_{\mathrm{CDD}}(t,\tau_1,\tau_2) = \sigma(t)\int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}\Phi\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)\right)}{v(t,s)}\right)ds \tag{6.39}$$

and Φ *is cumulative standard normal distribution function.*

Proof $F_{\mathrm{CDD}}(t,\tau_1,\tau_2)$ is Q martingale; hence, the proposition follows after a direct application of the Itô formula. We focus only on the part $dW(t)$ since the drift part is zero. First, note that $v(t,s)$ does not depend on $T(t)$ and that

$$m'\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)\right) = \frac{dm\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)\right)}{dT} = e^{\int_t^s \kappa(z)dz}.$$

Also, substituting $\Psi'(x) = \Phi(x)$, we have that

$$\frac{dF_{\mathrm{CDD}}}{dT} = \int_{\tau_1}^{\tau_2} v(t,s)\Psi\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)\right)}{v(t,s)}\right)ds$$

$$= \int_{\tau_1}^{\tau_2} v(t,s)\Psi'\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)\right)}{v(t,s)}\right)\frac{m'\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)\right)v(t,s)}{v^2(t,s)}ds$$

$$= \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}\Phi\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)\right)}{v(t,s)}\right)ds.$$

Hence, we have that

$$dF_{\mathrm{CDD}}(t,\tau_1,\tau_2) = \sigma(t)\int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}\Phi\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)\right)}{v(t,s)}\right)dsdW(t).$$

In Proposition 6.9, the term $\Sigma_{\mathrm{CDD}}(t,\tau_1,\tau_2)$ represents the term structure of the volatility of CDD futures. Hence, the price of a call option on a CDD future can be derived. From Proposition 6.9, the price of a CDD future option can be estimated.

Proposition 6.10 *The price at time $t \le \tau$ of a call option written on a* HDD *future with strike price K at exercise time $\tau \le \tau_1$ is*

$$C_{\mathrm{CDD}}(t,\tau,t_1,t_2) = e^{-r(\tau-t)}E_Q\left[\max\left(\int_{\tau_1}^{\tau_2} v(t,s)Z\big(t,s,\tau,\tilde{T}(t)\big)ds - K, 0\right)\right], \quad (6.40)$$

where

$$Z\big(t,s,\tau,\tilde{T}(t)\big) = \tilde{\Psi}\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t) + \int_t^{\tau}\sigma(u)\theta(u)e^{\int_u^s \kappa(z)dz}du + \Sigma(s,t,\tau)Y\right) \quad (6.41)$$

and

$$\tilde{\Psi}(t,s,x) = \Psi\left(\frac{m(t,s,x)}{v(t,s)}\right), \quad (6.42)$$

and

$$\Sigma^2(s,t,\tau) = \int_t^{\tau}\sigma^2(u)e^{2\int_u^s \kappa(z)dz}du \quad (6.43)$$

and Y is a standard normal random variable.

Proof The option price is given as

$$C_{\mathrm{CDD}}(t,\tau,t_1,t_2) = e^{-r(\tau-t)}E_Q[\max(F_{\mathrm{CDD}}(\tau,t_1,t_2)-K,0)|\mathbf{F}_t],$$

we have that

$$\begin{aligned} F_{\mathrm{CDD}}(\tau,t_1,t_2) &= \int_{\tau_1}^{\tau_2} v(t,s)\tilde{\Psi}\left(t,s,e^{\int_\tau^s \kappa(z)dz}\tilde{T}(\tau)\right)ds \\ &= \int_{\tau_1}^{\tau_2} v(t,s)\tilde{\Psi}\left(t,s,e^{\int_t^s \kappa(z)dz}\tilde{T}(t)+\int_t^\tau \sigma(u)\theta(u)e^{\int_u^s \kappa(z)dz}du\right. \\ &\quad \left.+\int_t^\tau \sigma(u)e^{\int_u^s \kappa(z)dz}dW(u)\right)ds. \end{aligned}$$

The Itô integral inside the expectation is independent of $\mathbf{F}_t$ and has variance $\int_t^\tau \sigma^2(u)e^{2\int_u^s \kappa(z)dz}du$. Taking the conditional expectation yields the result.

□

6.5 Pricing Under the Assumption of a Lévy Noise Process

Under the assumption of a Lévy motion as the driving noise process, the stochastic differential equation that describes the DAT is a generalization of the proposed model (6.7). Hence, the DATs follow a mean-reverting O–U process with time-varying speed of mean reversion and seasonal mean and variance and a Lévy driving noise process given by

$$dT(t) = dS(t) + \kappa(t)(T(t)-S(t))dt + \sigma(t)dL(t), \tag{6.44}$$

where $L(t)$ is Lévy noise. Applying the Itô formula for semimartingales (Ikeda and Watanabe 1981), the explicit solution of (6.44) is obtained:

$$T(t) = S(t) + e^{\int_0^t \kappa(u)du}(T(0)-S(0)) + e^{\int_0^t \kappa(u)du}\int_0^t \sigma(t)e^{-\int_0^s \kappa(u)du}dL(t). \tag{6.45}$$

As in the case of the BM, we derive the price of a CAT future to be

$$F_{\mathrm{CAT}}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\int_{\tau 1}^{\tau_2} \mathrm{T}(\tau)d\tau|\mathbf{F}_t\right].$$

Proposition 6.11 *The cumulative temperature over the time interval* $[\tau_1, \tau_2]$ *is given by*

$$\int_{\tau_1}^{\tau_2} T(t)dt = \int_{\tau_1}^{\tau_2} S(t)dt + \int_{\tau_1}^{\tau_2} e^{\int_0^t \kappa(z)dz}(T(0) - S(0))dt + \int_0^{\tau_1}\int_{\tau_1}^{\tau_2} \sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t) + \int_{\tau_1}^{\tau_2}\int_t^{\tau_2} \sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t). \quad (6.46)$$

Proof We have that

$$\int_{\tau_1}^{\tau_2} T(t)dt = \int_{\tau_1}^{\tau_2} S(t)dt + \int_{\tau_1}^{\tau_2} e^{\int_0^t \kappa(z)dz}(T(0) - S(0))dt + \int_{\tau_1}^{\tau_2}\int_0^t \sigma(t)e^{\int_s^t \kappa(z)dz}dL(t)dt.$$

Focusing on the last integral, we have that

$$\begin{aligned}
&\int_{\tau_1}^{\tau_2}\int_0^t \sigma(t)e^{\int_s^t \kappa(z)dz}dL(t)dt = \int_{\tau_1}^{\tau_2}\int_0^{\tau_2} 1_{[0,t]}\sigma(t)e^{\int_s^t \kappa(z)dz}dL(t)dt \\
&= \int_0^{\tau_2}\int_{\tau_1}^{\tau_2} 1_{[0,t]}\sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t) \\
&= \int_0^{\tau_1}\int_{\tau_1}^{\tau_2} 1_{[0,t]}\sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t) + \int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2} 1_{[0,t]}\sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t) \\
&= \int_0^{\tau_1}\int_{\tau_1}^{\tau_2} \sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t) + \int_{\tau_1}^{\tau_2}\int_t^{\tau_2} \sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t).
\end{aligned}$$

□

In the previous section, the Girsanov theorem was applied in order to find an equivalent probability measure Q. The Girsanov theorem is a special case of the Esscher transform when the distribution is a BM. In the case of a jump process, the Esscher transform is applied.

Let $\theta(t)$ to be a real-valued measurable and bounded function denoting the market price of risk and consider the stochastic process

$$Z(t) = \exp\left(\int_0^t \theta(s)L(s) - \int_0^t \varphi(\theta(s))ds\right), \quad (6.47)$$

where $\varphi(\lambda)$ is the logarithm of the moment-generating function of $L(t)$

$$\varphi(\lambda) = \ln \mathrm{E}[\exp(\lambda L(1))]. \quad (6.48)$$

We make the same assumptions as in Benth and Saltyte-Benth (2005) and Bellini (2005). We assume that the process $Z(t)$ is well defined under natural

exponential integrability conditions on the Lévy measure $l(dz)$, which we assume to hold. Then, the following proposition for the price of CAT futures follows:

Proposition 6.12 *The future prices* $F_{\mathrm{CAT}}(t,\tau_1,\tau_2)$ *at time* $t \le \tau_1 < \tau_2$ *written on* CAT *over the interval* $[\tau_1,\tau_2]$ *is*

$$\begin{aligned} F_{\mathrm{CAT}}(t,\tau_1,\tau_2) = &\int_{\tau_1}^{\tau_2} S(t)dt + \int_{\tau_1}^{\tau_2} e^{\int_0^t \kappa(z)dz}(T(0)-S(0))dt \\ &+ \int_0^t \int_{\tau_1}^{\tau_2} \sigma(u) e^{\int_s^u \kappa(z)dz} du dL(u) + \int_t^{\tau_2} \int_u^{\tau_2} \sigma(u) e^{\int_s^u \kappa(z)dz} du\varphi'(\theta(u))du \\ &- \int_t^{\tau_1} \int_u^{\tau_1} \sigma(u) e^{\int_s^u \kappa(z)dz} du\varphi'(\theta(u))du. \end{aligned} \tag{6.49}$$

Proof First, we prove that for a real-valued measurable and bounded function $f(t)$,

$$\mathrm{E}_Q\left[\int_t^\tau f(u)dL(u)|\mathrm{F}_t\right] = \int_t^\tau f(u)\varphi'(\theta(u))du. \tag{6.50}$$

The proof of (6.50) can be found in many studies. For reasons of completeness of this study, we reproduce the proof here. We follow the method presented in Benth and Saltyte-Benth (2005). First, note the following lemma:

$$E\left[\exp\left\{\int_s^t g(u)dL(u)\right\}\right] = \exp\left\{\int_s^t \varphi(g(u))du\right\} \tag{6.51}$$

if $g:[s,t]\to\mathbb{R}$ is a bounded and measurable function and the integrability condition of the Lévy measure holds. The proof of this lemma can be found in Benth and Saltyte-Benth (2004). Hence, we have that

$$\begin{aligned} \mathrm{E}_Q\left[\int_t^\tau f(u)dL(u)|\mathbf{F}_t\right] &= \mathrm{E}_Q\left[\int_t^\tau f(u)dL(u)\frac{Z(\tau)}{Z(t)}\right] \\ &= \exp\left(-\int_t^\tau \varphi(\theta(u))du\right)\frac{d}{d\lambda}\mathrm{E}_Q\left[\exp\left(\int_t^\tau \lambda f(u)+\theta(u)dL(u)\right)\right]_{\lambda=0} \\ &= \exp\left(-\int_t^\tau \varphi(\theta(u))du\right)\frac{d}{d\lambda}\exp\left(\int_t^\tau \varphi(\lambda f(u)+\theta(u))du\right) \\ &= \int_t^\tau f(u)\varphi'(\theta(u))du. \end{aligned}$$

Hence, (6.50) holds.

Next, the dynamics of the price of the CAT future

$$\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds|\mathbf{F}_t\right] = \mathrm{E}_Q\left[\int_t^{\tau_2} T(s)ds|\mathbf{F}_t\right] - \mathrm{E}_Q\left[\int_t^{\tau_1} T(s)ds|\mathbf{F}_t\right].$$

From (6.46) and the adaptivity property of the Lévy process, we have that

$$\begin{aligned}
\mathrm{E}_Q\left[\int_t^{\tau} T(s)ds|\mathbf{F}_t\right] &= \int_t^{\tau} S(u)du + \int_t^{\tau} e^{\int_0^u \kappa(z)dz}(T(0) - S(0))du \\
&\quad + \mathrm{E}_Q\left[\int_0^{\tau}\int_t^{\tau} \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u) + \int_t^{\tau}\int_u^{\tau} \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u)|\mathbf{F}_t\right] \\
&= \int_t^{\tau} S(u)du + \int_t^{\tau} e^{\int_0^u \kappa(z)dz}(T(0) - S(0))du \\
&\quad + \mathrm{E}_Q\left[\int_0^{t}\int_t^{\tau} \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u)|\mathbf{F}_t\right] \\
&\quad + \mathrm{E}_Q\left[\int_t^{\tau}\int_u^{\tau} \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u)|\mathbf{F}_t\right].
\end{aligned}$$

Hence, using the adaptivity property again and (6.50), we have that

$$\begin{aligned}
\mathrm{E}_Q\left[\int_t^{\tau} T(s)ds|\mathbf{F}_t\right] &= \int_t^{\tau} S(u)du + \int_t^{\tau} e^{\int_0^u \kappa(z)dz}(T(0) - S(0))du \\
&\quad + \int_0^{t}\int_t^{\tau} \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u) \\
&\quad + \int_t^{\tau}\int_u^{\tau} \sigma(u)e^{\int_s^u \kappa(z)dz}du\varphi'(\theta(u))du.
\end{aligned}$$

Substituting the above equation to the initial expectation yields the result.

□

As it was mentioned earlier, the Pacific Rim index is simply the average of the CAT index over the specific time period. Then, the arbitrage-free future price of a CAT contract at time $t \leq \tau_1 \leq \tau_2$ is given by

$$e^{-r(\tau_2 - t)}\mathrm{E}_Q\left[\int_{\tau 1}^{\tau_2} \frac{1}{\tau_2 - \tau_1}\mathrm{T}(\tau)d\tau - F_{\mathrm{PAC}}(t, \tau_1, \tau_2)|\mathbf{F}_t\right] = 0,$$

and since F_{PAC} is $\mathbf{F}_t$ adapted, we derive the price of a PAC future to be

$$F_{\mathrm{PAC}}(t, \tau_1, \tau_2) = \mathrm{E}_Q\left[\frac{1}{\tau_2 - \tau_1}\int_{\tau_1}^{\tau_2} T(s)ds|\mathbf{F}_t\right]. \tag{6.52}$$

Hence we conclude that

$$F_{\mathrm{PAC}}(t,\tau_1,\tau_2)=\frac{1}{\tau_2-\tau_1}F_{\mathrm{CAT}}(t,\tau_1,\tau_2). \tag{6.53}$$

Unfortunately, introducing the Lévy noise process prevents the calculation of option prices. In addition, finding closed form solutions for AccHDD and AccCDD futures and options including a Lévy process in the temperature stochastic differential equation is not possible. The problem arises from the fact that the class of generalized hyperbolic distributions is not closed under convolution (Bellini 2005). Alternatively, estimating the prices of weather derivatives under the Lévy assumption can be done numerically. One approach is by applying the Fourier transform (FT). In order to do so, it is necessary to know the distributional properties of the random variable $T(t)$. The unknown density $f_T(x)$ can be estimated by a Fourier approach of the following integral of the characteristic function $\Psi_T(\lambda)$:

$$f_T(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-isx}\Psi_T(s)ds. \tag{6.54}$$

Hence, if the characteristics function of the Lévy process is known, then option prices as well as futures on AccHDD and AccCDD can be estimated. This approach is analytically discussed in Carr and Madan (1999).

Proposition 6.13 *The characteristic function of $T(t)$ under the risk-neutral measure Q is given by*

$$\Psi_T(\lambda)=\mathrm{E}_Q[\exp\{i\lambda T(t)\}|\mathbf{F}_t]=\exp\{\Psi(\lambda)\}, \tag{6.55}$$

where

$$\begin{aligned}\Psi(\lambda)&=i\lambda S(t)+i\lambda e^{\int_s^t\kappa(z)dz}(T(s)-S(s))-\int_s^t\varphi(\theta(u))du\\&\quad+\int_s^t\varphi\left(i\lambda\sigma(u)e^{\int_u^t\kappa(z)dz}+\theta(u)\right)du.\end{aligned} \tag{6.56}$$

Proof We have that

$$\begin{aligned}&\mathrm{E}_Q[\exp\{i\lambda T(t)\}|\mathbf{F}_t]\\&=\mathrm{E}_Q\left[\exp\left\{i\lambda S(t)+i\lambda e^{\int_s^t\kappa(z)dz}(T(s)-S(s))+i\lambda\int_s^t\sigma(u)e^{\int_u^t\kappa(z)dz}dL(u)\right\}|\mathbf{F}_t\right]\\&=\exp\left\{i\lambda S(t)+i\lambda e^{\int_s^t\kappa(z)dz}(T(s)-S(s))\right\}\mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t\sigma(u)e^{\int_u^t\kappa(z)dz}dL(u)\right\}|\mathbf{F}_t\right].\end{aligned} \tag{6.57}$$

Focusing on the expectation, we have that

$$\begin{aligned}
&\mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz}dL(u)\right\}|\mathbf{F}_t\right]\\
&=\mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz}dL(u)\right\}\frac{Z(t)}{Z(s)}|\mathbf{F}_t\right]\\
&=\mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz}dL(u)+\int_s^t\theta(u)dL(u)-\int_s^t\varphi(\theta(u))du\right\}|\mathbf{F}_t\right]\\
&=\exp\left\{-\int_s^t\varphi(\theta(u))du\right\}\mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz}+\theta(u)dL(u)\right\}\right]\\
&=\exp\left\{-\int_s^t\varphi(\theta(u))du\right\}\exp\left\{\int_s^t\varphi\left(i\lambda\sigma(u)e^{\int_u^t \kappa(z)dz}+\theta(u)\right)du\right\}. \qquad (6.58)
\end{aligned}$$

From (6.55), (6.57), and (6.58) yields the result

$$\begin{aligned}
\Psi(\lambda)=i\lambda S(t)+i\lambda e^{\int_s^t \kappa(z)dz}(T(s)-S(s))-\int_s^t\varphi(\theta(u))du\\
+\int_s^t\varphi\left(i\lambda\sigma(u)e^{\int_u^t \kappa(z)dz}+\theta(u)\right)du,
\end{aligned}$$

where $\varphi(\cdot)$ is the moment-generating function of $L(1)$ and $i^2=-1$.

□

In the case of the generalized hyperbolic distribution (and hyperbolic distribution), the moment-generating function φ is known. Hence, the characteristic function $\Psi(\lambda)=\varphi(i\lambda)$ is also known. Now, the distribution of our model can be found by numerical inversion of the characteristic function. Hence, we can proceed on deriving the pricing formulas for the CDDs futures using a Lévy process:

$$\begin{aligned}
F_{\mathrm{CDD}}(t,\tau_1,\tau_2)&=\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\max(T(s)-c)ds|\mathbf{F}_t\right]\\
&=\int_{\tau_1}^{\tau_2}\mathrm{E}_Q[\max(T(s)-c)|\mathbf{F}_t]ds\\
&=\int_{\tau_1}^{\tau_2}\int_c^{+\infty}(x-c)f_T(x)dxds, \qquad (6.59)
\end{aligned}$$

where $f_T(x)$ is the density function of $T(t)$ under the risk-neutral measure Q conditional on $\mathbf{F}_t$ and it is given by (6.54). Similarly, the HDD future price is given by

$$F_{\mathrm{CDD}}(t,\tau_1,\tau_2)=\int_{\tau_1}^{\tau_2}\int_0^{65}(c-x)f_T(x)dxds. \qquad (6.60)$$

Practitioners often prefer easy-to-implement models than realistic ones. A classic example is the Black–Scholes equation. The above solution of the price of a CAT future is not easy to solve, and to calculate the above pricing formulas is not a straightforward process. Alternatively, the price of a future or an option contract on a temperature index can be estimated using numerical procedures.

6.6 The Market Price of Risk

The weather derivatives market is a classical incomplete market. In incomplete markets, the derivative contract cannot be replicated by the underlying index. Moreover, in incomplete markets, there are infinite equivalent martingale measures. Here, we limit our choices by selecting a parametric class of probabilities Q^{θ}, where $\theta(t)$ is square integrable function as it is assumed by equation. Since temperature is non-tradable, the market price of risk must be incorporated in the pricing model.

The parameter $\theta(t)$ is called the market price of risk. The choice of $\theta(t)$ actually captures the risk preference of the market participants. In other words, it reflects the trader's view on exposing themselves to risk (Benth 2004). The market price of risk, $\theta(t)$, was introduced by applying the Girsanov's theorem (or the Esscher transform). The change of measure of an asset's stochastic process is closely related to the concept of the market price of risk (Xu et al. 2008). Actually the drift rate of the asset's stochastic process is corrected by a parameter that reflects the market price of risk (Xu et al. 2008).

In most studies so far, the market price of risk was considered zero. However, recently many studies examine the market price of risk and found that it is different than zero.

Turvey (2005) proposed to estimate the market price of risk by using the capital asset pricing model. Cao and Wei (2004) and Richards et al. (2004) apply a generalized Lucas' (1978) equilibrium pricing model to study the market price of risk. In that framework, direct estimation of the weather risk's market price is avoided (Xu et al. 2008). Their findings indicate that the market price of risk associated with the temperature variable is significant. They also conclude that the market price of risk affects option values much more than forward prices, mainly due to the payoff specification. In Xu et al. (2008), an indifference pricing approach which is also based on utility maximization is proposed.

The most common approach is the one presented in Alaton et al. (2002), and it was followed by Bellini (2005), Benth et al. (2009), and Hardle and Lopez Cabrera (2009).

Alaton et al. (2002) suggest that the market price of risk can be estimated from the market data. More precisely, the market price of risk is derived as follows: we examine what value of the $\theta(t)$ gives a price from the theoretical model that fits the observable market price.

In Bellini (2005), the implicit market price of risk is estimated by comparing theoretical future prices, given in previous formulas, to the prices observed in the

market under the assumption of a Lévy noise process. Their results indicate that for four cities in use that market price of risk has always a negative sign while it was found not to be constant. Moreover, in Bellini (2005), the time dependence of the market price of risk is examined. It was found that there is a relation between $\theta(t)$ and its lag as well as with the number of available for trading future contracts.

In Hardle and Lopez Cabrera (2009), the implied market price of risk from Berlin was estimated. Their results indicate that the market price of risk for CAT derivatives is different from zero and shows a seasonal structure that increases as the expiration date of the temperature future increases. In a more recent paper, Benth et al. (2009) study the market price of risk in various Asian cities. The market price of risk was estimated by calibrating model prices. Their results indicate that the market price of risk for Asian temperature derivatives is different from zero and shows a seasonal structure that comes from the seasonal variance of the temperature process. Their empirical findings suggest that by knowing the formal dependence of the market price of risk on seasonal variation, one can infer the market price of risk for regions where weather derivatives market does not exist.

Similar, in Huang et al. (2008), a pricing method for temperature derivatives in Taiwan is presented. Since no active weather market exists in Taiwan, the parameter θ is approximated by a function of the market price of risk of the Taiwan Stock Exchange.

6.7 Conclusions

In this chapter, the pricing formulas for the weather derivatives on various temperature indices were presented. Assuming a normal distribution, the pricing formulas for the following indices were derived: CAT, AccHDD, AccCDD, and Pacific Rim. The appealing properties of the normal distributions allows for derivation of pricing formulas in both futures and options on the above indices. In order to find the pricing formulas, the Itô Lemma and the Girsanov theorem were applied.

Then, based on our results that the hyperbolic distribution provides a better fit to the residuals, a Lévy motion noise process was assumed. In this case, the pricing formulas for the CAT and Pacific Rims futures were presented. Under the assumption of a jump process, the Esscher transform and the Itô formula for semimartingales were applied. However, finding closed form solutions including a Lévy process in the temperature stochastic differential equation is not possible. Alternatively, estimating the prices of weather derivatives under the Lévy assumption can be done numerically. We provided a representation of the characteristic function of the temperature dynamics under the risk-neutral probability measure which is crucial for finding the density function necessary for pricing options and futures on AccHDD and AccCDD.

Finally, the importance of the market price of risk was discussed, and an estimation method was presented.

References

Alaton P, Djehince B, Stillberg D (2002) On modelling and pricing weather derivatives. Appl Math Finance 9:1–20

Bellini F (2005) The weather derivatives market: modelling and pricing temperature. Ph.D. Thesis, University of Lugano, Lugano

Benth FE (2003) On arbitrage-free pricing of weather derivatives based on fractional Brownian motion. Appl Math Finance 10:303–324

Benth FE (2004) Option theory with stochastic analysis. Springer, Berlin

Benth FE, Hardle WK, Lopez Cabrera B (2009) Pricing of Asian temperature risk. SFB649 working paper. Humboldt-Universitat zu Berlin, Berlin

Benth FE, Saltyte-Benth J (2004) The normal inverse Gaussian distribution and spot price modeling in energy markets. Int J Theor Appl Finance 7(2):177–192. doi:10.1142/S0219024904002360

Benth FE, Saltyte-Benth J (2005) Stochastic modelling of temperature variations with a view towards weather derivatives. Appl Math Finance 12(1):53–85

Benth FE, Saltyte-Benth J (2007) The volatility of temperature and pricing of weather derivatives. Quant Finance 7(5):553–561

Benth FE, Saltyte-Benth J, Koekebakker S (2008) Stochastic modelling of electricity and related markets, vol 11, Advance series on statistical science & applied probability. World Scientific, Singapore

Brockwell PJ, Marquardt T (2005) Levy-driven and fractionality integrated ARMA process with continuous time parameter. Statistica Sinica 15:477–494

Brody CD, Syroka J, Zervos M (2002) Dynamical pricing of weather derivatives. Quant Finance 2:189–198

Cao M, Wei J (2004) Weather derivatives valuation and market price of weather risk. J Future Markets 24(1):1065–1089

Carr M, Madan DB (1999) Option valuation using the fast fourier transform. J Comput Finance 2 (4):61–73

Davis M (2001) Pricing weather derivatives by marginal value. Quant Finance 1:1–4

Garman M, Blanco C, Erickson R (2000) Weather derivatives: instruments and pricing issues. Environ Finance

Geman H (1999) Insurance and weather derivatives. RISK Books, London

Hardle WK, Lopez Cabrera B (2009) Infering the market price of weather risk. SFB649 working paper. Humboldt-Universitat zu Berlin, Berlin

Huang H-H, Shiu Y-M, Lin P-S (2008) HDD and CDD option pricing with market price of weather risk for Taiwan. J Futures Markets 28(8):790–814

Ikeda N, Watanabe S (1981) Stochastic differential equations and diffusion processes. North-Holland/Kodansha, Tokyo

Jewson S, Brix A, Ziehmann C (2005) Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations. Cambridge University Press, Cambridge, UK

Platen E, West J (2005) A fair pricing approach to weather derivatives. Asia-Pacific Financial Markets 11:23–53

Richards TJ, Manfredo MR, Sanders DR (2004) Pricing weather derivatives. Am J Agricul Econ 4 (86):1005–1017

Turvey CG (2005) The pricing of degree-day weather options. Agricul Finance Rev 2005:59–85

Xu W, Odening M, Musshof O (2008) Indifference pricing of weather derivatives. Am J Agricul Econ 90(4):979–993

Zapranis A, Alexandridis A (2008) Modelling temperature time dependent speed of mean reversion in the context of weather derivative pricing. Appl Math Finance 15(4):355–386

Chapter 7
Using Meteorological Forecasts for Improving Weather Derivative Pricing

7.1 Introduction

Weather risk is unique in that it is highly localized, and despite the new technologies in meteorological science, it still cannot be predicted precisely and consistently. Due to the chaotic nature of the equations involved in predicting the behavior of the atmosphere, weather cannot be forecasted accurately more than few days in advance. However, even short forecasts are useful for contracts that have already started or that will start in the next few days.

The availability of historical weather data is central to the success of any weather risk management program (Banks 2002). Similarly, available and accurate weather forecasts can significantly improve both weather derivative pricing and weather risk management. Considerable advances have been made in the meteorological science. Satellite images have improved the ability of observing the state of the atmosphere and, combined with mathematical models, allow forecasters to produce weather forecasts and climate predictions.

Weather forecasting focuses on shorter maturity weather events while climate prediction centers on longer-term phenomena. Weather forecasting is typically divided into several ranges: very short-range forecasts with time horizon of 0–12 h, short forecasts that range from 12 to 72 h, and medium forecasts with time horizon that range from 3 to 5 days. Moreover, there are more extended forecasts that cover monthly periods. Long-range climate forecasts extend onto seasonal and interannual forecasts. Finally, climate prediction refers to periods of over 2 years (Banks 2002). In Table 7.1, the different forecasting categories, the time horizon of each category, and the typical events that are forecasted in each category are presented.

Forecasters employ different techniques in developing weather forecasts. Common approaches include the numerical weather prediction methods, ensemble forecasts, and probabilistic forecasts.

Despite the great advances in technology and the development of new models, the maximum forecast window for useful daily estimates of weather is limited to 5–7

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_7,

Table 7.1 Forecasting categories, time horizon, and typical events

Forecast category	Time horizon	Typical event
"Nowcast"	<2 h	Tornado strike
Very short range	12 h	Severe storm
Short-range forecast	12–72 h	Developing cold front
Medium-range forecast	3–10 days	Developing high-/low-pressure zone
Extended-range climate forecast	1 month	Anticipated rain shortfall
Long-range climate forecast	Season/1–2 years	Deviations in regional temperature ranges
Climate prediction	2+ years	Developing global warming patterns

Source: (Banks 2002)

days, rather than the 10–14 days that meteorologists believe is theoretically possible, due to practical limitations (Banks 2002). Although satellite observations and computing speed and power have significantly improved, further improvement in these areas will bring forecasting accuracy close to the theoretical level (Banks 2002).

In the perspective of weather derivative pricing and weather risk management, only few papers take into consideration the possible usefulness of weather forecasts. In Alaton et al. (2002), it is suggested that meteorological forecasts for short to medium range should be used. On the other hand, for long-term pricing, a general trend is suggested (Ritter et al. 2011).

In Zeng (2000a, b), a framework where seasonal forecasts are employed in the pricing of weather derivatives is presented. Three scenarios are assumed: the above, below, or near normal. Obtaining probabilities for each scenario, the derivatives are priced using a weighted HBA.

A similar method is followed in Yoo (2003). Again, three possible forecasting scenarios are assumed. For each scenario, a different O–U temperature process is estimated. Depending on the seasonal forecast, the parameters of the corresponding O–U are used in order to determine the future evolution of the temperature process and the price of the contract.

In using single and ensemble forecasts, Jewson and Caballero (2003) derive probabilistic weather forecasts to price derivatives which have already begun or will begin in the forecast period. Furthermore, they present two methods that combine probabilistic forecasts and climatological models to improve the estimates of the distribution of the index.

In a more recent paper, Dorfleitner and Wimmer (2010) estimate the temperature index including meteorological forecasts in their model. Then, they compare their results with the prices of monthly and seasonal contracts traded in the CME. Their results indicate that meteorological weather forecasts have to be included into the pricing process. However, their approach is accurate only if the forecasts reach into the accumulation period (Ritter et al. 2011).

Taylor and Buizza (2002, 2003) describe an "end-to-end" application of ensemble forecasts to predict electricity load and demand. In more recent papers, Taylor and Buizza (2004, 2006) have investigated the use of ensemble predictions in forecasting the density of temperature at five locations in the UK. Then, the

ensemble prediction was compared against an AR–GARCH model that is based on simple information. Their results indicate that the ensemble predictions outperformed the point forecast from the AR–GARCH models in all cases.

In Ritter et al. (2011), a daily modeling approach which also regards meteorological temperature forecasts is introduced in order to explain the market prices. Their results indicate that the inclusion of meteorological forecasts has a clear impact on the pricing of weather contracts. Since meteorological forecasts are more reliable in the short term, this effect is more significant near the expiration of the contract. On the other hand, Ritter et al. (2011) conclude that the value of meteorological forecasts declines with a longer forecast horizon.

The rest of the chapter is organized as follows. In Sect. 7.2, the numerical weather prediction approach is presented. In Sect. 7.3, the ensemble methods are described. The probabilistic and seasonal forecasts are presented in Sect. 7.4. In Sect. 7.5, approaches of weather derivative pricing using meteorological forecasts are presented. Finally, in Sect. 7.6, we conclude.

7.2 Numerical Weather Prediction

By collecting quantitative data about the current state of the atmosphere and using scientific understanding of atmospheric processes to project how the atmosphere will evolve, weather forecasts are made. The atmosphere is a complex dynamical system with many degrees of freedom (Taylor and Buizza 2003).

Numerical weather prediction applies mathematical models of the atmosphere and oceans and predicts the weather based on the current atmosphere state. The current state of atmosphere is described by studying different variables such as the spatial distribution of the wind, temperature, humidity, and surface pressure (Taylor and Buizza 2006).

The atmosphere is considered to be a fluid. Hence, in numerical weather prediction, the equations of fluid dynamics and thermodynamics are applied in order to estimate the state of the atmosphere at some time in the future.

First, the initial conditions, for example, the current state of the atmosphere, are estimated. The best estimate of the current state of the system usually is determined using observations from earth, ocean, and satellite platforms taken during the preceding 12 h (Taylor and Buizza 2006). Then, the initial conditions are entered in the equations and then meteorological forecasts are produced. More precisely, computer programs are used to produce meteorological forecasts at future times at a given location and altitude.

The forecasts are produced by solving the nonlinear partial differential equations that describe the dynamics of the atmosphere. However, the scale of some meteorological processes is too small or too complex to be explicitly included in numerical weather prediction models. In order to rectify this, parameterization is used. Examples of parameterization are the simulation of effects of clouds, radiation, air quality, humidity, or raindrops. The nonlinear partial differential equations cannot

be solved using analytical methods (Pielke 2002). These equations are solved using numerical methods. An exact solution is not obtained but rather an approximate one. Different models use different solution methods. In Strikwerda (2004), various methods for solving differential equations are described such as the finite difference methods that are used in global and regional models for all three dimensions or spectral methods that are used for the horizontal dimensions and finite difference methods in the vertical (Strikwerda 2004).

After solving the nonlinear partial differential equations, the state of the atmosphere is predicted at a short time into the future. Then, the parameters of the equations are reestimated to predict further time steps ahead. Different models use different time steps that range from 1 to 40 min (Bourchtein 2005; Michalakes et al. 2000). The forecast window of each meteorological center depends on the computational power and resources. The UKMET Unified Model is run 6 days into the future, while the European Centre for Medium-Range Weather Forecasts (ECMWF) run out to 10 days into the future, and the Global Forecast System model run by the Environmental Modeling Center is run 16 days into the future.[1]

7.3 Ensemble Forecasts

It is considered that the weather has a chaotic nature. Hence, small errors in the initial conditions can produce large errors in forecasting. In other words, the accuracy of the predictability is significantly reduced when the time horizon increases. On the other hand, despite the advances on meteorological modeling, the approximation of the atmospheric dynamics by a numerical model induces model errors to the forecasts because of imperfections in the model, such as the finite grid spacing. As a result, the above sources of uncertainty reduce the forecasting ability of existing models that are based on the estimate of the initial conditions (Taylor and Buizza 2003).

The most common approach to overcome this problem is the use of ensemble forecasts (Lynch 2008). Nowadays, ensemble forecasts are made at most of the major operational weather prediction facilities worldwide, including the National Centers for Environmental Prediction (US), the ECMWF, the United Kingdom Met Office, Meteo France, Environment Canada, the Japanese Meteorological Agency, the Bureau of Meteorology (Australia), the China Meteorological Administration, the Korea Meteorological Administration, and CPTEC (Brazil).[2]

Ensemble forecasting is a form of MC analysis. Ensemble forecasting is a numerical prediction method that is used to attempt to generate a representative sample of the possible future states of a dynamical system. Multiple numerical predictions are conducted using slightly different initial conditions that are all

[1] http://en.wikipedia.org/wiki/Numerical_weather_prediction

[2] http://en.wikipedia.org/wiki/Ensemble_forecasting

plausible given the past and current set of observations or measurements. Sometimes, the ensemble of forecasts may use different forecast models for different members, or different formulations of a forecast model. The frequency distribution of the different outcomes, which are known as ensemble members, provides an estimate of the PDF (Taylor and Buizza 2006).

The weather prediction model is both complex and high dimensional (Taylor and Buizza 2006). Hence, because the demand of producing forecast is fast and because of the existing computer power, some limitations are induced to ensemble forecasting. The initial conditions are not sampled as in a statistical simulation because this is not practical. The number of ensemble members is limited. In practice, forecasters try to guess a small number of perturbations (usually around 20) that they deem are most likely to yield distinct weather outcomes. The ECMWF uses a more sophisticated algorithm. It operates on one forecast started from the unperturbed best estimate of the atmosphere initial state plus 50 others generated by varying the initial conditions (Taylor and Buizza 2006). Ensemble predictions are generated using a lower resolution (horizontal grid spacing) than traditional single point forecasts (Taylor and Buizza 2003; Taylor and Buizza 2006).

7.4 Probabilistic Forecasts and Scenario Analysis

Probabilities are often used in the publication of precipitation in the form of probability of precipitation. For example, there might be 40% of precipitation in a particular area. This allows for a degree of uncertainty to be incorporated in the forecast. The probability of precipitation is given by

$$PoP = C \times A, \tag{7.1}$$

where C is the confidence that precipitation will occur somewhere in the forecast area and A the percent of the area that will receive measurable precipitation, if it occurs at all.

Similarly, probabilistic forecasts can be given for temperature. Temperature cannot be forecasted accurately more than 6–10 days in advance. However, reports are published for seasonal forecasts by the NOAA and the National Center of Environmental Forecasts (NCEF). These forecasts report the probabilities of deviation of the temperature from its historical normal up to 3–6 months ahead. These forecasts can be used on a hedging strategy to price a weather contract long before its starting date. For example, there is a possibility that an energy producer might want in autumn to hedge against adverse weather condition in winter.

Assume that the parameters p_A, p_B, and p_N are the probabilities that the temperature will be above, below, or near the climate normal during the next 3 months. We focus on the temperature in Berlin for the years 1951–2008. The first 57 years are used to estimate the historical mean and the last year to evaluate the scenario analysis and HBA methods.

Table 7.2 Mean and standard deviation of the HDD index for different historical lengths for Berlin

Years	Mean	St. D.
57	550.17	90.47
Last 50	549.62	94.93
Last 40	542.27	96.60
Last 30	538.52	100.19
Last 20	509.65	80.95
Last 10	512.81	82.22

St. D. = standard deviation

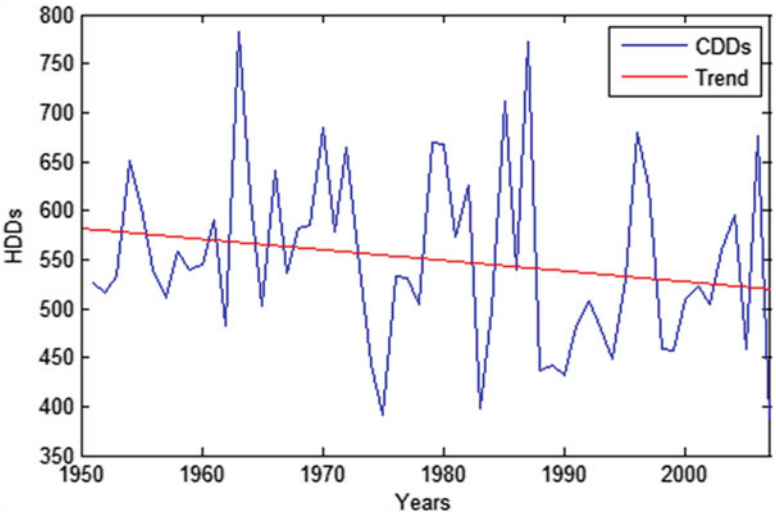

Fig. 7.1 The CDDs index and a linear trend for the years of 1951–2007 at Berlin

An energy producer wants to hedge against adverse weather condition for the period of January. Hence, he is interested in the HDD index. Studying the HDDs in the past years, the HBA suggests that the mean of the HDD index is 550.17 and the standard deviation is 90.47. Hence, a hedger that uses the HBA method will price the weather contract based on these two estimations.

Considering different lengths of historical datasets, both the mean and the standard deviation change. In Table 7.2, the mean and standard deviation for different lengths of historical data are presented. A warming trend is evident. It is clear that a decrease in the level of the HDDs is observed in the last years. This is also evident from Fig. 7.1 where the HDD index and a simple linear trend are presented.

Alternatively, suppose that the energy producer instead of the mean and the standard deviation of the HDD index also possesses seasonal forecasts that report the probabilities of deviation of the temperature from its historical normal. As an illustration example, we suppose that the probabilities for above, near, and below

Table 7.3 Forecasted **H**DD index using the HBA and scenario analysis methods

Method	Mean	St. D.
HBA	550.17	90.47
Scenario analysis	494.14	36.68
Actual	436.20	

St. D. = standard deviation

the climate normal are 0.7, 0.2, and 0.1 respectively. A slightly different approach of MC will be applied. First, it is assumed that the probabilities p_A, p_B, and p_N approximate the probabilities that the HDDs will be above, below, and near the historical normal. This is justified since the correlation between HDDs and temperature is close to one. Next, the historical observations of the HDD index are separated in the lower, middle, and higher values. Next, a bootstrapped sample is created sampling from the lower, middle, and higher part of the values of the HDD index using the probabilities p_A, p_B, and p_N. In other words, a biased MC simulation method is applied. Hence, if the bootstrapped sample has 10,000 simulations, 7,000 are drawn from the lower part, 2,000 from the middle part, and 1,000 from the upper part. Finally, the weighted average of the HDD index is estimated. Our results indicate that the weighted average of the HDDs is 494.14, which is closer to the actual value 436.20. Note that the classical HBA suggested that the HDD is 550.17. Also, the standard deviation is reduced and it is only 36.68. In Table 7.3, the forecasted HDD index using the HBA and scenario analysis methods is presented.

When someone is interested in the daily temperature, a similar procedure can be followed. Suppose that the temperature process is given by (5.15), the seasonal mean by (5.16), and the seasonal variance by (5.17). As it is described in Yoo (2003), the seasonal mean is a linear combination of the mean temperatures for each scenario and is given by

$$S(t) = p_A S_A(t) + p_N S_N(t) + p_B S_B(t), \tag{7.2}$$

where $S_A(t)$, $S_B(t)$, and $S_N(t)$ correspond to an above-, below-, or near-normal seasonal temperature. For each scenario, a different seasonal mean is estimated according to (5.16).

In order to estimate (7.2), the following procedure is followed. First, the available years of historical data are separated in above, below, and near normal. Then, in each set, the corresponding above-, below-, or near-normal seasonal mean is estimated as it was described in Chap. 5. Finally, the unconditional seasonal mean is estimated by weighted linear combination according to (7.2). The procedure for estimating the remaining parts of the temperature process and for forecasting the future evolution of the temperature process is the same as it was described in Chap. 5.

7.5 Meteorological Forecasts and Pricing

As it was shown in the previous chapter, the price of the weather derivatives depends on the expectation of future of the evolution of the temperature. If meteorological forecasts are not available, all uncertainty is included in the market price of risk (Yoo 2003). On the other hand, if meteorological forecasts are available, then the pricing of a weather derivative can be improved.

Depending on the length of the contract, the length of the available meteorological forecasts and the how many days before the start of the contract different approaches are followed.

In the first case, the valuation of the contract is conducted long before the starting day of the weather derivative. Hence, meteorological forecasts are not available for the period of life of the weather derivative.

To illustrate this, we suppose that there are available forecasts for today's DAT and forecasts for the next 10 days. Also, t denotes the current day, τ_1 denotes the day that the weather derivative starts, and τ_2 denotes the expiration day of the weather derivative. In our example, the contract starts at the first of July and has duration of 1 month. In the first case, a company wants to enter to a CAT future during the first days of May. Since the available forecasts are only for 10 days, they do not add any additional information to the statistical models. Hence, the procedure described in Chap. 5 is followed to produce statistical forecasts of the temperature and then valuate the CAT future contract using the formulas presented in Proposition 6.2 in Chap. 6.

In the second case, the valuation of the contract is conducted before the starting day, but meteorological forecasts are available during the life of contract but not until the expiration day.

Suppose that the company wants to buy a CAT future few days before the start of the contract, hence few days before the 1st of July. In this case, the inside the contract period valuation is used. At the 30th of June, the company has a forecast for the DAT of this day and for the next 10 days. Hence, the valuation formula is separated in two parts. The first one consists of the meteorological forecasts of the next 10 days and is deterministic. The second one is the stochastic part for the period of the 10th to the 31st of July. Hence, the valuation is given by

$$F_{CAT}(10,1,31) = \int_1^{10} T(s)ds + F_{CAT}(10,10,31). \tag{7.3}$$

Weather derivatives are path dependent. In order to evaluate a weather contract, simulations of the paths of the temperature are produced. In Fig. 7.2, five simulations of the temperature process are presented. In the first part of Fig. 7.2, meteorological forecasts are not included. Hence, simulations for the whole period of 31 days are conducted. On the second part, meteorological forecasts for the first 10 days are available. Hence, it is considered that the temperature for the first 10 days is known. Also, the forecast of the last day, the 10th of July, is used as a starting

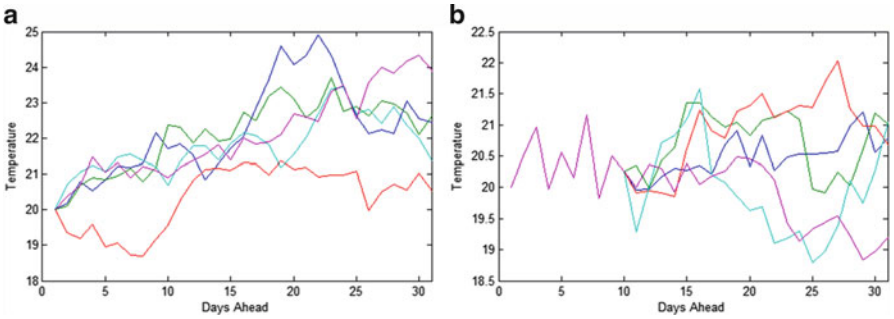

Fig. 7.2 Five sample paths of simulated temperature forecasts (**a**) without meteorological forecasts and (**b**) with meteorological forecasts

point for the simulations of the temperature paths for the rest of the period. Note that the two procedures result to different outcomes, hence to a different valuation of the contract.

The first 10 days are replaced by the meteorological forecasts. In general, it is expected that these forecasts are more accurate than the forecasts produced by statistical models.

The final case is to estimate the price of the future contract on the CAT index inside the contract period and when meteorological forecasts are available until the end of the life of the contract. Suppose that the company wants to enter a CAT future at the 21st of July. Also, the company has available forecasts for the 21st of July and for the next 10 days, till the 31st of July when the contract expires. This is the simplest case. The temperature of the first 20 days of the contract – the 1st to the 20th of July – is known, while the temperature of the last days is replaced by the meteorological forecasts. Hence, the price of the contract is given by the inside the contract period valuation formula presented in Chap. 6:

$$F_{CAT}(21,1,31) = \int_1^{21} T(s)ds + F_{CAT}(21,21,31). \tag{7.4}$$

However, the second part is replaced by the meteorological forecasts. Hence, we have that

$$F_{CAT}(21,1,31) = \int_1^{21} T(s)ds + \int_{21}^{31} T(s)ds = \int_1^{31} T(s)ds. \tag{7.5}$$

7.6 Conclusions

The availability of historical weather data is central to the success of any weather risk management program. Similarly, available and accurate weather forecasts can significantly improve both weather derivative pricing and weather risk management.

Forecasters employ different techniques in developing weather forecasts. Common approaches include the numerical weather prediction methods, ensemble forecasts, and probabilistic forecasts.

Numerical weather prediction applies mathematical models of the atmosphere and oceans and predicts the weather based on the current atmosphere state. It is considered that the weather has a chaotic nature. Hence, small errors in the initial conditions can produce large in forecasting. The most common approach to overcome this problem is the use of ensemble forecasts. Multiple numerical predictions are conducted using slightly different initial conditions that are all plausible given the past and current set of observations or measurements.

Probabilistic seasonal forecasts are reported for temperature. The probabilities that the temperature will be above, below, or near the climate normal during the next 3 months are available to risk managers. Since the dynamics of the climate system are chaotic, seasonal forecasts are necessarily less specific than weather forecasts. Using weighted average techniques, a better estimate of the temperature can be calculated using numerical models.

The accuracy of meteorological forecasts declines with a longer forecast horizon. Weather and climate forecasts only have value if they result in improved decision making.

Despite the great advances in technology and the development of new models, the maximum forecast window for useful daily estimates of weather is limited to 5–7 days. Although satellite observations and computing speed and power have significantly improved, further improvement in these areas will bring forecasting accuracy close to the theoretical level.

References

Alaton P, Djehince B, Stillberg D (2002) On modelling and pricing weather derivatives. Appl Math Finance 9:1–20

Banks E (ed) (2002) Weather risk management: markets, products and applications. Finance and Capital Markets/Palgrave Macmillan, Basingstoke

Bourchtein A (2005) Semi-Langranian implicit-explicit two-time-level scheme for numerical weather prediction. In: Sunderam VS, van Albada GD, Sloot PMA, Dongarra JJ (eds) 5th international conference computational science ICCS 2005, Springer, Atlanta, 22–25 May 2005, p 1089

Dorfleitner G, Wimmer M (2010) The pricing of temperature futures at the Chicago Mercantile Exchange. J Bank Finance. doi:10.1016/j.bankfin.2009.12.004

Jewson S, Caballero R (2003) The use of weather forecasts in the pricing of weather derivatives. Meteorol Appl 10:377–389. doi:10.1017/S1350482703001099

Lynch P (2008) The origins of computer weather prediction and climate modeling. J Comput Phys 227(7):3431–3444

Michalakes J, Chen S, Dudhia J, Hart L, Klemp J, Middlecoff J, Skamarock W (2000) Development of next-generation regional weatehr research and forecast model. In: Zwieflhofer W, Kreitz N (eds) 9th ECMWF workshop on the use of high performance computing in meteorology, Reading, 13–17 Nov 2000. Developments in teracomputing

Pielke RA (ed) (2002) Mesoscale meteorological modeling, vol 78. Academic, San Diego

Ritter M, Musshoff O, Odening M (2011) Meteorological forecasts and the pricing of temperature futures. J Derivatives 19(2):45–60. doi:10.3905/jod.2011.19.2.045
Strikwerda JC (2004) Finite difference schemes and partial differential equations. SIAM, USA
Taylor JW, Buizza R (2002) Neural network load forecasting with weather ensemble predictions. IEEE Trans Power Syst 17(3):626–632
Taylor JW, Buizza R (2004) A comparison of temperature density forecasts from GARCH and atmospheric models. J Forecasting 23:337–355
Taylor JW, Buizza R (2006) Density forecasting for weather derivative pricing. Int J Forecasting 22:29–42
Taylor JW, Buizza R (2003) Using weather ensemble predictions in electricity demand forecasting. Int J Forecasting 19(1):57–70. doi:10.1016/s0169-2070(01)00123-6
Yoo S (2003) Weather derivatives and seasonal forecast. Department of Applied Economics and Management, Cornell University. http://www.card.iastate.edu/faculty/profiles/bruce_babcock/Shiyong_Yoo_papers/WD_2.pdf. Accessed Jan 2003
Zeng L (2000a) Pricing weather derivatives. J Risk Finance 1(3):72–78
Zeng L (2000b) Weather derivatives and weather insurance: concept, application and analysis. Bull Am Meteorol Soc 81(9):2075–2082

Chapter 8
The Effects of the Geographical and Basis Risk

8.1 Introduction

In Hull (2003), the basis risk is defined as the difference between the spot and future price. In other words, it is the risk to a hedger arising from uncertainty about the basis at a future time. In the sense of weather derivatives, basis risk has a different definition. Weather risk is unique in that it is highly localized, and despite great advances in meteorological science, it still cannot be predicted precisely and consistently. Risk managers often face unique basis risks arising from both the choice of weather station where a derivative contract is written, as well as the relationship between the hedged volume and the underlying weather index (Manfredo and Richards 2009). We will refer to the first as spatial or geographical basis risk while to the second as basis risk.

The exchange-traded weather derivatives eliminated the default risk while at the same time the liquidity and the transparency increased. On the other hand, investors who wish to trade weather derivatives outside the list of the traded cities in CME face a spatial risk.

Geographical basis risk or spatial risks results from the distance between the hedging company and the site at which the weather measurement takes place. Geographical basis risk can reach critical levels in some cases (Rohrer 2004). As the distance between a hedging company and the measurement weather station of the weather derivative increases, the demand for weather derivative decreases (East 2005; Edwards and Simmons 2004).

It is expected that spatial risk will always be positive. However, Woodard and Garcia (2008) shows that weather derivatives from a variety of stations around the hedging company can improve the hedging effectiveness. Using nonlocal derivatives for a weather variable that are highly spatially correlated, the hedging strategy obtained may be as good as the one obtained using locally derived contracts (Woodard and Garcia 2008).

In many studies, energy and weather are considered highly correlated. Hence, companies from the energy sector are extensively using weather derivatives to

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_8,

hedge both the price and volumetric risk of energy demand (Gabbi and Zanotti 2005; Henley and Peirson 1998; Pirrong and Jermakyan 2008). Moreover, weather derivatives are used for the valuation of gas and CO_2 emissions contracts (Bataller et al. 2006; Zanotti et al. 2003; Geman 1999)

However, these two variables, energy and temperature (or any other weather variable), are not perfectly correlated. The payoff of the weather derivative depends on the weather index, and it is unlikely that the payoff will compensate exactly for the money lost due to weather (Jewson et al. 2005). As a result, a risk is induced on the hedging strategy, called basis risk. As the correlation between the weather index and the financial loss increases, it is expected that the basis risk will decrease.

Rohrer (2004) mentions another source of basis risk, called time basis risk. In analogy to commodity markets, Rohrer (2004) defines time basis risk as the difference between the period of exposure and period of reference of the weather derivatives. However, this risk can be easily overcome by using a combination of weather derivatives with different maturities. Hence, we will not further study the time basis risk.

However, a lack of understanding of weather basis risk has contributed to liquidity problems (Manfredo and Richards 2009). The study and understanding of spatial and basis risk will draw new participants to the weather market.

The rest of the chapter is organized as follows. In Sect. 8.2, the notion of the geographical/spatial risk and how can affect the weather risk management is presented. In Sect. 8.2.1, a spatial model for temperature is introduced. The basis risk and its impact on the weather risk management are presented in Sect. 8.3. Finally, in Sect. 8.4 we conclude.

8.2 Weather Risk Management and the Geographical/Spatial Risk

It is often cited that one of the major problems facing weather-based derivatives is that of the weather basis risk. However, only few studies take into consideration the impact of basis risk. Geographical basis risk results from the distance between the hedging company and the site at which the weather measurement takes place.

Although the list of contracts in CME and the locations that these contracts are traded is constantly expanding, it is not possible to cover every location in the world. For example, in the USA, the CME trades temperature derivatives in 24 states.

As a result, the question of weather hedging for the firms in the remaining states arises.

Geographical basis risk can reach critical levels in some cases (Rohrer 2004). As the distance between a hedging company and the measurement weather station of the weather derivative increases, the demand for weather derivative decreases (East 2005; Edwards and Simmons 2004). However, great differences

Table 8.1 Correlation of temperature (original and deseasonalized) measured in five meteorological stations in Germany for 2000–2009

	Berlin-Dahlem	Bremen	Munich	Hannover	Berlin-Tempelhof
Berlin-Dahlem	1.0000	0.9699	0.9388	0.9753	0.9983
Bremen	0.8845	1.0000	0.9249	0.9924	0.9699
Munich	0.7473	0.7141	1.0000	0.9361	0.9397
Hannover	0.9110	0.9719	0.7636	1.0000	0.9764
Berlin-Tempelhof	0.9933	0.8812	0.7501	0.9127	1.0000

The upper diagonal part of the matrix corresponds to the original time series, while the lower diagonal part of the matrix refers to the deseasonalized time series

can be observed even in small distances. Often, in regions like coastal or mountainous areas, microclimates are created (Brockett et al. 2005; Manfredo and Richards 2009).

One solution is the OTC market. In the OTC market, a firm eliminates the basis risk. However, measurement and monitoring may be costly. At the same time, the firm has to develop a fair price for the contract. In addition, the OTC market offers limited liquidity. At the same time, in OTC market, there is always the danger of default or credit risk. Finally, usually it is very hard to close your position in a contract bought in OTC market.

Another solution to reduce the effects of geographical risk is the use of basis derivatives. Basis derivatives are written on the difference between the indexes on two different meteorological stations. For example, on one station that weather derivatives are traded on CME, and one station more close to the location of the firm that want to hedge the weather exposure (Considine 2000). Basis weather derivatives reduce the basis risk; however, create some credit risk as basis derivatives are OTC contracts (Brockett et al. 2005).

Trading a weather derivative in an exchange offers liquidity and fair priced contracts. At the same time, the exchange eliminates the default risk.

In the rest of the section, we will examine the geographical/spatial risk between five meteorological stations in Germany. Two of them are located in Berlin. The first one in Dahlem, and the second one is in Tempelhof. The distance between these two stations is less than 7 km. The remaining three stations are located in Bremen, Munich, and Hannover. The distances between theses cities range from 100 km between Bremen and Hannover to 584 km between Bremen and Munich. All distances were measured in straight lines between the cities. Hence, the spatial risk will be examined in cases of large and small distances.

First, we examine the correlation of the temperature. In Table 8.1, the correlation of the temperature time series in the five meteorological stations is presented. The upper part of Table 8.1 corresponds to the original time series, while the lower part refers to the detrended and deseasonalized temperature series. It is clear from Table 8.1 that the correlation is very high. Even between Bremen and Munich where the distance is 620 km, the correlation is 0.9249, while the correlation between the two stations in Berlin is 0.9983. However, as it was shown on the previous chapters, temperature is dominated by a strong seasonal component.

Table 8.2 Correlation of the HDD's for January measured in five meteorological stations in Germany for 2000–2009

	Berlin-Dahlem	Bremen	Munich	Hannover	Berlin-Tempelhof
Berlin-Dahlem	1.0000	0.9762	0.9290	0.9848	0.9984
Bremen	0.9762	1.0000	0.9160	0.9932	0.9686
Munich	0.9290	0.9160	1.0000	0.9390	0.9236
Hannover	0.9848	0.9932	0.9390	1.0000	0.9783
Berlin-Tempelhof	0.9984	0.9686	0.9236	0.9783	1.0000

Table 8.3 Correlation of precipitation measured in five meteorological stations in Germany for 2000–2009

	Berlin-Dahlem	Bremen	Munich	Hannover	Berlin-Tempelhof
Berlin-Dahlem	1.0000	0.3597	0.1635	0.3938	0.9178
Bremen	0.3597	1.0000	0.1375	0.6353	0.3812
Munich	0.1635	0.1375	1.0000	0.1444	0.1806
Hannover	0.3938	0.6353	0.1444	1.0000	0.3962
Berlin-Tempelhof	0.9178	0.3812	0.1806	0.3962	1.0000

It is expected that the seasonal mean will affect the correlation between the temperature series. Hence, we next examine the correlation of the detrended and deseasonalized temperature between the five stations. The results are presented in the lower part of Table 8.1. In this case, the correlation between Bremen and Munich is significantly reduced to 0.7141. On the other hand, the correlation between the stations in Berlin and Hannover remain over 0.90. In addition, the correlation between Berlin-Dahlem and Berlin-Tempelhof is almost unaffected since from 0.9983 it slightly reduced to 0.9933.

One might argue that it is more interesting, from the practitioners point of view, to examine the correlation of the temperature indices instead of the temperature itself. Table 8.2 presents the correlation of the HDD's for January measured in the five meteorological stations in Germany. More precisely, the HDD's over January for 10 years, 2000–2009, was estimated in each city and then the correlation was calculated. Again, close meteorological stations have higher correlation, while the level of correlation drops as one moves farther away. However, it is clear that regions of high correlation extend out to for hundred of kilometers. More precisely, the correlation ranges from 0.9160 between Bremen and Munich to 0.994 between Dahlem and Tempelhof.

The previous example shows that temperature risk can be hedged using nearby stations to the location where the end user has temperature exposure. However, this is not the case for the remaining weather variables. Since precipitation, snowfall, rainfall, and wind speed are local phenomena, it is expected to have lower correlation. Table 8.3 presents the correlation of precipitation between the five meteorological stations in Germany. The correlation was estimated for the years 2000–2009.

Table 8.4 Monthly temperature correlation between Berlin-Tempelhof and four meteorological stations in Germany for 2000–2009

	Berlin-Dahlem		Bremen		Munich		Hannover	
	Original	Deseason	Original	Deseason	Original	Deseason	Original	Deseason
Jan	0.9964	0.9965	0.7666	0.7759	0.6852	0.6871	0.8556	0.8613
Feb	0.9971	0.9968	0.7738	0.7068	0.7987	0.7872	0.8538	0.8112
Mar	0.9913	0.9943	0.7654	0.8609	0.7505	0.8110	0.7496	0.8528
Apr	0.9881	0.9865	0.4730	0.4766	0.1685	0.3222	0.6211	0.6391
May	0.9786	0.9718	0.7757	0.6701	0.6700	0.6263	0.8846	0.8350
Jul	0.9935	0.9897	0.9055	0.8728	0.4578	0.3726	0.9428	0.9245
Jun	0.9946	0.9950	0.8830	0.8863	0.7903	0.7887	0.9318	0.9344
Aug	0.9860	0.9853	0.6996	0.6743	0.7743	0.7775	0.8082	0.8008
Sep	0.9769	0.9650	0.8037	0.6919	0.7207	0.6566	0.8801	0.8322
Oct	0.9958	0.9942	0.8411	0.8247	0.8943	0.8368	0.9279	0.9248
Nov	0.9970	0.9984	0.7308	0.8650	0.6891	0.8220	0.7888	0.8903
Dec	0.9985	0.9984	0.8872	0.8730	0.8056	0.7951	0.9377	0.9294
Mean	0.9912	0.9893	0.7755	0.7649	0.6838	0.6903	0.8485	0.8530
StD	0.0073	0.0108	0.1150	0.1243	0.1941	0.1739	0.0949	0.0827

Original: Original temperature series
Deseason: Deseasonalized temperature series
StD: Standard deviation

The smallest correlations are observed between distant cities as it was expected. Hence, the correlation between Munich and Tempelhof, Dahlem, Hannover, and Bremen is only 0.1806, 0.1635, 0.1444, and 0.1375, respectively. On the other hand, the correlation between the two stations in Berlin is 0.9178. Even if the distance between these two stations is only 7 km, we observe that the correlation for precipitation is significant smaller than in the case of temperature.

Hence, our analysis so far shows that measurements from local phenomena such as snowfall, rainfall or precipitation, and wind speed significantly differ even in distances of few kilometers. The sum of rainfall for the first 3 months of 2000 at Berlin-Dahlem is 175.7 mm, while it is 192.2 mm for the same period at Berlin-Tempelhof. In other words, even if the two stations have a distance of only 7 km, the difference of the index is about 10 %. Hence, the outcome of a hedging strategy will provide 10 % less revenues.

Next, we will further investigate the spatial risk in temperature derivatives. In this section, we will examine the correlation of temperature in monthly basis. The data from the 2000 to 2009 period are grouped in monthly basis, and then the correlation of each meteorological station to a reference station is estimated. The reference station is the Berlin-Tempelhof station. Table 8.4 presents the monthly temperature correlation between Berlin-Tempelhof and the remaining four stations. Table 8.4 presents the results from the original and the deseasonalized data. A closer inspection of Table 8.4 reveals that only the correlation between Dahlem and Tempelhof is almost constant. On the other hand, the correlation for the remaining stations significantly changes over different months with great variations. Moreover, comparing the results from Tables 8.1 and 8.4, we conclude that the

Table 8.5 Correlation between the temperature and the precipitation in five meteorological stations in Germany for the period 2000–2009

	Temperature				
Precipitation	Berlin-Dahlem	Bremen	Munich	Hannover	Berlin-Tempelhof
Berlin-Dahlem	0.0707	0.0677	0.0877	0.0699	0.0687
Bremen	0.0869	0.0687	0.0960	0.0818	0.0839
Munich	0.1087	0.0827	0.0939	0.0848	0.1063
Hannover	0.0505	0.0375	0.0646	0.0394	0.0484
Berlin-Tempelhof	0.0615	0.0588	0.0810	0.0626	0.0587

correlation is significantly smaller when it is estimated in monthly basis rather than in the case of estimating the correlation for the whole 10-year period.

Since temperature derivatives are usually written on a period of a month or a season, an analysis in monthly basis is more of an interest since estimating the correlation for the whole period may provide an overestimation of the correlation. Hence, the results from Table 8.4 indicate that spatial risk can be significant and difficult to estimate even in the case of temperature.

Finally, the correlation between the precipitation and the temperature is examined. Table 8.5 presents the correlation between precipitation and temperature for the five meteorological stations in Germany for the period 2000–2009. Intuitively, a negative correlation between these two variables is expected. However, the results presented in Table 8.5 indicate a very low positive correlation between these two weather variables.

It is clear that for an effective hedging strategy, a hedger will buy a weather derivate with the smallest geographical basis risk (Manfredo and Richards 2009). Our previous analysis indicates that the contracts written on a meteorological that is closest to the firm is the best choice. However, in some cases, a weather derivative written on multiple nearby stations is more appropriate (Martin et al. 2001).

8.2.1 A Spatial Model for Temperature

The previous example shows that temperature risk can be hedged using nearby stations to the location where the end user has temperature exposure. On the other hand, precipitation is much localized, and there is very little correlation even between neighboring cities. Hence, in this section, we focus on building a spatial model for temperature. The new model is a modification of the model presented in Chap. 5.

First, we assume that the temperature on each station is given by the model proposed in Chap. 5. Hence, at location i, the temperature time series at time t is denoted by $T^i(t)$. Hence, the mean reverting O–U stochastic model for temperature at location i is given by

$$dT^i(t) = dS^i(t) + \kappa^i(t)\big(T^i(t) - S^i(t)\big)dt + \sigma^i(t)dB^i(t), \tag{8.1}$$

where $S^i(t)$ is the seasonal mean of temperature at meteorological station i and the seasonal variance is denoted by $\sigma^{2,i}$. Both $S^i(t)$ and $\sigma^{2,i}$ are the same as presented in Chap. 5. For simplicity, we assume that $B^i(t)$ is BM process. However, as it was presented in the Chap. 5, there are cases where the hyperbolic distribution provides a better fitting to the residuals. In such cases, the $B^i(t)$ is replaced by a Lévy motion-driven process. We also assume that $B^i(t)$ is independent in time, and the correlation between two difference locations i, j is given by

$$\rho(i,j) = corr\big\{B^i(t), B^j(t)\big\}, \tag{8.2}$$

for all $i, j \in D$. We assume that the cross-correlation in time and space is equal to zero.

The estimating procedure is similar as the one described in Chap. 5 where only one location was considered. The linear trends and the seasonal mean can be estimated simultaneously for all locations. Next, as in the case of only one location, the autoregressive models can be estimated. Finally, the seasonal variance is estimated from the data.

8.3 Weather Risk Management and the Basis Risk

It is estimated that nearly 30 % of the US economy and 70 % of the US companies are affected by weather (CME 2005). More precisely, agriculture, construction, energy, entertainment, governments, insurance, manufacturing, offshore, retailing, and transportation are sectors of the economy that are heavily affected by the weather. Companies from these sectors can use weather derivatives to hedge against weather exposure. However, the quantification of the weather impact in each company must be estimated carefully in order to have an effective hedging strategy.

The concept of basis is less clear when weather derivatives are used to hedge volumetric risks (e.g., crop yields) associated with unfavorable weather conditions (Rohrer 2004; Manfredo and Richards 2009). The payoff of a weather derivative depends on a weather index and not on the actual amount of money lost due to weather; it is unlikely that the payoff will compensate exactly for the money lost. The potential for such a difference is known as basis risk. In general, basis risk is smallest when the financial loss is highly correlated with the weather and when contracts of the optimum size and structure, based on the optimum location, are used for hedging. For a company deciding how to hedge its risk, there is usually a trade-off between basis risk and the price of the weather hedge (Jewson et al. 2005). It is ultimately the behavior of the basis that determines the effectiveness of a hedge (Manfredo and Richards 2009). In practice, it may be difficult to find a counterpart

willing to offer the index or the combination of indexes best suited for the exposure to be hedged. Moreover, in most cases, there is a residual volume risk not related to weather.

Similarly, in agriculture sector, crop yields decrease due to unfavorable temperature and rainfall conditions during the critical growing periods (Manfredo and Richards 2009; Turvey 2001). It is clear that the correlation between the weather index and the yield is important. For example, weather risk can affect both the quantity and the quality of the grapes destined for winemaking (Zara 2010). The correlation itself depends on the definition of the index and the considered product. Second, the quantification of the weather–yield relationship is subject to estimation errors (Odening and Musshoff 2007). Temperature and precipitation are among the most important factors contributing to yield variability. However, establishing a link between yields and weather is not a simple task. Part of the problems stems from the tendency of this relationship to be localized and crop dependent (Vedenov and Barnett 2004).

In the case of hedging crop yields with weather derivatives, basis risk not only depends on the spatial risk but also on the relationship between yields and weather conditions. This relationship is not always linear (Manfredo and Richards 2009).

In Turvey (2001), a simple model that relates rainfall and crop yield is presented. The model is based on the Cobb-Douglas production function (Cobb and Douglas 1928). Hence, the crop yield is given by

$$Y = AR^{\alpha}H^{\beta}, \tag{8.3}$$

where Y is the crop yield, R is cumulative daily rainfall during the critical growing period, H is the growing degree days – similar to HDDs but with different base temperature – and α and β are the output elasticities of rainfall and growing degree days.

Previous studies have shown that electricity consumption is mostly affected by temperature (Engle et al. 1992; Li and Sailor 1995; Sailor and Munoz 1997; Valor et al. 2001; Henley and Peirson 1998; Peirson and Henley 1994; Pirrong and Jermakyan 2008; Stoft 2002; Zanotti et al. 2003). In Valor et al. (2001) and Moral-Carcedoa and Vicens-Otero (2005), the daily air temperature and electricity load in Spain was investigated. In Engle et al. (1986) and (1992), temperature was used as an explanatory variable to model electricity consumption.

Hence, an energy producer can hedge the profit of the electricity consumption using temperature indexes. For example, a simple model is given by

$$Y = a + bP + cT + \varepsilon, \tag{8.4}$$

where Y is the profit of the energy company for a month, P is the average energy prices for the month, T corresponds to the relevant temperature index, for example, HDDs or CDDs for the month, and ε is the error term.

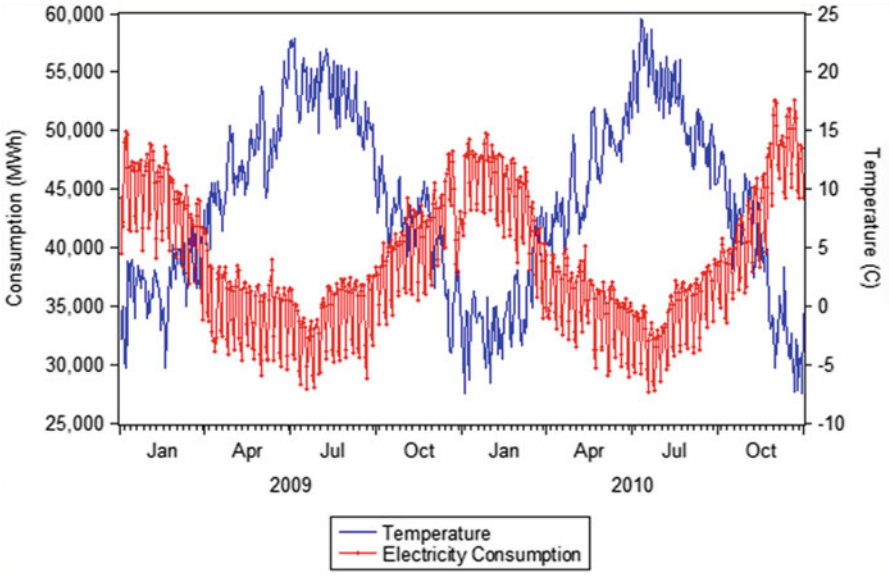

Fig. 8.1 Electricity consumption and temperature in Copenhagen

A similar but more advance framework is proposed by Engle et al. (1992) and Pardo et al. (2002). This framework incorporates various dummy variables to model the day and the monthly seasonality of the electricity demand. Hence, the electricity consumption was modeled by

$$E_t = c_1 + a_1 t + \beta_1 HDD_t + \gamma_1 CDD_t + \sum_{i=2}^{7} \delta_{1i} W_{it} + \omega_1 H_t + \kappa_1 H_{t-1} + \sum_{j=2}^{12} \lambda_{1j} M_{jt} + \varepsilon_{1t}. \quad (8.5)$$

The electricity demand depends on CDDs and HDDs. Moreover, six dummy variables are introduced to capture the day effect of the electricity consumption. The index i represents all days in the week except the base day Monday, and W_{it} is 1 if t belongs to i and 0 otherwise. The index j indicates the month, and M_{jt} is 1 if t belongs to j and 0 otherwise, and January is the base month. The parameter H_t is 1 if t is a holiday and zero otherwise. Similarly, H_{t-1} is 1 if t corresponds to a day following a holiday and 0 otherwise.

Figure 8.1 shows the relationship between the electricity consumption and the temperature in Copenhagen. Similarly, Fig. 8.2 shows the relationship between the electricity consumption and the HDDs in Copenhagen. It is clear that these two variables are highly correlated. The correlation between the consumption and the temperature or the HDD's in Copenhagen is estimated to be -0.83 and 0.84, respectively. Hence, an energy provider can use weather derivatives to hedge a significant part of its volumetric risk. This volumetric risk arises from unfavor

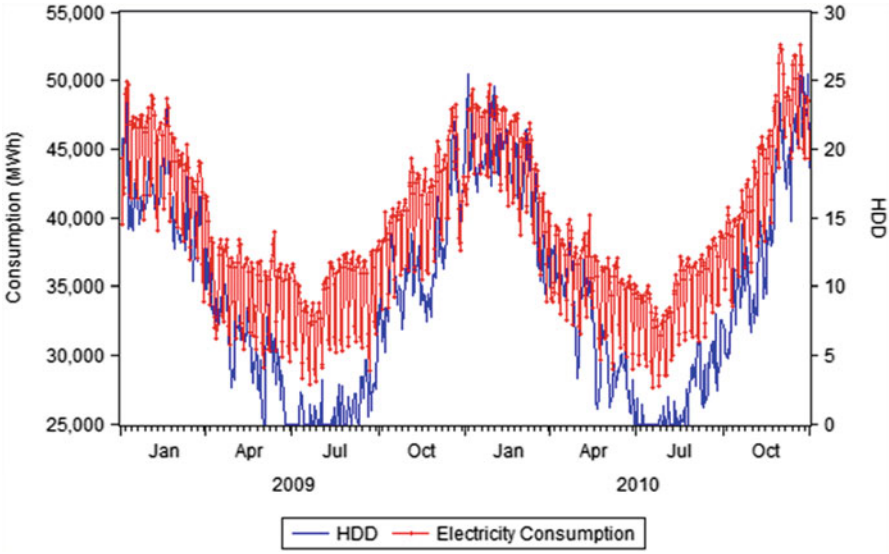

Fig. 8.2 Electricity consumption and HDD in Copenhagen

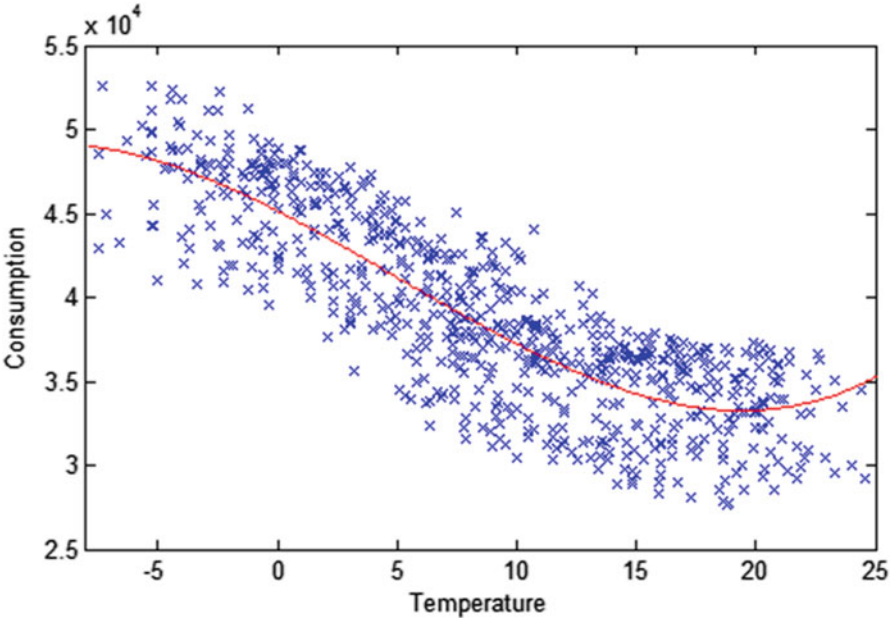

Fig. 8.3 Scatter plot between electricity consumption and temperature in Copenhagen

weather conditions. In the case of an energy provider, both the price and the volume of sales are affected. A closer inspection of Figs. 8.1 and 8.2 reveals that the lower the temperature, the more electricity is consumed for heating. On the other hand, as

the temperature increases beyond a "normal comfortable" level, the electricity consumption starts to increase again. In other words, electricity consumption increases at both low and high temperatures (Engle et al. 1986). Observing Fig. 8.3, we conclude that the electricity consumption is minimized when the temperature is between 17 °C and 20 °C. This is further confirmed by Fig. 8.2 where the electricity consumption and HDDs are plotted. It is clear that the electricity consumption almost perfectly follows the HDDs.

8.4 Conclusions

In this chapter, the concept of basis risk was presented in the context of weather derivatives. Unlike in the case of classical financial markets, the notion of basis risk is different for weather derivatives. Risk managers often face unique basis risks arising from both the choice of weather station where a derivative contract is written, as well as the relationship between the hedged volume and the underlying weather index. The first one is referred as spatial or geographical basis risk while to the second as basis risk.

So far, basis risk has not sufficiently studied. Hence, investors who lack of understanding of weather basis risk avoid the use of weather derivative for hedging purposes. As a result, basis risk contributes to liquidity problems in the weather market (Manfredo and Richards 2009).

In our analysis, first, the correlation of temperature and precipitation between five cities in Germany was studied. The correlation coefficients for temperature between the five German cities are very high even in the cases of distant cities. Similarly, the correlation coefficients for temperature-based indexes are also high. Next, we examined the changes in the correlation coefficients in monthly basis. Our results indicate that the correlation coefficients exhibit significant variation over the months. Hence, one must be very careful when estimating the basis risk for his hedging strategy.

Next, the correlation of precipitation levels between various meteorological stations was examined. Our results indicate that, unlike temperature, precipitation is localized, and the correlations coefficients are very low and significantly differ even when the distance is very small. Finally, the correlation between precipitation and temperature was examined. We found that the correlation coefficients are very small. Finally, a spatial model for temperature was presented. The new model is a system of stochastic equations. Each equation represents each meteorological station. The parameters of the system of equations are simultaneously estimated.

In the second part of the chapter, the basis risk between temperature and electricity was estimated. Our results indicate that temperature and electricity are highly correlated. Also, the electricity consumption increases when we have a cold winter or a hot summer. In other words, the electricity consumption is smaller when the temperature is between 17° and 20°.

References

Bataller MM, Tornero AT, Micó EV (2006) CO_2 prices, energy and weather. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=913964. Accessed July 2006

Brockett PL, Wang M, Yang C (2005) Weather derivatives and weather risk management. Risk Manage Insur Rev 8(1):127–140

CME (2005) An introduction to CME weather products. http://www.cme.com/edu/res/bro/cmeweather. Accessed Jan 2007

Cobb CW, Douglas PH (1928) A theory of production. Am Econ Rev 18:139–165

Considine G (2000) Introduction to weather derivatives. http://www.cme.com/files/weather.pdf. Accessed Jan 2006

East M (2005) Issues of geographical basis risk in weather derivatives for Australian wheat farmers. In: 49th annual conference of the Australian Agricultural & Resource Economics Society, Coffs Harbour, February 2005

Edwards M, Simmons P (2004) Preliminary results for the measurement of willingness to pay for climate derivatives. In: 48th annual conference of the Australian Agricultural & Resource Economics Society, Melbourne, February 2004

Engle RF, Granger CWJ, Rice J, Weiss A (1986) Semiparametric estimates of the relation between weather and electricity sales. J Am Stat Assoc 81(394):310–320

Engle RF, Mustafa C, Rice J (1992) Modelling peak electricity demand. J Forecasting 11:241–251

Gabbi G, Zanotti G (2005) Climate variables and weather derivatives: gas demand, temperature and the cost of weather for a gas supplier. http://www.efmaefm.org/efma2005/papers/285-zanotti_paper.pdf. Accessed 30 Oct

Geman H (1999) Insurance and weather derivatives. RISK Books, London

Henley A, Peirson J (1998) Residential energy demand and the interaction of price and temperature: British experimental evidence. Energy Econ 20:157–171

Hull CJ (2003) Option, futures and other derivatives, 5th edn. Prentice Hall, New Jersey

Jewson S, Brix A, Ziehmann C (2005) Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations. Cambridge University Press, Cambridge, UK

Li X, Sailor DJ (1995) Electricity use sensitivity and climate and climate change. World Resour Rev 3:334–346

Manfredo MR, Richards TJ (2009) Hedging with weather derivative: a role for options in reducing basis risk. Appl Financial Econ 19(2):87–97

Martin WS, Barnett JB, Coble HK (2001) Developing and pricing precipitation insurance. J Agricul Resour Econ 26(1):261–274

Moral-Carcedoa J, Vicens-Otero J (2005) Modelling the non-linear response of Spanish electricity demand to temperature variations. Energy Econ 27:477–494

Odening M, Musshoff OXW (2007) Analysis of rainfall derivatives using daily precipitation models: opportunities and pitfalls. Agricul Finance Rev 67:135–156

Pardo A, Meneu V, Valor E (2002) Temperature and seasonality influences on Spanish electricity load. Energy Econ 24(1):55–70. doi:10.1016/s0140-9883(01)00082-2

Peirson J, Henley A (1994) Electricity load and temperature issues in dynamic specification. Energy Econ 16:235–243

Pirrong C, Jermakyan M (2008) The price of power: the valuation of power and weather derivatives. J Bank Finance 32(12):2520–2529

Rohrer MB (2004) The relevance of basis risk in the weather derivatives market. In: Constantino M, Brebbia CA (eds) Computational finance and its applications. WIT Press, Southampton, UK, pp 67–76

Sailor DJ, Munoz R (1997) Sensitivity of electricity and natural gas consumption to climate in the USA – meteorology and results for eight states. Energy Int J 22:987–998

Stoft S (2002) Power system economics. Wiley, New York

Turvey CG (2001) Weather derivatives for specific event risks in agriculture. Rev Agricul Econ 23 (2):333–351

Valor E, Meneu V, Caselles V (2001) Daily air temperature and electricity load in Spain. Am Meteorol Soc 40:1413–1421

Vedenov DV, Barnett BJ (2004) Efficiency of weather derivatives as primary crop insurance instruments. J Agricul Resour Econ Assoc 29(3):387–403

Woodard JD, Garcia P (2008) Basis risk and weather hedging effectiveness. Agri Finance Rev 68:111–124

Zanotti G, Gabbi G, Laboratore D (2003) Climate variables and weather derivatives: gas demand, temperature and seasonality effects in the Italian case http://ssrn.com/abstract=488745. Accessed Jan 2003

Zara C (2010) Weather derivatives in the wine industry. Int J Wine Bus Res 22(3):222–237

Chapter 9
Pricing the Power of Wind

9.1 Introduction

In this book, the various aspects of weather derivative have been presented. So far, we have focused on modeling and pricing temperature derivatives. In this chapter, we focus on wind derivatives. A model for the dynamics of the wind-generating process using a nonparametric nonlinear wavelet network is presented. Moreover, the proposed methodology is compared against alternative methods, proposed in prior studies, and the pricing equations for wind futures are provided.

The notional value of the traded wind-linked securities is around $36 million indicating a large and growing market (WRMA 2009). However, after the close of the US Future Exchange, wind derivatives are traded in the Chicago Climate Futures Exchange and in the OTC. The demand from these derivatives exists. However, investors hesitate to enter into wind contracts. The main reasons of the slow growth of the wind market compared to temperature contracts are the difficulty in accurately modeling wind and the challenge to find a reliable model for valuing related contracts. As a result, there is a lack of reliable valuation framework that makes financial institutions reluctant to quote prices over these derivatives.

The aim of this chapter is to model and price wind derivatives. Wind derivatives are standardized products that depend only on the daily average wind speed measured by a predefined meteorological station over a specified period and can be used by wind (and weather in general)-sensitive business such as wind farms, transportation companies, construction companies, and theme parks to name a few. The financial contracts that are traded are based on the simple daily average wind speed index, and this is the reason that we choose to model only the dynamics of the daily average wind speeds. The revenues of each company have a unique dependence and sensitivity to wind speeds. Although wind derivatives and weather derivatives can hedge a significant part of the weather risk of the company always, some basis risk will still exist which must be hedged from each company separately. This can be done either by defining a more complex wind index or by taking an additional hedging position.

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_9,

Wind is free, renewable, and environmentally friendly source of energy (Billinton et al. 1996). While the demand for electricity is closely related to the temperature, the electricity produced by a wind farm is dependent on the wind conditions. The risk exposure of the wind farm depends on the wind speed and the wind direction and in some cases on the wind duration of the wind speed at certain level. However, modern wind turbines include mechanisms that allow turbines to rotate on in the appropriate wind direction (Caporin and Pres 2010). However, the underlying wind indices do not account for the duration of the wind speed at certain level but rather, usually, measure the average daily wind speed. Hence, the parameter of the duration of the wind speed at certain level is not considered in our daily model. Hence, the risk exposure of a wind farm can be analyzed by quantifying only the wind speed. On the other hand, companies like wind farms that its revenues depend on the duration effect can use an additional hedging strategy that includes this parameter. This can be done by introducing a second index that measures the duration. A similar index for temperature is the frost day index.

Many different approaches have been proposed so far for modeling the dynamics of the wind speed process. The most common is the generalized ARMA approach. There have been a number of studies on the use of linear ARMA models to simulate and forecast wind speed in various locations (Saltyte-Benth and Benth 2010; Billinton et al. 1996; Caporin and Pres 2010; Castino et al. 1998; Daniel and Chen 1991; Huang and Chalabi 1995; Kamal and Jafri 1997; Martin et al. 1999; Tol 1997; Torres et al. 2005). In Kavasseri and Seetharaman (2009), a more sophisticated ARFIMA model was used. Most of these studies did not consider in detail the accuracy of the wind speed forecasts (Huang and Chalabi 1995). On the other hand, Ailliot et al. (2006) apply an AR with time-varying coefficients to describe the space–time evolution of wind fields. In Benth and Saltyte-Benth (2009), a stochastic process called CAR model is introduced in order to model and forecast daily wind speeds. Finally, in Nielsen et al. (2006), various statistical methods were presented for short-term wind speed forecasting. Sfetsos (2002) argues about the use of linear or meteorological models since their prediction error is not significantly lower than the elementary persistent method. Alternatively, some studies use space–state models to simultaneously fit the speed and the direction of the wind (Castino et al. 1998; Cripps et al. 2005; Haslett and Raftery 1989; Martin et al. 1999; Tolman and Booij 1998; Tuller and Brett 1984).

Alternatively to the linear models, artificial intelligence was applied in wind speed modeling and forecasting. In Alexiadis et al. (1998), Barbounis et al. (2006), Beyer et al. (1994), More and Deo (2003), Sfetsos (2000), Mohandes et al. (1998), and Sfetsos (2002), neural networks were applied in order to model the dynamics of the wind speed process. In Mohandes et al. (2004), support vector machines were used while in Pinson and Kariniotakis (2003), and fuzzy neural networks were applied.

Depending on the application, wind modeling is based on hourly (Ailliot et al. 2006; Castino et al. 1998; Daniel and Chen 1991; Kamal and Jafri 1997; Martin et al. 1999; Sfetsos 2000, 2002; Torres et al. 2005; Yamada 2008), daily (Benth and Saltyte-Benth 2009; Billinton et al. 1996; Caporin and Pres 2010; Huang and Chalabi 1995; More and Deo 2003; Tol 1997), weekly, or monthly basis (More

and Deo 2003). When the objective is to hedge against electricity demand and production, hourly modeling is used, while for weather derivative pricing, the daily method is used. More rarely, weekly or monthly modeling is used in order to estimate monthly wind indexes. Since we want to focus on weather derivative pricing, the daily modeling approach is followed; however, the proposed method can be easily adapted in hourly modeling too.

Wind speed modeling is much more complicated than temperature modeling since wind has a direction and is greatly affected by the surrounding terrain such as building and trees (Jewson et al. 2005). However, in Benth and Saltyte-Benth (2009), it is shown that wind speed dynamics share a lot of common characteristics with the dynamics of temperature derivatives as it was found on Benth and Saltyte-Benth (2007), Zapranis and Alexandridis (2008, 2009, 2011), and Alexandridis (2010). In this context, we use a mean-reverting Ornstein–Uhlenbeck stochastic process to model the dynamics of the wind speed dynamics where the innovations are driven by a Brownian motion. The statistical analysis reveals seasonality in the mean and variance. In addition, we use a novel approach to model the autocorrelation of the wind speeds. More precisely, a WN is applied in order to capture accurately the autoregressive characteristics of the wind speeds.

The evaluation of the proposed methodology against alternative modeling procedures proposed in prior studies indicates that WNs can accurately model and forecast the dynamics and the evolution of the speed of the wind. The performance of each method was evaluated in-sample as well as out-of-sample and for different time periods.

The rest of the chapter is organized as follows: in Sect. 9.2, a statistical analysis of the wind speed dynamics is presented. In Sects. 9.2.1 and 9.2.2, a linear ARMA model and a nonlinear nonparametric WN is applied, respectively. The evaluation of the studied models is presented in Sect. 9.2.3. In Sect. 9.3, we derive the pricing formulas for future derivatives written on the wind index. Finally, in Sect. 9.4, we conclude.

9.2 Modeling the Daily Average Wind Speed

In this section, we derive empirically the characteristics of the daily average wind speed (DAWS) dynamics in New York, USA. The data were collected from NOAA[1] and correspond to DAWSs. The wind speed is measured in 0.1 knots. The measurement period is between 1st January 1988 and 28th February 2008. The first 20 years are used for the estimation of the parameters, while the remaining 2 months are used for the evaluation of the performance of the proposed model. In order for each year to have the same number of observations, the 29th of February

[1] http://www.noaa.gov

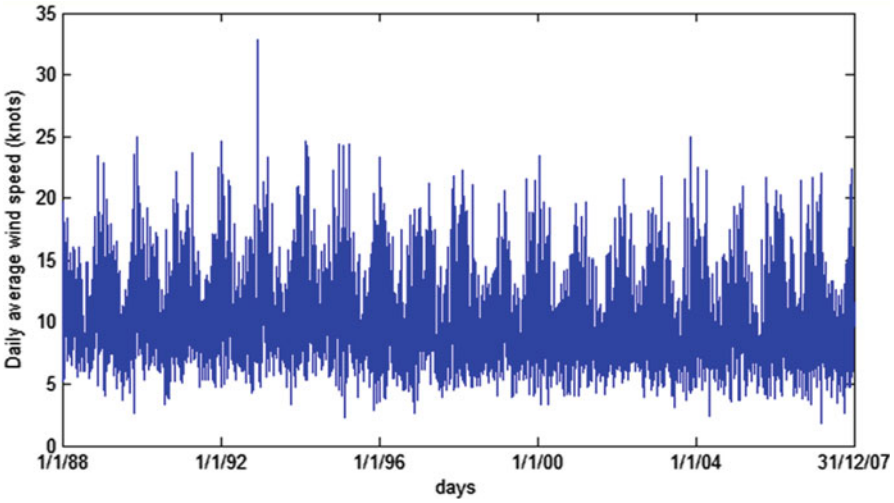

Fig. 9.1 Daily average wind speed for New York

Table 9.1 Descriptive statistics of the wind in New York

	Mean	Median	Max	Min	St. dev.	Skew	Kurt	J–B	*p* value
Original	9.91	9.3	32.8	1.8	3.38	0.96	4.24	1595.41	0
Transformed	2.28	2.3	3.6	0.6	0.34	0.00	3.04	0.51	1

J–B Jarque–Bera statistic, *P value p* values of the J–B statistic

is removed from the data resulting to 7,359 data points. The time series is complete without any missing values.

In Fig. 9.1, the DAWSs for the first 20 years are presented. A closer inspection of Fig. 9.1 reveals seasonality. The descriptive statistics of the in-sample data are presented in Table 9.1. The values of the data are always positive and range from 1.8 to 32.8 with mean around 9.91. The descriptive statistics of the DAWSs indicate that there is a strong positive kurtosis and skewness, while the normal hypothesis is rejected based on the Jarque–Bera statistic. The same conclusion can be reached observing the first part of Fig. 9.2 where the histogram of the DAWSs is represented. It is clear that the density of the DAWSs has positive skewness and excess kurtosis. Hence, the distribution of DAWSs deviates significantly from the normal, and it is not symmetrical. In literature, the Weibull or Rayleigh (which is a special case of the Weibull) distributions were proposed (Benth and Saltyte-Benth 2009; Saltyte-Benth and Benth 2010; Brown et al. 1984; Celik 2004; Daniel and Chen 1991; Garcia et al. 1998; Justus et al. 1978; Kavak Akpinar and Akpinar 2005; Nfaoui et al. 1996; Torres et al. 2005; Tuller and Brett 1984). In addition, some studies propose the use of the lognormal distribution (Benth and Saltyte-Benth 2009; Garcia et al. 1998) or the Chi-square (Dorvlo 2002). Finally, in Jaramillo and Borja (2004), a bimodal Weibull and Weibull distribution are used. However, empirical studies favor the use of the Weibull distribution (Celik 2004; Tuller and Brett 1984).

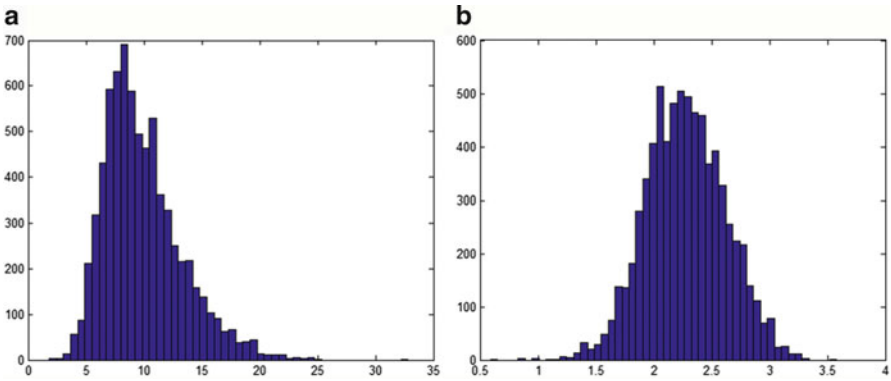

Fig. 9.2 Histogram of the (**a**) original and (**b**) Box–Cox transformed data

A closer inspection of part (a) of Fig. 9.2 reveals that the DAWSs in New York follow a Weibull distribution with scale parameter $\lambda = 11.07$ and shape parameter $k = 3.04$. Following Benth and Saltyte-Benth (2009), Brown et al. (1984), Daniel and Chen (1991), in order to symmetrize the data, the Box–Cox transform is applied. The Box–Cox transformation is given by

$$W^{(l)} = \begin{cases} \dfrac{W^l - 1}{l} & l \neq 0 \\ \ln(W) & l = 0 \end{cases} \tag{9.1}$$

where $W^{(l)}$ is the transformed data. The parameter l is estimated by maximizing the log-likelihood function. Note that the log transform is a special case of the Box–Cox transform with $l = 0$. The optimal l of the Box–Cox transform for the DAWSs in New York is estimated to be 0.014. In the second part of Fig. 9.2, the histogram of the transformed data can be found, while the second row on Table 9.1 shows the descriptive statistics of the transformed data.

The DAWSs exhibit a clear seasonal pattern which is preserved in the transformed data. The same conclusion can be reached by examining the ACF of the DAWS in the first part of Fig. 9.3. In Benth and Saltyte-Benth (2009), Saltyte-Benth and Benth (2010), and Caporin and Pres (2010), the seasonality was captured by series of sinusoids. As in Zapranis and Alexandridis (2008, 2009, 2011) and as it was presented in the previous chapters for the case of temperature process, the seasonal effects are modeled by a truncated Fourier series given by

$$S(t) = a_0 + b_0 t + \sum_{i=1}^{I_1} a_i \sin(2\pi i(t - f_i)/365) + \sum_{j=1}^{J_1} b_j \sin\big(2\pi j\big(t - g_j\big)/365\big). \tag{9.2}$$

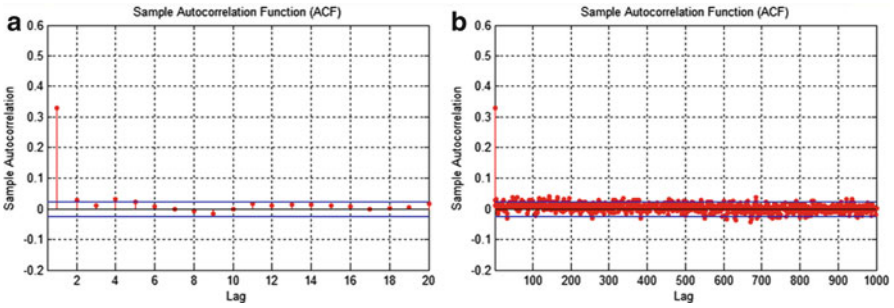

Fig. 9.3 The autocorrelation function of the transformed DAWSs in New York (**a**) before and (**b**) after removing the seasonal mean

Table 9.2 Estimated parameters of the seasonal component

a_0	b_0	a_1	f_1	b_1	g_1
2.3632	−0.000024	0.0144	827.81	0.1537	28.9350

In addition, we examine the data for a linear trend representing the global warming or the urbanization around the meteorological station. First, we quantify the trend by fitting a linear regression to the DAWS data. The regression is statistically significant with intercept $a_0 = 2.3632$ and $b_0 = -0.000024$ indicating a slight decrease in the DAWSs. Next, the seasonal periodicities are removed from the detrended data. The remaining statistically significant parameters of (9.2) with $I_1 = J_1 = 1$ are presented in Table 9.2. As it is shown on the second part of Fig. 9.3, the seasonal mean was successfully removed. The same conclusion was reached in previous studies for daily models for both temperature and wind (Alexandridis 2010; Zapranis and Alexandridis 2008, 2009, 2011; Benth et al. 2009; Benth and Saltyte-Benth 2005, 2007, 2009; Benth et al. 2007).

9.2.1 The Linear ARMA Model

In literature, various methods for studying the statistical characteristics of the wind speed, in daily or hourly measurements, were proposed. However, the majority of the studies utilize variations of the general ARMA model (Ailliot et al. 2006; Billinton et al. 1996; Brett and Tuller 1991; Daniel and Chen 1991; Huang and Chalabi 1995; Kamal and Jafri 1997; Lei et al. 2009; Nfaoui et al. 1996; Rehman and Halawani 1994; Torres et al. 2005). In this chapter, we will first estimate the dynamics of the detrended and deseasonalized DAWSs process using a general ARMA model, and then we will compare our results with a WN.

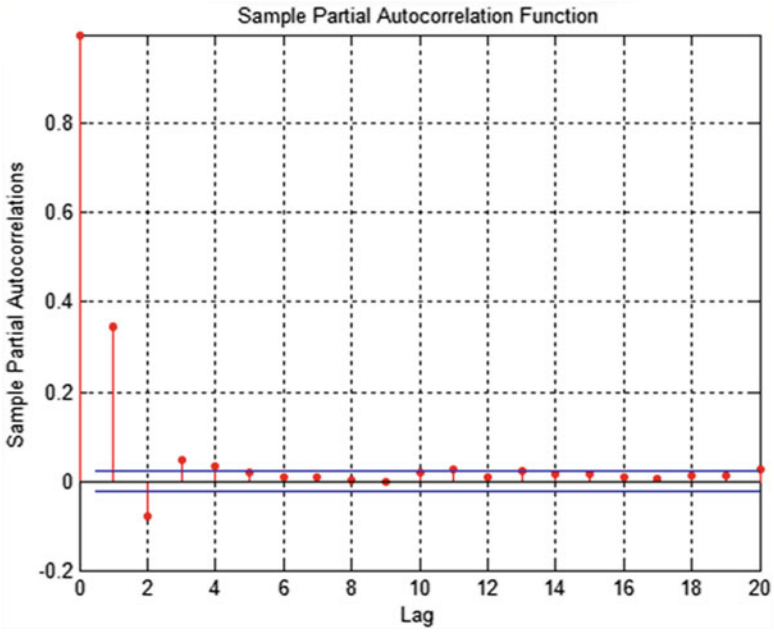

Fig. 9.4 The partial autocorrelation function of the detrended and deseasonalized DAWS in New York

We define the detrended and deseasonalized DAWS as

$$\tilde{W}^{(l)}(t) = W^{(l)}(t) - S(t). \tag{9.3}$$

The dynamics of $\tilde{W}^{(l)}(t)$ are modeled by an O–U stochastic process:

$$d\tilde{W}^{(l)}(t) = \kappa \tilde{W}^{(l)}(t)dt + \sigma(t)dB_t. \tag{9.4}$$

First, in order to select the correct ARMA model, we examine the ACF of the detrended and deseasonalized DAWS. A closer inspection of the second part of Fig. 9.3 reveals that the 1st, 2nd, and the 4th lags are significant. On the other hand, by examining the PACF in Fig. 9.4, we conclude that the first four lags are necessary to model the autoregressive effects of the wind speed dynamics.

In order to find the correct model, we estimate the log-likelihood function (LLF) and the Akaike information criterion (AIC). Consistent with the PACF, both criteria suggest that an AR (4) model is adequate for modeling the wind process since they were minimized when a model with four lags was used. The estimated parameters and the corresponding p values are presented in Table 9.3. It is clear that the three first parameters are statistically very significant since their p value is less than 0.05.

Table 9.3 Estimated parameters of the linear AR (4) model

Parameter	AR (1)	AR (2)	AR (3)	AR (4)
Value	0.3617	−0.0999	0.0274	0.0216
P value	0.0000	0.0000	0.0279	0.0657

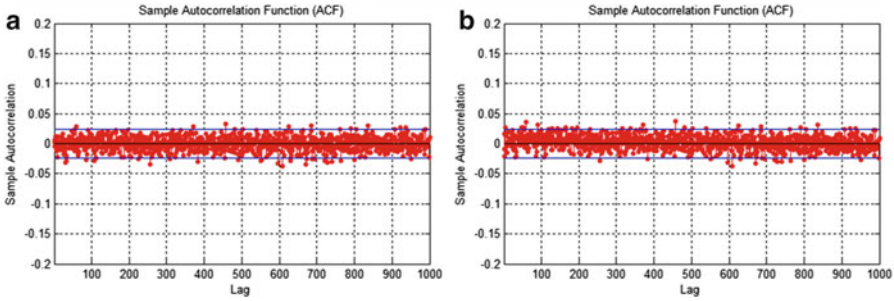

Fig. 9.5 Autocorrelation function of the residuals of (**a**) the linear model and (**b**) the WN

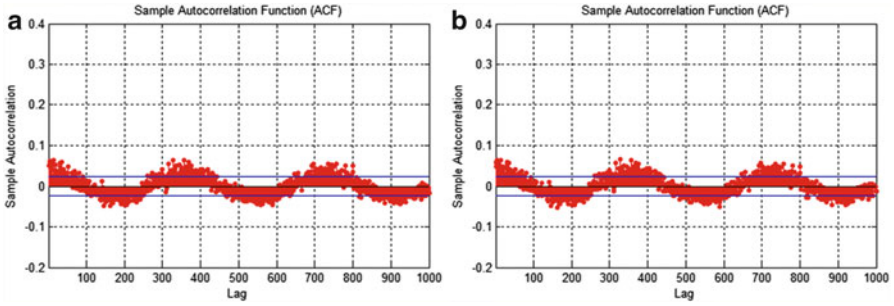

Fig. 9.6 Autocorrelation function of the squared residuals of (**a**) the linear model and (**b**) the WN

The parameter of the 4th lag is statistically significant with p value 0.0657. The AIC for this model is 0.46852, while the LLF is −1,705.14.

Observing the residuals of the AR model in the first part of Fig. 9.5, we conclude that the autocorrelation was successfully removed. However, the ACF of the squared residuals indicates a strong seasonal effect in the variance of the wind speed as it is shown in Fig. 9.6. The same conclusion was reached in previous studies for daily models for both temperature and wind (Alexandridis 2010; Benth et al. 2009; Benth and Saltyte-Benth 2005, 2007, 2009; Benth et al. 2007; Zapranis and Alexandridis 2008, 2009, 2011). Following the similar procedure that was described in the previous chapters for the temperature, we model the seasonal variance with a truncated Fourier series:

$$\sigma^2(t) = c_0 + \sum_{i=1}^{I_2} c_i \sin(2\pi it/365) + \sum_{j=1}^{J_2} d_j \sin(2\pi jt/365). \tag{9.5}$$

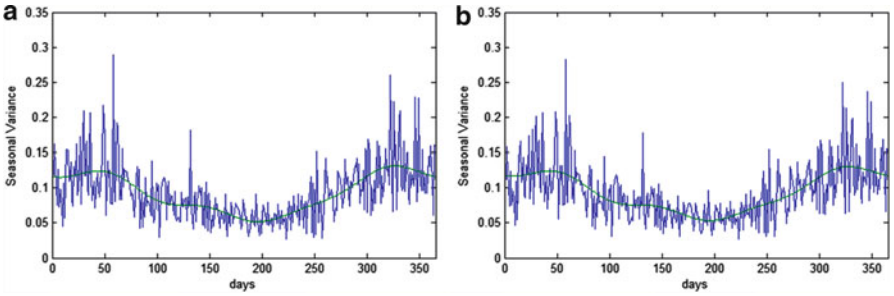

Fig. 9.7 Empirical and fitted seasonal variance of (**a**) the linear model and (**b**) the WN

Table 9.4 Estimated parameters of the seasonal variance in the case of the linear model

c_0	c_1	c_2	c_3	c_4	d_1	d_2	d_3	d_4
0.0932	0.000032	−0.0041	0.0015	−0.0028	0.0358	−0.0025	−0.0048	−0.0054

Table 9.5 Descriptive statistics of the residuals for the linear AR (4) model

Var	Mean	St. dev.	Max	Min	Skew	Kur	JB	*P* value	KS	*P* value	LBQ	P value
Noise	0	1	3.32	−5.03	−0.09	3.03	10.097	0.007	1.033	0.2349	8.383	0.989

St. dev. standard deviation, *JB* Jarque–Bera statistic, *KS* Kolmogorov–Smirnov statistic, *LBQ* Ljung–Box Q-statistic

Note that we assume that the seasonal variance is periodic and repeated every year, that is, $\sigma^2(t+365) = \sigma^2(t)$ where $t = 1, \ldots, 7359$. The empirical and the fitted seasonal variances are presented in Fig. 9.7, while in Table 9.4, the estimated parameters of (9.5) are presented. Non-surprisingly, the variance exhibits the same characteristics as in the case of temperature (Alexandridis 2010; Zapranis and Alexandridis 2008; Benth and Saltyte-Benth 2007). More precisely, the seasonal variance is higher in the winter and early summer, while it reaches its lower values during the summer period.

Finally, the descriptive statistics of the final residuals are examined. A closer inspection of Table 9.5 shows that the autocorrelation has successfully removed as indicated by the Ljung–Box Q-statistic. In addition, the distribution of the residuals is very close to the normal distribution as it is shown on the first part of Fig. 9.8; however, small negative skewness exists. More precisely, the residuals have mean 0 and standard deviation of 1. In addition, the kurtosis is 3.03 and the skewness is −0.09.

9.2.2 *Wavelet Networks for Wind Speed Modeling*

In this section, WNs are used in the transformed, detrended, and deseasonalized wind speed data in order to model the daily dynamics of wind speeds in New York.

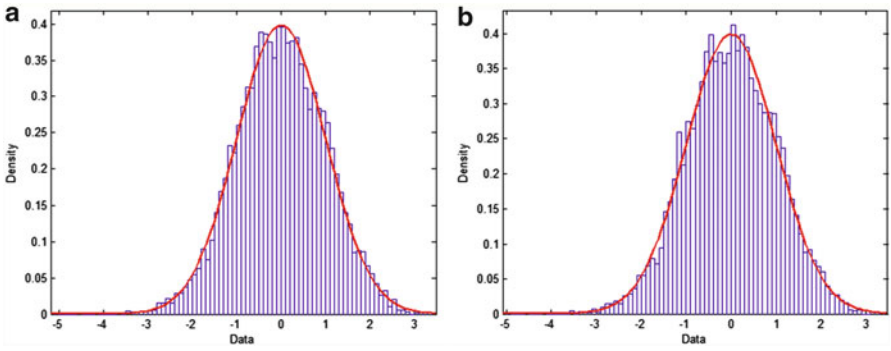

Fig. 9.8 Empirical and fitted normal distribution of the final residuals of the WN

Motivated by the waveform of the data, we expect a wavelet function to better fit the wind speed. In addition, it is expected that the nonlinear form of the WN will provide more accurate representation of the dynamics of the wind speed process both in-sample and out-of-sample.

The structure and the mathematical expressions of a WN are presented analytically in Appendices A and B, while in Alexandridis (2010), detailed explanation of how to use WNs in model identification problems is described. Since WNs are nonlinear tools, criteria like AIC or LLF cannot be used. Hence, in this section, WNs will be used in order to select the significant lags, to select the appropriate network structure, to train a WN in order to learn the dynamics of the wind speeds, and finally, to forecast the future evolution of the wind speeds.

The algorithm developed by Alexandridis (2010) simultaneously estimates the correct number of lags that must be used in order to model the wind speed dynamics and the architecture of the WN by using a recurrent algorithm. An illustration of the model identification algorithm is presented in Appendix A.

Our backward elimination algorithm examines the contribution of each available explanatory variable to the predictive power of the WN. First, the prediction risk of the WN is estimated as well as the statistical significance of each variable. If a variable is statistically insignificant, it is removed from the training set, and the prediction risk and the new statistical measures are estimated. The algorithm stops if all explanatory variables are significant. Hence, in each step of our algorithm, the variable with the larger p value greater than 0.1 will be removed from the training set of our model. After each variable removal, a new architecture of the WN will be selected and a new WN will be trained. However, the correctness of the decision of removing a variable must be examined. This can be done either by examining the prediction risk or the $\bar{R}^2$. If the new prediction risk is smaller than the new prediction risk multiplied by a threshold, then the decision of removing the variable was correct. If the prediction risk increased more than the allowed threshold, then the variable was reintroduced back to the model. We set this threshold at 5%. In this study, the selected statistical measure is the SBP proposed by Moody and Utans (1992). Previous analysis in Alexandridis (2010) indicates that the SBP fitness

Table 9.6 Variable selection with backward elimination in New York

Step	Variable to remove (lag)	Variable to enter (lag)	Variables in model	Hidden units (parameters)	*N*/*P* ratio	Empirical loss	Prediction risk
–			7	1 (23)	317.4	0.0467	0.0938
1	7	–	6	1 (20)	365.0	0.0467	0.0940
2	5	–	5	1 (17)	429.4	0.0467	0.0932
3	6	–	4	2 (23)	317.4	0.0467	0.0938
4	4	-	3	2 (18)	405.6	0.0468	0.0937

The algorithm concluded in four steps. In each step, the following are presented: which variable is removed, the number of hidden units for the particular set of input variables and the parameters used in the wavelet network, the ratio between the parameters and the training patterns, the empirical loss, and the prediction risk

criterion was found to significantly outperform alternative criteria in the variable selection algorithm. The SBP quantifies the effect on the empirical loss of replacing a variable by its mean. Analytical description of the SBP is given in Alexandridis (2010), Zapranis and Refenes (1999), and Moody and Utans (1992). In each step, the SBP and the corresponding p value are calculated. For analytical explanation of each step of the algorithm, we refer to Alexandridis (2010).

The proposed variable selection framework will be applied on the transformed, detrended, and deseasonalized wind speeds in New York in order to select the length of the lag series. The target values of the WN are the DAWSs. The explanatory variables are lagged versions of the target variable. The relevance of a variable to the model is quantified by the SBP criterion which was introduced in Moody and Utans (1992). Initially, the training set contains the dependent variable and seven lags. The analysis in the previous section indicates that a training set with seven lags will provide all the necessary information of the ACF of the detrended and deseasonalized DAWSs. Hence, the training set consists of 7 inputs, 1 output, and 7,293 training pairs.

Table 9.6 summarizes the results of the model identification algorithm for New York. Both the model selection and the variable selection algorithms are included in Table 9.6. The algorithm concluded in four steps and the final model contains only three variables, that is, three lags. The prediction risk for the reduced model is 0.0937 while for the original model was 0.0938. On the other hand, the empirical loss slightly increased from 0.0467 for the initial model to 0.0468 for the reduced model indicating that the explained variability (unadjusted) slightly decreased. Finally, the complexity of the network structure and number of parameters were significantly reduced in the final model. The initial model needed one hidden unit (HU) and seven inputs. Hence, 23 parameters were adjusted during the training phase. Hence, the ratio of the number of training pairs n to the number of parameters p was 317.4. In the final model, only two HU and three inputs were used. Hence, only 18 parameters were adjusted during the training phase, and the ratio of the number of training pairs n to the number of parameters p was 405.6.

The proposed algorithm suggests that a WN needs only three lags to extract the autocorrelation from the data while the linear model needed four lags. A closer inspection of Table 9.6 reveals that the WN with three and four lags have the same

Table 9.7 Estimated parameters of the seasonal variance in the case of the WN

c_0	c_1	c_2	c_3	c_4	d_1	d_2	d_3	d_4
0.0935	−0.000020	−0.0034	0.0014	−0.0026	0.0353	−0.0016	−.0042	−0.0052

Table 9.8 Descriptive statistics of the residuals for the WN model

Var	Mean	St. dev	Max	Min	Skew	Kur.	JB	*P* value	KS	*P* value	LBQ	*P* value
noise	0	1	3.32	−4.91	−0.08	3.04	8.84	0.0043	0.927	0.3544	13.437	0.858

St. dev. standard deviation, *JB* Jarque–Bera statistic, *KS* Kolmogorov–Smirnov statistic, *LBQ* Ljung–Box Q-statistic

predictive power in-sample and out-of-sample. Hence, we chose the simpler model. Our model is similar to an AR (3) model with time-varying parameters.

Examining the second part of Fig. 9.5, we conclude that the autocorrelation was successfully removed from the data; however, the seasonal autocorrelation in the squared residuals is still present as it is shown in Fig. 9.6. We will remove the seasonal autocorrelation using (9.5). The estimated parameters are presented in Table 9.7, and as it was expected, their values are similar to those of the case of the linear model. In Fig. 9.7, the empirical and the fitted seasonal variance is presented. Again, the same conclusions are reached for the seasonal variance. The variance is higher at winter period, while it reaches its minimum during the summer period.

Finally, examining the final residuals of the WN model, we observe that the distribution of the residuals is very close to the normal distribution as it is shown in Fig. 9.8, while the autocorrelation was successfully removed from the data. In addition, we observe an improvement in the distributional statistics in contrast to the case of the linear model. The distributional statistics of the residuals are presented in Table 9.8.

Concluding, the distributional statistics of the residuals indicate that in in-sample, the two models can accurately represent the dynamics of the DAWSs; however, an improvement is evident when a nonlinear nonparametric WN is used.

9.2.3 Forecasting Daily Average Wind Speeds

In this section, our proposed model will be validated out-of-sample. In addition, the performance of our model will be tested against two models, first, against the linear model previously described and, second, against the simple persistent method usually referred as benchmark. The linear model is the AR (4) model described in the previous section. The persistent method assumes that today's and tomorrow's DAWSs will be equal, that is, $W^*(t+1) = W(t)$ where the W^* indicates the forecasted value.

Table 9.9 Out-of-sample comparison. One month

	Persistent	AR (4)	WN
Md.AE	2.3000	2.2147	2.1081
MAE	3.3000	2.5547	2.5026
Max AE	8.2000	7.9217	7.7590
SSE	507.9300	328.9947	320.2573
RMSE	4.0478	3.2577	3.2142
NMSE	1.5981	1.0351	1.0076
MSE	16.3848	10.6127	10.3309
MAPE	0.3456	0.2744	0.2680
SMAPE	0.3233	0.2570	0.2518
POCID	47%	80%	80%
IPOCID	33%	33%	37%
POS	100%	100%	100%

Md. AE median absolute error, *MAE* mean absolute error, *Max AE* maximum absolute error, *SSE* sum of squared errors, *RMSE* root-mean-square error, *NMSE* normalized mean square error, *MSE* mean square error, *MAPE* mean absolute percentage error, *SMAPE* symmetric MAPE, *POCID* position of change in direction, *IPOCID* independent POCID, *POS* position of sign

The three models will be used for forecasting DAWSs for two different periods. Usually, wind derivatives are written for a period of a month. Hence, DAWSs for 1 and 2 months will be forecasted. The out-of-sample dataset correspond to the period from January 1 to February 28, 2008, and were not used for the estimation of the linear and nonlinear models. Note that our previous analysis reveals that the variance is higher in the winter period indicating that it is more difficult to forecast accurately DAWS for these two months.

In Table 9.9, the performance of the three methods when the forecast window is 1 month is presented. Various error criteria are estimated like the mean, median, and max. AE; the mean square error (MSE); the POCID; and the IPOCID. As it is shown on Table 9.9, our proposed method outperforms both the persistent and the AR (4) model. The AR (4) model performs better than the naïve persistent method; however, all error criteria are improved when a nonlinear WN is used. The MSE is 16.3848 for the persistent method, 10.6127 for the AR (4) model, and 10.3309 for the WN. In addition, our model can predict more accurately the movement of the wind speed since the POCID is 80% for the WN and the AR (4) models, while it is only 47% for the persistent method. Moreover, the IPOCID is 37% for the proposed model, while it is only 33% for the other two methods.

In order to compare our model directly with the linear method, we estimate a linear AR (3) model. However, our proposed methodology still outperforms the linear method.

Next, the three forecasting methods are evaluated in 2 months day-ahead forecasts. The results are similar and presented in Table 9.10. The proposed WN outperforms the other two methods. Only the max. AE and the POCID are slightly smaller when the AR (4) model is used. However, the IPOCID is 38% for both methods. Also, our results indicate that the persistent method produces significantly

Table 9.10 Out-of-sample comparison. Two months

	Persistent	AR (4)	WN
Md.AE	2.4000	2.7981	2.6589
MAE	3.3678	2.8126	2.7976
Max AE	11.2000	7.9345	8.0194
SSE	1054.3500	706.1806	702.4437
RMSE	4.2273	3.4596	3.4505
NMSE	1.4110	0.9450	0.9400
MSE	17.8703	11.9692	11.9058
MAPE	0.3611	0.3014	0.3001
SMAPE	0.3289	0.2798	0.2782
POCID	45%	71%	69%
IPOCID	36%	38%	38%
POS	100%	100%	100%

Md. AE median absolute error, *MAE* mean absolute error, *Max AE* maximum absolute error, *SSE* sum of squared errors, *RMSE* root-mean-square error, *NMSE* normalized mean square error, *MSE* mean square error, *MAPE* mean absolute percentage error, *SMAPE* symmetric MAPE, *POCID* position of change in direction, *IPOCID* independent POCID, *POS* position of sign

Table 9.11 Estimation of the cumulative rainfall index for 1 and 2 months using an AR (4) model, WN, and historical burn analysis

	AR (4)	WN	HBA	Actual
1 month	305.1	312.7	345.5	311.2
2 months	579.5	591.1	658.3	600.6

worse forecasts. Finally, the WN and the linear AR (3) model are compared with first to show better forecasting ability.

Our results indicate that the WN can forecast the evolution of the dynamics of the DAWSs, and hence, they constitute an accurate tool for wind derivatives pricing.

In order to have a better insight of the performance of each method, the cumulative average wind speed (CAWS) index is calculated. Since we are interested in weather derivatives, one common index is the sum of the daily rainfall index over a specific period. In Table 9.11, the estimation of three methods is presented. More precisely, the WN, the AR (4), as well as the HBA methods are compared. The HBA is a simple statistical method that estimates the performance of the index over the specific period the previous years and it is often used in the industry. In other words, it is the average of 20 years of the index over the period of January and February, and it serves as a benchmark.

The final row of Table 9.11 presents the actual values of the cumulative rainfall index. An inspection of Table 9.11 reveals that the WN significantly outperforms the other two methods. For the first case, where forecasts for 1 month ahead are estimated, the forecast of the CAWS index using WN is 312.7, while the actual index is 311.2. On the other hand, the forecast using the AR (4) model is 305.1. However, when the forecast period increases, the forecast of the AR (4) model

significantly deviates. More precisely, for the second case, the forecast of the WN is 591.1 while the actual index is 600.6 and the AR (4) forecast is 579.5. Finally, we have to mention that the WN uses less information than the AR (4) model, since in the case of WN, only the information of three lags is used.

Since we are interested in wind derivatives and the valuation of wind contracts, next, an illustration of the performance of each method using a theoretical contract is presented. A common wind contract has a tick size of 0.1 knots and pays 20$ per tick size. Hence, for the case of a 1-month contract, the AR (4) method underestimates the contract size for 1,200$, while the WN overestimates the contract for 300$ only. Similarly, for the case of a 2-month contract, the AR (4) method underestimates the contract size for 4,220$, while the WN underestimates the contract for 1,900$.

Incorporating meteorological forecasts can lead to a potentially significant improvement of the performance of the proposed model. Meteorological forecasts can be easily incorporated in both the linear and the WN models previously presented. A similar approach was followed for temperature derivatives by Dorfleitner and Wimmer (2010) for temperature derivatives. However, this method cannot be always applied. Despite great advances in meteorological science, weather still cannot be predicted precisely and consistently, and forecasts beyond 10 days are not considered accurate (Wilks 2011). If the day that the contract is traded is during or close to the life of the derivative (during the period that wind measurements are considered), the meteorological forecasts can be incorporated in order to improve the performance of the methods. However, very often, weather derivatives are traded long before the start of the life of the derivative. More precisely, very often, weather derivatives are traded months or even a season before the starting day of the contract. In this case, meteorological forecasts cannot be used.

9.3 Pricing Wind Derivatives

In this section, the pricing formulas for wind derivatives are presented under the assumption of a normal driving noise process. The analysis performed in the previous section indicates that the assumption that the final residuals, after dividing out the seasonal variance, follow a normal distribution is justified.

When the market is complete, a unique risk-neutral probability measure $Q \sim P$ can be obtained, where P is the real-world probability measure. This change of measure turns the stochastic process into a martingale. Hence, financial derivatives can be priced under the risk-neutral measure by the discounted expectation of the derivative payoff.

The same implications that we faced in the pricing of temperature derivatives appear also in the pricing of wind derivatives. The wind market is an incomplete market. The underlying weather derivative cannot be stored or traded. Moreover, the market is relatively illiquid. In principle (extended), risk-neutral valuation can still be carried out in incomplete markets.

The method that was used in order to proceed in temperature derivative pricing will be followed also in this section. The change of measure from the real world to the risk-neutral world under the dynamics of a BM can be performed using the Girsanov's theorem.

The statistical analysis indicates that the transformed DAWSs can be modeled by a mean-reverting O–U process where the speed of mean reversion variable is a function of time:

$$dW_t^{(l)} = S(t) + a(t)\left(W_{t-1}^{(l)} - S(t-1)\right)dt + \sigma(t)dB_t, \tag{9.6}$$

where $S(t)$ is the seasonal function, $\sigma(t)$ is the seasonal variance which is bounded by zero, $a(t)$ is the speed of mean reversion, and B_t is the driving noise process.

Using the Girsanov's theorem, under the risk-neutral measure Q, we have that

$$dB_t^{\theta} = dB_t - \theta(t), \tag{9.7}$$

where $\theta(t)$ is the market price of risk and

$$\int_0^T \theta^2(t)dt < \infty. \tag{9.8}$$

Hence, applying Itô formula on (9.6) and (9.7), the solution of the transformed DAWS under the risk-neutral measure Q is given by

$$\begin{aligned} W_t^{(l)} = S(t) + e^{\int_0^t a(z)dz}\left(W_0^{(l)} - S(0)\right) + e^{\int_0^t a(z)dz}\int_0^t \sigma(s)\theta(s)e^{-\int_0^s a(z)dz}ds \\ + e^{\int_0^t a(z)dz}\int_0^t \sigma(s)e^{-\int_0^s a(z)dz}dB_t^{\theta}. \end{aligned} \tag{9.9}$$

The proposed model is an extension of the CAR (p) introduced by Brockwell and Marquardt (2005) and applied by Benth and Saltyte-Benth (2009) in wind derivative pricing. Hence, we follow a similar pricing approach presented in Benth and Saltyte-Benth (2009).

The transformed, detrended, and deseasonalized DAWS $\tilde{W}_t^{(l)} = W_t^{(l)} - S(t)$ are normally distributed with mean

$$\mu_{\theta}(t, s, \tilde{W}_t^{(l)}) = e^{\int_t^s a(z)dz}\tilde{W}_s^{(l)} + e^{\int_t^s a(z)dz}\int_t^s \sigma(u)\theta(u)e^{-\int_t^u a(z)dz}du, \tag{9.10}$$

and variance

$$\Sigma^2(t,s) = e^{2\int_t^s a(z)dz} \int_t^s \sigma^2(s) e^{-2\int_t^u a(z)dz} du. \tag{9.11}$$

The market price of wind risk is necessary in pricing wind derivatives. However, in order to estimate θ, the actual prices of derivatives are required. Since the shutdown of the US Future Exchange, wind derivatives are traded only in the Chicago Climate Futures Exchange and in the OTC market, and as a result, it is hard to obtain market data. Hence, it is very difficult to estimate the market price of wind risk. However, the trading volume of wind derivatives is increasing every year (WRMA 2010), and it is expected that wind derivatives will be soon included in the listed products of the CME.

A solution to this problem is presented by Benth et al. (2009) where they study the market price of risk for temperature derivatives in various Asian cities. The market price of risk was estimated by calibrating model prices. Their results indicate that the market price of risk for Asian temperature derivatives is different from zero and shows a seasonal structure that comes from the seasonal variance of the temperature process. Their empirical findings suggest that by knowing the formal dependence of the market price of risk on seasonal variation, one can infer the market price of risk for regions where weather derivative market does not exist.

Similarly, in Huang et al. (2008), a pricing method for temperature derivatives in Taiwan is presented. Since no active weather market exists in Taiwan, the parameter θ is approximated by a function of the market price of risk of the Taiwan Stock Exchange.

9.3.1 The Cumulative Average Wind Speed Index

In this section, we derive the pricing equation for the CAWS index. Similar to the CAT index, the CAWS index is the sum of the DAWSs over a specific period $[\tau_1, \tau_2]$, and it is given by

$$\text{CAWS} = \int_{\tau_1}^{\tau_2} W(s)ds. \tag{9.12}$$

Our aim is to give a mathematical expression for the CAWS future price. If Q is the risk-neutral probability and r is the constant compounding interest rate, then the arbitrage-free future price of a CAT contract at time $t \leq \tau_1 < \tau_2$ is given by

$$e^{-r(\tau_2 - t)} \mathrm{E}_Q \left[\int_{\tau_1}^{\tau_2} W(s)ds - F_{\text{CAWS}}(t, \tau_1, \tau_2) | \mathbf{F}_t \right] = 0, \tag{9.13}$$

and since F_{CAWS} is $\mathbf{F}_t$ adapted, we derive the price of a CAT futures to be

$$F_{\text{CAWS}}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} W(s)ds|\mathbf{F}_t\right]. \tag{9.14}$$

To derive the future price, we must calculate the conditional expectation of $W(s)$ given F_t, for $s \geq t$. This is done in the following Lemma, first presented in Benth and Saltyte-Benth (2009). For reasons of completeness, we reproduce this Lemma here.

Lemma 9.1 *Let $0 \leq t \leq s \leq T$, then for $l \in (0,1]$, it holds that*

$$E_Q[W(s)|\mathbf{F}_t] = M_{1/l}\Big(1 + l\Big(S(s) + \mu_\theta\Big(t,s,\tilde{W}_t^{(l)}\Big), l^2\Sigma^2(t,s)\Big)\Big), \tag{9.15}$$

where $M_k(a,b^2)$ is the k^{th} moment of a normal random variable with mean a and variance b^2 and $\mu_\theta\Big(t,s,\tilde{W}_t^{(l)}\Big)$ and $\Sigma^2(t,s)$ are given by (9.10) and (9.11), respectively.

Proof From (9.1) and (9.9), we have that the wind speed at time s given F_t can be represented (for $l \neq 0$) as

$$W(s) = \Big[l\Big(S(s) + \mu_\theta\Big(t,s,\tilde{W}_t^{(l)}\Big) + \Sigma(t,s)Z\Big) + 1\Big]^{\frac{1}{l}},$$

where Z is a standard normally distributed random variable independent of $\mathbf{F}_t$. Further, $\mu_\theta\Big(t,s,\tilde{W}_t^{(l)}\Big)$ is $\mathbf{F}_t$ measurable. Hence, the result follows from a direct calculation. The lognormal case $l = 0$ follows similarly.

□

Hence, the arbitrage-free price of the CAWS index easily follows from Lemma 9.1.

Proposition 9.1 *The arbitrage-free price of* CAWS *index at time $t \leq \tau_1 < \tau_2$ is given by*

$$F_{\text{CAWS}}(t,\tau_1,\tau_2) = \int_{\tau_1}^{\tau_2} \Big(M_{1/l}\Big(1 + l\Big(S(s) + \mu_\theta\Big(t,s,\tilde{W}_t^{(l)}\Big), l^2\Sigma^2(t,s)\Big)\Big)\Big)ds, \tag{9.16}$$

where $M_k(a,b^2)$ is the k^{th} moment of a normal random variable with mean a and variance b^2 and $\mu_\theta\Big(t,s,\tilde{W}_t^{(l)}\Big)$ and $\Sigma^2(t,s)$ are given by (9.10) and (9.11), respectively.

Proof We have from (9.14) that

$$F_{\text{CAWS}}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} W(s)ds|\mathbf{F}_t\right],$$

and using Itô's isometry, we can interchange the expectation and the integral

$$\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} W(s)ds|\mathbf{F}_t\right] = \int_{\tau_1}^{\tau_2} \mathrm{E}_Q[W(s)|\mathbf{F}_t]ds,$$

and from a direct application from Lemma 9.1, we have that

$$F_{\text{CAWS}}(t,\tau_1,\tau_2) = \int_{\tau_1}^{\tau_2} M_{1/l}\left(1 + l\left(S(s) + \mu_\theta\left(t,s,\tilde{W}_t^{(l)}\right), l^2\Sigma^2(t,s)\right)\right)ds. \qquad \square$$

9.3.2 The Nordix Wind Speed Index

In this section, we derive the pricing equations for the Nordix wind speed index. The Nordix wind speed index is the index that the US Future Exchange used to settle wind derivatives. The Nordix index is given by

$$I(\tau_1,\tau_2) = 100 + \sum_{s=\tau_1}^{\tau_2}(W(s) - w_{20}(s)), \tag{9.17}$$

and measures the daily wind speed deviations from the mean of the past 20 years over a period $[\tau_1,\tau_2]$.

The result of Lemma 9.1 is applied again to derive the price of a Future Nordix wind speed index.

Proposition 9.2 *The arbitrage-free price of Nordix wind future speed index at time* $t \le \tau_1 < \tau_2$ *is given by*

$$\begin{aligned} F_{NWI}(t,\tau_1,\tau_2) = 100 &- \sum_{s=\tau_1}^{\tau_2} w_{20}(s) \\ &+ \sum_{s=\tau_1}^{\tau_2} M_{1/l}\left(1 + l\left(S(s) + \mu_\theta(t,s,\tilde{W}_t^{(l)}\right), l^2\Sigma^2(t,s)\right), \end{aligned} \tag{9.18}$$

where $M_k(a,b^2)$ *is the* k^{th} *moment of a normal random variable with mean a and variance* b^2.

Proof If Q is the risk-neutral probability and r is the constant compounding interest rate, then the arbitrage-free future price of a Nordix wind speed index contract at time $t \le \tau_1 < \tau_2$ is given by

$$e^{-r(\tau_2-t)}\mathrm{E}_Q\left[100-\sum_{s=\tau_1}^{\tau_2}w_{20}(s)+\sum_{s=\tau_1}^{\tau_2}W(s)-F_{NWI}(t,\tau_1,\tau_2)|\mathbf{F}_t\right]=0,$$

and since $F_{NWI}(t,\tau_1,\tau_2)$ is $\mathbf{F}_t$ adapted, we derive the price of a Nordix wind future index to be

$$F_{NWI}(t,\tau_1,\tau_2)=100-\sum_{s=\tau_1}^{\tau_2}w_{20}(s)+\mathrm{E}_Q\left[\sum_{s=\tau_1}^{\tau_2}W(s)|\mathbf{F}_t\right].$$

Applying the Lemma 4.1 from Benth and Saltyte-Benth (2009), we find the explicit solution for the price of the Nordix wind future index:

$$\begin{aligned}F_{NWI}(t,\tau_1,\tau_2)&=100-\sum_{s=\tau_1}^{\tau_2}w_{20}(s)\\&+\sum_{s=\tau_1}^{\tau_2}M_{1/l}\left(1+l\left(S(s)+\mu_\theta(t,s,\tilde{W}_t^{(l)}\right),l^2\Sigma^2(t,s)\right),\end{aligned}$$

where $M_k(a,b^2)$ is the k^{th} moment of a normal random variable with mean a and variance b^2.

□

9.4 Conclusions

In this chapter, DAWSs from New York were studied. Our analysis revealed strong seasonality in the mean and variance. The DAWSs were modeled by a mean-reverting Ornstein–Uhlenbeck process in the context of wind derivative pricing. In this study, the dynamics of the wind-generating process are modeled using a nonparametric nonlinear WN. Our proposed methodology was compared in-sample and out-of-sample against two methods often used in prior studies. The characteristics of the wind speed process are very similar to the process of daily average temperatures.

Our method is validated in a 2-month-ahead out-of-sample forecast period. Moreover, the various error criteria produced by the WN are compared against the linear AR model and the persistent method. Results show that the WN outperforms the other two methods, indicating that WNs constitute an accurate model for forecasting DAWSs. More precisely, the WN's forecasting ability is stronger in both samples. Testing the fitted residuals of the WN, we observe that the distribution of the residuals is very close to the normal. Also, the WN needed only the information of the past 3 days, while the linear method suggested a model with four lags. Finally, we provided the pricing equations for wind futures of the Nordix

index. Although we focused on DAWSs, our model can be easily adapted in hourly modeling.

The results in this chapter are preliminary and can be further analyzed. More precisely, alternative methods for estimating the seasonality in the mean and in the variance can be developed. Alternative methods could improve the fitting to the original data as well as the training of the WN.

Also, it is important to test the largest forecasting window of each method. Since meteorological forecasts of a window larger than few days are considered inaccurate, this analysis will suggest the best model according to the desired forecasting interval.

Finally, a large-scale comparison must be conducted. Testing the proposed methods as well as more sophisticated models, like general ARFIMA or GARCH, in various meteorological stations will provide a better insight in the dynamics of the DAWS as well as in the predictive ability of each method.

References

Ailliot P, Monbet V, Prevosto M (2006) An autoregressive model with time-varying coefficients for wind fields. Envirometrics 17:107–117

Alexandridis A (2010) Modelling and pricing temperature derivatives using wavelet networks and wavelet analysis. Ph.D. Thesis, University of Macedonia, Thessaloniki

Alexiadis MC, Dokopoulos PS, Sahsamanoglou HS, Manousaridis IM (1998) Short-term forecasting of wind speed and related electrical power. Solar Energy 63(1):61–68

Barbounis TG, Theocharis JB, Alexiadis MC, Dokopoulos PS (2006) Long-term wind speed and power forecasting using local recurrent neural network models. IEEE Trans Energy Convers 21 (1):273–284

Benth FE, Saltyte-Benth J (2005) Stochastic modelling of temperature variations with a view towards weather derivatives. Appl Math Finance 12(1):53–85

Benth FE, Saltyte-Benth J (2007) The volatility of temperature and pricing of weather derivatives. Quant Finance 7(5):553–561

Benth FE, Saltyte-Benth J (2009) Dynamic pricing of wind futures. Energy Econ 31:16–24

Benth FE, Saltyte-Benth J, Koekebakker S (2007) Putting a price on temperature. Scand J Stat 34:746–767

Benth FE, Hardle WK, Lopez Cabrera B (2009) Pricing of Asian temperature risk. SFB649 Working Paper. Humboldt-Universitat zu Berlin, Berlin

Beyer HG, Degner T, Hausmann J, Hoffmann M, Rujan P (1994) Short-term prediction of wind speed and power output of a wind turbine with neural networks. In: 2nd European congress on intelligent techniques and soft computing, Aachen, 20–23 Sept 1994

Billinton R, Chen H, Ghajar R (1996) Time-series models for reliability evaluation of power systems including wind energy. Microelectron Reliab 36(9):1253–1261

Brett AC, Tuller SE (1991) The autocorrelation of hourly wind speed observations. J Appl Meteorol 30(6):823–833. doi:10.1175/1520-0450(1991)030<0823:TAOHWS>2.0.CO;2

Brockwell PJ, Marquardt T (2005) Levy-driven and fractionality integrated ARMA process with continuous time parameter. Stat Sin 15:477–494

Brown BG, Katz RW, Murphy AH (1984) Time-series models to simulate and forecast wind speed and wind power. J Clim Appl Meteorol 23:1184–1195

Caporin M, Pres J (2010) Modelling and forecasting wind speed intensity for weather risk management. Comput Stat Data Anal. doi:doi:10.1016/j.csda.2010.06.019

Castino F, Festa R, Ratto CF (1998) Stochastic modelling of wind velocities time-series. J Wind Eng Ind Aerodyn 74–76:141–151
Celik AN (2004) A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. Renew Energy 29(4):593–604. doi:10.1016/j.renene.2003.07.002
Cripps E, Nott D, Dunsmuir WTM, Wikle C (2005) Space-time modelling of Sydney Harbour winds. Aust N Z J Stat 47(1):3–17. doi:10.1111/j.1467-842X.2005.00368.x
Daniel AR, Chen AA (1991) Stochastic simulation and forecasting of hourly average wind speed sequences in Jamaica. Solar Energy 46(1):1–11
Dorfleitner G, Wimmer M (2010) The pricing of temperature futures at the Chicago Mercantile Exchange. J Bank Finance. doi:10.1016/j.bankfin.2009.12.004
Dorvlo ASS (2002) Estimating wind speed distribution. Energy Convers Manage 43 (17):2311–2318. doi:10.1016/s0196-8904(01)00182-0
Garcia A, Torres JL, Prieto E, de Francisco A (1998) Fitting wind speed distributions: a case study. Solar Energy 62(2):139–144. doi:10.1016/s0038-092x(97)00116-3
Haslett J, Raftery AE (1989) Space-time modelling with long-memory dependence: assessing Ireland's wind power resource. J R Stat Soc Ser C (Appl Stat) 38(1):1–50
Huang Z, Chalabi ZS (1995) Use of time-series analysis to model and forecast wind speed. J Wind Eng Ind Aerodyn 56:311–322
Huang H-H, Shiu Y-M, Lin P-S (2008) HDD and CDD option pricing with market price of weather risk for Taiwan. J Futures Mark 28(8):790–814
Jaramillo OA, Borja MA (2004) Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case. Renew Energy 29(10):1613–1630. doi:10.1016/j.renene.2004.02.001
Jewson S, Brix A, Ziehmann C (2005) Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations. Cambridge University Press, Cambridge, UK
Justus CG, Hargraves WR, Mikhail A, Graber D (1978) Methods for estimating wind speed frequency distributions. J Appl Meteorol (United States) 17(3):350–385
Kamal L, Jafri YZ (1997) Time-series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. Solar Energy 61(1):23–32
Kavak Akpinar E, Akpinar S (2005) A statistical analysis of wind speed data used in installation of wind energy conversion systems. Energy Convers Manage 46(4):515–532. doi:10.1016/j.enconman.2004.05.002
Kavasseri RG, Seetharaman K (2009) Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 34(5):1388–1393. doi:10.1016/j.renene.2008.09.006
Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z (2009) A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 13(4):915–920. doi:10.1016/j.rser.2008.02.002
Martin M, Cremades LV, Santabarbara JM (1999) Analysis and modelling of time-series of surface wind speed and direction. Int J Climatol 19:197–209
Mohandes MA, Rehman S, Halawani TO (1998) A neural networks approach for wind speed prediction. Renew Energy 13(3):345–354. doi:10.1016/s0960-1481(98)00001-9
Mohandes MA, Halawani TO, Rehman S, Hussain AA (2004) Support vector machines for wind speed prediction. Renew Energy 29(6):939–947. doi:10.1016/j.renene.2003.11.009
Moody JE, Utans J (1992) Principled architecture selection for neural networks: applications to corporate bond rating prediction. In: Refenes AP (ed) Neural networks in the capital markets. Wiley, Chichester/New York
More A, Deo MC (2003) Forecasting wind with neural networks. Marine Struct 16:35–49
Nfaoui H, Buret J, Sayigh AAM (1996) Stochastic simulation of hourly average wind speed sequences in Tangiers (Morocco). Solar Energy 56(3):301–314. doi:10.1016/0038-092x(95)00103-x
Nielsen TS, Madsen H, Nielsen HA, Pinson P, Kariniotakis G, Siebert N, Marti I, Lange M, Focken U, Bremen LV, Louka G, Kallos G, Galanis G (2006) Short-term wind power

forecasting using advanced statistical methods. Paper presented at the European wind energy conference, EWEC, Athens
Pinson P, Kariniotakis GN (2003) Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment. In: Power Tech conference proceedings, 2003 IEEE, vol 2, Bologna, 23–26 June 2003, p 8
Rehman S, Halawani TO (1994) Statistical characteristics of wind in Saudi Arabia. Renew Energy 4(8):949–956. doi:10.1016/0960-1481(94)90229-1
Saltyte-Benth J, Benth FE (2010) Analysis and modelling of wind speed in New York. J Appl Stat 37(6):893–909
Sfetsos A (2000) A comparison of various forecasting techniques applied to mean hourly wind speed time-series. Renew Energy 21:23–35
Sfetsos A (2002) A novel approach for the forecasting of mean hourly wind speed time series. Renew Energy 27:163–174
Tol RSJ (1997) Autoregressive conditional heteroscedasticity in daily wind speed measurements. Theor Appl Climatol 56:113–122
Tolman HL, Booij N (1998) Modeling wind waves using wavenumber-direction spectra and a variable wavenumber grid. Global Atmos Ocean Syst 6:295–309
Torres JL, Garcia A, De Blas M, De Francisco A (2005) Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Solar Energy 79:65–77
Tuller SE, Brett AC (1984) The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis. J Clim Appl Meteorol 23(1):124–134
Wilks DS (2011) Statistical methods in the atmospheric sciences, vol 100, 3rd edn, International geophysics series. Academic, Oxford, UK
WRMA (2009) Celebrating 10 years of weather risk industry growth. http://www.wrma.org/pdf/WRMA_Booklet_%20FINAL.pdf. Accessed Aug 2009
WRMA (2010) Weather derivatives volume plummets. www.wrma.org/pdf/weatherderivatives-volumeplummets.pdf. Accessed Jan 2010
Yamada Y (2008) Simultaneous optimization for wind derivatives based on prediction errors. In: American control conference, Washington, DC, 11–13 June 2008. pp 350–355
Zapranis A, Alexandridis A (2008) Modelling temperature time dependent speed of mean reversion in the context of weather derivative pricing. Appl Math Finance 15(4):355–386
Zapranis A, Alexandridis A (2009) Weather derivatives pricing: modelling the seasonal residuals variance of an Ornstein-Uhlenbeck temperature process with neural networks. Neurocomputing 73:37–48
Zapranis A, Alexandridis A (2011) Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing. Neural Comput Appl 20(6):787–801. doi:10.1007/s00521-010-0494-1
Zapranis A, Refenes AP (1999) Principles of neural model identification, selection and adequacy: with applications to financial econometrics. Springer, London/New York

Chapter 10
Precipitation Derivatives

10.1 Introduction

As it is already shown on the previous chapter, the modeling procedure of daily average temperature and daily wind speed is similar. Both weather variables are sharing common characteristics and as a result were modeled by a mean-reverting stochastic process with seasonality in mean and variance. In this chapter we focus on a different type of weather contracts, namely, precipitation derivatives. The first objective of this chapter is to present a daily model for the underlying weather variable, that is, precipitation. The second objective is to use the daily model in order to derive a valuation framework for precipitation derivatives. Although rainfall and snowfall are two different weather variables and contracts are traded separately, as it is shown in this chapter, they share a lot of common characteristics. Hence, the same model can be used for both of them. From now on we will refer to both of them as precipitation.

Currently, CME trades snowfall futures and options for the following US cities: Boston, New York, Chicago, Minneapolis, and Detroit. Contracts are offered for trading on a monthly or seasonal basis from November through April. For futures, the contract size is \$500 the CME snowfall index. The snowfall in CME is measured in inches, and the corresponding index is the cumulative snowfall over a period measured in inches. The accuracy of the index is 0.1 in., and 1 in. corresponds to 1 index point. Finally, the tick size is 0.1 points in the respective index. The options are European type written on futures on the snowfall index.

Similarly, CME trades rainfall futures and options for the following US cities: Chicago, Dallas, Des Moines, Detroit, Jacksonville, Los Angeles, New York, Portland, Raleigh, and Kansas. Contracts are offered for trading on a monthly or seasonal basis from March to October. For futures, the contract size is \$500 the CME rainfall index. The rainfall in CME is measured in inches, and the corresponding index is the cumulative rainfall over a period measured in inches. The accuracy of the index is 0.1 in., and 1 in. corresponds to 1 index point. Finally,

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8_10,

the tick size is 0.1 points in the respective index. The options are European type written on futures on the rainfall index.

Probably, agriculture is the most depended sector of the economy to precipitation. For a successful crop, precipitation must occur in particular times and at particular amounts. Although government programs are used in order to secure farmers against adverse weather condition, they are plagued with moral hazard and adverse selection problems. Further, federal crop insurance policies all contain deductibles that leave growers with some exposure to losses associated with extreme weather (Martin et al. 2001). Hydroelectric energy producers, construction, and retail clothing are some of the sectors that are interested in precipitation derivatives. Also, governments are interested in using precipitation derivative in order to quantify the risk of floods and droughts or snow removal. Airports have begun to purchase snow insurance to reduce the cost associated with an exceptionally snowy winter. In addition, restaurants, theme parks, beach resorts, golf courses, automobile insurers, tourism, and entertainment are sectors of the economy that are heavily depended in rainfall and snowfall.

Common underlying weather variables for weather derivatives are temperature, rainfall, and wind. As it is already mentioned, the majority of weather market is dominated by temperature derivatives. However, there is evidence of a growing demand for precipitation weather derivatives especially by farmers, tourism industry, and hydroelectric producers. The demand from these derivatives exists. However, investors hesitate to enter into precipitation contracts. The main reasons of the slow growth of the market compared to temperature contracts are the difficulty in accurately modeling precipitation and the challenge to find a reliable model for valuing related contracts (Cao et al. 2004; Goncu 2011). As a result, there is a lack of reliable valuation framework that makes financial institutions reluctant to quote prices over these derivatives (Goncu 2011).

Unlike temperature, precipitation is harder to model. First, precipitation is not a continuous variable. Rainfall is a binary event. In each, day may or may not be observed precipitation. Moreover, precipitation evolves much more irregularly and unevenly than temperature changes (Stowasser 2012).

In Dischel (2000) and Cao et al. (2004), various difficulties are presented in modeling precipitation. These difficulties include difficulty in accurate measurements. Since most measurement methods depend on the collection of raindrops, the accuracy of the collection accuracy can be affected by local winds or extreme weather and can contain melted snow or hail. Alternatively, radar measurements can be used. Satellites can detect low-altitude atmospheric water content and have excellent spatiotemporal resolution. However, there are two drawbacks on satellites measurements. The current geographic coverage is limited while the historical record is short (Little et al. 2009). Secondly, precipitation is highly localized, and basis risk is present for precipitation contracts for sites that are relative away from the measurement site. Purchasers of weather derivatives are generally exposed to some degree of geographical basis risk (Martin et al. 2001); however, rainfall geographical risk is greater than in the case of temperature. This is already demonstrated in chapter 8. Finally, the statistical properties of precipitation

are complex. Hence, it is difficult to select an appropriate distribution to describe precipitation data.

Two settings exist for modeling precipitation amounts, single-site (Goncu 2011) and multisite models (Wilks 1998, 1999). Single-site models try to model the dynamics of precipitation in a single site, while multisite models are multifractal cascades to describe rainfall (Carmona and Diko 2005).

Precipitation can be forecasting using meteorological models which seek to capture the dynamics of the large-scale atmospheric processes controlling precipitation (Little et al. 2009). Meteorological approaches are of limited use to this book for two reasons, namely, their complexity and limited forecasting horizon. These facts seem to suggest a preference for a statistical approach. The literature on statistically forecasting precipitation includes a variety of distributional approaches. Statistical models use purely statistical techniques to fit the rainfall data to well-known distribution types with little emphasis on underlying physical processes (Coles et al. 2003). Since we want a process to describe the underlying dynamics of precipitation, a stochastic process-based model is applied. Stochastic process-based models try to describe the rainfall behavior by a small set of physically meaningful parameters driving a stochastic process (Carmona and Diko 2005).

In some recent studies, precipitation was modeled by artificial neural networks (Feng and Kitzmiller 2006; Williams 1998; Valverde Ramírez et al. 2005). The framework in these studies was similar. A set of meteorological variables like temperature, humidity, and wind were used in order to forecast possible precipitation amounts.

In this chapter, we follow a similar approach presented in Wilks 2011. Precipitation modeling is separated in two stages. In the first step, a Markov chain process is applied in order to model the frequency of precipitation. Then, a distribution is fitted to the data in order to model the magnitude of the precipitation conditional on the likelihood of a rainy day. In density forecasting a distribution over all possible future rainfall events, rather than a single-point forecast that misrepresents these uncertainties. Density forecasts are particularly flexible, allowing the calculation of the probability of any event of interest, such as the probability of occurrence of rain or of extreme rainfall above any threshold.

The rest of the chapter is organized as follows: in Sect. 10.2, an introduction in precipitation modeling is presented. More precisely, in Sect. 10.2.1, annual modeling of precipitation is presented, while in Sects. 10.2.2 and 10.2.3, monthly and daily modeling is discussed. A daily rainfall model is described in 10.3. More precisely, a process to describe the frequency of precipitation is presented in 10.3.1. A two-state, first-order Markov chain is described in 10.3.1.1, while in 10.3.1.2, a higher-order Markov chain is presented. In Sect. 10.3.1.3, a method for selecting the appropriate order of a Markov chain is presented, while in 10.3.1.4, the presented framework was applied in real data. In Sect. 10.3.1.5, a method for estimating time-varying transition probabilities is described. The magnitude modeling procedure is described in Sect. 10.3.2. More precisely, the gamma distribution is described in Sect. 10.3.2.1, while the exponential and the mixed

exponential distributions are described in Sects. 10.3.2.2 and 10.3.2.3, respectively. An example of magnitude modeling in read data is presented in Sect. 10.3.2.4. A pricing framework for precipitation derivatives is presented in 10.4. More precisely, in Sect. 10.4.1, indifference pricing methodology is presented, while in Sect. 10.4.2, some limitations of this framework are discussed. In Sect. 10.4.3, the hedging effectiveness using rainfall derivative is presented. An alternative pricing methodology using MC simulations is described in Sect. 10.4.4. Finally, in Sect. 10.5, we conclude.

10.2 Precipitation Modeling

In this section, we will investigate the concept of precipitation modeling. More precisely, we will investigate modeling precipitation in annual, monthly, and daily bases. Our dataset consists of daily rainfall observations in Berlin-Tempelhof and Berlin-Dahlem. We have collected daily observations from 1st of January 1951 until 31st of December 2010, resulting to 21,000 values from each station. The data were collected from ECAD.[1] The daily precipitation is measured from 06:00 in the morning till 06:00 the next morning. The precipitation is measured in 0.1 mm. Precipitation under 0.1 mm is considered as an event of nonprecipitation. In other words, the precipitation in this case is labeled as zero. The meteorological station at Tempelhof is located in +52:28:07 latitude and +013:24:14 longitude and in 48 m height. The meteorological station at Tempelhof is located in +52:27:50 latitude and +013:18:06 longitude and in 51 m height. The distance between the two stations is less than 7 km in a straight line. Finally, the dataset is complete without missing or suspicious values.

10.2.1 Annual Rainfall

As it was presented in Moreno (2002), a water company is interested in annual amounts of precipitation. Annual amounts are the period basis on which water companies usually make their plans for water storage (Moreno 2002).

In Fig. 10.1, the yearly precipitation magnitude in Berlin-Tempelhof and Berlin-Dahlem is presented. A closer inspection of Fig. 10.1 reveals that there are years where the difference in the rainfall magnitude is significant between the two meteorological stations. More precisely, the absolute relative difference varies from 0.17% in 1967 to 23.58% in 1959, while the average absolute relative difference for the last 60 years is 6.25%. Hence, even if the distance between these two meteorological stations is only 7 km, there are cases where the difference

[1] http://eca.knmi.nl/

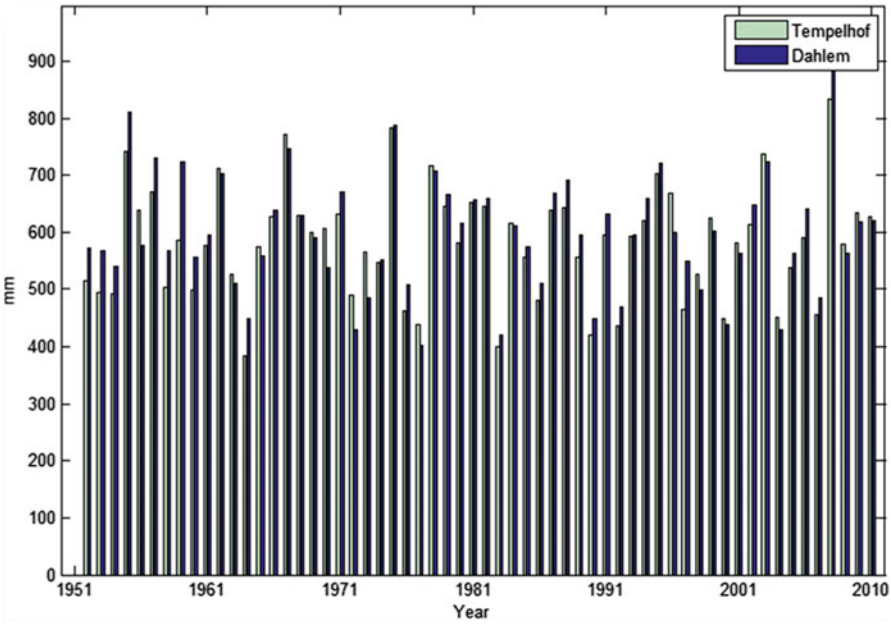

Fig. 10.1 Yearly precipitation magnitude in Berlin-Tempelhof and Berlin-Dahlem

in the annual magnitude of rainfall is significant. Precipitation is a localized weather variable even when it is measured in annual averages where small errors are canceled out. As a result, basis risk is significant and can lead to large errors in hedging and valuation strategies.

The relation between the Tempelhof and Dahlem is presented in Fig. 10.2. Moreover, a linear fitting is presented also with $R^2 = 83.83\%$. The correlation of the annual precipitation magnitude between the two stations is 91%.

10.2.2 Monthly Rainfall

In Leobacher and Ngare (2010), monthly precipitation modeling was proposed. More precisely, the precipitation is modeled in monthly basis by a Markov-gamma process with seasonality. In Fig. 10.3, the average cumulative monthly precipitation in Berlin-Tempelhof and Berlin-Dahlem is presented. Although the monthly precipitation pattern is similar, the absolute relative error is up to 10% in August, while the average absolute relative error is 3.05%. In Fig. 10.4, the correlation of the monthly precipitation is presented. A closer inspection of Fig. 10.4 reveals that the correlation is higher from October till April and it is over 95% with an exception of February where it is around 92%. On the other hand, the correlation is significantly smaller between May and September with values ranging from 86% to 89%.

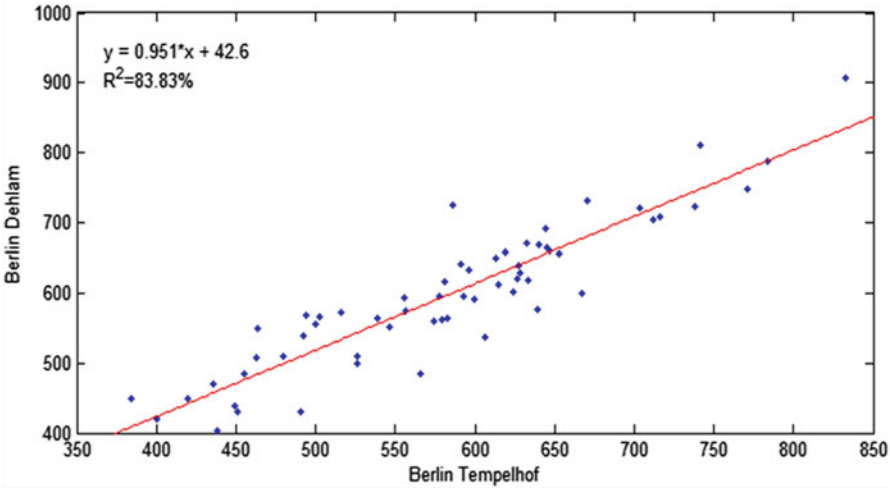

Fig. 10.2 Correlation between yearly magnitude in Berlin-Tempelhof and Berlin-Dahlem

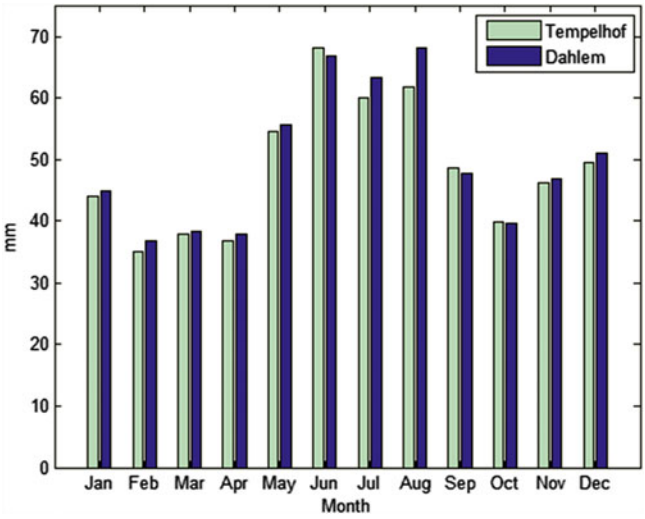

Fig. 10.3 Monthly precipitation magnitude in Berlin-Tempelhof and Berlin-Dahlem

A closer inspection of Fig. 10.3 reveals that rainfall is more intense in summer than it is in winter. More precisely, in winter, rainfall occurs more often; however, it has small intensity. On the other hand, during summer, rainfall occurs in a form of short intense storms. Moreover, these storms last for few minutes and usually affect only a very small area. This is the main reason why correlation is significantly lower during the summer period.

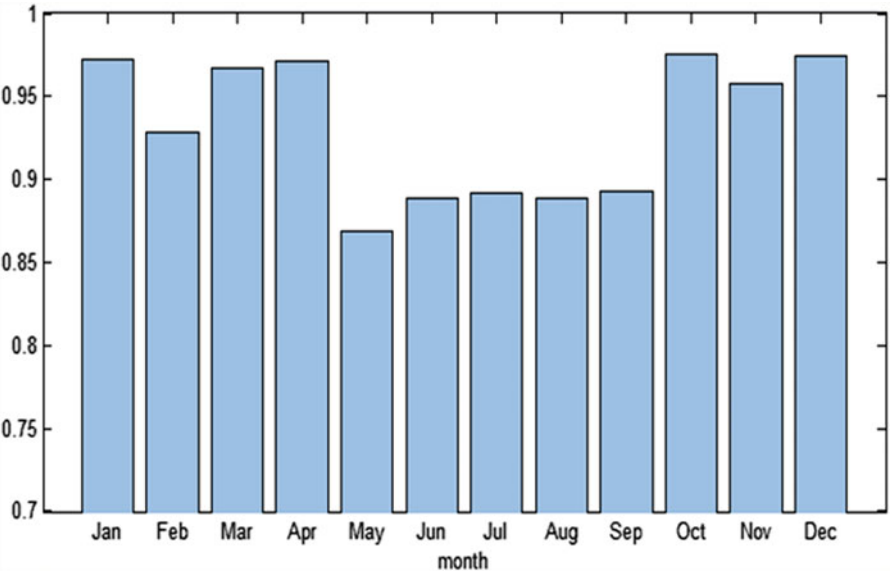

Fig. 10.4 Correlation of monthly precipitation between Berlin-Tempelhof and Berlin-Dahlem

As a result, for hedging purposes one must be very careful. Even in the case of small distances, the localization of rainfall can induce large error in the hedging strategy or in the derivative valuation process. In addition as it is shown from Fig. 10.4, the correlation between the two meteorological stations in Berlin is not constant but it rather has a strong seasonal variation.

10.2.3 Daily Rainfall

Daily precipitation modeling is more difficult than modeling rainfall in annual or yearly basis. Precipitation is not a continuous variable. In other words, there is possibility to occur several days without any precipitation. However, a model for daily precipitation provides more flexibility and more information regarding the dynamics that govern the frequency and amount processes of precipitation. After a daily model is estimated, it can be used in order to derive any rainfall index in daily, monthly, or yearly basis.

In Fig. 10.5, the average daily rainfall magnitude for the period between 1951 and 2010 for Berlin-Tempelhof and Berlin-Dahlem is presented. The correlation between monthly magnitude in Berlin-Tempelhof and Berlin-Dahlem is presented in Fig. 10.6. A closer inspection of Fig. 10.6 reveals the existence of really big differences between the two metrological differences. Hence, in the case of daily rainfall protection, the reference site upon which a contract is based should be extremely close to the location to be insured (Moreno 2002). As it is mentioned in

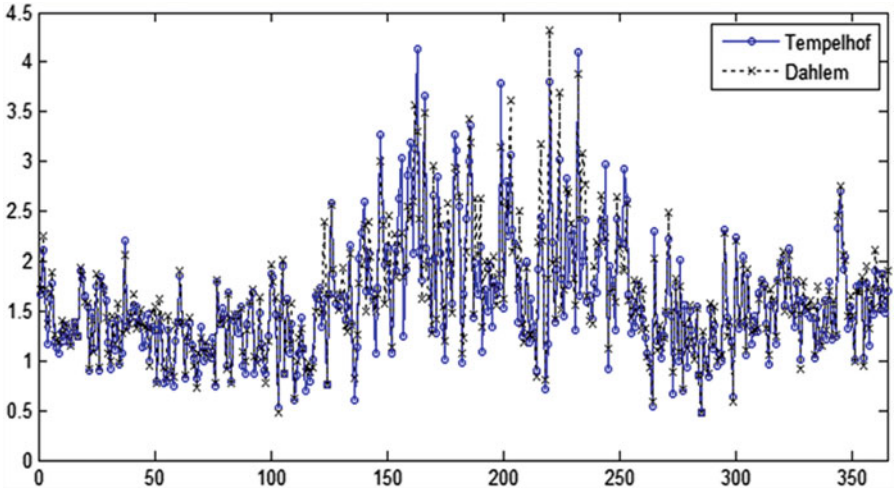

Fig. 10.5 Average daily rainfall magnitude for the period between 1951 and 2010 for Berlin-Tempelhof and Berlin-Dahlem

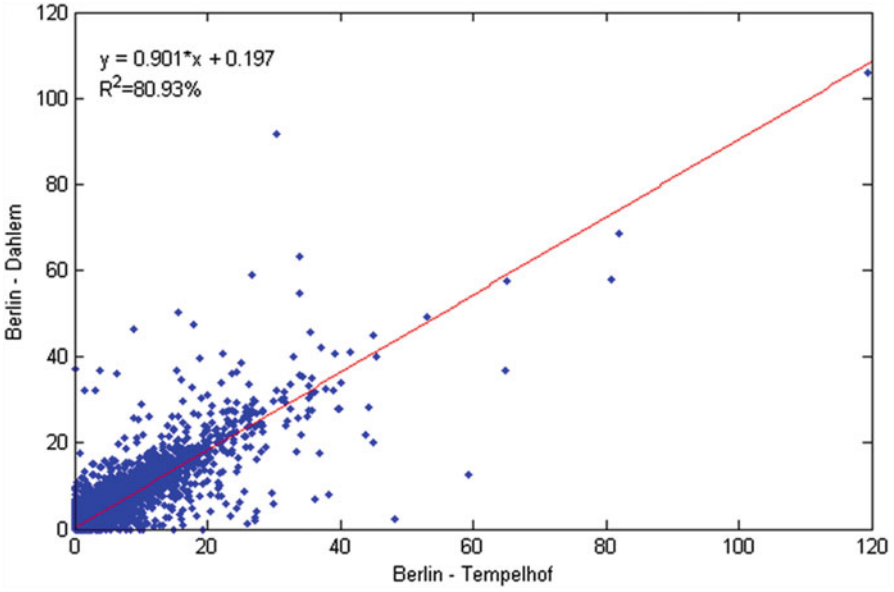

Fig. 10.6 Correlation between monthly magnitude in Berlin-Tempelhof and Berlin-Dahlem

Moreno (2002), for longer periods, proxy sites can be considered with caution. Moreover, if the required protection on a site is daily, the use of several meteorological stations surrounding the location should be considered (Moreno 2002).

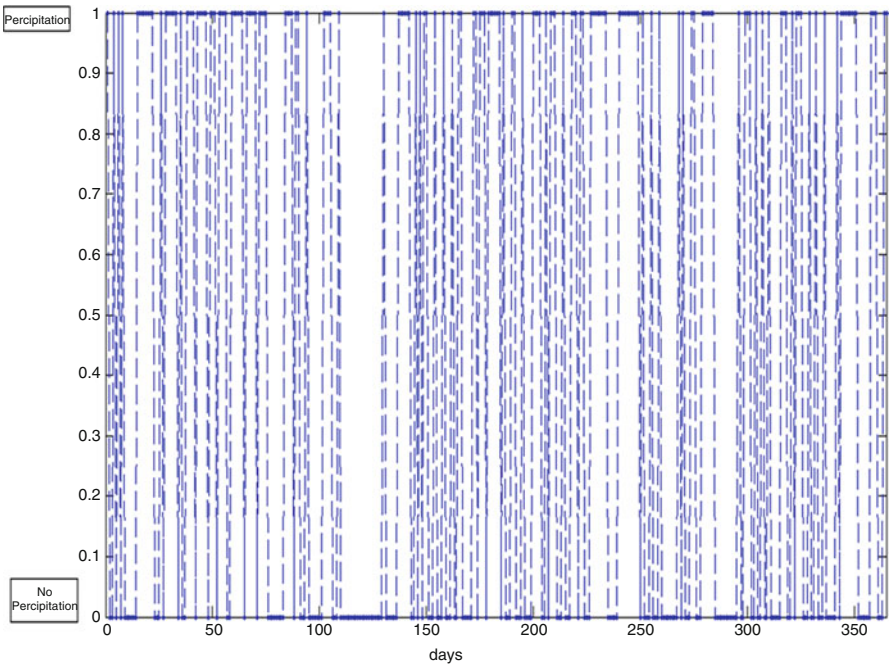

Fig. 10.7 Rain frequency pattern in Berlin-Tempelhof for the year 2000

10.3 A Daily Rainfall Process

As it is already mentioned, a daily model provides more flexibility, since every precipitation index can be built in easily and more information regarding the dynamics that govern the frequency and amount processes of precipitation. In addition, the accuracy of a daily model is expected to be higher since a large amount of data is used in order to calibrate it (Stowasser 2012). Finally, it is possible to incorporate weather and seasonal forecasts. However, small misspecifications in the model can lead to large pricing errors in the valuation of precipitation derivatives. In Musshoff et al. (2006), the basic characteristics of a precipitation model are presented. A precipitation model should capture the following characteristics of daily rainfall:

- The probability of rainfall occurrence obeys a seasonal pattern.
- The sequence of wet and dry days follows an autoregressive process.
- The amount of precipitation on a wet day varies with the season.
- The volatility of the amount of rainfall also changes seasonally.

In Fig. 10.7, the rain frequency pattern in Berlin-Tempelhof for the year 2000 is presented.

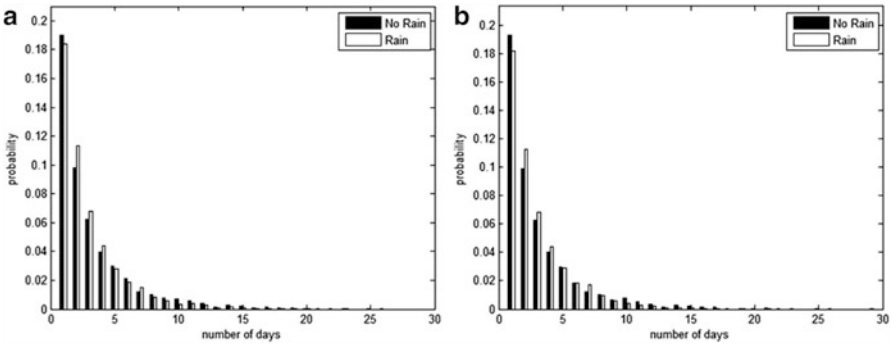

Fig. 10.8 Historical distribution of the length of rainy (*white*) and dry (*black*) days in (**a**) Berlin-Tempelhof and (**b**) Berlin-Dahlem

Investigating the data from Berlin for the last 60 years for both meteorological stations, we conclude that there is evidence of absence of a linear trend. In addition a linear trend was fitted to the precipitation occurrence days only. Again, the slope of the linear trend is not significantly different than zero. The same conclusion is reached when only the last 10 years were examined.

Inspecting Fig. 10.7, we conclude that rainfall occurred more frequently in winter and midsummer while on the other hand precipitation occurred more rarely during the late spring and early summer and during the late summer and early autumn periods. Finally, in Fig. 10.8 the historical distribution of the length of consecutive rainy and dry days in Tempelhof and Dahlem is presented.

10.3.1 Frequency Modeling

As it is already mentioned, precipitation is a different variable in contrast to wind or temperature. In order to develop a model for the daily average rainfall, first we must understand with which probability rainfall occurs. Figure 10.7 detects if there is any persistence of rainfall. Figure 10.7 represents the events of "precipitation" or "no precipitation" on a daily basis during the year of 2000. Consistent with previous studies, Figure 10.7 reveals that a rainy day is more likely to be followed by a rainy day. Similarly, a dry day is more likely to be followed by a dry day than a rainy one. Hence, the probability that it rains is conditional on the past.

10.3.1.1 A Two-State, First-Order Markov Chain

A Markov chain model is described by two properties. The first is the values that the variable can take. This is known as the state of the chain. In the case of

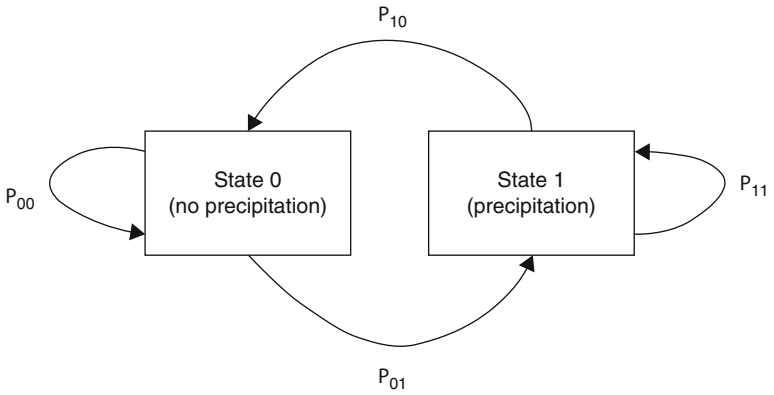

Fig. 10.9 A two-state, first-order Markov chain for precipitation occurrence

precipitation, there are two states. The random variable X_t is defined to describe the two states, "precipitation" $(X_t = 1)$ and "no precipitation" $(X_t = 0)$:

$$X_t = \begin{cases} 0 \text{ precipitation} \\ 1 \text{ no precipitation.} \end{cases} \tag{10.1}$$

The second property is the order of the chain. The order defines the number of previous values that are necessary to determine the state-to-state transition probabilities. Hence, the simplest Markov model is a two-state, first-order Markov model used in Cao et al. (2004), Odening et al. (2007), Goncu (2011), and Moreno (2002) (Fig. 10.9). In this case, the Markovian property can be expressed as

$$\Pr\{X_{t+1}|X_t, X_{t-1}, \ldots, X_1\} = \Pr\{X_{t+1}|X_t\}. \tag{10.2}$$

A two-state, first-order Markov chain is described by four transition probabilities. The transition probabilities are conditional probabilities for the state time at time $t + 1$ given the state at time t:

$$p_{00} = \Pr\{X_t = 0|X_{t-1} = 0\} \tag{10.3}$$

$$p_{01} = \Pr\{X_t = 1|X_{t-1} = 0\} \tag{10.4}$$

$$p_{10} = \Pr\{X_t = 0|X_{t-1} = 1\} \tag{10.5}$$

$$p_{11} = \Pr\{X_t = 1|X_{t-1} = 1\}. \tag{10.6}$$

Hence, (10.3) and (10.4) define the conditional probability distribution for the value of the variable at time $t + 1$ given that $X_t = 0$ at time t. Similarly, (10.5) and (10.6) define the conditional probability distribution for the value of the variable at time $t + 1$ given that $X_t = 1$ at time t.

It is crucial to estimate the transition probabilities given by (10.3, 10.4, 10.5 and 10.6). However, it is not necessary to estimate all four parameters since there is some redundant information given by (10.3, 10.4, 10.5 and 10.6) (Wilks 2011). More precisely, we have that $p_{00} + p_{01} = 1$ and that $p_{10} + p_{11} = 1$. Hence, we can focus on estimating only two parameters, one of each pair of the transition probabilities. For example, we focus on p_{01} (the probability that precipitation will occur tomorrow given that precipitation did not occur today) and p_{11} (the probability that precipitation will occur tomorrow given that precipitation did occur today). Estimations of the transition probabilities can be obtained using the MLEs:

$$\hat{p}_{01} = \frac{n_{01}}{n_{00} + n_{01}} \tag{10.7}$$

$$\hat{p}_{11} = \frac{n_{11}}{n_{10} + n_{11}}, \tag{10.8}$$

where n_{01} is the historical count of wet days that followed dry days and n_{00} is the number of transitions from state 0 (dry day) to state 0 (dry day).

As it is suggested by Wilks (2011), the estimation of parameters of (10.7) and (10.8) for a two-state, first-order Markov chain is equivalent to fitting two Bernoulli distributions with $N = 1$. The first Bernoulli distribution describes the behavior of the points in the time series that follow the state 0 while the second one describes the behavior of the points in the time series that follow the state 1.

Due to the properties of the Markov chain, various properties for the time series can be computed. These properties are described by the transition probabilities. First, we estimated the stationary probabilities. The stationary probabilities are the long-run relative frequencies of the events corresponding to a two-state Markov chain (Wilks 2011). In the case of precipitation, the stationary probability π_1 describes the unconditional probability of precipitation and it is given by

$$\pi_1 = \frac{p_{01}}{1 + p_{01} - p_{11}} \tag{10.9}$$

and for state 0, the stationary probability is given by $\pi_0 = 1 - \pi_1$. According to Wilks (2011), the serial correlation that often exhibits the precipitation time series results to $p_{01} < \pi_1 < p_{11}$. This means that the conditional probability of a dry day followed by a wet dry is less than the unconditional probability of a wet day which in turn is less than the conditional probability of a wet day followed by a wet day (Wilks 2011).

As it is already presented, precipitation time series exhibit strong autocorrelation. In the sense of the transition probabilities, the lag-1 autocorrelation is given by

$$r_1 = p_{11} - p_{01}. \tag{10.10}$$

Relation (10.10) reveals that as the value of r_1 increases, the difference between p_{11} and p_{01} increases. Hence, it follows that a wet day is more likely to be followed by a wet day. Similarly, the lag-k autocorrelation can be easily calculated. Due to

the Markovian property the lag-k autocorrelation is simply the lag-1 autocorrelation at the power of k:

$$r_k = (r_1)^k. \tag{10.11}$$

10.3.1.2 A Higher-Order Markov Chain

As it is already mentioned in the previous section, in a first-order Markov chain the transition probabilities depend only on the state of the variable in the previous time period. Similarly, an m-order Markov chain is one where the transition probabilities depend on the states of the variable in the previous m time periods only (Stowasser 2012). The transition probabilities are conditional probabilities for the state time at time $t + 1$ given the states at time $t, t - 1, \ldots, t - m$. Hence, the Markovian property is expressed as

$$\Pr\{X_{t+1}|X_t, X_{t-1}, \ldots, X_1\} = \Pr\{X_{t+1}|X_t, X_{t-1}, \ldots, X_{t-m}\}. \tag{10.12}$$

For example, in a second-order Markov chain the transition probabilities depend on the states at lags of both one and two time periods. Hence, the transition probability for a second-order Markov chain can be defined as

$$p_{hij} = \{X_{t+1} = j|X_t = i, X_{t-1} = h\}. \tag{10.13}$$

Note that in the case of precipitation, there are only two states ($s = 2$). As a result the subscripts h, i, and j take the values 0 (no precipitation) and 1 (precipitation).

As in the case of the first-order Markov chain, the transition probabilities are estimated by the historical transition counts that are obtained from the data. However, since the state at the next time period depends on more previous time steps, more transition probabilities must be estimated. In general, for an s-state, m-order Markov chain there are s^{m+1} different transition probabilities. It is clear that the number of the possible transition probabilities increases exponentially with the order, m.

For a two-state, second-order Markov chain where $s = m = 2$, the number of transition probabilities that must be estimated is 8. In Table 10.1 the different possible outcomes as well as the historical counts for these outcomes for a two-state, second-order Markov chain are presented. The historical counts are determined by examining consecutive groups of three historical values. In the general case, these groups consists of m + 1 data points. In order to simplify the equation complexity, the notation from Wilks (2011) is adapted were $n_{00*} = n_{000} + n_{001}$. Hence, for a second-order Markov chain the conditional transition probabilities are obtained from the conditional relative frequencies of the transition counts:

$$\hat{p}_{hij} = \frac{n_{hij}}{n_{hi*}}. \tag{10.14}$$

Table 10.1 Arrangement of the $2^{2+1} = 8$ transition counts for the two-state, second-order Markov chain

X_{t-1}	X_t	$X_{t+1} = 0$	$X_{t+1} = 1$	Marginal Totals
0	0	n_{000}	n_{001}	$n_{00\bullet} = n_{000} + n_{001}$
0	1	n_{010}	n_{011}	$n_{01\bullet} = n_{010} + n_{011}$
1	0	n_{100}	n_{101}	$n_{10\bullet} = n_{100} + n_{101}$
1	1	n_{110}	n_{111}	$n_{11\bullet} = n_{110} + n_{111}$

Note that if only two states exist, as in the case of precipitation, then the two-state, second-order Markov chain conditional transition probabilities that are given by (10.14) result to

$$\hat{p}_{000} = \frac{n_{000}}{n_{00*}} = \frac{n_{000}}{n_{000} + n_{001}}. \tag{10.15}$$

Similarly, the higher-order transition probabilities are obtained. For example, the conditional transition probabilities for a third-order Markov chain are given by

$$\hat{p}_{ghij} = \frac{n_{ghij}}{n_{ghi*}} = \frac{n_{ghij}}{n_{ghi0} + n_{ghi1}}. \tag{10.16}$$

10.3.1.3 Selecting the Order of the Markov Chain

In order to determine the order of the Markov chain, two criteria can be used. The first one is the AIC proposed by Akaike (1974), and the second one is the Bayesian information criterion (BIC) proposed by Schwarz (1978). The AIC were proposed to estimate the order of the Markov chain by Wilks (2011), while the BIC by Schoof and Pryor (2008), Stowasser (2012), and Wilks (2011).

Both criteria are based on the log-likelihood function. In Wilks (2011), the log-likelihoods for s-state Markov chains of order 0, 1, 2, and 3 are presented:

$$L_0 = \sum_{j=0}^{s-1} n_j \ln\left(\hat{p}_j\right) \tag{10.17}$$

$$L_1 = \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} n_{ij} \ln\left(\hat{p}_{ij}\right) \tag{10.18}$$

$$L_2 = \sum_{h=0}^{s-1} \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} n_{hij} \ln\left(\hat{p}_{hij}\right) \tag{10.19}$$

and

$$L_3 = \sum_{g=0}^{s-1} \sum_{h=0}^{s-1} \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} n_{ghij} \ln\left(\hat{p}_{ghij}\right). \tag{10.20}$$

Table 10.2 The Historical counts and the transition probabilities for a two-state, first-order Markov chain for the two meteorological stations in Berlin

	Berlin-Tempelhof		Berlin-Dahlem	
	Historical counts	Probability	Historical counts	Probability
State 0 to state 0	8,025	0.6932	7,885	0.6867
State 0 to state 1	3,551	0.3068	3,598	0.3133
State 1 to state 0	3,550	0.3439	3,597	0.3453
State 1 to state 1	6,774	0.6561	6,820	0.6547

The extension to higher-order Markov chains can be easily obtained similarly.

The two criteria try to find the model with best fit given the minimum parameters. In other words, there is a penalty function depending on the number of the parameters used to fit the model. The AIC criterion for order m is given by

$$\mathrm{AIC}(m) = -2L_m + 2s^m(s-1), \tag{10.21}$$

while the BIC is given by

$$\mathrm{BIC}(m) = -2L_m + s^m \ln(n). \tag{10.22}$$

The only difference on the two criteria is the penalty function. However, the use of the BIC in general is more preferable (Stowasser 2012; Wilks 2011).

10.3.1.4 An Example of Selecting the Order of the Markov Chain

In this section, the BIC and AIC will be used in order to select the appropriate order of a Markov chain that describes the precipitation in Berlin.

For simplicity, we start with a two-state, first-order Markov chain. First, the historical counts for each event are counted. For Berlin-Tempelhof, we have that $n_{00} = 8025$, $n_{01} = 3551$, $n_{10} = 3550$, and $n_{11} = 6774$. Hence, from (10.7) and (10.8), the conditional transition probabilities can be calculated. Hence, we have that $\hat{p}_{00} = 0.6932, \hat{p}_{01} = 0.3068, \hat{p}_{10} = 0.3439$, and $\hat{p}_{11} = 0.6561$. In Table 10.2 the transition probabilities and the historical counts for a two-state, first-order Markov chain for Berlin-Tempelhof and Berlin-Dahlem are presented.

Next, the stationary probabilities and the lag-1 autocorrelation can be estimated. Following (10.9), we have that π_1 is 0.4715 and 0.4753 for Tempelhof and Dahlem, respectively. Moreover, $\pi_2 = 1 - \pi_1$ is 0.5285 and 0.5247 for Tempelhof and Dahlem, respectively. The lag-1 autocorrelation, r_1, in terms of the transition probabilities is given by (10.10), and it is 0.3494 and 0.3414 for Tempelhof and Dahlem, respectively.

In order to calculate the AIC and BIC, first, the log-likelihood function must be estimated. Since our starting model is a two-state, first-order Markov order, (10.18) is used. Hence, for Tempelhof we have that $L_1 = -13781$, while for Dahlem $L_1 = -13853$. Once the log-likelihood function is estimated, the AIC and BIC criteria can be derived applying (10.21) and (10.22), respectively.

Table 10.3 Estimation of AIC and BIC for Berlin-Tempelhof and Berlin-Dahlem for different orders of the Markov chain

	Tempelhof		Dahlem	
m	AIC	BIC	AIC	BIC
0	30,290	30,298	30,310	30,318
1	27,565	27,581	27,710	27,726
2	27,428	27,460	27,578	**27,610**
3	27,387	**27,451**	27,552	27,616
4	**27,382**	27,509	**27,546**	27,674
5	27,389	27,645	27,552	27,808
6	27,419	27,931	27,585	28,096

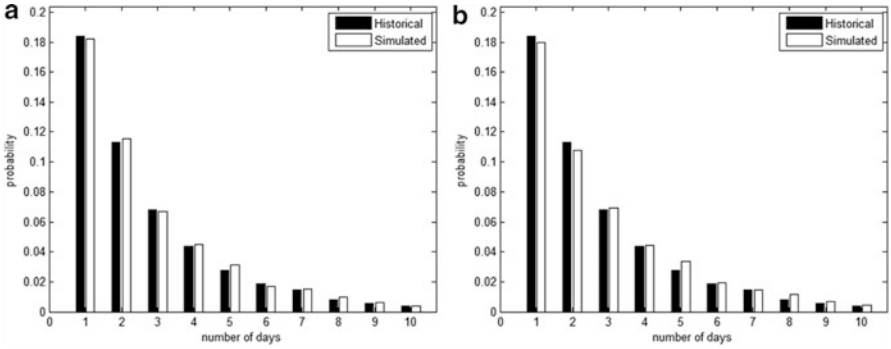

Fig. 10.10 Historical (*black*) and simulated (*white*) frequency of rainy days in (**a**) Berlin-Tempelhof and (**b**) Berlin-Dahlem. The simulated data produced by a third- and a second-order Markov chain, respectively

In Table 10.3, the AIC and BIC criteria for Berlin-Tempelhof and Berlin-Dahlem for different orders of the Markov chain are presented. A closer inspection of Table 10.3 reveals that AIC tends to propose higher-order models than BIC. Moreover, BIC proposed a third-order model for Tempelhof and a second-order model for Dahlem.

Finally, Fig. 10.10 shows the historical and the simulated frequency of rainy days for the two meteorological stations in Berlin. The simulated data produced by Markov chain on which order was suggested by the BIC criterion. Hence, a third-order Markov chain was used for Tempelhof while a second order was used for Dahlem. A closer inspection of Fig. 10.10 reveals that, for both stations, the historical and simulated probabilities of up to 10 consecutive days of rain are very close. As a result these models can be used in order to simulate the frequency process of precipitation in the two under inspection meteorological stations.

10.3.1.5 Time-Varying Transition Probabilities

The framework presented in the previous section is very elegant and easy in implementation. However, it has the disadvantage of not taking into account

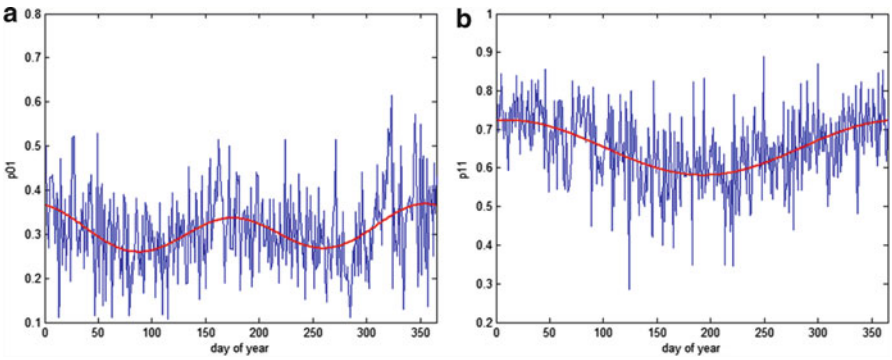

Fig. 10.11 Empirical (*thick line*) and historical transition probabilities (**a**) $\hat{p}_{01}$ and (**b**) $\hat{p}_{11}$ for Berlin-Tempelhof

seasonal variation of rainfall. For some countries this disadvantage may even render that model useless (Leobacher and Ngare 2010).

In order to overcome this drawback, in this section, an algorithm that estimates the daily and monthly transition probabilities is presented. First, the amounts of precipitation are grouped into 365 groups, comprising 60 observations each (each group corresponds to a single day of the year). Then, starting from the first group (1^{st} of January) for each observation, the historical counts are calculated. Then, transition probability for that day can be estimated. Finally, the same procedure is repeated for the remaining groups. That is, we assume that the transition probabilities vary seasonally and are repeated every year, that is, $\hat{p}_{11}(t + 365) = \hat{p}_{11}(t)$.

In Fig. 10.11, the estimated daily transition probabilities $\hat{p}_{01}$ and $\hat{p}_{11}$ for Berlin-Tempelhof are presented. For simplicity, the transition probabilities $\hat{p}_{01}$ and $\hat{p}_{11}$ in Fig. 10.11 are produced by applying a two-state, first-order Markov chain. It is clear, a seasonal component is presented. More precisely, a yearly seasonal component is present in the $\hat{p}_{11}$ series, while a biannual seasonal component is present in the $\hat{p}_{01}$ series. The seasonality in the transition probabilities can be easily estimated by fitting a truncated Fourier series:

$$\hat{p}(t) = c + \sum_{i=1}^{I} c_i \sin(2i\pi t/365) + \sum_{j=1}^{J} d_j \cos(2j\pi t/365). \tag{10.23}$$

In this example, a second-order $(I = J = 2)$ truncated Fourier series was used in order to fit the seasonal component of $\hat{p}_{01}$ while a first order $(I = J = 1)$ for $\hat{p}_{11}$. The results for the meteorological station in Berlin-Dahlem are similar. In general, as I and J increase, (10.23) produces more cycles in the seasonality pattern of the transition probabilities. However, larger I and J reduce the estimation accuracy (Cao et al. 2004).

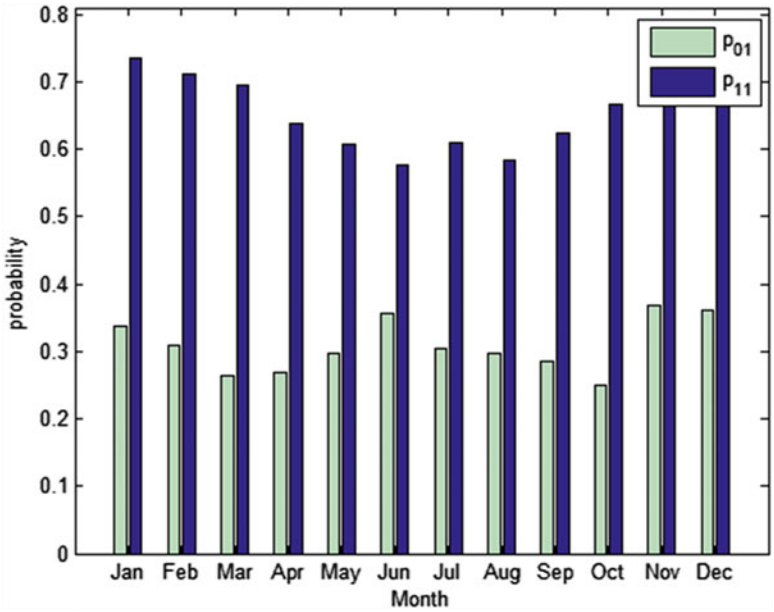

Fig. 10.12 Monthly transition probabilities $\hat{p}_{01}$ and $\hat{p}_{11}$ for Berlin-Tempelhof

Alternatively, the transition probabilities can be estimated in monthly basis. Figure 10.12 presents the monthly transition probabilities $\hat{p}_{01}$ and $\hat{p}_{11}$ for Berlin-Tempelhof. As it was expected, the form of the seasonal components is preserved.

The advantage of this method is that it allows us to build a multi-order Markov chain model. More precisely, the transition probabilities can be used to estimate the log-likelihood function and then AIC or BIC criteria for each month. Hence, it is possible to apply a different Markov chain model in order to simulate the frequency process in each month. This method has the potential of additional accuracy since a different Markov chain is applied in each month of the year and potentially can capture various dynamics of the rainfall process that a constant model may be not able to do so.

10.3.2 Magnitude Modeling

The second part of the modeling process is to estimate the magnitude of the precipitation conditional on the fact that it rains on that particular day. This is usually done by fitting a distribution to the data.

In order to indentify a proper distribution to fit the data, first, the histogram of the daily rainfall is examined. In part (a) of Fig. 10.13, the histogram of the daily

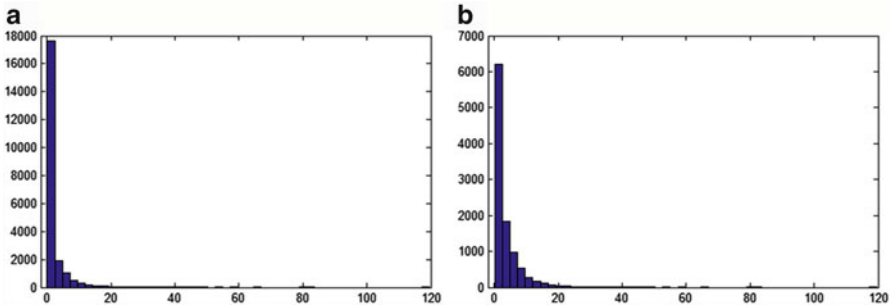

Fig. 10.13 Histogram of the (**a**) daily precipitations in Berlin-Tempelhof and (**b**) daily precipitation excluding zeros

precipitation in Berlin-Tempelhof from 1/1/1951 to 31/12/2010 is presented. A closer inspection of Fig. 10.13 reveals that the distribution of the precipitation is asymmetric, skewed to the right, and with only positive values.

However, almost 50% of the data have the same value, zero indicating days of no precipitation. In the second part of Fig. 10.13, the histogram of the precipitation levels, excluding the days of no precipitation, is presented. A closer inspection of part (b) of Fig. 10.13 reveals that the distribution remains asymmetric and positive skewed.

Different distributions with a nonnegative domain have been proposed to fit the historical precipitation data in literature so far. A gamma distribution is proposed by Stowasser (2012), Cao et al. (2004), Leobacher and Ngare (2010), Williams (1998), Goncu (2011), Vedenov and Miranda (2001), Martin et al. (2001), and Wilks (1999, 2011) while a mixed exponential distribution is proposed by Cao et al. (2004), Goncu (2011), and Wilks (1999). This distribution has the advance to better represent extreme events. Similarly, in Odening et al. (2007), a beta distribution is found to provide the better fit in data, while in Odening et al. (2007) and Musshoff et al. (2006), a Weibull distribution is used. A kernel density estimation method is applied in Goncu (2011) and Cao et al. (2004).

Since events of extreme rainfall are often reported, Goncu (2011) and Carmona and Diko (2005) propose a Markov chain that incorporates jumps driven by a Poisson process. Finally, Coles et al. (2003) propose the generalized extreme value family of distributions for annual rainfall and a Pareto distribution for daily rainfall in order to capture events of extreme rainfall in Venezuela.

10.3.2.1 The Gamma Distribution

Given that we are on a wet day (i.e., $X_t = 1$), the amount of rainfall Y_t is modeled as a random variable that follows a particular distribution. It is clear from the literature that a widely used choice is the gamma distribution. The gamma distribution is often used to describe the distribution of various atmospheric variables such as

precipitation amounts or wind speeds (Wilks 2011). The gamma distribution exhibits the same characteristics as the distribution of precipitation amounts. It is positive skewed, asymmetric, and bounded on the left by zero. The gamma distribution is defined by its PDF which is given by

$$f_G(x) = \frac{(x/\beta)^{\alpha-1} e^{(-x/\beta)}}{\beta\Gamma(\alpha)} \quad x, \alpha, \beta > 0, \tag{10.24}$$

where α is the shape parameter and β is the scale parameter. $\Gamma(\alpha)$ is a gamma function evaluated at α, and it is given by

$$\Gamma(\alpha) = \int_0^{+\infty} t^{a-1} e^{-t} dt. \tag{10.25}$$

From (10.24), it is clear that the PDF of a gamma distribution is fully defined by the parameters α and β. The simplest method to estimate the scale and shape parameters is to use the method of moments. The mean of the gamma distributions is given by $\alpha\beta$, while the variance by $\alpha\beta^2$. Hence, we have that

$$\hat{\alpha} = \frac{\bar{x}^2}{s^2} \tag{10.26}$$

and

$$\hat{\beta} = \frac{s^2}{\bar{x}} = \frac{\bar{x}}{\hat{\alpha}}, \tag{10.27}$$

where s^2 is the estimated variance and $\bar{x}$ is the estimated mean. However, this method can yield poor results from small values of the shape parameter (Wilks 2011). Alternatively, the parameters can be estimated using the maximum likelihood estimation method.

Following the method presented in Wilks (2011), the maximum likelihood estimator can be estimated by first calculating the sample statistic:

$$D = \ln(\bar{x}) - \frac{1}{n} \sum_{i=1}^{n} \ln(x_i). \tag{10.28}$$

Then, the shape parameter is given by

$$\hat{\alpha} = \frac{1 + \sqrt{1 + 4D/3}}{4D} \tag{10.29}$$

and the scale parameter by

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}}. \tag{10.30}$$

However, precipitation contains a lot of zero values. In this case a different approach in estimating the maximum likelihood estimator is followed. Suppose that the given dataset contains N_0 data points recorded as zeros and N_W is the number of nonzero data points. Then, the likelihood function is given by

$$\begin{aligned} L(x;\alpha,\beta) &= \prod_{i=1}^{N_0} G(\alpha,\beta;C) \prod_{i=1}^{N_w} f_G(\alpha,\beta;x_i) \\ &= [G(\alpha,\beta;C)]^{N_0} \prod_{i=1}^{N_w} f_G(\alpha,\beta;x_i), \end{aligned} \tag{10.31}$$

where C denotes the censoring level and the CDF $G(\alpha,\beta,C)$ is given by

$$G(\alpha,\beta,C) = \int_0^C f_G(\alpha,\beta;x_i)dx = \mathrm{P}[x_i \le C]. \tag{10.32}$$

In the case where $N_0 = 0$, we obtain the same maximum likelihood estimators. In the case where $N_0 \neq 0$, the log-likelihood function is given by

$$\begin{aligned} \Lambda(x;\alpha,\beta) &= N_0 \ln[G(\alpha,\beta;C)] - N_w[\alpha \ln(\beta) + \ln(\Gamma(\alpha))] \\ &\quad + (\alpha-1)\sum_{i=1}^{N_w} \ln(x_i) - \frac{1}{\beta}\sum_{i=1}^{N_w} x_i. \end{aligned} \tag{10.33}$$

The amount of censored values in (10.33) depends on the censoring level C. The maximization of (10.33) for the parameters α and β can be computed numerically.

10.3.2.2 The Exponential Distribution

The exponential distribution is a special case of the gamma distribution. It is obtained by setting $\alpha = 1$. The PDF of the exponential distribution is given by

$$f_{EXP}(x) = \frac{1}{\beta} e^{(-x/\beta)} \quad x \ge 0, \tag{10.34}$$

while the CDF is defined by

$$F_{EXP}(x) = 1 - e^{(-x/\beta)}. \tag{10.35}$$

The exponential distribution describes the time between the events in a Poisson process, that is, a process which events occur continuously and independently at a constant average rate. The mean of the distribution is given by β and the variance by β^2. The main application area is in studies of lifetime. The exponential distribution is special because of its utility in modeling events that occur randomly over time. As a result it is often used in rainfall modeling.

10.3.2.3 The Mixed Exponential Distribution

Alternatively, to the gamma distribution, a mixed exponential distribution is often used in order to describe the distribution of the precipitation amount. The mixed exponential is a weighted combination of two simple exponential distributions and inherits their properties (Odening et al. 2007). This distribution has the advantage of a better representation of extreme events.

The mixed exponential distribution is defined by the following PDF:

$$f_{\mathrm{MIX}}(x) = \frac{\alpha}{\beta_1} e^{-\frac{x}{\beta_1}} + \frac{1-\alpha}{\beta_2} e^{-\frac{x}{\beta_2}} \quad 0 \leq \alpha \leq 1, \ 0<\beta_1<\beta_2, \tag{10.36}$$

and the CDF is defined by

$$F_{\mathrm{MIX}}(x) = \alpha e^{-\frac{x}{\beta_1}} + (1-\alpha) e^{-\frac{x}{\beta_2}}. \tag{10.37}$$

The maximum likelihood estimators for three parameters α, β_1 and β_2 can only evaluated numerically.

10.3.2.4 An Example of Magnitude Modeling

In this section, a gamma distribution will be used in order to simulate the daily average rainfall given the probability that the particular day is a rainy day in Berlin-Tempelhof.

Applying (10.33) to the rainfall dataset from Tempelhof, the two parameters that define the gamma distribution can be estimated. The estimated shape parameter is $\alpha = 0.2576$ while the estimated scale parameter is $\beta = 6.2152$. A closer inspection of Fig. 10.14 reveals that the simulation underestimated the daily average rainfall over the summer period. This is clear in both parts of Fig. 10.14 that correspond to the daily and to the cumulative monthly rainfall. Among the three distributions – gamma, exponential and a mixed exponential – the gamma distribution provided the best results.

Finally, in Table 10.4 the estimation of the first four moments are presented using the HBA, a 3rd order Markov chain with gamma distribution and a 3rd order Markov chain with exponential distribution. Again, the gamma distribution

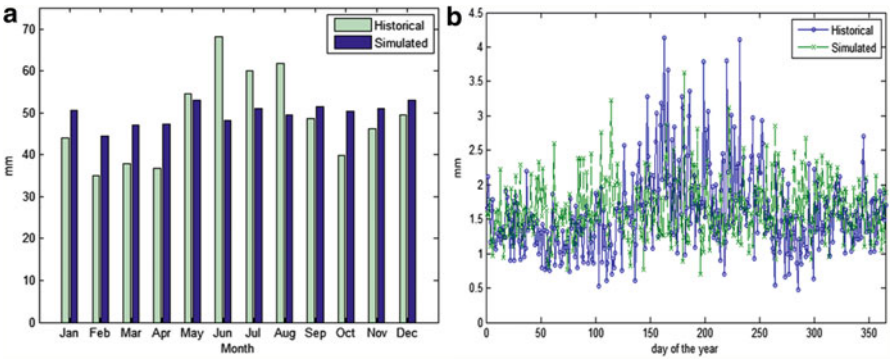

Fig. 10.14 Historical and simulated (**a**) monthly and (**b**) daily rainfall in Berlin-Tempelhof. The simulated rainfall magnitude produced by a gamma distribution with $\alpha = 0.2576$ and $\beta = 6.2152$

Table 10.4 Comparison of the gamma, exponential, and HBA models in estimating the descriptive statistics

	Actual	HBA	Gamma	Exp
Mean	0.67419	1.4328	0.65409	0.63468
Std	1.3178	0.32567	0.42555	0.29224
Kurtosis	6.5895	2.01570	4.61450	3.31100
Skewness	2.1807	0.17269	1.08990	0.52240

Std standard deviation, *HBA* historical burn analysis, *Gamma* a 3rd-order Markov model with gamma distribution, *Exp* a 3rd-order Markov model with exponential distribution

provides the best results. However, although a very close estimation of the mean is produced, all models underestimate the standard deviation, the skewness and the kurtosis. This is a common pitfall of daily models often reported in literature (Musshoff et al. 2006; Dubrovsky et al. 2004).

10.4 Pricing Precipitation Derivatives

In this section, a pricing framework for rainfall derivatives will be presented. First, we define the cumulative rainfall (CR) index. The CR for a period $[\tau_1, \tau_2]$ is defined by

$$\mathrm{CR} = \sum_{t=\tau_1}^{\tau_2} Y(t). \tag{10.38}$$

The CR index is simply the sum of the rainfall over the period $[\tau_1, \tau_2]$. If Q is the risk-neutral probability and r is the constant compounding interest rate, then the arbitrage-free future price of a CR contract at time $t \leq \tau_1 < \tau_2$ is given by

$$e^{-r(\tau_2-t)}\mathrm{E}_Q\left[\sum_{t=\tau_1}^{\tau_2}(Y(t)-F_{\mathrm{CR}}(t,\tau_1,\tau_2))|\mathbf{F}_t\right]=0 \tag{10.39}$$

and since F_{CR} is F_t adapted, we derive the price of a CR futures to be

$$F_{\mathrm{CR}}(t,\tau_1,\tau_2)=\mathrm{E}_Q\left[\sum_{t=\tau_1}^{\tau_2}Y(t)|\mathbf{F}_t\right]. \tag{10.40}$$

Unfortunately, we cannot pursue the same methodology that we followed in the derivation of the pricing formulas for temperature and wind futures contracts. In this chapter a different method will be followed, namely, indifference pricing. The indifference pricing framework will be applied in order to derive theoretical derivative prices for the CR index.

10.4.1 Indifference Pricing for Rainfall Derivatives

In this section, the indifference pricing approach will be applied. More precisely, the method proposed by Brockett et al. (2006) and Xu et al. (2008) will be followed. Note that this framework is different than the one presented in previous chapters in order to derive arbitrage-free prices for temperature and wind derivatives.

We assume that the market consists from two participants, a seller and a buyer. Furthermore, it is assumed a two-date economy where at $t=0$ both the buyer and seller try to optimize their wealth at time T. Finally, no trading between these two dates is allowed.

The financial market consists of two assets, a risky asset with a random gross rate of return r at date 1 and a savings account with a gross risk-free rate of return r_f, as well as some random weather indexes on which weather derivatives can be written.

The maximization of the wealth of an investor depends on a utility function. Here, the negative exponential utility function is assumed. The negative exponential utility function is given by

$$u(X)=-\exp(-\lambda X), \tag{10.41}$$

where X are the revenues and λ is the absolute risk aversion parameter. The investor wants to maximize the mean variance of his terminal wealth at date $T=1$ by using weather derivatives.

We will focus on the seller, but the framework for the buyer is similar. The seller has initial wealth x_s. In the first scenario, a strategy without the weather derivative is considered. The amount a_s is invested to the risk capital market portfolio while the remaining amount is invested to a risk-free asset. Hence, the value of the portfolio without a weather derivative, X_s^{wo}, at time T is given by

$$X_s^{wo} = (x_s - a_s)q_f + a_s q_s, \tag{10.42}$$

where

$$q_f = 1 + r_f, \; q_s = 1 + r_s, \tag{10.43}$$

where r_f and r_s denote the return of the risk-free asset and the return of capital market investment.

In the second scenario, the seller can additionally sell k shares of the weather contract for a price F_s. In this case, the value of the portfolio with a weather derivative, X_s^w, at time T is given by

$$X_s^w = (x_s - a_s + kF_s)q_f + a_s q_s - kW_T(CR). \tag{10.44}$$

Similar equation can be derived for the buyer. Hence, the value of the portfolio for the buyer without and with a weather derivative is given by

$$X_b^{wo} = (x_b - a_b)q_f + a_b q_b \tag{10.45}$$

$$X_b^w = (x_b - a_b - kF_b)q_f + a_b q_b + kW_T(CR), \tag{10.46}$$

where $q_b = 1 + r_b$ and r_b is the return of production or yield in the case of a farmer.

The indifference price is obtained when the seller is indifferent in which of the two strategies to follow. In other words, we have that

$$\sup_{a_s} \mathrm{E}\left[u\left(X_s^{wo}\right)\right] = \sup_{a_s} \mathrm{E}\left[u\left(X_s^{w}\right)\right]. \tag{10.47}$$

By replacing the expected utility by the certainty equivalent given by

$$CE = \mathrm{E}[X] - \frac{\lambda}{2}\sigma^2(X) \tag{10.48}$$

and following Xu et al. (2008), we have that

$$\sup_{a_s}\left(\mathrm{E}\left[X_s^{wo}\right] - \frac{\lambda_s}{2}\sigma^2\left(X_s^{wo}\right)\right) = \sup_{a_s}\left(\mathrm{E}\left[X_s^{w}\right] - \frac{\lambda_s}{2}\sigma^2\left(X_s^{w}\right)\right). \tag{10.49}$$

Hence, explicit expressions for the CE of the wealth at time T, without and with the weather derivative, are obtained:

$$CE^{wo} = x_s q_f + a_s\left(\mathrm{E}[q_s] - q_f\right) - \frac{\lambda_s}{2}a_s^2\sigma_{q_s}^2 \tag{10.50}$$

$$CE^w = (x_s + kF_s)q_f + a_s\big(\mathrm{E}[q_s] - q_f\big) - k\mathrm{E}[W] - \frac{\lambda_s}{2}a_s^2\sigma_{q_s}^2 - \frac{\lambda_s}{2}k^2\sigma_W^2 + \lambda_s a_s k\mathrm{cov}(q_s, W). \tag{10.51}$$

If we differentiate (10.50) and (10.51) with respect to a_s and set the derivatives equal to zero, we get that the optimal a_s which is given by

$$a_s^{wo,*} = \frac{\mathrm{E}[q_s] - q_f}{\lambda_s\sigma_{q_s}^2} \tag{10.52}$$

$$a_s^{w,*} = \frac{\mathrm{E}[q_s] - q_f + \lambda_s k\mathrm{cov}(q_s, W)}{\lambda_s\sigma_{q_s}^2}. \tag{10.53}$$

Then we replace (10.52) and (10.53) to (10.50) and (10.51), respectively. The indifference pricing approach assumes that (10.50) and (10.51) must be equal. Hence, the price of the weather derivative for the seller, F_s, can be obtained:

$$F_s = \frac{1}{q_f}(\mathrm{E}[W] + \pi_s), \tag{10.54}$$

where

$$\pi_s = -\frac{\lambda_s}{2}k\sigma_W^2\left(\rho_{q_s,W}^2 - 1\right) - \frac{\sigma_W}{\sigma_{q_s}}\big(\mathrm{E}[q_s] - q_f\big)\rho_{q_s,W}^2 \tag{10.55}$$

and $\rho_{q_s,W}$ measures the correlation between the return of the market portfolio and the payoff of the derivative. Similarly, the price of the weather derivative for the buyer, F_b, can be derived:

$$F_b = \frac{1}{q_f}(\mathrm{E}[W] + \pi_b), \tag{10.56}$$

where

$$\pi_b = \frac{\lambda_b}{2}k\sigma_W^2\left(\rho_{q_b,W}^2 - 1\right) - \frac{\sigma_W}{\sigma_{q_s}}\big(\mathrm{E}[q_s] - q_f\big)\rho_{q_b,W}^2. \tag{10.57}$$

10.4.2 Limitations of Indifference Pricing Method

Expected utility maximization is a powerful tool for deciding whether or not to accept a project at a given time and scale. An investor will accept a project if and only if it increases her expected utility. In the framework of weather derivatives,

indifference pricing, and expected utility, maximization was proposed by Brockett et al. (2006), Xu et al. (2008), Odening et al. (2007), Richards et al. (2004), Carmona and Diko (2005), Leobacher and Ngare (2010), Edwards and Simmons (2004), and Simmons et al. (2007).

It is clear that the indifference pricing methodology is very simple and easy to follow. Moreover, in Sect. 10.4.1, a general framework was developed. Note that we did not define exactly the payoff of the weather derivative, W. In other words, this framework can be followed for any weather derivative written on any weather index and is not limited only to the CR index.

However, there is a series of limitations and drawbacks that an investor should take into account. Utility functions are too much preference dependent and sensitive to the selection of the risk aversion parameter (Carr et al. 2001). In other words, the results are sensitive to the risk aversion parameter λ while at the same time extra care must be taken when a utility function is selected. In our case the negative exponential utility function was used. However, other functions can be applied, like the mean variance proposed by Brockett et al. (2006). Hence, the utility function is quite difficult to describe exactly as the real utility function of the investor.

In Carr et al. (2001), strong criticism over the utility maximization theory is applied. It is mentioned that utility maximization has had limited acceptance in practice. This is due to the difficulty in specifying the required inputs to the optimization problem. These inputs include the current endowment, the joint stochastic process over all assets, and the utility function over all certain wealth levels (Carr et al. 2001). Unfortunately, the maximization of the utility function is sensitive to these three inputs. Hence, the difficulty to fully and accurately specify these parameters renders the methodology potentially useless (Carr et al. 2001).

Moreover, using temperature or rainfall forecasts with a utility function to estimate the demand curve for the derivate reduces the proposed methodology to simply using the forecasts (Campbell and Diebold 2005; Oetomo and Stevenson 2005).

10.4.3 Hedging Effectiveness

In this section, we will investigate if it is attractive for a farmer to buy a weather derivative. Farmers are interested in knowing if and to what extend the existing yield risk can be eliminated by holding this security. The risk reduction that can be attributed to weather insurance is measured by comparing the revenue distribution of a production activity or a whole farm with and without possessing the weather derivative. Farmers usually are interested in smoothing their income since yield and product prices are highly volatile.

Assume that there is linear relationship between the production Q and the CR index. Then, the production function that gives the amount of the produced product depending on the rainfall index is given by

$$Q(CR) = \kappa CR + \omega, \tag{10.58}$$

where $Q(\mathrm{CR})$ gives the amount of the product depending on the outcome of the cumulative rainfall index. Hence, the amount that the risk is reduced by the inclusion of a weather derivative into the farmer's portfolio can be estimated by comparing the distribution of the farmer's revenues with and without the corresponding weather derivative.

Following Xu et al. (2008) and Stowasser (2012), the present value of the revenues, R, of a farmer who produces Q and holds a position of the weather derivative on the CR index, $W(\mathrm{CR})$, with price F is defined by

$$R = (Q(CR) \cdot C + W(CR))e^{-rT} - F, \tag{10.59}$$

where C is the product price, which, for simplicity, is assumed to be constant. It can be easily shown that without a position in the rainfall derivative, the terms W and F vanish. Obviously, the risk-reducing potential of any weather insurance depends on the correlation between the weather index and the considered product.

10.4.4 Monte Carlo Simulation

Alternative to the indifference pricing method, the MC technique can be used. The MC is applied in order to numerically estimate the expected value in (10.40) without any simplifying assumptions. The main objective is to simulate different rainfall scenarios over the period $[\tau_1, \tau_2]$. For the period $[\tau_1, \tau_2]$, a scenario of the daily rainfall is simulated. Each day is determined if it is a rainy or a dry day by using a Markov chain model (the model that was estimated from our historical data). Then, in the case of a rainy day, the appropriate rainfall amount is generated by the gamma distribution. This method is repeated n times. In our case, we choose $n = 10,000$. Finally, the CR index is calculated for the period $[\tau_1, \tau_2]$ by averaging each scenario. Given that the representation of the rainfall is well defined by our model, the forecasted CR index will be very close to the real one and hence the discounted payoff of the futures CR derivatives (10.40) can be easily calculated.

10.5 Conclusions

In this chapter, we focused on precipitation derivatives. Although rainfall and snowfall are two different weather variables and contracts are traded separately, as it is shown in this chapter, they share a lot of common characteristics. Hence, the same model can be used for both of them. Unlike other weather variables that were studied in this book so far, precipitation is not a continuous variable.

The first objective of this chapter was to present a daily model for the underlying weather variable, that is, precipitation. The modeling procedure was separated in

two steps. In the first step, a Markov chain model was used in order to model the frequency of the precipitation occurrence. In this chapter three different types of Markov chain models were presented. The first one is the classic Markov chain model with constant transition probabilities. The other two models assume seasonal and daily transition probabilities than have better potential in capturing the dynamics of the frequency process of precipitation.

In the second step, gamma distribution was selected as an appropriate distribution to fit the precipitation amount conditioned on a rainy day. Alternatively, the exponential and the mixed exponential distributions were presented.

The second objective was to use the daily model in order to derive a valuation framework for precipitation derivatives. An indifference pricing methodology was proposed. In this framework, an investor tries to maximize its utility function. A negative exponential utility function was proposed.

References

Akaike H (1974) New look at statistical-model identification. IEEE Trans Automat Contr AC19 (6):716–723

Brockett PL, Wang M, Yang C, Zou H (2006) Portfolio effects and valuation of weather derivatives. Financ Rev 41:55–76

Campbell SD, Diebold FX (2005) Weather forecasting for weather derivatives. J Am Stat Assoc 100:6–16

Cao M, Li A, Wei J (2004) Precipitation modeling and contract valuation: a frontier in weather derivatives. J Altern Invest 7(2):92–99

Carmona R, Diko P (2005) Pricing precipitation based derivatives. Int J Theor Appl Finance 8 (7):959–988

Carr P, Geman H, Madan DB (2001) Pricing and hedging in incomplete markets. J Financ Econ 62:131–167

Coles S, Pericchi LR, Sisson S (2003) A fully probabilistic approach to extreme rainfall modeling. J Hydrol 273:35–50

Dischel B (2000) Seeding a rain market. Environ Finance (September):2–4

Dubrovsky M, Buchtele J, Zalud Z (2004) High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modelling. Clim Change 63:145–179

Edwards M, Simmons P (2004) Preliminary results for the measurement of willingness to pay for climate derivatives. In: 48 annual conference of the Australian Agricultural & Resource Economics Society, Melbourne, Australia, February 2004

Feng Y, Kitzmiller D (2006) A short-range quantitative precipitation forecast algorithm using back-propagation neural network approach. Adv Atmos Sci 23(3):405–414. doi:10.1007/s00376-006-0405-7

Goncu A (2011) Modeling and pricing precipitation-based weather derivatives. Financ Math Appl 1(1):9–18

Leobacher G, Ngare P (2010) On modelling and pricing rainfall derivatives with seasonality. Appl Math Finance 18(1):71–91

Little MA, McSharry PE, Taylor JW (2009) Generalized linear models for site-specific density forecasting of U.K. daily rainfall. Mon Weather Rev 137(3):1029–1045. doi:10.1175/2008mwr2614.1

Martin WS, Barnett JB, Coble HK (2001) Developing and pricing precipitation insurance. J Agric Resour Econ 26(1):261–274

Moreno M (2002) Rain risk. Speedwell Weather Derivatives, London

Musshoff O, Odening M, Xu W (2006) Modeling and hedging rain risk. In: American Agricultural Economics Association annual meeting, Long Beach, California, 23–26 July 2006

Odening M, Musshoff O, Xu W (2007) Analysis of rainfall derivatives using daily precipitation models: opportunities and pitfalls. Agric Finance Rev 67:135–156

Oetomo T, Stevenson M (2005) Hot or cold? A comparison of different approaches to the pricing of weather derivatives. J Emerg Mark Finance 4(2):101–133

Richards TJ, Manfredo MR, Sanders DR (2004) Pricing weather derivatives. Am J Agric Econ 4(86):1005–1017

Schoof JT, Pryor SC (2008) On the proper order of Markov chain model for daily precipitation occurrence in the contiguous United States. J Appl Meteorol Climatol 47(9):2477–2486. doi:10.1175/2008jamc1840.1

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Simmons P, Edwards M, Byrne J (2007) Willingness to pay for weather derivatives by Australian wheat farmers. In: European Association of Agricultural Economists, 101st seminar, Berlin, 5–6 July 2007

Stowasser M (2012) Modelling rain risk: a multi-order Markov chain model approach. J Risk Finance 13(1):45–60

Valverde Ramírez MC, de Campos Velho HF, Ferreira NJ (2005) Artificial neural network technique for rainfall forecasting applied to the São Paulo region. J Hydrol 301 (1–4):146–162. doi:10.1016/j.jhydrol.2004.06.028

Vedenov DV, Miranda MJ (2001) Rainfall insurance for midwest crop production. American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association)

Wilks DS (1998) Multisite generalization of a daily stochastic precipitation generation model. J Hydrol 210(1–4):178–191. doi:10.1016/s0022-1694(98)00186-3

Wilks DS (1999) Multisite downscaling of daily precipitation with a stochastic weather generator. Clim Res 11:125–136

Wilks DS (2011) Statistical methods in the atmospheric sciences, vol 100, 3rd edn, International geophysics series. Academic, Oxford, UK

Williams PM (1998) Modelling seasonality and trends in daily rainfall data. Paper presented at the proceedings of the 1997 conference on advances in neural information processing systems, vol 10, Denver

Xu W, Odening M, Musshof O (2008) Indifference pricing of weather derivatives. Am J Agric Econ 90(4):979–993

Appendix A: Wavelet Networks for Weather Derivatives Modeling

A.1 Introduction

Wavelet networks (WNs) are a new class of networks that combine the classic sigmoid NNs and the WA. WNs have been used with great success in a wide range of applications. In this appendix, we present a complete statistical model identification framework in order to apply WNs in weather derivative modeling and pricing. Model identification can be separated in two parts, variable significance testing and model selection. The first will be used in order to indentify the statistical significant lags of a weather process like temperature or wind, while the later will be used in order to build the optimal WNs that will be used in order to model the dynamics of the underlying weather process. More precisely, the following subjects were thoroughly presented: the structure of a WN, methods to train a WN, initialization algorithms, variable significance and variable selection algorithms, a model selection method, and finally methods to construct confidence and prediction intervals.

In Pati and Krishnaprasad (1993), it has been demonstrated that it is possible to construct a theoretical formulation of a feedforward NN in terms of wavelet decompositions. WNs were proposed by Zhang and Benveniste (1992) as an alternative to feedforward NNs which would alleviate the aforementioned weaknesses associated with each method. The WNs are a generalization of radial basis function networks (RBFN). WNs are one hidden layer networks that use a wavelet as an activation function, instead of the classic sigmoidal family. It is important to mention here that the multidimensional wavelets preserve the "universal approximation" property that characterizes NNs. The nodes (or wavelons) of WNs are the wavelet coefficients of the function expansion that have a significant value. In Bernard et al. (1998), various reasons were presented in why wavelets should be used instead of other transfer functions. In particular, firstly, wavelets

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8,

have high compression abilities, and secondly, computing the value at a single point or updating the function estimate from a new local measure involves only a small subset of coefficients.

WNs have been used in a variety of applications so far, that is. in short-term load forecasting (Bashir and El-Hawary 2000; Benaouda et al. 2006; Gao and Tsoukalas 2001; Ulugammai et al. 2007; Yao et al. 2000), in time-series prediction (Cao et al. 1995; Chen et al. 2006; Cristea et al. 2000), signal classification and compression (Kadambe and Srinivasan 2006; Pittner et al. 1998; Subasi et al. 2005), signal denoising (Zhang 2007), static, dynamic (Allingham et al. 1998; Oussar and Dreyfus 2000; Oussar et al. 1998; Pati and Krishnaprasad 1993; Postalcioglu and Becerikli 2007; Zhang and Benveniste 1992), and nonlinear modeling (Billings and Wei 2005), nonlinear static function approximation, (Jiao et al. 2001; Szu et al. 1992; Wong and Leung 1998), to mention the most important. In Khayamian et al. (2005), WN were even proposed as a multivariate calibration method for simultaneous determination of test samples of copper, iron, and aluminum.

In contrast to classical "sigmoid NNs," WNs allow for constructive procedures that efficiently initialize the parameters of the network. Using wavelet decomposition, a "wavelet library" can be constructed. In turn, each wavelon can be constructed using the best wavelet of the wavelet library. The main characteristics of these procedures are (1) convergence to the global minimum of the cost function and (2) initial weight vector into close proximity of the global minimum, and as a consequence drastically reduced training times (Zhang 1997; Zhang and Benveniste 1992). In addition, WNs provide information for the relative participation of each wavelon to the function approximation and the estimated dynamics of the generating process. Finally, efficient initialization methods will approximate the same vector of weights that minimize the loss function each time.

In Zapranis and Alexandridis (2008) and Zapranis and Alexandridis (2009), we give a concise treatment of wavelet theory. For a complete theoretical background on wavelets and wavelet analysis, we refer to Daubechies (1992) and Mallat (1999). Here the emphasis is in presenting the theory and mathematics of wavelet neural networks.

The rest of the chapter is organized as follows: in Sect. A.2, we present the WN. More precisely, in Sect. A.2.1, the structure of a WN is described. In Sect. A.2.2, various initialization methods were described. In Sect. A.2.3, a training method of the WN is presented, and in Sect. A.2.4, the stopping conditions of the training are described. A model selection algorithm is described in Sect. A.3, and a model selection algorithm without training is presented in Sect. A.3.1. Next, a criterion for selecting significant variables is presented in Sect. A.4, while a variable selection algorithm is analytically presented in Sect. A.4.1. In Sect. A.5, methods to estimate the model and variance uncertainty are described. In Sect. A.5.1, a framework for constructing confidence intervals is presented, while in Sect. A.5.2, a framework for constructing prediction intervals is presented. Finally, in Sect. A.6, we conclude.

A.2 Wavelet Neural Networks for Weather Variables Modeling

A.2.1 Structure of a Wavelet Network

A WN usually has the form of a three-layer network. The lower layer represents the input layer, the middle layer is the hidden layer, and the upper layer is the output layer.

In the input layer, the explanatory variables are introduced to the WN. The hidden layer consists of the hidden units (HUs). The HUs are often referred as wavelons, similar to neurons in the classical sigmoid NNs. In the hidden layer, the input variables are transformed to dilated and translated version of the mother wavelet. Finally, in the output layer, the approximation of the target values is estimated.

The idea of a WN is to adapt the wavelet basis to the training data. Hence, the wavelet estimator is expected to be more efficient than a sigmoid NN, (Zhang 1993).

In this book, we implement a multidimensional WN with a linear connection between the wavelons and the output. Moreover, in order for the model to perform well in the presence of linearity, we use direct connections from the input layer to the output layer. Hence, a network with zero HUs is reduced to the linear model.

The structure of a single hidden-layer feedforward WN is given in Fig. A.1. The network output is given by the following expression:

$$g_\lambda(\mathbf{x};\mathbf{w}) = \hat{y}(\mathbf{x}) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \Psi_j(\mathbf{x}) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i \tag{A.1}$$

In that expression, $\Psi_j(x)$ is a multidimensional wavelet which is constructed by the product of m scalar wavelets, x is the input vector, m is the number of network inputs, λ is the number of HUs, and w stands for a network weight. The multidimensional wavelets are computed as follows:

$$\Psi_j(\mathbf{x}) = \prod_{i=1}^{m} \psi(z_{ij}) \tag{A.2}$$

where ψ is the mother wavelet and

$$z_{ij} = \frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}} \tag{A.3}$$

In the above expression, $i = 1, \ldots, m$, $j = 1, \ldots, \lambda + 1$ and the weights w correspond to the translation ($w_{(\xi)ij}^{[1]}$) and the dilation ($w_{(\zeta)ij}^{[1]}$) factors. The complete

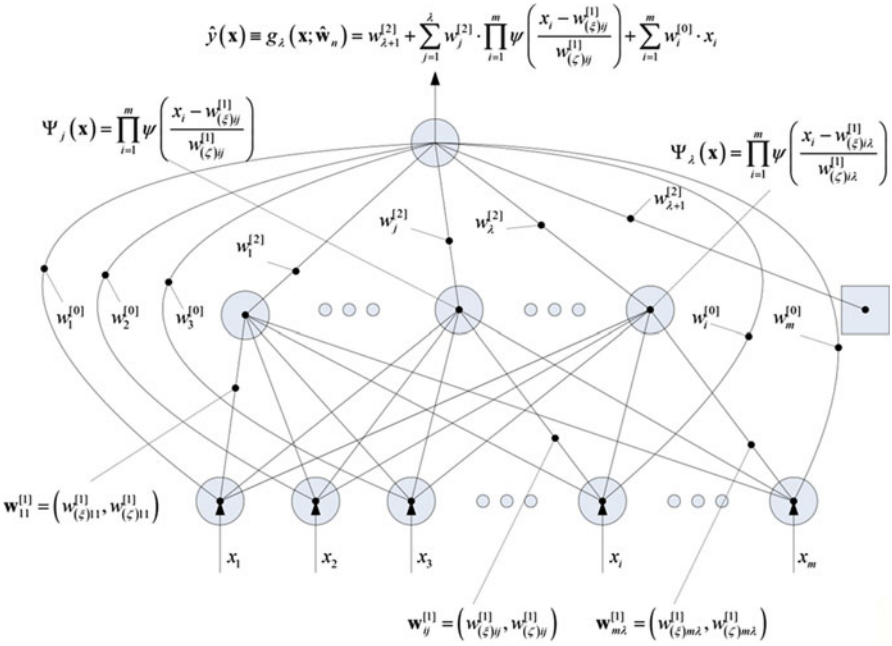

Fig. A.1 A feedforward wavelet neural network

vector of the network parameters comprises $w = \left(w_i^{[0]}, w_j^{[2]}, w_{\lambda+1}^{[2]}, w_{(\xi)ij}^{[1]}, w_{(\zeta)ij}^{[1]}\right)$. These parameters are adjusted during the training phase.

In bibliography, three mother wavelets are usually suggested, the Gaussian derivative; the second derivative of the Gaussian, the so-called Mexican Hat; and the Morlet wavelet.

The selection of the mother wavelet depends on the application and is not limited to the above choices. The activation function can be a wavenet (orthogonal wavelets) or a wave frame (continuous wavelets). Following Zhang (1994), Billings and Wei (2005), and Becerikli et al. (2003), we use as a mother wavelet the Mexican Hat function which proved to be useful and to work satisfactorily in various applications and is given by

$$\psi(z_{ij}) = (1 - z_{ij}^2)e^{-\frac{1}{2}z_{ij}^2} \tag{A.4}$$

A.2.2 Initialization of the Parameters of the Network

In WNs, in contrast to NNs that use sigmoid functions, selecting initial values of the dilation and translation parameters randomly may not be suitable (Oussar et al.

1998). A wavelet is a waveform of effectively limited duration that has an average value of zero and localized properties; hence, a random initialization may lead to wavelons with a value of zero. Random initialization may result to wavelons that do not represent the dynamics of the underlying weather variable.

Training algorithms like gradient descent with random initialization are inefficient (Zhang 1993), since random initialization affects the speed of training and may lead to a local minimum of the loss function (Postalcioglu and Becerikli 2007).

Utilizing the information that can be extracted by the WA from the input weather dataset, the initial values of the parameters w of the network can be selected in an efficient way. Efficient initialization will result to less iterations in the training phase of the network and training algorithms that will avoid local minimums of the loss function in the training phase. Finally, efficient initialization methods will approximate the same vector of weights that minimize the loss function each time.

Various methods have been proposed for an optimized initialization of the wavelet parameters. Recent studies proposed complex methods that utilize the information extracted by the WA (Oussar and Dreyfus 2000; Zhang 1997; Xu and Ho 2002; Wong and Leung 1998; Oussar et al. 1998; Kan and Wong 1998). These methods are not optimal but a trade-off between optimality and efficiency (He et al. 2002).

The implementation of these methods can be summed in the following three steps:

1. Construct a library W of wavelets.
2. Remove the wavelets that their support does not contain any sample points of the training data.
3. Rank the remaining wavelets and select the best wavelet regressors.

In the first step, the wavelet library can be constructed either by an orthogonal wavelet or a wavelet frame (He et al. 2002; Postalcioglu and Becerikli 2007). By determining an orthogonal wavelet basis, the WN is simultaneously constructed. However, in order to generate an orthogonal wavelet basis, the wavelet function has to satisfy strong restrictions (Daubechies 1992; Mallat 1999). In addition, the fact that orthogonal wavelets cannot be expressed in closed form constitutes them inappropriate for applications of function approximation or process modeling (Oussar and Dreyfus 2000).

On the other hand, constructing wavelet frames is very easy and can be done by translating and dilating the selected mother wavelet. The results from Gao and Tsoukalas (2001) indicate that a family of compactly supported non-orthogonal wavelets is more appropriate for function approximation. Due to the fact that a wavelet family can contain a large number of wavelets, it is more convenient to use a truncated wavelet family than an orthogonal wavelet basis (Zhang 1993).

However, constructing a WN using wavelet frames is not a straightforward process. The wavelet library may contain a large number of wavelets since only the input data were considered in the construction of the wavelet frame. In order to construct a WN, the "best" wavelets must be selected. However, arbitrary

truncations may lead to large errors (Xu and Ho 2005). In the second step, Zhang (1993) proposes to remove the wavelets that have very few training patterns in their support. Alternatively, in Cannon and Slotine (1995), magnitude-based methods were used to eliminate wavelets with small coefficients.

In the third step, the remaining wavelets are ranked, and the wavelets with the highest rank are used for the construction of the WN.

In Zhang (1994) and Zhang (1997), three alternative methods were proposed in order to reduce and rank the wavelets in the wavelet library: residual based selection (RBS), stepwise selection by orthogonalization (SSO), and backward elimination (BE).

In the framework of RBS, first, the wavelet that best fits the output data is selected. Then the wavelet that best fits the residual of the fitting of the previous stage is repeatedly selected. RBS is considered as a very simple method but not an effective one (Juditsky et al. 1994). However, if the wavelet candidates reach a very large number, computational efficiency is essential and the RBS method may be used (Juditsky et al. 1994). In Kan and Wong (1998) and Wong and Leung (1998), the RBS algorithm was used for the synthesis of a WN. In Xu and Ho (2002), a modified version of the RBS algorithm was used. More precisely, an orthogonalized residual based selection (ORBS) algorithm is proposed for the initialization of the WN. The ORBS method combines both the RBS and the orthogonalized least squares method. In this way, high efficiency is obtained, while relatively low computational burden is maintained.

The SSO method is an extension of the RBS first proposed by Chen et al. (1989) and Chen et al. (1991). In order to initialize the WN, the following procedure is followed: first, the wavelet which best fits the output data is selected. Then the wavelet that best fits the residual of the fitting of the previous stage together with the previous selected wavelet is repeatedly selected. In other words, the SSO considers the interaction or the non-orthogonality of the wavelets. The selection of the wavelets is performed using the modified Gram–Schmidt algorithm that has better numerical properties and is computationally less expensive than the ordinary Gram–Schmidt algorithm (Zhang 1997). SSO is considered to have good efficiency while it is not computationally expensive. In Oussar and Dreyfus (2000), an algorithm similar to SSO was proposed.

In contrast to previous methods, the BE starts the regression by selecting all the available wavelets from the wavelet library. Then the wavelet that contributes the least in the fitting of the training data is repeatedly eliminated. The drawback of BE is that it is computationally expensive, but it is considered to have good efficiency. In this book, the BE method is the preferred initialization framework that it is used in order to model the temperature and the wind.

All methods described above are used just for the initialization of the dilation and translation parameters. Then the network is further trained in order to obtain the vector of the parameters $w = \hat{\mathbf{w}}_n$ which minimizes the cost function.

It is clear that additional computational burden is added in order to initialize efficiently the WN. However, the efficient initialization significantly reduces the

training phase; hence, the total amount of computations is significantly smaller than in a network with random initialization.

A.2.3 Training a Wavelet Network with Back-Propagation

After the initialization phase, the network is further trained in order to find the weights which minimize the cost function.

In our implementation, the ordinary back-propagation (BP) was used. BP is probably the most popular algorithm used for training WNs (Oussar and Dreyfus 2000; Oussar et al. 1998; Postalcioglu and Becerikli 2007; Zhang 1997, 2007; Zhang and Benveniste 1992; Fang and Chow 2006; Jiao et al. 2001). Ordinary BP is less fast but also less prone to sensitivity to initial conditions than higher order alternatives (Zapranis and Refenes 1999).

The basic idea of BP is to find the percentage of contribution of each weight to the error. The error e_p for pattern p is simply the difference between the target (y_p) and the network output (E_p). By squaring and multiplying by ½, we take the pairwise error which is used in network training:

$$E_p = \frac{1}{2}\left(y_p - \hat{y}_p\right)^2 = \frac{1}{2}e_p^2 \tag{A.5}$$

The weights of the network were trained to minimize the mean quadratic cost function (or loss function):

$$L_n = \frac{1}{n}\sum_{p=1}^{n} E_p = \frac{1}{2n}\sum_{p=1}^{n} e_p^2 = \frac{1}{2n}\sum_{p=1}^{n}\left(y_p - \hat{y}_p\right)^2 . \tag{A.6}$$

Other functions can be used instead of (A.6); however, the mean quadratic cost function is the most commonly used. The network is trained until a vector of weights $w = \hat{\mathbf{w}}_n$ that minimizes the proposed cost function is found. The previous solution corresponds to a training sample of size n. Computing the parameter vector $\hat{\mathbf{w}}_n$ is always done by iterative methods. At each iteration t, the derivative of the loss function with respect to the network weights is calculated. Then, the updating of the parameters is performed by the following (delta) learning rule:

$$w_{t+1} = w_t - \eta\frac{\partial L_n}{\partial w_t} + \kappa(w_t - w_{t-1}) \tag{A.7}$$

where η is the learning rate and it is constant. The complete vector of the network parameters comprises $w = \left(w_i^{[0]}, w_{(\xi)ij}^{[1]}, w_{(\zeta)ij}^{[1]}, w_j^{[2]}, w_{\lambda+1}^{[2]}\right)$.

A constant momentum term, defined by κ, is induced which increases the training speed and helps the algorithm to avoid oscillations. The learning rate and momentum speed take values between 0 and 1. The choice of the learning rate and the momentum depends on the application and the training sample. Usually, values between 0.1 and 0.4 are used.

The partial derivative of the cost function with respect to a weight w is given by

$$\begin{aligned}\frac{\partial L}{\partial w} &= \frac{1}{2n}\sum_{p=1}^{n}\frac{\partial E_p}{\partial w} = \frac{1}{2n}\sum_{p=1}^{n}\frac{\partial E_p}{\partial \hat{y}_p}\frac{\partial \hat{y}_p}{\partial w} \\ &= \frac{1}{n}\sum_{p=1}^{n} -\left(y_p - \hat{y}_p\right)\frac{\partial \hat{y}_p}{\partial w} = \frac{1}{n}\sum_{p=1}^{n} -e_p\frac{\partial \hat{y}_p}{\partial w}\end{aligned} \tag{A.8}$$

The partial derivatives with respect to each parameter, $\frac{\partial \hat{y}_p}{\partial w}$, and with respect to each input variable, $\frac{\partial \hat{y}_p}{\partial x_i}$, are presented in Appendix B.

A.2.4 Stopping Conditions for Training

After the initialization phase of the network parameters w, the weights $w_i^{[0]}$ and $w_j^{[2]}$ and parameters $w_{(\xi)ij}^{[1]}$ and $w_{(\zeta)ij}^{[1]}$ are trained during the learning phase for approximating the target function. A key decision related to the training of a WN is when the weight adjustment should end. Under the assumption that the WN contains the number of wavelets that minimizes the prediction risk, the training is stopped when one of the following criteria is met: the cost function reaches a fixed lower bound or the variations of the gradient or the variations of the parameters reach a lower bound or the number of iterations reaches a fixed maximum, whichever is satisfied first. In our implementation, the fixed lower bound of the cost function, of the variations of the gradient, and of the variations of the parameters was set to 10^{-5}.

A.3 Model Selection for Weather Time-Series Modeling

In this section, we describe the model selection procedure. One of the most crucial steps is to identify the correct topology of the network. A desired WN architecture should contain as few HUs as necessary while at the same time it should explain as much variability of the training data as possible. A network with less HUs than needed would not be able to learn the underlying function that describes the weather

process, while selecting more HUs than needed will result to an over-fitted model. Therefore, an algorithm to select the appropriate WN model for a given problem is necessary to be derived.

The minimum prediction risk (MPR) principle can be applied (Efron and Tibshirani 1993; Zapranis and Refenes 1999). The idea behind MPR is to estimate the out-of-sample performance of incrementally growing networks. Assuming that the explanatory variables $\mathbf{x}$ were correctly selected and remain fixed, then the model selection procedure is the following: the procedure starts with a fully connected network with 0 HUs. The WN is trained, and then the prediction risk is estimated. Then, iteratively a new HU is added to the network. The new WNs are trained, and the new prediction risk is estimated at each step. The number of HUs that minimizes the prediction risk is the appropriate number of HUs that should be used for the construction of the WN.

The prediction risk measures the generalization ability of the network. More precisely, the prediction risk of a network $g_\lambda(\mathbf{x}; \hat{\mathbf{w}}_n)$ is the expected performance of the network on new data that were not introduced during the training phase and is given by

$$P_\lambda = E\left[\frac{1}{n}\sum_{p=1}^{n}\left(y_p^* - \hat{y}_p^*\right)^2\right] \tag{A.9}$$

where $\left(\mathbf{x}_p^*, y_p^*\right)$ are the new observations that have not been used in the construction of the network $g_\lambda(\mathbf{x}; \hat{\mathbf{w}}_n)$ and $\hat{y}_p^*$ is the network output using the new observations, $g_\lambda(\mathbf{x}^*; \mathbf{w})$.

However, finding a statistical measure that estimates the prediction risk is not a straightforward procedure. The use of sampling methods such as bootstrap and cross-validation can be employed since they do not depend on any assumptions regarding the model (Efron and Tibshirani 1993). The only assumption made by sampling methods is that the data are a sequence of independent and identically distributed variables. Another advantage of bootstrap and cross-validation is their robustness. In contrast to sampling methods, alternative methods like the generalized cross-validation (GCV) and the final prediction error (FPE) require a roughly correct model to obtain the estimate of the noise variance.

The bootstrap and the v-fold cross-validation approaches are analytically described in Efron and Tibshirani (1993). It is known that the simple estimation of the bootstrap approach is not very accurate (Efron and Tibshirani 1993). Hence, we estimate the improved estimation of the prediction risk following the suggestion of Efron and Tibshirani (1993). The number of new samples B is usually over 30 (Aczel 1993; Efron and Tibshirani 1993). It is clear that as the number of new samples B increases, the bootstrap method becomes more accurate but also more computationally expensive. Cross-validation is an another standard tool for estimating the prediction error that makes an efficient use of the available

information (Efron and Tibshirani 1993). The v-fold cross-validation is applied as described in Efron and Tibshirani (1993).

A.3.1 Model Selection Without Training

In Zhang (1997), the estimation of the preferred information criteria is performed after the initialization stage of the network. More precisely, in the SSO and RBS, the preferred information criteria are evaluated after the selection of each wavelet in the initialization stage. Similarly, when the BE algorithm is used, the preferred information criteria are evaluated after the elimination of each wavelet in the initialization stage. Since the initialization of the WN is very good, as presented in the previous section, the initial approximation is expected to be very close to the target function. Hence, a good approximation of the prediction risk is expected to be obtained. The same idea can also be applied when the BS or the CV are used. The above procedure is significantly less computational expensive.

However, the above procedure is similar to early stopping techniques. Usually early stopping techniques suggest a network with more HUs than necessary, though the network is not fully trained to avoid over-fitting (Samarasinghe 2006), while they do not work satisfactorily in complex problems (Samarasinghe 2006).

Since sampling techniques are computationally expensive methods, the FPE criterion can be used initially. Then the BS or the CV methods can be used in +/−5 HU around the HUs proposed by FPE in order to define the best network topology.

A.4 Selecting the Significant Lags of the Weather Variables

In real problems it is important to determine correctly the independent variables. In most problems there is a little information about the relationship of any explanatory variable with the dependent variable. As a result, unnecessary independent variables are included in the model, reducing its predictive power. Similarly, in the case of temperature and wind modeling, the length of the lag series must be chosen. Hence, the lags that contribute statistically significant to our model must be selected. Since WN are nonlinear, nonparametric tools, methods derived from linear models like the partial autocorrelation function cannot be used.

In linear models, in order to determine if a coefficient, and as a result an input variable, is significant, the t-stats or the p values of each coefficient are examined. Applying the previous method in WNs is not a straightforward process since the coefficients (weights) are estimated iteratively and each variable contribute to the output of the WN linearly through the direct connections and nonlinearly through the HUs.

In this book, the sensitivity based pruning (SBP) proposed by Moody and Utans (1994) is used in order to define the statistical significant lags that must be used in temperature and wind modeling. The SBP method quantifies a variable's relevance to the model by the effect on the empirical loss of the replacement of that variable by its mean. The SBP is given by

$$SBP(x_j) = L_n(\mathbf{x}; \hat{\mathbf{w}}_n) - L_n(\bar{\mathbf{x}}^{(j)}; \hat{\mathbf{w}}_n) \tag{A.10}$$

where

$$\bar{\mathbf{x}}^{(j)} = \left(x_{1,t}, x_{2,t}, \ldots, \bar{x}_j, \ldots, x_{m,t}\right) \tag{A.11}$$

and

$$\bar{x}_j = \frac{1}{n}\sum_{t=1}^{n} x_{j,t} \tag{A.12}$$

Additional criteria can be used like the ones presented in Dimopoulos et al. (1995) and Alexandridis (2010). For additional information on the SBP criterion presented above, we refer to Alexandridis (2010) and Zapranis and Refenes (1999).

A.4.1 An Algorithm for Selecting the Significant Lags in Modeling of Weather Processes

In order to statistically test whether a variable is insignificant and can be removed for the training dataset or not, the distribution of the criterion presented in the previous section is needed. Without the distribution of the preferred measure of relevance, it is not clear if the effects of the variable x_i on y are statistically significant (Zapranis and Refenes 1999). More precisely, the only information obtained by criteria described in the previous section is how sensitive is the dependent variable to small perturbations of the independent variable. It is clear that the smaller the value of the preferred criterion, the less significant is the corresponding variable. However, there is no information if this variable should be removed from the model or not. In other words, we cannot define the correct number of lags that should be used in order to model the dynamics of the temperature or wind processes.

In order to approximate asymptotically the distribution of the measures of relevance, we use the bootstrap method. More precisely, a number of bootstrapped training samples can be created by the original training dataset. The idea is to estimate the preferred criterion on each bootstrapped sample. If the number of the bootstrapped samples is large, then a good approximation of the empirical distribution of the criterion is expected to be achieved. Obtaining an approximation of the

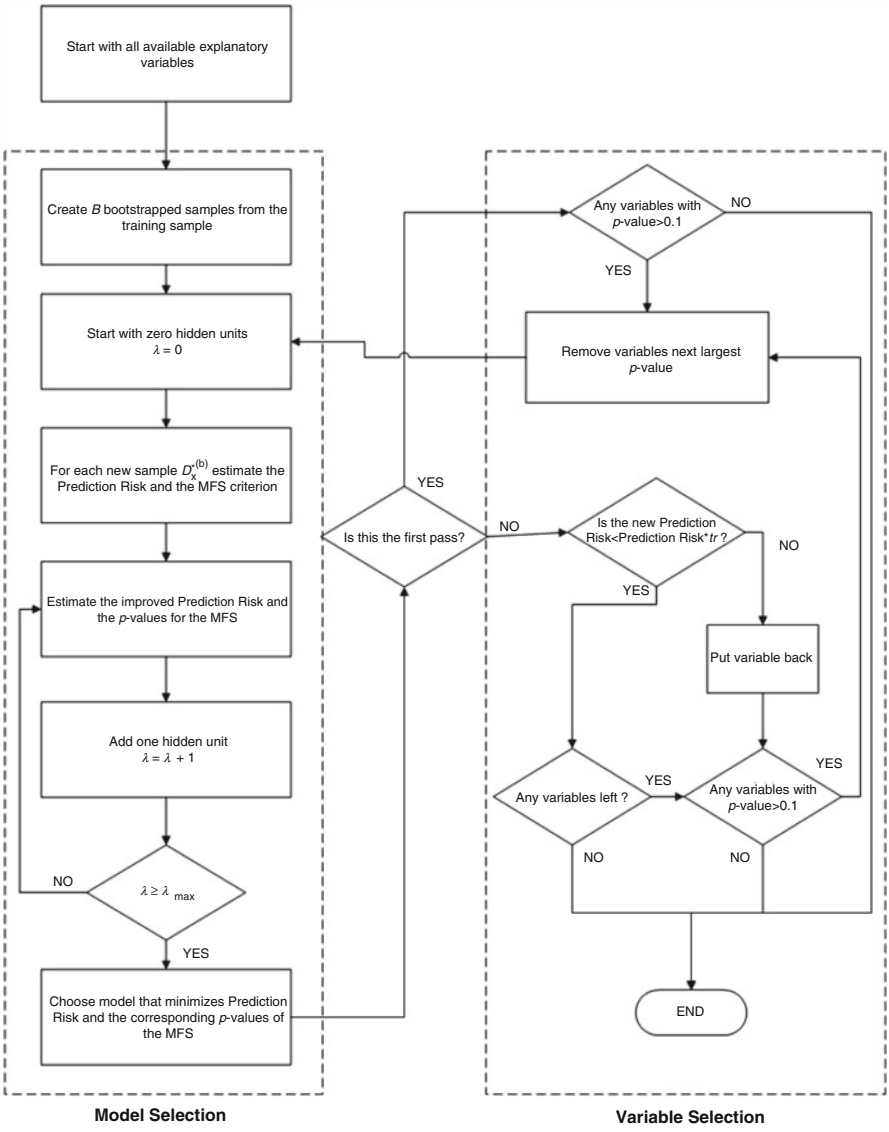

Fig. A.2 Model identification. Model selection and variable selection algorithms

empirical distributions, confidence intervals and hypothesis tests can be constructed for the value of the criterion. The variable selection algorithm is analytically explained bellow and is illustrated in Fig. A.2.

The procedure is the following: the algorithm starts with the training sample that consists of all available explanatory variables.

The first step is to create B bootstrapped training samples from the original dataset.
The second step is to identify the correct topology of the WN following the procedure described in the previous section and estimate the prediction risk.
The third step is to estimate the preferred measure of relevance for each explanatory variable for each one of the B bootstrapped training samples.
The fourth step is to calculate the p values of the measure of relevance.
The fifth step is to test if any explanatory variables have a p value greater than 0.1 If variables with a p value greater than 0.1 exist, then the variable with the largest p value is removed from the training dataset else the algorithm stops.
The sixth step is to estimate the prediction risk and the new p values of the reduced model. If the new estimated prediction risk is smaller than the prediction risk multiplied by a threshold (usually 1.05), then the decision of removing the variable was correct and we return to the fifth step.

If the new prediction risk is greater than the new prediction risk multiplied by a threshold (usually 1.05), then the decision of removing the variable was wrong, and the variable must be reintroduced to the model. In this case the variable with the next largest p value which is also greater than 0.1 is removed from the training sample, and we return to step six. If the remaining variables have p values smaller than 0.1, then the algorithm stops.

In order to have a good estimation of the prediction risk as well as an approximation of the distribution of the measure of relevance, a large number of bootstrapped samples B are needed. As B increases, the accuracy of the algorithm also increases but also increases the computational burden. In Zapranis and Refenes (1999), two different bootstrap methods were presented, the local bootstrap and the parametric sampling, which are significantly less computationally expensive. However, the bootstrapped samples may significantly differ from the original sample. Hence, applying local bootstrap or parametric sampling may lead to wavelets outside their effective support, that is, wavelets with value of zero, since wavelets are local functions with limited duration. In addition, in contrast to the case of NNs, the asymptotic distribution of the weights of a WN is not known. These observations constitute both local bootstrap and parametric sampling inappropriate for WNs.

Alternatively new samples from training patterns can be constructed. This can be done by applying bootstrap from pairs and train a WN for each sample. Since, the initialization of a WN is very good, this procedure is not of a prohibited computational cost.

A.5 Modeling the Uncertainty of the Weather Predictions

In the previous sections, a framework were a WN can efficiently be constructed, initialized, and trained was presented. In this section, this framework is expanded by presenting two methods for estimating confidence and prediction intervals. The

output of the WN is the approximation of the underlying function $f(\mathbf{x})$ obtained from the noisy data. In many applications and especially in finance, risk managers may be more interested in predicting intervals for future movements of the underlying function $f(\mathbf{x})$ than simply point estimates. Similarly, in weather risk management, confidence and prediction intervals can be used in order to derive bounds for the possible outcomes of the evolution of the weather variables and adjust the weather risk management strategies analogously.

In weather datasets, the training patterns usually are inaccurate since they contain noise or they are incomplete due to missing observations. Especially financial time series as well as temperature time series are dominated by these characteristics. As a result, the validity of the predictions of our model (as well as of any other model) is questioned. The uncertainty that results from the data contributes to the total variance of the prediction, and it is called the data noise variance, σ_{ε}^2 (Papadopoulos et al. 2000; Breiman 1996; Carney et al. 1999; Heskes 1997).

On the other hand, presenting to a trained network new data that were not introduced to the WN during the training phase, additional uncertainty is introduced to the predictions. Since the training set consists of a finite number of training pairs, the solution $\hat{\mathbf{w}}_n$ is likely not to be valid in regions not represented in the training sample (Papadopoulos et al. 2000). In addition, the iterative algorithm that is applied to train a network often results to local minima of the loss function. This source of uncertainty that arises from misspecifications in model or parameter selection as well as from limitation of the training algorithm contributes also to the total variance of the prediction, and it is called the model variance, σ_m^2 (Papadopoulos et al. 2000).

The model variance and the data noise variance are assumed to be independent. The total variance of the prediction is given by the sum of two variances:

$$\sigma_p^2 = \sigma_m^2 + \sigma_{\varepsilon}^2 \tag{A.13}$$

If the total variance of a prediction can be estimated, then it is possible to construct confidence and prediction intervals. The rest of the section is dedicated to this purpose. We assume that the total variance of the prediction is not constant and is given by

$$\sigma_p^2(\mathbf{x}) = \sigma_m^2(\mathbf{x}) + \sigma_{\varepsilon}^2(\mathbf{x}) \tag{A.14}$$

Two of the most often cited methods is the bagging (Breiman 1996) and balancing method (Carney et al. 1999; Heskes 1997). In this section, we adapt these two methods in order to construct confidence and prediction intervals under the framework of WNs. A framework similar to the one presented in Carney et al. (1999) to estimate the total prediction variance, σ_p^2, and construct confidence and prediction intervals is adapted.

A.5.1 Confidence Intervals

To generate confidence intervals, the distribution of the accuracy of the network prediction to the true underlying function is needed. In other words, the variance of the distribution of

$$f(\mathbf{x}) - \hat{y} \equiv f(\mathbf{x}) - g_{\lambda}(\mathbf{x}, \hat{\mathbf{w}}_n) \tag{A.15}$$

must be estimated.

The model variance σ_m^2 will be estimated using the balancing method proposed by Heskes (1997) and Carney et al. (1999).

First, $B = 200$ new random samples with replacement are created from the original training sample. Each new sample is used to train a new WN with the same topology as the original one, $g_{\lambda}(\mathbf{x}^{(*i)}; \hat{\mathbf{w}}^{(*i)})$, where $(*i)$ indicates i^{th} the i^{th} bootstrapped sample and $\hat{\mathbf{w}}^{(*i)}$ is the solution of the i^{th} bootstrapped sample. Then each new network is evaluated using the original training sample $\mathbf{x}$. Next, the average output of the B networks is estimated by

$$g_{\lambda,avg}(\mathbf{x}) = \frac{1}{B}\sum_{i=1}^{B} g_{\lambda}\left(\mathbf{x}; \hat{\mathbf{w}}^{(*i)}\right) \tag{A.16}$$

It is assumed that the WN produces an unbiased estimate of the underlying function $f(\mathbf{x})$. This means that the distribution of $P\left(f(\mathbf{x})|g_{\lambda,avg}(\mathbf{x})\right)$ is centered on the estimate $g_{\lambda,avg}(\mathbf{x})$ (Carney et al. 1999; Heskes 1997; Zapranis and Livanis 2005). Since the WN is not an unbiased estimator (as any other model), it assumed that the bias component arising from the WN is negligible in comparison to the variance component (Carney et al. 1999; Zapranis and Livanis 2005). Finally, if we assume that the distribution of $P\left(f(\mathbf{x})|g_{\lambda,avg}(\mathbf{x})\right)$ is normal, then the model variance can be estimated by

$$\hat{\sigma}_m^2(\mathbf{x}) = \frac{1}{B-1}\sum_{i=1}^{B}\left(g_{\lambda}\left(\mathbf{x}; \hat{\mathbf{w}}^{(*i)}\right) - g_{\lambda,avg}(\mathbf{x})\right)^2 \tag{A.17}$$

In order to construct confidence intervals, the distribution of $P\left(g_{\lambda,avg}(\mathbf{x})|f(\mathbf{x})\right)$ is needed. Since the distribution of $P\left(f(\mathbf{x})|g_{\lambda,avg}(\mathbf{x})\right)$ is assumed to be normal, then the "inverse" distribution $P\left(g_{\lambda,avg}(\mathbf{x})|f(\mathbf{x})\right)$ is also normal. However, this distribution is unknown. Alternatively it is empirically estimated by the distribution of $P\left(g_{\lambda}(\mathbf{x})|g_{\lambda,avg}(\mathbf{x})\right)$ (Carney et al. 1999; Zapranis and Livanis 2005). Then the confidence intervals are given by

$$g_{\lambda,avg}(\mathbf{x}) - t_{\frac{a}{2}}\hat{\sigma}_m \leqslant f(\mathbf{x}) \leqslant g_{\lambda,avg}(\mathbf{x}) + t_{\frac{a}{2}}\hat{\sigma}_m \tag{A.18}$$

where $t_{\frac{a}{2}}$ can be found in a student's t table and $1 - a$ is the desired confidence level.

However, the estimator of the model variance, $\hat{\sigma}_m^2$, given by (A.17) is known to be biased (Carney et al. 1999); as a result, wider confidence intervals will be produced. Carney et al. (1999) proposed a balancing method to improve the model variance estimator.

The B bootstrapped samples are divided in M groups. More precisely the 200 ensemble samples are divided in 8 groups of 25 samples each. Next, the average output of each group is estimated:

$$\zeta = \left\{ g_{\lambda,avg}^{(i)}(\mathbf{x}) \right\}_{i=1}^{M} \tag{A.19}$$

The model variance is not estimated just by the M ensemble output since this estimation will be highly volatile (Carney et al. 1999). In order to overcome this, a set of $P = 1000$ bootstraps of the values of ζ are created:

$$Y = \left\{ \zeta_j^* \right\}_{j=1}^{P} \tag{A.20}$$

where

$$\zeta_j^* = \left\{ g_{\lambda,avg}^{(*j1)}(\mathbf{x}), g_{\lambda,avg}^{(*j2)}(\mathbf{x}), \ldots, g_{\lambda,avg}^{(*jM)}(\mathbf{x}) \right\} \tag{A.21}$$

is a bootstrapped sample of ζ. Then the model variance is estimated on each one of these sets by

$$\hat{\sigma}_j^{2*}(\mathbf{x}) = \frac{1}{M} \sum_{k=1}^{M} \left(g_{\lambda,avg}^{(*jk)}(\mathbf{x}) - g_{\lambda,avg}^{j}(\mathbf{x}) \right)^2 \tag{A.22}$$

where

$$g_{\lambda,avg}^{j}(\mathbf{x}) = \frac{1}{M} \sum_{k=1}^{M} g_{\lambda,avg}^{(*jk)}(\mathbf{x}) \tag{A.23}$$

Then the average model variance is estimated by taking the average of all $\hat{\sigma}_j^{2*}(\mathbf{x})$:

$$\hat{\sigma}_m^2(\mathbf{x}) = \frac{1}{P} \sum_{j=1}^{P} \hat{\sigma}_j^{2*}(\mathbf{x}) \tag{A.24}$$

This procedure is not computationally expensive since there is no need to train new networks. Since a good estimator of the model variance is obtained, the improved confidence intervals using the balancing methods are given by

$$g_{\lambda,avg}(\mathbf{x}) - z_{\frac{a}{2}}\hat{\sigma}_m \leqslant f(\mathbf{x}) \leqslant g_{\lambda,avg}(\mathbf{x}) + z_{\frac{a}{2}}\hat{\sigma}_m \tag{A.25}$$

where $z_{\frac{a}{2}}$ can be found in a standard Gaussian distribution table and $1-a$ is the desired confidence level.

A.5.2 Prediction Intervals

To generate prediction intervals, the distribution of the accuracy of the network prediction to target values is needed. In other words, the variance of the distribution of

$$y - \hat{y} \equiv y - g_{\lambda}(\mathbf{x}, \hat{\mathbf{w}}_n) \tag{A.26}$$

must be estimated.

In order to construct prediction intervals, the total variance of the prediction, σ_p^2, must be estimated. As it was presented earlier, the total variance of the prediction is the sum of the model variance and the data noise variance. In the previous section, a method for estimating the model variance was presented. Here we emphasize on a method for estimating the data noise variance.

In order to estimate the noise variance σ_{ε}^2, maximum likelihood methods are used. First, the initial WN, $g_{\lambda}(\mathbf{x}; \hat{\mathbf{w}}_n)$, is estimated and the solution $\hat{\mathbf{w}}_n$ of the loss function is found. Since it is assumed that the estimated WN is a good approximation of the unknown underlying function, the vector $\hat{\mathbf{w}}_n$ is expected to be very close to the true vector $\mathbf{w}_0$ that minimizes the loss function. Hence, the noise variance can be approximated by a second WN, $f_v(\mathbf{x}; \hat{\mathbf{u}}_n)$, where the squared residuals of the initial WN are used as target values (Satchwell 1994). In the second WN, $f_v(\mathbf{x}; \hat{\mathbf{u}}_n)$, v is the number of HUs and $\hat{\mathbf{u}}_n$ is the estimated vector of parameters that minimizes the loss function of the second WN. Since it is assumed that the estimated WN is a good approximation of the unknown underlying function, the vector $\hat{\mathbf{u}}_n$ is expected to be very close to the true vector $\mathbf{u}_0$ that minimizes the loss function. Hence, the following cost function is minimized in the second network:

$$\sum_{i=1}^{n} \left\{ \left(g_{\lambda}(\mathbf{x}_i; \mathbf{w}_0) - y_i\right)^2 - f_v(\mathbf{x}_i; \mathbf{u}_0) \right\}^2 \tag{A.27}$$

and for a new set of observations, $\mathbf{x}^*$, that were not used in the training:

$$\hat{\sigma}_{\varepsilon}^2(\mathbf{x}^*) \approx f_v(\mathbf{x}^*; \mathbf{u}_0) \tag{A.28}$$

This technique assumes that the residual errors are caused by variance alone (Carney et al. 1999). In order to estimate the noise variance, data that were not used

in the training of the bootstrapped sample should be used. One way to do this is to divide the dataset in training and a validation set. However, leaving out these test patterns is a waste of data (Heskes 1997). Alternatively, an unbiased estimation of the output of the WN, $\hat{y}_{ub}(\mathbf{x})$, can be approximated by

$$\hat{y}_{ub}(\mathbf{x}) = \sum_{i=i}^{B} q_i^m \hat{y}_i(\mathbf{x}) \Big/ \sum_{i=i}^{B} q_i^m \tag{A.29}$$

where q_i^m is 0 if pattern m appears on the i^{th} bootstrap sample and 1 otherwise. Constructing the new network $f_v(\mathbf{x};\mathbf{u})$, we face the problem of model selection again. Using the methodology described in the previous section, the correct number of v HUs is selected. Usually, 1 or 2 HUs are enough to model the residuals. In finding the estimator of the noise variance, the prediction intervals can be constructed:

$$g_{\lambda,avg}(\mathbf{x}^*) - z_{\frac{a}{2}}\hat{\sigma}_p(\mathbf{x}^*) \leqslant f(\mathbf{x}^*) \leqslant g_{\lambda,avg}(\mathbf{x}^*) + z_{\frac{a}{2}}\hat{\sigma}_p(\mathbf{x}^*) \tag{A.30}$$

where $z_{\frac{a}{2}}$ can be found in a standard Gaussian distribution table and $1-a$ is the desired confidence level.

A.6 Conclusions

In this appendix, a complete statistical model identification framework in order to apply WNs in weather derivative modeling and pricing was presented. Model identification can be separated in two parts, variable significance testing and model selection. The first was used in order to indentify the statistical significant lags of a weather process like temperature or wind, while the latter was used in order to build the optimal WNs that will be used in order to model the dynamics of the underlying weather process. Finally, methods for constructing confidence and prediction intervals were presented.

Although a framework for selecting an appropriate model was presented, the adequacy of the final model must be further tested. This is usually done by examining the residuals by various criteria in order to indentify if the residuals satisfy the initial conditions of the proposed model. However, the selection of these criteria depends on the nature of the underlying function and the assumptions made while building the model.

References

Aczel AD (1993) Complete business statistics, 2nd edn. Irwin, Homewood

Alexandridis A (2010) Modelling and pricing temperature derivatives using wavelet networks and wavelet analysis. Ph.D. thesis, University of Macedonia, Thessaloniki

Allingham D, West M, Mees AI (1998) Wavelet reconstruction of nonlinear dynamics. Int J Bifurcat Chaos 8(11):2191–2201

Bashir Z, El-Hawary ME (2000) Short term load forecasting by using wavelet neural networks. In: The proceedings of Canadian conference on electrical and computer engineering, Halifax, pp 163–166

Becerikli Y, Oysal Y, Konar AF (2003) On a dynamic wavelet network and its modeling application. Lecture notes in computer science, vol 2714. Springer, Berlin, pp 710–718

Benaouda D, Murtagh G, Starck J-L, Renaud O (2006) Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting. Neurocomputing 70:139–154

Bernard C, Mallat S, Slotine J-J (1998) Wavelet interpolation networks. In: The proceedings of ESANN '98, Bruges, 22–24 Apr 1998, pp 47–52

Billings SA, Wei H-L (2005) A new class of wavelet networks for nonlinear system identification. IEEE Trans Neural Netw 16(4):862–874

Breiman I (1996) Bagging predictors. Mach Learn 24:123–140

Cannon M, Slotine J-JE (1995) Space-frequency localized basis function networks for nonlinear system estimation and control. Neurocomputing 9:293–342

Cao L, Hong Y, Fang H, He G (1995) Predicting chaotic time series with wavelet networks. Phys D 85:225–238

Carney JG, Cunningham P, Bhagwan U (1999) Confidence and prediction intervals for neural network ensembles. In: IJCNN'99, Washington, DC

Chen Y, Billings SA, Luo W (1989) Orthogonal least squares methods and their application to non-linear system identification. Int J Contr 50:1873–1896

Chen Y, Cowan C, Grant P (1991) Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans Neural Netw 2:302–309

Chen Y, Yang B, Dong J (2006) Time-series prediction using a local linear wavelet neural wavelet. Neurocomputing 69:449–465

Cristea P, Tuduce R, Cristea A (2000) Time series prediction with wavelet neural networks. In: The proceedings of 5th seminar on neural network applications in electrical engineering, Belgrade, 25–27 Sept 2000, pp 5–10

Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia

Dimopoulos Y, Bourret P, Lek S (1995) Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process Lett 2(6):1–4

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, New York

Fang Y, Chow TWS (2006) Wavelets based neural network for function approximation. Lecture notes in computer science, vol 3971. Springer, Berlin/New York, pp 80–85

Gao R, Tsoukalas HI (2001) Neural-wavelet methodology for load forecasting. J Intell Robot Syst 31:149–157

He Y, Chu F, Zhong B (2002) A hierarchical evolutionary algorithm for constructing and training wavelet networks. Neural Comput Appl 10:357–366

Heskes T (1997) Practical confidence and prediction intervals. In: Mozer MC, Jordan MI, Petsche T (eds) Advances in neural information processing systems, vol 9. MIT Press, Cambridge, pp 176–182

Jiao L, Pan J, Fang Y (2001) Multiwavelet neural network and its approximation properties. IEEE Trans Neural Networks 12(5):1060–1066

Juditsky A, Zhang Q, Delyon B, Glorennec P-Y, Benveniste A (1994) Wavelets in identification – wavelets, splines, neurons, fuzzies: how good for identification? Technical report, INRIA

Kadambe S, Srinivasan P (2006) Adaptive wavelets for signal classification and compression. Int J Electron Commun 60:45–55

Kan K-C, Wong KW (1998) Self-construction algorithm for synthesis of wavelet networks. Electron Lett 34:1953–1955

Khayamian T, Ensafi AA, Tabaraki R, Esteki M (2005) Principal component-wavelet networks as a new multivariate calibration model. Anal Lett 38(9):1447–1489

Mallat SG (1999) A wavelet tour of signal processing. Academic, San Diego

Moody JE, Utans J (1994) Architecture selection strategies for neural networks: Application to corporate bond rating prediction. In: Refenes AP (ed) Neural Networks in the Capital Markets. New York, John Wiley & Sons

Oussar Y, Dreyfus G (2000) Initialization by selection for wavelet network training. Neurocomputing 34:131–143

Oussar Y, Rivals I, Presonnaz L, Dreyfus G (1998) Training wavelet networks for nonlinear dynamic input output modelling. Neurocomputing 20:173–188

Papadopoulos G, Edwards JP, Murray FA (2000) Confidence estimation methods for neural networks: a comparison. In: ESANN, Bruges, Belgium, 26–28 August 2000, pp 75–80

Pati YC, Krishnaprasad PS (1993) Analysis and synthesis of feedforward neural networks using discrete affine wavelet transforms. IEEE Trans Neural Netw 4(1):73–85

Pittner S, Kamarthi SV, Gao Q (1998) Wavelet networks for sensor signal classification in flank wear assessment. J Intell Manuf 9:315–322

Postalcioglu S, Becerikli Y (2007) Wavelet networks for nonlinear system modelling. Neural Comput Appl 16:434–441

Samarasinghe S (2006) Neural networks for applied sciences and engineering. Taylor & Francis, New York

Satchwell C (1994) Neural networks for stochastic problems: more than one outcome for the input space. In: NCAF conference, Aston University, Birmingham, UK, September 1994

Subasi A, Alkan A, Koklukaya E, Kiymik MK (2005) Wavelet neural network classification of EEG signals by using AR model with MLE pre-processing. Neural Netw 18:985–997

Szu H, Telfer B, Kadambe S (1992) Neural network adaptive wavelets for signal representation and classification. Opt Eng 31:1907–1916

Ulugammai M, Venkatesh P, Kannan PS, Padhy NP (2007) Application of bacterial foraging technique trained artificial and wavelet neural networks in load forecasting. Neurocomputing 70:2659–2667

Wong K-W, Leung AC-S (1998) On-line successive synthesis of wavelet networks. Neural Process Lett 7:91–100

Xu J, Ho DWC (2002) A basis selection algorithm for wavelet neural networks. Neurocomputing 48:681–689

Xu J, Ho DWC (2005) A constructive algorithm for wavelet neural networks. In: Proceedings of the first international conference, ICNC 2005, Changsha, China, 27–29 Aug 2005. Lecture notes in computer science, vol 3610. Springer, Berlin/Heidelberg, pp 730–739

Yao SJ, Song YH, Zhang LZ, Cheng XY (2000) Wavelet transform and neural networks for short-term electrical load forecasting. Energy Conv Manage 41:1975–1988

Zapranis A, Alexandridis A (2008) Modelling temperature time dependent speed of mean reversion in the context of weather derivative pricing. Appl Math Finance 15(4):355–386

Zapranis A, Alexandridis A (2009) Weather derivatives pricing: modelling the seasonal residuals variance of an Ornstein-Uhlenbeck temperature process with neural networks. Neurocomputing 73:37–48

Zapranis A, Refenes AP (1999) Principles of neural model identification, selection and adequacy: with applications to financial econometrics. Springer, London/New York

Zapranis AD, Livanis E (2005) Prediction intervals for neural network models. Paper presented at the proceedings of the 9th WSEAS international conference on computers, Athens

Zhang Q (1993) Regressor selection and wavelet network construction. Technical report, INRIA

Zhang Q (1994) Using wavelet network in nonparametric estimation. Technical report, INRIA

Zhang Q (1997) Using wavelet network in nonparametric estimation. IEEE Trans Neural Netw 8 (2):227–236

Zhang Z (2007) Learning algorithm of wavelet network based on sampling theory. Neurocomputing 71(1):224–269

Zhang Q, Benveniste A (1992) Wavelet networks. IEEE Trans Neural Netw 3(6):889–898

Appendix B: First-Order Derivatives of the Wavelet Network w.r.t the Network Weights

This appendix contains the derivations of the various derivatives of the first order needed for the calculation of the updates of the wavelet network weights during the training procedure. The loss function is the normal quadratic given by

$$L_n = \frac{1}{n}\sum_{p=1}^{n} E_p = \frac{1}{2n}\sum_{p=1}^{n} e_p^2 = \frac{1}{2n}\sum_{p=1}^{n}\left(y_p - \hat{y}_p\right)^2$$

where n is the number of training pairs. The output of the wavelet network is given by

$$g_\lambda(\mathbf{x};\mathbf{w}) = \hat{y}(\mathbf{x}) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \Psi_j(\mathbf{x}) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i$$

The partial derivatives with respect to each parameter, $\frac{\partial \hat{y}_p}{\partial w}$, and with respect to each input variable, $\frac{\partial \hat{y}_p}{\partial x_i}$, are presented here.

B.1 Partial Derivatives w.r.t. the Bias Term

$$\frac{\partial \hat{y}_p}{\partial w_{\lambda+1}^{[2]}} = 1$$

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8,

B.2 Partial Derivatives w.r.t. the Direct Connections

$$\frac{\partial \hat{y}_p}{\partial w_i^{[0]}} = x_i \quad i = 1, \ldots, m$$

B.3 Partial Derivatives w.r.t. the Linear Connections Between the Wavelets and the Output

$$\frac{\partial \hat{y}_p}{\partial w_j^{[2]}} = \Psi_j(\mathbf{x}) \quad j = 1, \ldots, \lambda$$

B.4 Partial Derivatives w.r.t. the Translation Parameters

$$\begin{aligned}
\frac{\partial \hat{y}_p}{\partial w_{(\xi)ij}^{[1]}} &= \frac{\partial \hat{y}_p}{\partial \Psi_j(\mathbf{x})} \cdot \frac{\partial \Psi_j(\mathbf{x})}{\partial \psi\left(z_{ij}\right)} \cdot \frac{\partial \psi\left(z_{ij}\right)}{\partial z_{ij}} \cdot \frac{\partial z_{ij}}{\partial w_{(\xi)ij}^{[1]}} \\
&= w_j^{[2]} \cdot \psi\left(z_{1j}\right) \cdots \psi'\left(z_{ij}\right) \cdots \psi\left(z_{mj}\right) \cdot \frac{-1}{w_{(\zeta)ij}^{[1]}} \\
&= -\frac{w_j^{[2]}}{w_{(\zeta)ij}^{[1]}} \psi\left(z_{1j}\right) \cdots \psi'\left(z_{ij}\right) \cdots \psi\left(z_{mj}\right)
\end{aligned}$$

B.5 Partial Derivatives w.r.t. the Dilation Parameters

$$\begin{aligned}
\frac{\partial \hat{y}_p}{\partial w^{[1]}_{(\zeta)ij}} &= \frac{\hat{y}_p}{\partial \Psi_j(\mathbf{x})} \cdot \frac{\partial \Psi_j(\mathbf{x})}{\partial \psi\left(z_{ij}\right)} \cdot \frac{\partial \psi\left(z_{ij}\right)}{\partial z_{ij}} \cdot \frac{\partial z_{ij}}{\partial w^{[1]}_{(\zeta)ij}} \\
&= w^{[2]}_j \cdot \psi\left(z_{1j}\right) \cdots \psi'\left(z_{ij}\right) \cdots \psi\left(z_{mj}\right) \cdot \frac{x_i - w^{[1]}_{(\xi)ij}}{{w^{[1]}_{(\zeta)ij}}^2} \\
&= -\frac{w^{[2]}_j}{w^{[1]}_{(\zeta)ij}} \frac{x_i - w^{[1]}_{(\xi)ij}}{{w^{[1]}_{(\zeta)ij}}^2} \psi\left(z_{1j}\right) \cdots \psi'\left(z_{ij}\right) \cdots \psi\left(z_{mj}\right) \\
&= -\frac{w^{[2]}_j}{w^{[1]}_{(\zeta)ij}} z_{ij} \psi\left(z_{1j}\right) \cdots \psi'\left(z_{ij}\right) \cdots \psi\left(z_{mj}\right) \\
&= z_{ij} \frac{\partial \hat{y}_p}{\partial w^{[1]}_{(\xi)ij}}
\end{aligned}$$

B.6 Partial Derivatives w.r.t. the Input Variables

$$\begin{aligned}
\frac{\partial \hat{y}_p}{\partial x_i} &= w^{[0]}_i + \frac{\sum_{j=1}^{\lambda} w^{[2]}_j \partial \Psi_j(\mathbf{x})}{\partial \psi\left(z_{ij}\right)} \cdot \frac{\partial \psi\left(z_{ij}\right)}{\partial z_{ij}} \cdot \frac{\partial z_{ij}}{\partial x_i} \\
&= w^{[0]}_i + \sum_{j=1}^{\lambda} w^{[2]}_j \cdot \psi\left(z_{1j}\right) \cdots \psi'\left(z_{ij}\right) \cdots \psi\left(z_{mj}\right) \cdot \frac{1}{w^{[1]}_{(\zeta)ij}} \\
&= w^{[0]}_i + \sum_{j=1}^{\lambda} \frac{w^{[2]}_j}{w^{[1]}_{(\zeta)ij}} \cdot \psi\left(z_{1j}\right) \cdots \psi'\left(z_{ij}\right) \cdots \psi\left(z_{mj}\right) \\
&= w^{[0]}_i - \sum_{j=1}^{\lambda} \frac{\partial \hat{y}_p}{\partial w^{[1]}_{(\xi)ij}}
\end{aligned}$$

Index

A.K. Alexandridis and A.D. Zapranis, *Weather Derivatives: Modeling and Pricing Weather-Related Risk*, DOI 10.1007/978-1-4614-6071-8,