Modeling and Pricing the European Temperature in the Context of Weather Derivative Pricing

A. Alexandridis
Lecturer in Finance
University of Kent
School of Mathematics, Statistics and Applied Science
Cornwallis Building
Canterbury, Kent
CT2 7NF
UK
E-mail: A.Alexandridis@kent.ac.uk

A. Zapranis
Associate Professor
Department of Accounting and Finance
University of Macedonia of Economics and Social Studies
156 Egnatia St.
P.O. 54006, Thessaloniki
Greece
E-mail: zapranis@uom.gr

Abstract. The purpose of this study is to develop a model that accurately describes the dynamics of the daily average temperature. The statistical properties of the daily average temperatures will be examined in order to propose a process that exhibits the same behavior. A framework using wavelet analysis for accurately selecting the seasonal parts of the mean and variance was presented. In addition a novel approach using wavelet networks was applied in selecting the number of the lags of the speed of mean reversion. Our model was evaluated and compared in-sample and out-of-sample in seven locations against models previously proposed in literature. Our results indicate that the proposed model significantly outperforms alternative methods. In order to obtain a better understanding of the distributions of the residuals we expanded our analysis by fitting additional distributions to the residuals. Our findings suggest that the hyperbolic distribution provides a slightly better fit than the normal distribution. Finally, the pricing formulas for various temperature derivates were presented under the assumption of a normal and a hyperbolic distribution.

Keywords: Weather derivatives, temperature derivatives, forecasting, wavelet networks

Mathematics Subject Classification (2000): 91G20·91G80·60G15·60G20

JEL Classification: G12·G13·G17

1. Introduction

The purpose of this study is to develop a model that accurately describes the dynamics of the Daily Average Temperature (DAT) in the context of weather derivative pricing. A statistical analysis of the DAT in various European cities that weather derivatives are actively traded will be performed. The statistical properties of the DATs will be examined in order to propose a process that exhibits the same behavior. Then, the estimated model will be used in order to derive the pricing formulas for the weather derivatives on various temperature indices.

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.

According to (Hanley, 1999) and (Challis, 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). Today the weather market is one of the fasted developing markets. In 2004, the notional value of Chicago Mercantile Exchange (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 Over-The-Counter (OTC) market was still more active than the exchange, so the bid-ask spreads were quite large.

Early methods such as the Actuarial Method or the Historical Burn Analysis (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, (Jewson, Brix, & Ziehmann, 2005).

Alternatively, one can directly model the corresponding index, namely “index modeling”, such as the Heating Degree Day (HDD) index, the Cooling Degree Day (CDD) index, the Cumulative Average Temperature (CAT) index, the Accumulated HDDs AccHDDs index or the Accumulated CDDs (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 & Wimmer, 2010; Geman & Leonardi, 2005) and (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 modelling 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 & Jewson, 2002; Caballero, Jewson, & Brix, 2002; Campbell & Diebold, 2005; Cao, Li, & Wei, 2004; Cao & Wei, 1999, 2000, 2003, 2004; Carmona, 1999; Franses, Neele, & van Dijk, 2001; Jewson & Caballero, 2003a, 2003b; Moreno, 2000; Roustant, Laurent, Bay, & Carraro, 2003a, 2003b; Svec & Stevenson, 2007; Taylor & Buizza, 2002, 2004) and (Tol, 1996)make use of a general ARMA framework.

On the other hand a wide range of studies suggest a temperature diffusion stochastic differential equation, (Alaton, Djehince, & Stillberg, 2002; Bellini, 2005; Benth, 2003; Benth & Saltyte-Benth,

2005, 2007; Benth, Saltyte-Benth, & Koekebakker, 2007, 2008; Bhowan, 2003; Brody, Syroka, & Zervos, 2002; Dischel, 1998a, 1998b, 1999; Dornier & Queruel, 2000; Geman & Leonardi, 2005; Hamisultane, 2006a, 2006b, 2007, 2008; McIntyre & Doherty, 1999; Oetomo & Stevenson, 2005; Richards, Manfredo, & Sanders, 2004; Schiller, Seidler, & Wimmer, 2008; Torro, Meneu, & Valor, 2003; Yoo, 2003; Zapranis & Alexandridis, 2006, 2007, 2008, 2009b, 2009c) and (Zapranis & Alexandridis, 2011). 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 this study 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 previous studies such as (Bellini, 2005; Benth & Saltyte-Benth, 2005, 2007; Benth, et al., 2007; Zapranis & Alexandridis, 2008, 2009b, 2009c) a stochastic process is selected for describing the temperature process. The stochastic process will be build upon the statistical properties found on the seven DAT time series.

In (Zapranis & Alexandridis, 2008) a non-linear time depended speed of mean reversion variable was modeled by a neural network (NN). More precisely wavelet analysis was used in order to identify the trend and the seasonal part of the temperature signal and then a NN was used in the detrended and deseasonalized series. Deducting the form of the seasonal mean and variance was based on observing the wavelet decomposition. In this paper we expand the initial framework presented in (Zapranis & Alexandridis, 2008) by combining these two steps.

In this study wavelet analysis (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 non-parametrically 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). In addition we expand the analysis of (Zapranis & Alexandridis, 2008) by applying a novel procedure first presented in (A. Alexandridis, 2010) in order to estimate the length of the lag series in each city. Then, our proposed model will be evaluated and compared against other models previously proposed in literature in 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 standard BM is investigated.

The rest of the chapter is organized as follows. In section 2 the data is described and examined. In section 3 a model for the DATs is proposed based on the results of the data examination. Next, in section 3.1 WA is used in order to identify the statistical significant seasonal components of the DAT.

A WN is constructed in order to model the detrended and deseasonalized DAT in section 4. More precisely in section 4.1 the training data set is constructed by selecting the significant lags of the DATs. In section 4.2 the topology of the WN is selected while in section 4.3 the WN is initialized and trained in order to model a nonlinear autoregressive (AR) model. Next, in section 4.4 the statistical properties of the time depended mean reversion function is examined. Next, in section 5 WA is used in order to identify model seasonal variance that exists in the residuals. In section 6 the distributional statistics of residuals after removing the seasonal variance are examined. Moreover our proposed model is compared in-sample against two popular models previously proposed in literature. Next, the residuals are tested under the assumption of a Lévy motion driving noise process in section 6.2. In section 7 an evaluation of our model out-of-sample is performed. In section 8 the pricing formulas of the weather derivatives are presented. More precisely, in section 8.1 the temperature derivatives traded on the exchange are presented. In section 8.2 and 8.3 the pricing formulas under the assumption of a normal and a hyperbolic distribution respectively are derived while in section 8.4 the market price of risk is discussed. Finally, in section 9 we conclude.

2. Data description

For accurate pricing and efficient weather risk management the weather data must be both of adequate amount and highly quality, (Dunis & Karalis, 2003). For this study we obtained data for the European cities that are traded in CME. 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. The data corresponding to the European cities were provided by the European Climate Assessment & Dataset[1] (ECAD). The weather index 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 (^oC). European weather contracts traded on the CME use the same measurement for the temperature. Precision with which temperature in the ECAD is measured is 0.1 ^oC. 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 18615 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 & Karalis, 2003) different methods for filling the missing data were described. In this study 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 is replaced by the average of these two parameters.

The above procedure is very easy in implementation. More precisely a normal average is obtained which is balanced by the temporal temperature conditions around the missing values. However, in some cities (for example Rome) there are consecutive missing values. In this case the missing values are filled using the simple average over the years for the particular day.

Due to space limitations only the data from Berlin are graphically presented. The results of the remaining cities are similar. More precisely the DAT and the empirical distribution, the mean, the standard deviation, the skewness and the kurtosis of the DAT in Berlin are presented in Figure 1.

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

In part (a) of Figure 1 the DATs for Berlin for the period 1/1/1991-31/12/2000 are presented. A closer inspection of Figure 1 reveals a seasonal cycle of one 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. The mean temperature ranges from 6.49 in Oslo to 15.57 in Rome. 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. 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 degrees for Rome and Madrid while it is over 40 degrees 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 performed. In all cities the normality is strongly rejected by a Jarque- Bera (JB) test. 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. Part (b) of Figure 1 presents the empirical distributions of the DAT in Berlin. 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 daily mean, standard deviation, skewness and kurtosis of DAT were estimated. The mean of the DAT, $T_{avy,t}$, was estimated using only observations for each particular day t. In part (c) of Figure 1 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. The mean DAT in Amsterdam fluctuates from $.1.9^{o}C$. to $19.7^{o}C$. Similarly, in Berlin the mean temperature fluctuates from $-1.2^{o}C$ to $22.4^{o}C$, from $4.9^{o}C$ to $28.3^{o}C$ in Madrid, from $-5.7^{o}C$ to $18.8^{o}C$ in Oslo, from $3.1^{o}C$ to $.23^{o}C$. in Paris, from $6^{o}C$ to $27^{o}C$ in Rome and from $-4.2^{o}C$ to $19.5^{o}C$ in Stockholm.

Next the standard deviation of the DAT is estimated. In part (d) of Figure 1 the standard deviation for Berlin is presented. 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 & Saltyte-Benth, 2005, 2007; Benth, et al., 2007; Zapranis & Alexandridis, 2008, 2009c).

Part (e) of Figure 1 presents the estimated skewness for each day t for the seven cities. Figure 1 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 can be found in part (f) of Figure 1. Figure 1 does not reveal any seasonal pattern of the kurtosis. On the other hand it is clear that for all cities the kurtosis have small deviations around two with many upwards 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. 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 or substantial values are estimated.

Next 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. The Hurst exponent was estimated using iterative methods described in (Koutsouyiannis, 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 or an unsubstantial value will be estimated, (Bellini, 2005).

Two unit root test were performed in the DAT for the seven cities. Each time-series is tested for unit root using an Augmented Dickey-Fuller (ADF) test. The ADF performed using the Schwartz information criterion in order to select the optimal number of lagged values. Our results reveal 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. More precisely the largest value for the ADF statistic in -3.7785 in Rome while the critical value at 5% significance level is -2.8621. Moreover the *p*-values are almost zero for the seven cities.

In order to obtain a more powerful test, the (Kwiatkowski, Phillips, Schmidt, & Shin, 1992) (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 non-stationary (a unit root exists). The optimal bandwidth number was estimated using the Newery-West method. The largest KPSS statistic is obtained in Amsterdam with value of 0.0844. The KPSS statistic has a value smaller than the critical value 0.463 for all cities, hence the null hypothesis, that the time-series is stationary, cannot be rejected for all cities.

3. A model for the daily average temperature: a Gaussian Ornstein-Uhlenbeck process with lags and time-varying mean-reversion

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, et al., 2004; Cao & Wei, 1999, 2000, 2003) 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 section our results confirm the above assumptions. It is known that temperature follows a predicted cycle. As it was expected and it is shown on Figure 1 a strong cycle of one 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 Figure 1. In other words it is not possible to observe temperatures of $20^{o}C$ in winter in Oslo. Additionally, 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, industrial buildings and cooling units. Hence, urbanization around a weather station results to an increment in the observed measurements of temperature. Finally, observing Figure 1 it is clear that the temperature volatility is higher in winter than in summer.

Following (Zapranis & Alexandridis, 2008, 2011) and (A. Alexandridis, 2010) a model that describes the temperature dynamics is given by a Gaussian mean-reverting O-U process defined as follows:

$$dT(t) = dS(t) + \kappa(t)\big(T(t) - S(t)\big)dt + \sigma(t)dB(t) \tag{1}$$

where $T(t)$ is the average daily temperature, $\kappa(t)$ 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 & Queruel, 2000) the term $dS(t)$ should be added for a proper mean-reversion towards the historical mean, $S(t)$.

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; i.e. the difference of today and tomorrows 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 $\kappa(t)$.

In (Benth & Saltyte-Benth, 2007) the historical mean is captured by a simple sinusoid while the seasonal variance captured by a Fourier series arbitrary truncated. In this study 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:

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

$$\sigma^2(t) = c + \sum_{i=1}^{I_2} a_i \sin(2p_i' \; t\cos(2 + \pi\sum_{j=1}^{J_2} a_j 365) \quad p_j' \; t \tag{3}$$

$$Trend_t = a + bt \tag{4}$$

In order to identify the terms I_1, p_i in (2) and I_2, J_2, p_i' in (3) we decompose the temperature series using a wavelet transform.

Finally, the driving noise process $B(t)$ is modeled by a standard Brownian motion (BM). In Figure 2 the histogram of the first difference of the DAT and the normal distribution (solid line) is presented. A

closer inspection of Figure 2 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.

3.1 Identifying and removing the trend and the seasonal mean using wavelet analysis

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

In order to justify the structure of the seasonal part of the temperature and to identify the terms I_1, p_i in (2) the temperature series is decomposed using a wavelet transform. (Lau & Weng, 1995) confirmed seasonalities in the temperature series with a period significantly greater than one year. (Lau & 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: inter-annual (2-5 yrs), inter-decadal (10-12 yrs, 20-25yrs and 40-60 yrs) and century (~180 yrs) scales. This conclusion was also reached in (Zapranis & Alexandridis, 2006, 2008, 2009b, 2009c).

In this study 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. For a more detailed exposition on the mathematical aspects of wavelets refer to (Daubechies, 1992; Mallat, 1999).

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 Table 1 and Table 2. Figure 3 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 Figure 3, in all approximations, a_j. Moreover a series of cycles affects the dynamics of temperature. As expected an one 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, 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 (detail d_1) reflect the noise part of the time series. A closer inspection of the noise part reveals seasonalities, which will be extracted later. Hence, modelling the historical seasonal mean by (2) is justified by the results of the previous analysis.

Panel *A* of Table 2 reports all the cycles than can be found on the temperature dynamics using WA for the seven cities. In Table 2 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 1 shows the estimated parameters a and b of the linear trend represented by (4). All *p*-values are smaller than 0.05 suggesting that a trend exists and it is statistical 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 $0.9^{o}C$ in Madrid to $1.9^{o}C$ in Stockholm the last 10 years. Subtracting the trend form the original data we obtain the de-trended 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 (2). 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 one 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 2. In Table 2 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.

4. Using Wavelet Networks On Detrended and Deseasonalized Daily Average Temperatures

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 (A. Alexandridis, 2010) and (Antonis Alexandridis & Zapranis, 2011) 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 backward elimination method and will be trained using the back-propagation method, (Antonis Alexandridis & Zapranis, 2011). Here we present only the basic notion of WN. For an analytical study on WN refer to (Antonis Alexandridis & Zapranis, 2011).

In Figure 4 the structure and the mathematical expressions of a WN are presented. 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:

$$g_\lambda(\mathbf{x};\mathbf{w}) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \Psi_j(\mathbf{x}) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i \qquad (5)$$

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:

$$\Psi_j(x) = \prod_{i=1}^{m} \psi(z_{ij}) \tag{6}$$

where

$$z_{ij} = \frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}} \tag{7}$$

The mother wavelet is given by the Mexican Hat function:

$$\psi(\alpha) = \left(1 - \alpha^2\right) e^{-\frac{1}{2}a^2} \tag{8}$$

Following (Benth & Saltyte-Benth, 2007) and (Zapranis & Alexandridis, 2008) a discrete approximation of (1) is obtained and is given by:

$$\Delta T(t) \approx \Delta S(t) + \kappa\left(T(t-1) - S(t-1)\right) + \sigma(t)\Delta B(t) \tag{9}$$

Next, by setting

$$T(t) = T(t) - S(t) \tag{10}$$

we have that

$$T(t) = aT(t-1) + \sigma(t)\varepsilon(t) \tag{11}$$

with

$$a = 1 + \kappa \tag{12}$$

Similarly, when the speed of mean reversion is a function of time we have that

$$T(t) = a(t-1)T(t-1) + \sigma(t)\varepsilon(t) \tag{13}$$

$$a(t) = 1 + \kappa(t) \tag{14}$$

The detrended and deseasonalized temperature series, $T(t)$, can be modeled with an AR(1) process with a zero constant term, as shown in (13). 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 & 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 (13). 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). However, (Zapranis & Alexandridis, 2008) estimated daily values of the variable

$a(t)$ for the city of Paris using a non-parametric non-linear NN. Their results indicate strong time dependence in the daily values of $a(t)$.

In this section, we address that issue, by using a WN to estimate non-parametrically relationship (13) and then estimate a as a function of time. In addition, we propose a novel approach for selecting the number of lags in the temperature process first presented by (A. Alexandridis, 2010). Hence a series of speed of mean reversion parameters, $a_i(t)$, are estimated. 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 $a_i(t)$. 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 $a_i(t)$ is quite significant after all. In addition it is expected that the waveform of the WN will provide a better fit to the DATs that are governed by seasonalities and periodicities.

Previous studies, (Alaton, et al., 2002; Bellini, 2005; Benth & Saltyte-Benth, 2005, 2007; Zapranis & Alexandridis, 2008, 2009c), 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 & Leonardi, 2005). Using WNs the generalized version of (13) is estimated nonlinearly and non-parametrically, that is:

$$T(t+1) = \phi\left(T(t), T(t-1), \ldots\right) + e(t) \tag{15}$$

where

$$e(t) = \sigma(t)\varepsilon(t) \tag{16}$$

Model (15) 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 & Saltyte-Benth, 2007) and (Zapranis & Alexandridis, 2008). However, the length of the lag series must be selected. Since the WN is a non-parametrical nonlinear estimator results from the autocorrelation function (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 (A. Alexandridis, 2010) is applied in order to determine the number of significant lags in each city.

4.1 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 number of lags 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. Focusing on 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 (A. Alexandridis, 2010) will be applied. In this framework first a network with a lot of information - .i.e. a large number of lags - is build and then gradually it is optimized by removing the parts that do

not contribute to the predictive power of the WN. Initially, the training set contains the dependent variable and 7 lags. Hence, the training set consists of 7 inputs, 1 output and 3643 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 by (Moody & Utans, 1992). The analysis presented in (A. Alexandridis, 2010) indicates that the SBP fitness 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. In order to approximate the empirical distribution of the SBP the bootstrap method is applied as it was described in (A. Alexandridis, 2010). Using the empirical distribution of the SBP, hypothesis tests can be constructed.

The 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%.

Table 3 summarizes the results of the model identification algorithm for Berlin. Both the model selection and variable selection algorithms are included in Table 3. The algorithm concluded in 4 steps and the final model contains only 3 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 were significantly reduced in the final model. The initial model needed 5 HUs and 7 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 the full model the last four variables have *p*-value greater than 0.1 while the 6th lag has a *p*-value of 0.8826 strongly indicating a “not significant” variable. In the first step by removing X_6 from the model the *p*-value of X_5 became 0 while for X_7 and X_4 the *p*-values became 0.5700 and 0.1403 respectively. At step 2, X_7, which had the largest *p*-value=0.5700 at the previous step, was removed from the model. 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. At step 3 only X_4 has a *p*-value greater than 0.1. Finally, at step 4 the three remaining variables have all *p*-value equal to zero.

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 8th and the 11th 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 5 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 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 4 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.

4.2 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 was presented in (A. Alexandridis, 2010; Zapranis & Alexandridis, 2009a). For simplicity we refer only to results from Berlin. The results of the remaining cities are similar. 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 minimizes 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 5 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 5 HUs. On other words the 5 higher ranking wavelets should be selected form the wavelet basis in order to construct the WN. The prediction risk at the initial model with only one HU is almost the same as in the model with 5 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 2 HUs are used. Similarly, at steps two, three and four the prediction risk is minimized when only one HU is used. Additional HUs does not improve the fitting or the predictive power of the model. It is clear that the prediction risk is minimized when one HU is used and then it increases almost monotonically. Table 4 presents the appropriate HUs for the construction of the final WN for each city. Our results indicate that a very simple model with only 1 HU is adequate to fit the DATs in the seven cities of our analysis.

4.3 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 method is used to initialize the WN. A wavelet basis is constructed by scanning the 4 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 5 sample points of the training data in their support are removed. 76 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 backward elimination method first introduced in (Zhang, 1997) the wavelets are ranked in order of significance. The

wavelets in the wavelet library are ranked as follows: the backward elimination 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. The initialization is very good and the WN converged after only 19 iterations. The training stopped when the minimum velocity, 10^{-5}, of the training algorithm was reached. The results in the remaining cities are similar.

In Table 4 various fitness criteria of the seven WNs corresponding to the seven cities are presented. A closer inspection of Table 4 reveals that the WNs fit the DATs reasonable 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 4 the Prediction of Sign (POS) as well 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. For a detailed explanation of the fitness criteria we refer to (A. Alexandridis, 2010) and (Zapranis & Refenes, 1999). 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.

4.4 The wavelet neural networks approach: time dependent mean reversion parameter

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 non-parametrically 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 (14).

Using a WN relation (15) is estimated non-parametrically. Once we have the estimator of the underlying function ϕ, then the daily values of a can be computed as follows:

$$\alpha_1(t) = dT(t+1) / dT(t) = d\phi / dT \qquad (17)$$

The analytic expression for derivative of the WN w.r.t. the input variable $d\phi / dT$ can be found in (A. Alexandridis, 2010). We estimate $\phi(\ )$ non-parametrically with a WN, $g(\ )$.

For Berlin the daily values of $a(t)$ (3,647 values) are depicted in Figure 5. Because in Berlin there are 3 significant lags, there are three mean-reverting functions, $a_i(t)$. The corresponding frequency histograms are given in Figure 6. The graphs for all cities are very similar. The relevant statistics of $a_i(t)$ for all cities are presented in Table 5. 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; i.e. 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 Figure 5 and Figure 6, 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 Figure 6. A closer inspection of Table 5 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 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 $a_3(t)$ 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 Figure 6 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 & Hartigan, 1985). The null hypothesis test is that $a_i(t)$ follow a unimodal distributions versus the alternative that $a_i(t)$ follows a bimodal distribution. 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 & Alexandridis, 2008) indicate that $a_1(t)$ follows a bimodal distribution in Paris. However in (Zapranis & Alexandridis, 2008) only one lag is used in order to estimate model (15) which may have a strong impact on the structure and values of $a_1(t)$.

Moreover, Figure 5 may suggest seasonalities in the structure of $a_i(t)$. The ACF of $a_i(t)$ is shown in Figure 7. A seasonality of a half year can be shown in the ACF. Also the first 35 lags are statistically important and positive 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 (15) can be obtained. In the next section the residuals $e(t)$ will be examined.

5. Identifying and removing the seasonal variance using wavelet analysis

In this section the residuals of the WN will be examined. The initial hypothesis for the residuals $e(t)$ of model (15) was that they follow the normal distribution. However a closer inspection of the noise part of the wavelet decomposition of Berlin's DAT (Figure 3) reveals seasonalities.

The mean value of the residuals is very close to zero for all cities 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 by the Kolmogorov-Smirnov test or the Kolmogorov distance. 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. 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 Figure 8. However, a closer inspection of the ACF reveals a seasonal component in the residuals in Madrid and Rome.

Previous studies identified the existence of seasonal variance in the residuals of either the linear or the nonlinear AR model, (Benth & Saltyte-Benth, 2005, 2007; Zapranis & Alexandridis, 2008, 2009c). 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 Figure 9. By squaring the residuals the seasonal pattern in the ACF is clear in every city as it is shown in Figure 9.

As it was mentioned earlier the seasonal variance is modeled by equation (3). Since for the residuals $e(t)$ of the nonlinear AR model it is true that $e(t) = \sigma(t)\varepsilon(t)$ 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:

$$\sigma^2(t+365) = \sigma^2(t) \tag{18}$$

where $t = 1,...,3650$.

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.

In Figure 10 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 (18). Moreover a half-year cycle (a_6 and d_7) as well as a seasonal cycle exist (d_6). Hence, in (3) 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 6 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 equation (3).

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 6 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 Madrid only 3 parameters were needed in order to fit and remove the seasonal variance, while in Amsterdam and Stockholm 6 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 Figure 11. 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. Moreover an increase in variance is observed during May.

6. Testing the residuals after dividing out the seasonal variance

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 & Saltyte-Benth, 2007).

First, the ACF of the residuals after dividing out the seasonal variance is examined. Figure 12 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 7 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 7. 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 7 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. Similar, 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:

$$H = d + \frac{1}{2} \tag{19}$$

The Hurst exponent takes values in the interval (0,1). For $1/2 < H < 1$ the process has long memory, for $0 < H < 1/2$ the process has short memory while for $H = 1/2$ the BM is retrieved, (Bellini, 2005). In Table 8 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 (Koutsouyiannis, 2003) is followed in order to estimate the Hurst exponent.

Results from Table 8 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 while the largest Hurst exponent was observed in Oslo with value of 0.5201. The above results indicate the absence of fractionality characteristics in the dynamics of the temperature process. Therefore, the assumption of a BM instead of Fractional BM 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. 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.

6.1 In sample comparison – distributional 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 & Saltyte-Benth, 2007). For simplicity we name the two models as the Alaton and the Benth model respectively.

In Table 9 the estimated parameters from Alaton model are presented while in Table 10 the descriptive statistics of the residuals can be found. In Table 9 only the statistical significant parameters at significance level 5% are reported. A closer inspection of Table 10 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 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 11 the estimated parameters from the Benth model are presented while in Table 12 the descriptive statistics of the residuals can be found. In Table 11 only the statistical significant parameters at significance level 5% are reported. A closer inspection of Table 12 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 & 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 & 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.

(Zapranis & 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.

6.2 Testing the residuals under the Lévy motion assumption

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 & Saltyte-Benth, 2005) and (Bellini, 2005) used a Lévy process as the driving noise process. In particular, (Benth & Saltyte-Benth, 2005) used a generalized hyperbolic distribution. In (Bellini, 2005) an 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 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). In Table 13 the estimated Kolmogorov distance and the Anderson-Darling statistics are presented for four distributions: normal, hyperbolic, NIG and stable. From Table 13 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 Kolmogorv-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.

7. Evaluating the temperature model out-of-sample

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. In this study 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.

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. 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, 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. 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. Again the results for the HDD and the CAT index are the same.

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, the forecasts of the Benth model deteriorate as the forecast window increases.

Table 14 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 enhance 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.

8. Pricing Temperature Derivatives

The analysis that performed in the previous chapter indicates that assuming a normal distribution is justified. In general the normal distributions fits 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 the Pacific Rim.

When the market is complete, then a unique risk-neutral probability measure $Q \sim P$ can 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, Odening, & Musshof, 2008). However, in incomplete markets a unique price cannot be obtained using the no-arbitrage assumption. In other words, under every measure Q all assets are martingales after discounting.

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 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's Lemma when a BM is considered or the Itô Formula for semimartingales when a Lévy motion is considered.

8.1 Temperature Derivatives Traded On the CME

The list of traded contracts on the weather derivatives market is extensive and constantly evolving. Chicago Mercantile Exchange (CME) offers various weather futures and options contracts. They are index-based products geared to average seasonal and monthly weather in 46 cities[2] around the world - 24 in the U.S., 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 one day. The value of a CAT index for the time interval $[\tau_1, \tau_2]$ is given by the following expression:

$$CAT = \int_{\tau_1}^{\tau_2} T(s)ds \tag{20}$$

where the temperature is measured in degrees Celsius. In London one CAT index futures 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 moths: 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 temperature is below a base temperature, and a CDD is the number of degrees by which the daily temperature is above the base temperature, i.e.

Daily HDD = max(0, base temperature – daily average temperature)

Daily CDD = max(0, daily average temperature – base temperature)

The base temperature is usually 65 degrees Fahrenheit in the USA and 18 degrees Celsius 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. 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:

$$PAC = \frac{1}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} T(s)ds \tag{21}$$

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, in (Alaton, et al., 2002) 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 Fractional BM. In a more recent paper (Benth & Saltyte-Benth, 2005) estimate the prices of a CAT future and option contracts under the assumption of a Lévy noise process. More precisely (Benth & Saltyte-Benth, 2005) propose that the residuals follow the generalized hyperbolic distribution. Similarly, (Bellini, 2005) presents the pricing of HDDs and CDDs contracts under the assumption of a Lévy noise process where the residuals follow the hyperbolic distributions. More recently, (Benth & 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.,

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

2008) the temperature dynamics were modeled by a Continuous Autoregressive (CAR(p)) process first introduced by (Brockwell & Marquardt, 2005). Under the normality assumption, pricing formulas for the CAT, HDDs and CDDs indices were presented. In (Zapranis & Alexandridis, 2008) the price of CAT futures were derived when the speed of mean is 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) price weather derivatives by marginal value using a modified Black-Scholes equation while (Platen & West, 2005) suggest a fair pricing approach based on an equilibrium method. On the other hand, (Garman, Blanco, & Erickson, 2000) introduces MC to price weather derivatives while (Xu, et al., 2008) apply an indifference pricing approach for weather derivatives that are traded OTC.

Our results indicate that the proposed model significantly outperforms alternative methods in the sense of forecasting, in sample and out-of-sample. It follows that the assumption of a constant mean-reversion parameter introduces significant error in the pricing of weather derivatives. In this study we present 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's Lemma.

Proposition 1. If the DAT follows a mean reverting O-U process with time varying speed of mean reversion and seasonal mean and variance:

$$dT(t) = dS(t) + \kappa(t)\big(T(t) - S(t)\big)dt + \sigma(t)dB(t)$$

an explicit solution can be derived from the Itô formula:

$$T(t) = S(t) + e^{\int_0^t \kappa(u)du}\big(T(0) - S(0)\big) + e^{\int_0^t \kappa(u)du}\int_0^t \sigma(s)e^{-\int_0^s \kappa(u)du}dB(s) \qquad (22)$$

Proof. Let us rewrite (1) as

$$dT(t) = \kappa(t)T(t)dt + \sigma(t)dB(t)$$

where $T(t) = \big(T(t) - S(t)\big)$. Then, the proof follows by a direct application of the Itô's Lemma.

□

8.2 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, AccHDDs, AccCDDs and Pacific Rim indices are derived.

8.2.1 CAT and Pacific Rim Futures and Options

Our aim is to give a mathematical expression for the CAT futures price. The weather derivative market it is a classical incomplete market. 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 $t \le \tau_1 < \tau_2$ is given by

$$e^{-r(\tau_2 - t)} \mathrm{E}_Q \left[\int_{\tau_1}^{\tau_2} \mathrm{T}(\tau) d\tau - F_{CAT}(t, \tau_1, \tau_2) \,|\, \mathfrak{F}_t \right] = 0 \tag{23}$$

and since F_{CAT} is $\mathfrak{F}_t$ adapted we derive the price of a CAT futures to be

$$F_{CAT}(t, \tau_1, \tau_2) = \mathrm{E}_Q \left[\int_{\tau_1}^{\tau_2} \mathrm{T}(\tau) d\tau \,|\, \mathfrak{F}_t \right] \tag{24}$$

Using Girsanov's theorem, under the equivalent measure *Q*, we have

$$W(t) = B(t) - \int_0^t \theta(u) du \tag{25}$$

or equivalently

$$dW(t) = dB(t) - \theta(t) dt \tag{26}$$

and note that $\sigma(t)$ is bounded away from zero. Hence, by combining equations (1) and (26) the stochastic process of the temperature in the risk-neutral probability Q^θ is

$$dT(t) = dS(t) + \left(\kappa(t) \left(T(t) - s(t) \right) + \sigma(t) \theta(t) \right) dt + \sigma(t) dW(t) \tag{27}$$

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 equation (27) under Q^θ is

$$\begin{aligned} T(t) = S(t) + e^{\int_0^t \kappa(u) du} \left(T(0) - S(0) \right) + e^{\int_0^t \kappa(u) du} \int_0^t \sigma(s) \theta(s) e^{-\int_0^s \kappa(u) du} ds \\ + e^{\int_0^t \kappa(u) du} \int_0^t \sigma(s) e^{-\int_0^s \kappa(u) du} dW(s) \end{aligned} \tag{28}$$

The proof of equation (28) is similar to the proof of Proposition 1. Note *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 (28) in (24) we find the price of a future contract on the CAT index at time *t*, where $t \le \tau_1 < \tau_2$.

Proposition 2. The CAT future price for $t \le \tau_1 < \tau_2$ is given by

$$F_{CAT}(t, \tau_1, \tau_2) = \mathrm{E}_Q \left[\int_{\tau_1}^{\tau_2} T(s) ds \,|\, \mathfrak{F}_t \right] = \int_{\tau_1}^{\tau_2} S(s) ds + I_1 + I_2 \tag{29}$$

where

$$I_1 = \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z) dz} T(t) ds \tag{30}$$

$$I_2 = \int_t^{\tau_1} \int_{\tau_1}^{\tau_2} e^{\int_0^s \kappa(z)dz} \sigma(u)\theta(u) e^{\int_u^0 \kappa(z)dz} dsdu + \int_{\tau_1}^{\tau_2} \int_u^{\tau_2} e^{\int_0^s \kappa(z)dz} \sigma(u)\theta(u) e^{\int_u^0 \kappa(z)dz} dsdu \tag{31}$$

Proof. From equations (24) and (28) we have

$$F_{CAT}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds \,|\mathcal{F}_t\right] = \int_{\tau_1}^{\tau_2} S(s)ds + \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds \,|\mathcal{F}_t\right]$$

and using Itô's isometry we can interchange the expectation and the integral

$$\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds \,|\mathcal{F}_t\right] = \int_{\tau_1}^{\tau_2} \mathrm{E}_Q\left[T(s) \,|\, \mathcal{F}_t\right] ds$$

$$= \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz} T(t)ds + \int_{\tau_1}^{\tau_2} \int_t^s \sigma(u)\theta(u) e^{\int_u^s \kappa(z)dz} duds$$

$$= I_1 + I_2$$

Hence,

$$I_1 = \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz} T(t)ds$$

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 2 gives the price of a futures CAT at time $t \le \tau_1 < \tau_2$. In other words the price of a futures CAT before the contract period. Hence, (29) 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 3. The CAT futures price for $\tau_1 \le t \le \tau_2$ is given by

$$F_{CAT}(t,\tau_1,\tau_2) = \int_{\tau_1}^{t} T(s)ds + F_{CAT}(t,t,\tau_2) \tag{32}$$

Proof. We have that the CAT futures price is

$$\begin{aligned} F_{CAT}(t,\tau_1,\tau_2) &= \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds \mid \mathfrak{F}_t\right] \\ &= \mathrm{E}_Q\left[\int_{\tau_1}^{t} T(s)ds + \int_{t}^{\tau_2} T(s)ds \mid \mathfrak{F}_t\right] \\ &= \int_{\tau_1}^{t} T(s)ds + \mathrm{E}_Q\left[\int_{t}^{\tau_2} T(s)ds \mid \mathfrak{F}_t\right] \\ &= \int_{\tau_1}^{t} T(s)ds + F_{CAT}(t,t,\tau_2) \end{aligned}$$

Note that the first term is known at time t since it refers to past temperatures while the second term is stochastic.

□

Similar, the in-period pricing formulas of the remaining indices can be easily extracted from the pricing formulas of the out-of-period valuation. Following the notation of (Benth & Saltyte-Benth, 2007) the dynamics of the CAT futures price under Q is given in the following proposition.

Proposition 4. The dynamics of $F_{CAT}(t,\tau_1,\tau_2)$ under the risk-neutral measure Q is

$$dF_{CAT}(t,\tau_1,\tau_2) = \Sigma_{CAT}(t,\tau_1,\tau_2)dW(t) \tag{33}$$

where

$$\Sigma_{CAT}(t,\tau_1,\tau_2) = \sigma(t)\int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz} ds \tag{34}$$

Proof. $F_{CAT}(t,\tau_1,\tau_2)$ 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

$$\frac{dF_{CAT}}{dT} = \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz} ds$$

hence,

$$dF_{CAT}(t,\tau_1,\tau_2) = \sigma(t)\int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz} ds dW(t)$$

□

Using Proposition 4 the price of call option written on CAT futures can be estimated.

Proposition 5. The price at time $t \le \tau$ of a call option written on a CAT futures with strike price K at exercise time $\tau \le \tau_1$ is

$$\begin{aligned} C_{CAT}(t,\tau,\tau_1,\tau_2) = e^{-r(\tau-t)} \Big\{ & \big(F_{CAT}(t,\tau_1,\tau_2) - K\big)\Phi\big(d(t,\tau,\tau_1,\tau_2)\big) \\ & + \Phi'\big(d(t,\tau,\tau_1,\tau_2)\big)\int_t^\tau \Sigma_{CAT}^2(t,\tau_1,\tau_2)ds \Big\} \end{aligned} \tag{35}$$

where

$$d(t,\tau,\tau_1,\tau_2) = \frac{F_{CAT}(t,\tau_1,\tau_2) - K}{\sqrt{\Sigma_{t,\tau}^2}} \tag{36}$$

and

$$\Sigma_{t,\tau}^2 = \int_t^\tau \Sigma_{CAT}^2(t,\tau_1,\tau_2)ds \tag{37}$$

and Φ is the cumulative standard normal distribution function.

Proof. The option price by definition is given by

$$C_{CAT}(t,\tau,\tau_1,\tau_2) = e^{-r(\tau-t)}\mathrm{E}_Q\left[\max\big(F_{CAT}(\tau,\tau_1,\tau_2) - K, 0\big) \mid \mathfrak{F}_t\right]$$

From Proposition 4 we have that the Q dynamics of the futures price can be written as

$$F_{CAT}(\tau,\tau_1,\tau_2) = F_{CAT}(t,\tau_1,\tau_2) + \int_t^\tau \Sigma_{CAT}(s,\tau_1,\tau_2)dW(s)$$

From this it follows that $F_{CAT}(\tau,\tau_1,\tau_2)$ conditioned on $F_{CAT}(t,\tau_1,\tau_2)$ follows the normal distribution with mean $F_{CAT}(t,\tau_1,\tau_2)$ and variance given by

$$\int_t^\tau \Sigma_{CAT}^2(s,\tau_1,\tau_2)ds$$

Hence,

$$\begin{aligned} & E_Q\left[\max\big(F_{CAT}(\tau,\tau_1,\tau_2) - K, 0\big) \mid \mathfrak{F}_t\right] \\ & \qquad = E_Q\left[\max\left(F_{CAT}(t,\tau_1,\tau_2) + \int_t^\tau \Sigma_{CAT}(s,\tau_1,\tau_2)dW(s) - K, 0\right) \mid \mathfrak{F}_t\right] \end{aligned}$$

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 $t \le \tau_1 < \tau_2$ is given by:

$$e^{-r(\tau_2-t)}\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\frac{1}{\tau_2-\tau_1}T(\tau)d\tau - F_{PAC}(t,\tau_1,\tau_2)\,|\,\mathfrak{F}_t\right]=0 \tag{38}$$

and since F_{PAC} is $\mathfrak{F}_t$ adapted we derive the price of a PAC futures to be

$$F_{PAC}(t,\tau_1,\tau_2)=\mathrm{E}_Q\left[\frac{1}{\tau_2-\tau_1}\int_{\tau_1}^{\tau_2}T(s)ds\,|\,\mathfrak{F}_t\right] \tag{39}$$

Observing equations (24) and (39) we conclude that:

$$F_{PAC}(t,\tau_1,\tau_2)=\frac{1}{\tau_2-\tau_1}F_{CAT}(t,\tau_1,\tau_2) \tag{40}$$

and, similarly, that the price of call option written on a PAC futures is given by:

$$C_{PAC}(t,\tau_1,\tau_2)=\frac{1}{\tau_2-\tau_1}C_{CAT}(t,\tau_1,\tau_2) \tag{41}$$

8.2.2 HDD and CDD 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

$$AccHDD=\int_{\tau_1}^{\tau_2}\max\left(c-T(s),0\right)ds \tag{42}$$

$$AccCDD=\int_{\tau_1}^{\tau_2}\max\left(T(s)-c,0\right)ds \tag{43}$$

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 $t\le\tau_1<\tau_2$ is given by:

$$e^{-r(\tau_2-t)}\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\max\left(0,c-T(\tau)\right)d\tau - F_{HDD}(t,\tau_1,\tau_2)\,|\,\mathfrak{F}_t\right]=0 \tag{44}$$

and since F_{HDD} is $\mathfrak{F}_t$ adapted we derive the price of a HDD futures to be

$$F_{HDD}(t,\tau_1,\tau_2)=\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\max\left(c-T(\tau),0\right)d\tau\,|\,\mathfrak{F}_t\right] \tag{45}$$

Similarly, we have that the arbitrage free futures price of a CDD contract at time $t\le\tau_1<\tau_2$ is given by:

$$F_{CDD}(t,\tau_1,\tau_2)=\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\max\left(T(\tau)-c,0\right)d\tau\,|\,\mathfrak{F}_t\right] \tag{46}$$

Observing equations (24), (45) and (46) we have the following proposition.

Proposition 6. The CDD, HDD and CAT prices are linked by the following relation:

$$F_{HDD}(t,\tau_1,\tau_2) = c(\tau_2-\tau_1) - F_{CAT}(t,\tau_1,\tau_2) + F_{CDD}(t,\tau_1,\tau_2) \tag{47}$$

Proof. We have that

$$\max(c-T(\tau),0) = c - T(\tau) + \max(T(\tau)-c,0)$$

Hence, by replacing the above relation to (45) we have that

$$\begin{aligned}
F_{HDD}(t,\tau_1,\tau_2) &= \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(c-T(\tau),0)\,d\tau \mid \mathfrak{F}_t\right] \\
&= \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \left(c-T(\tau)+\max(T(\tau)-c,0)\right)d\tau \mid \mathfrak{F}_t\right] \\
&= \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} c\,d\tau \mid \mathfrak{F}_t\right] - \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(\tau)d\tau \mid \mathfrak{F}_t\right] + \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(T(\tau)-c,0)d\tau \mid \mathfrak{F}_t\right] \\
&= c(\tau_2-\tau_1) - F_{CAT}(t,\tau_1,\tau_2) + F_{CDD}(t,\tau_1,\tau_2)
\end{aligned}$$

□

Proposition 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 7. The CDD future price for $0 \le t \le \tau_1 < \tau_2$ is given by

$$\begin{aligned}
F_{CDD}(t,\tau_1,\tau_2) &= \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(T(s)-c,0)\,ds \mid \mathfrak{F}_t\right] = \\
&= \int_{\tau_1}^{\tau_2} v(t,s)\Psi\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)\right)}{v(t,s)}\right)ds
\end{aligned} \tag{48}$$

where,

$$m(t,s,e^{\int_t^s \kappa(z)dz}T(t)) = S(s) + e^{\int_t^s \kappa(z)dz}T(t) + e^{\int_t^s \kappa(z)dz}\int_t^s \sigma(u)\theta(u)e^{-\int_t^u \kappa(z)dz}du - c \tag{49}$$

$$v^2(t,s) = e^{2\int_t^s \kappa(z)dz}\int_t^s \sigma^2(u)e^{-2\int_t^u \kappa(z)dz}du \tag{50}$$

and $\Psi(x) = x\Phi(x) + \Phi'(x)$ where Φ is the cumulative standard normal distribution function.

Proof. From equations (45) and (28) we have that:

$$F_{CDD}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(T(s)-c,0)\,ds \mid \mathfrak{F}_t\right]$$

and using Itô's Isometry we can interchange the expectation and the integral

$$\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max(T(s)-c,0)\,ds \mid \mathfrak{F}_t\right] = \int_{\tau_1}^{\tau_2} \mathrm{E}_Q\left[\max(T(s)-c,0) \mid \mathfrak{F}_t\right]ds$$

$T(s)$ is normally distributed under the probability measure Q with mean and variance given by:

$$E_Q\left[T(s)\,|\,\mathcal{F}_t\right] = S(s) + e^{\int_t^s \kappa(z)dz}T(t) + e^{\int_t^s \kappa(z)dz}\int_t^s \sigma(u)\theta(u)e^{-\int_t^u \kappa(z)dz}du$$

$$Var_Q\left[T(s)\,|\,\mathcal{F}_t\right] = e^{2\int_t^s \kappa(z)dz}\int_t^s \sigma^2(u)e^{-2\int_t^u \kappa(z)dz}du$$

Hence, $T(s)-c$ is normally distributed with mean given by $m(t,s,e^{\int_t^s \kappa(z)dz}T(t))$ and variance given by $v^2(t,s)$ and the proposition follows by standard calculations using the properties of the normal distribution.

□

Proposition 7 gives the price of a futures CDD at time $t \le \tau_1 < \tau_2$. In other words the price of a futures CDD before the contract period. Hence, (48) 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 8. The CDD future price for $\tau_1 \le t < \tau_2$ is given by

$$F_{CDD}(t,\tau_1,\tau_2) = \int_{\tau_1}^t \max\left(T(s)-c,0\right)ds + F_{CDD}(t,t,\tau_2) \tag{51}$$

Proof. We have that the futures price of a CDD is given by

$$\begin{aligned}
F_{CDD}(t,\tau_1,\tau_2) &= \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\max\left(T(s)-c,0\right)ds\,|\,\mathcal{F}_t\right] \\
&= \mathrm{E}_Q\left[\int_{\tau_1}^{t}\max\left(T(s)-c,0\right)ds + \int_{t}^{\tau_2}\max\left(T(s)-c,0\right)ds\,|\,\mathcal{F}_t\right] \\
&= \int_{\tau_1}^{t}\max\left(T(s)-c,0\right)ds + \mathrm{E}_Q\left[\int_{t}^{\tau_2}\max\left(T(s)-c,0\right)ds\,|\,\mathcal{F}_t\right] \\
&= \int_{\tau_1}^{t}\max\left(T(s)-c,0\right)ds + F_{CDD}(t,t,\tau_2)
\end{aligned}$$

Note that the first term is known at time t since it refers to past temperatures while the second term is stochastic.

□

Following the notation of (Benth, et al., 2007) and (Benth, et al., 2008) the dynamics of the CDD futures price under Q is given in the following proposition.

Proposition 9. The dynamics of $F_{CDD}(t,\tau_1,\tau_2)$ for $0 \le t \le \tau_1$ under Q is given by

$$dF_{CDD}(t,\tau_1,\tau_2) = \Sigma_{CDD}(t,\tau_1,\tau_2)dW(t) \tag{52}$$

where

$$\Sigma_{CDD}(t,\tau_1,\tau_2) = \sigma(t)\int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz}\Phi\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)\right)}{v(t,s)}\right)ds \tag{53}$$

and Φ is cumulative standard normal distribution function

Proof. $F_{CDD}(t,\tau_1,\tau_2)$ 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

$$m'\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)\right)=\frac{dm\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)\right)}{dT}=e^{\int_t^s \kappa(z)dz}$$

Also, substituting $\Psi'(x)=\Phi(x)$ we have that

$$\begin{aligned}\frac{dF_{CDD}}{dT}&=\int_{\tau_1}^{\tau_2}v(t,s)\Psi\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)\right)}{v(t,s)}\right)ds\\&=\int_{\tau_1}^{\tau_2}v(t,s)\Psi'\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)\right)}{v(t,s)}\right)\frac{m'\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)\right)v(t,s)}{v^2(t,s)}ds\\&=\int_{\tau_1}^{\tau_2}e^{\int_t^s \kappa(z)dz}\Phi\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)\right)}{v(t,s)}\right)ds\end{aligned}$$

Hence, we have that

$$dF_{CDD}(t,\tau_1,\tau_2)=\sigma(t)\int_{\tau_1}^{\tau_2}e^{\int_t^s \kappa(z)dz}\Phi\left(\frac{m\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)\right)}{v(t,s)}\right)dsdW(t)$$

□

In Proposition 9 the term $\Sigma_{CDD}(t,\tau_1,\tau_2)$ represents the term structure of the volatility of CDD futures. Hence, the price of a call option on a CDD futures can be derived. From Proposition 9 the price of a CDD futures option can be estimated.

Proposition 10. The price at time $t\le\tau$ of a call option written on a HDD futures with strike price K at exercise time $\tau\le\tau_1$ is

$$C_{CDD}(t,\tau,\tau_1,\tau_2)=e^{-r(\tau-t)}E_Q\left[\max\left(\int_{\tau_1}^{\tau_2}v(t,s)Z\left(t,s,\tau,T(t)\right)ds-K,0\right)\right] \qquad (54)$$

where

$$Z\left(t,s,\tau,T(t)\right)=\Psi\left(t,s,e^{\int_t^s \kappa(z)dz}T(t)+\int_t^\tau\sigma(u)\theta(u)e^{\int_u^s \kappa(z)dz}du+\Sigma(s,t,\tau)Y\right) \qquad (55)$$

and

$$\Psi(t,s,x)=\Psi\left(\frac{m(t,s,x)}{v(t,s)}\right) \tag{56}$$

and

$$\Sigma^2(s,t,\tau)=\int_t^\tau \sigma^2(u)e^{2\int_u^s \kappa(z)dz}du \tag{57}$$

and Y is a standard normal random variable.

Proof. The option price is given as

$$C_{CDD}(t,\tau,\tau_1,\tau_2)=e^{-r(\tau-t)}E_Q\left[\max\left(F_{CDD}(\tau,\tau_1,\tau_2)-K,0\right)|\,\mathfrak{F}_t\right]$$

we have that

$$\begin{aligned}F_{CDD}(\tau,\tau_1,\tau_2)&=\int_{\tau_1}^{\tau_2}v(t,s)\Psi\left(t,s,e^{\int_\tau^s\kappa(z)dz}T(\tau)\right)ds\\&=\int_{\tau_1}^{\tau_2}v(t,s)\Psi\left(t,s,e^{\int_t^s\kappa(z)dz}T(t)+\int_t^\tau\sigma(u)\theta(u)e^{\int_u^s\kappa(z)dz}du\right.\\&\quad\left.+\int_t^\tau\sigma(u)e^{\int_u^s\kappa(z)dz}dW(u)\right)ds\end{aligned}$$

The Itô integral inside the expectation is independent of $\mathfrak{F}_t$ 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.

□

8.3 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 (1). 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:

$$dT(t)=dS(t)+\kappa(t)\big(T(t)-S(t)\big)dt+\sigma(t)dL(t) \tag{58}$$

where $L(t)$ is Lévy noise. Applying the Itô formula for semimartingales, (Ikeda & Watanabe, 1981), the explicit solution of (58) is obtained:

$$T(t)=S(t)+e^{\int_0^t\kappa(u)du}\big(T(0)-S(0)\big)+e^{\int_0^t\kappa(u)du}\int_0^t\sigma(t)e^{-\int_0^s\kappa(u)du}dL(t) \tag{59}$$

As in the case of the BM we derive the price of a CAT futures to be

$$F_{CAT}(t,\tau_1,\tau_2)=\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\mathrm{T}(\tau)d\tau\,|\,\mathfrak{F}_t\right]$$

Proposition 11. The cumulative temperature over the time interval $[\tau_1,\tau_2]$ is given by:

$$\int_{\tau_1}^{\tau_2} T(t)dt = \int_{\tau_1}^{\tau_2} S(t)dt + \int_{\tau_1}^{\tau_2} e^{\int_0^t \kappa(z)dz}\left(T(0)-S(0)\right)dt + \int_0^{\tau_1}\int_{\tau_1}^{\tau_2} \sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t) + \int_{\tau_1}^{\tau_2}\int_t^{\tau_2} \sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t) \tag{60}$$

Proof. We have that

$$\int_{\tau_1}^{\tau_2} T(t)dt = \int_{\tau_1}^{\tau_2} S(t)dt + \int_{\tau_1}^{\tau_2} e^{\int_0^t \kappa(z)dz}\left(T(0)-S(0)\right)dt + \int_{\tau_1}^{\tau_2}\int_0^t \sigma(t)e^{\int_s^t \kappa(z)dz}dL(t)dt$$

Focusing on the last integral we have that

$$\int_{\tau_1}^{\tau_2}\int_0^t \sigma(t)e^{\int_s^t \kappa(z)dz}dL(t)dt = \int_{\tau_1}^{\tau_2}\int_0^{\tau_2} 1_{[0,t]}\sigma(t)e^{\int_s^t \kappa(z)dz}dL(t)dt$$

$$= \int_0^{\tau_2}\int_{\tau_1}^{\tau_2} 1_{[0,t]}\sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t)$$

$$= \int_0^{\tau_1}\int_{\tau_1}^{\tau_2} 1_{[0,t]}\sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t) + \int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2} 1_{[0,t]}\sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t)$$

$$= \int_0^{\tau_1}\int_{\tau_1}^{\tau_2} \sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t) + \int_{\tau_1}^{\tau_2}\int_t^{\tau_2} \sigma(t)e^{\int_s^t \kappa(z)dz}dtdL(t)$$

□

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

$$Z(t) = \exp\left(\int_0^t \theta(s)L(s) - \int_0^t \varphi\left(\theta(s)\right)ds\right) \tag{61}$$

where $\varphi(\lambda)$ is the logarithm of the moment generating function of $L(t)$

$$\varphi(\lambda) = \ln \mathrm{E}\left[\exp\left(\lambda L(1)\right)\right] \tag{62}$$

We make the same assumptions as in (Benth & 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 12. The futures prices $F_{CAT}(t,\tau_1,\tau_2)$ at time $t \le \tau_1 < \tau_2$ written on CAT over the interval $[\tau_1,\tau_2]$ is

$$F_{CAT}(t,\tau_1,\tau_2) = \int_{\tau_1}^{\tau_2} S(t)dt + \int_{\tau_1}^{\tau_2} e^{\int_0^t \kappa(z)dz}\left(T(0)-S(0)\right)dt + \int_0^t\int_{\tau_1}^{\tau_2} \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u) + \int_t^{\tau_2}\int_u^{\tau_2} \sigma(u)e^{\int_s^u \kappa(z)dz}du\varphi'\left(\theta(u)\right)du - \int_t^{\tau_1}\int_u^{\tau_1} \sigma(u)e^{\int_s^u \kappa(z)dz}du\varphi'\left(\theta(u)\right)du \tag{63}$$

Proof. First we prove that for a real-valued measurable and bounded function $f(t)$

$$\mathrm{E}_Q\left[\int_t^\tau f(u)dL(u)\,|\,\mathfrak{F}_t\right]=\int_t^\tau f(u)\varphi'\big(\theta(u)\big)du \tag{64}$$

The proof of (64) 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 & Saltyte-Benth, 2005). First note the following lemma:

$$E\left[\exp\left\{\int_s^t g(u)dL(u)\right\}\right]=\exp\left\{\int_s^t \varphi\big(g(u)\big)du\right\} \tag{65}$$

if $g:[s,t]\to$ 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 & Saltyte-Benth, 2004). Hence, we have that:

$$\begin{aligned}
\mathrm{E}_Q\left[\int_t^\tau f(u)dL(u)\,|\,\mathfrak{F}_t\right]&=\mathrm{E}_Q\left[\int_t^\tau f(u)dL(u)\frac{Z(\tau)}{Z(t)}\right]\\
&=\exp\left(-\int_t^\tau \varphi\big(\theta(u)\big)du\right)\frac{d}{d\lambda}\mathrm{E}_Q\left[\exp\left(\int_t^\tau \lambda f(u)+\theta(u)dL(u)\right)\right]_{\lambda=0}\\
&=\exp\left(-\int_t^\tau \varphi\big(\theta(u)\big)du\right)\frac{d}{d\lambda}\exp\left(\int_t^\tau \varphi\big(\lambda f(u)+\theta(u)\big)du\right)\\
&=\int_t^\tau f(u)\varphi'\big(\theta(u)\big)du
\end{aligned}$$

Hence, (64) holds.

Next, the dynamics of the price of the future CAT

$$\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds\,|\mathfrak{F}_t\right]=\mathrm{E}_Q\left[\int_t^{\tau_2} T(s)ds\,|\mathfrak{F}_t\right]-\mathrm{E}_Q\left[\int_t^{\tau_1} T(s)ds\,|\mathfrak{F}_t\right]$$

From equation (60) and the adaptivity property of the Lévy process we have that

$$\begin{aligned}
\mathrm{E}_Q\left[\int_t^\tau T(s)ds\,|\,\mathfrak{F}_t\right]&=\int_t^\tau S(u)du+\int_t^\tau e^{\int_0^u \kappa(z)dz}\big(T(0)-S(0)\big)du\\
&\quad+\mathrm{E}_Q\left[\int_0^t\int_t^\tau \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u)+\int_t^\tau\int_u^\tau \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u)\,|\,\mathfrak{F}_t\right]\\
&=\int_t^\tau S(u)du+\int_t^\tau e^{\int_0^u \kappa(z)dz}\big(T(0)-S(0)\big)du\\
&\quad+\mathrm{E}_Q\left[\int_0^t\int_t^\tau \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u)\,|\,\mathfrak{F}_t\right]\\
&\quad+\mathrm{E}_Q\left[\int_t^\tau\int_u^\tau \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u)\,|\,\mathfrak{F}_t\right]
\end{aligned}$$

Hence, using the adaptivity property again and equation (64) we have that

$$\mathrm{E}_Q\left[\int_t^\tau T(s)ds \mid \mathfrak{F}_t\right] = \int_t^\tau S(u)du + \int_t^\tau e^{\int_0^u \kappa(z)dz}\big(T(0)-S(0)\big)du$$
$$+ \int_0^t\int_t^\tau \sigma(u)e^{\int_s^u \kappa(z)dz}dudL(u)$$
$$+ \int_t^\tau\int_u^\tau \sigma(u)e^{\int_s^u \kappa(z)dz}du\varphi'\big(\theta(u)\big)du$$

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 $t \le \tau_1 \le \tau_2$ is given by:

$$e^{-r(\tau_2-t)}\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2}\frac{1}{\tau_2-\tau_1}\mathrm{T}(\tau)d\tau - F_{PAC}(t,\tau_1,\tau_2) \mid \mathfrak{F}_t\right] = 0$$

and since F_{PAC} is $\mathfrak{F}_t$ adapted we derive the price of a PAC futures to be

$$F_{PAC}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\frac{1}{\tau_2-\tau_1}\int_{\tau_1}^{\tau_2}T(s)ds \mid \mathfrak{F}_t\right] \tag{66}$$

Hence we conclude that:

$$F_{PAC}(t,\tau_1,\tau_2) = \frac{1}{\tau_2-\tau_1}F_{CAT}(t,\tau_1,\tau_2) \tag{67}$$

Unfortunately, introducing the Lévy noise process prevents the calculation of option prices. In addition, finding closed form solutions for AccHDDs and AccCDDs 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 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 $\psi_T(\lambda)$

$$f_T(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-isx}\psi_T(s)ds \tag{68}$$

Hence, if the characteristics function of the Lévy process is known then option prices as well futures on AccHDDs and AccCDDs can be estimated. This approach is analytically discussed in (Carr & Madan, 1999).

Proposition 13. The characteristic function of $T(t)$ under the risk neutral measure Q is given by:

$$\psi_T(\lambda) = \mathrm{E}_Q\left[\exp\{i\lambda T(t)\} \mid \mathfrak{F}_t\right] = \exp\{\Psi(\lambda)\} \tag{69}$$

where

$$\Psi(\lambda) = i\lambda S(t) + i\lambda e^{\int_s^t \kappa(z)dz}\left(T(s) - S(s)\right) - \int_s^t \varphi\left(\theta(u)\right)du + \int_s^t \varphi\left(i\lambda\sigma(u)e^{\int_u^t \kappa(z)dz} + \theta(u)\right)du \tag{70}$$

Proof. We have that:

$$\begin{aligned}
&\mathrm{E}_Q\left[\exp\left\{i\lambda T(t)\right\} \mid \mathcal{F}_t\right] = \\
&= \mathrm{E}_Q\left[\exp\left\{i\lambda S(t) + i\lambda e^{\int_s^t \kappa(z)dz}\left(T(s) - S(s)\right) + i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz}dL(u)\right\} \mid \mathcal{F}_t\right] \\
&= \exp\left\{i\lambda S(t) + i\lambda e^{\int_s^t \kappa(z)dz}\left(T(s) - S(s)\right)\right\}\mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz}dL(u)\right\} \mid \mathcal{F}_t\right]
\end{aligned} \tag{71}$$

Focusing on the expectation we have that:

$$\begin{aligned}
&\mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz}dL(u)\right\} \mid \mathcal{F}_t\right] \\
&= \mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz}dL(u)\right\}\frac{Z(t)}{Z(s)} \mid \mathcal{F}_t\right] \\
&= \mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz}dL(u) + \int_s^t \theta(u)dL(u) - \int_s^t \varphi\left(\theta(u)\right)du\right\} \mid \mathcal{F}_t\right] \\
&= \exp\left\{-\int_s^t \varphi\left(\theta(u)\right)du\right\}\mathrm{E}_Q\left[\exp\left\{i\lambda\int_s^t \sigma(u)e^{\int_u^t \kappa(z)dz} + \theta(u)dL(u)\right\}\right] \\
&= \exp\left\{-\int_s^t \varphi\left(\theta(u)\right)du\right\}\exp\left\{\int_s^t \varphi\left(i\lambda\sigma(u)e^{\int_u^t \kappa(z)dz} + \theta(u)\right)du\right\}
\end{aligned} \tag{72}$$

From (69), (71) and (72) yields the result

$$\Psi(\lambda) = i\lambda S(t) + i\lambda e^{\int_s^t \kappa(z)dz}\left(T(s) - S(s)\right) - \int_s^t \varphi\left(\theta(u)\right)du + \int_s^t \varphi\left(i\lambda\sigma(u)e^{\int_u^t \kappa(z)dz} + \theta(u)\right)du$$

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:

$$F_{CDD}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \max\left(T(s)-c\right) ds \mid \mathcal{F}_t\right]$$
$$= \int_{\tau_1}^{\tau_2} \mathrm{E}_Q\left[\max\left(T(s)-c\right) \mid \mathcal{F}_t\right] ds \qquad (73)$$
$$= \int_{\tau_1}^{\tau_2}\int_{c}^{+\infty} \left(x-c\right) f_T(x) dx ds$$

where $f_T(x)$ is the density function of $T(t)$ under the risk neutral measure Q conditional on $\mathcal{F}_t$ and it is given by (68). Similarly, the HDD future price is given by

$$F_{CDD}(t,\tau_1,\tau_2) = \int_{\tau_1}^{\tau_2}\int_{0}^{65} \left(c-x\right) f_T(x) dx ds \qquad (74)$$

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.

8.4 Market price of risk

The weather derivatives market is a classical incomplete market. Since, temperature is non-tradable the market price of risk must be incorporated in the pricing model. 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, contradicting the assumption of (Hull, 2003).

(Turvey, 2005) proposed to estimate the market price of risk by using the capital asset pricing model. (Cao & 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, Hardle, & Lopez Cabrera, 2009) and (Hardle & 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 futures 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 with the number of available for trading future contracts

In (Hardle & 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 on can infer the market price of risk for regions where weather derivative market does not exist.

Unfortunately, we do not possess any market data like futures or option prices. Hence, we cannot proceed on estimating and analyzing the market price of risk. Once, market data is available then the market price of risk can be easily estimated using the approach described in (Alaton, et al., 2002).

9. Conclusions

In this study, several temperature time series were studied in order to develop a model that describes the temperature evolution in the context of weather derivative pricing. 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. First, a novel approach for estimating the number of lags of the nonlinear AR model was applied. Then, by computing the derivative $dT(t+1)/dT(t)$ 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 category of models consists of 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.

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.

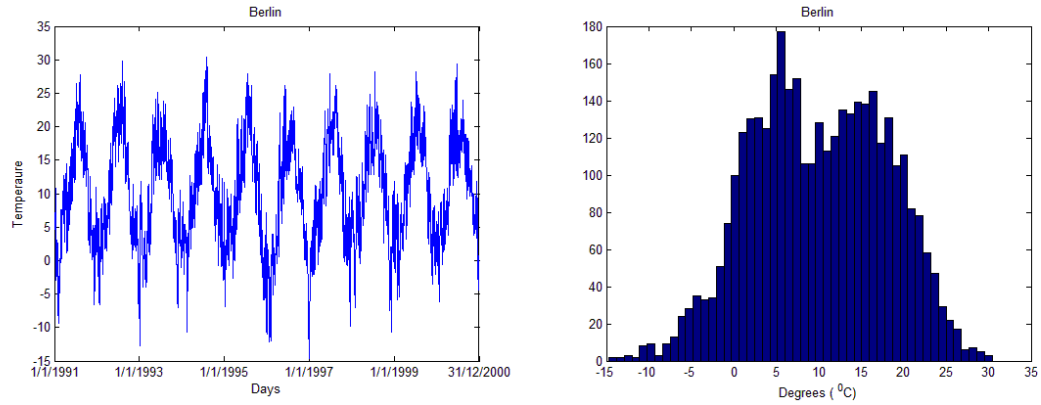
In order to obtain a better understanding of the distributions of the residuals we expanded our analysis by fitting additional distributions. Of the four distributions (normal, hyperbolic, NIG, stable) the hyperbolic distribution provides a slightly better fit than the normal distribution. However, introducing

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.

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, AccHDDs, AccCDDs and the Pacific Rim. The appealing properties of the normal distributions allows for derivation of pricing formulas in both futures and options on the above indices.

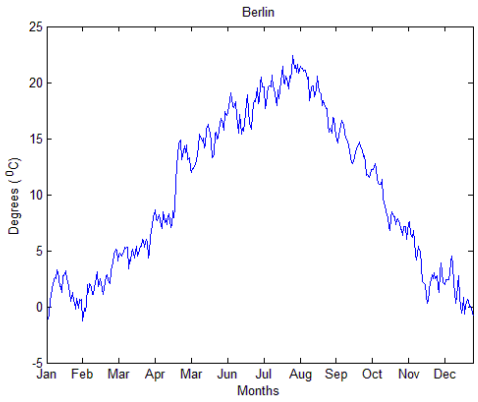
Then, based on our results that the hyperbolic distribution provides a better fit to the residuals, a Lévy motion noise process was assumed. In this case the pricing formulas for the CAT and Pacific Rims futures were presented. We provided a representation of the characteristic function of the temperature dynamics under the risk-neutral probability measure which is crucial for finding the density function necessary for pricing options and futures on AccHDDs and AccCDDs.

Finally, the importance of the market price of risk was discussed and an estimation method was presented.

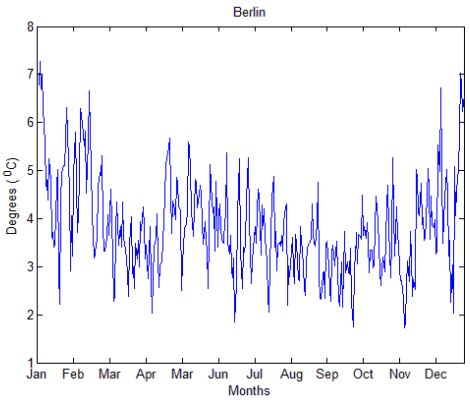


(a)

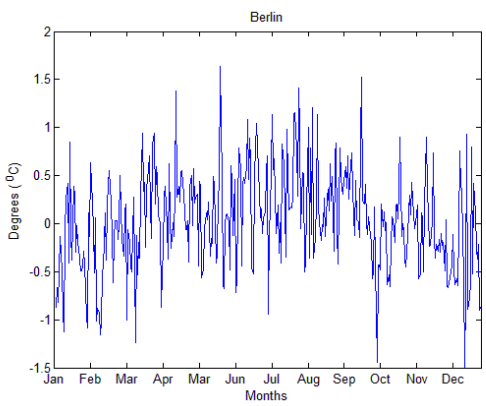
(b)



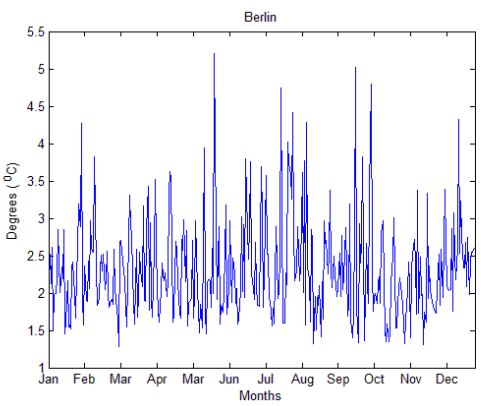
(c)



(d)



(e)



(f)

Figure 1. The (a) daily average temperature, (b), the empirical distribution, (c), the mean, (d), the standard deviation, (e), the skewness and, (e), the kurtosis of the DAT in Berlin.

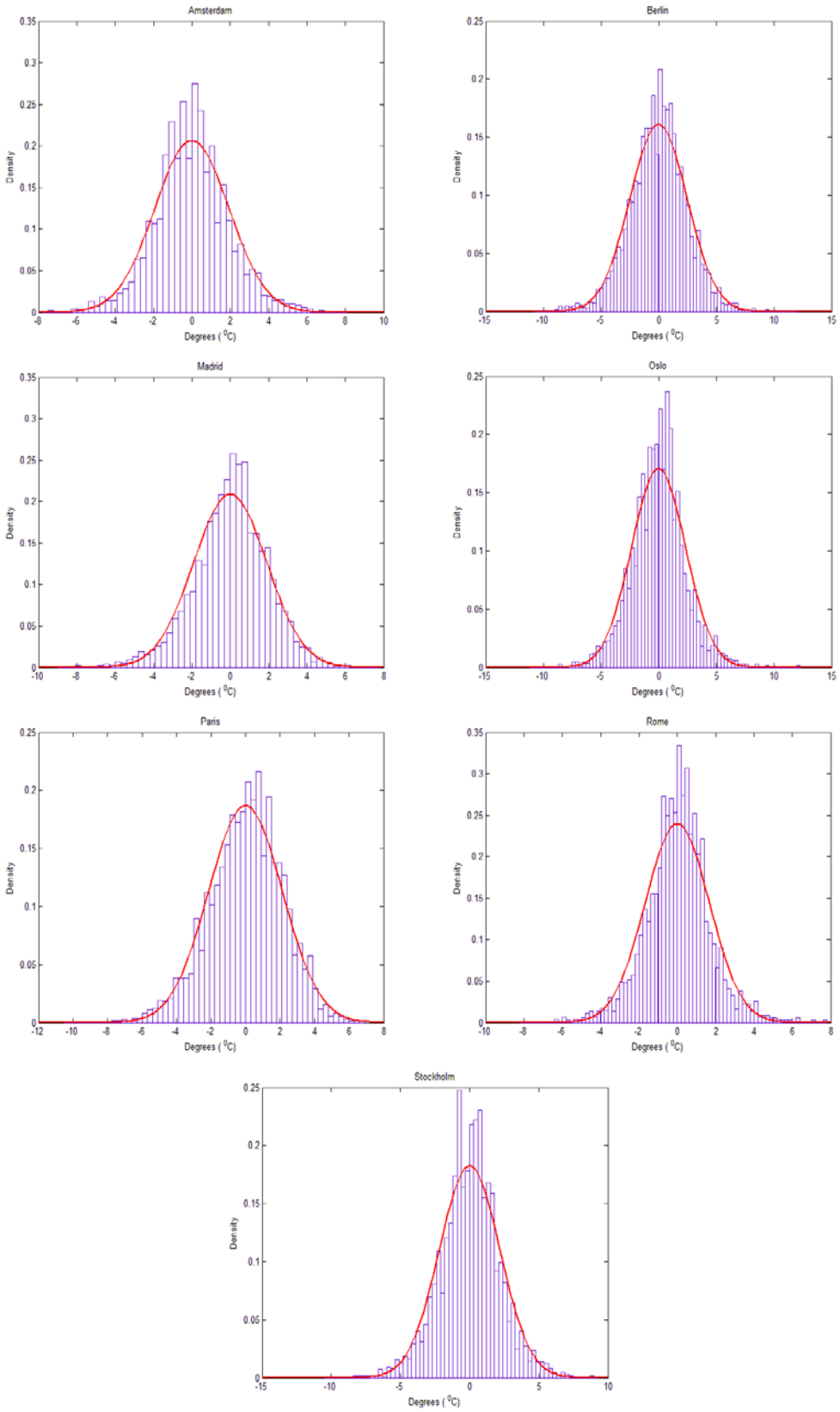


Figure 2. Empirical and normal distribution (solid line) of the first difference of the daily average temperature of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

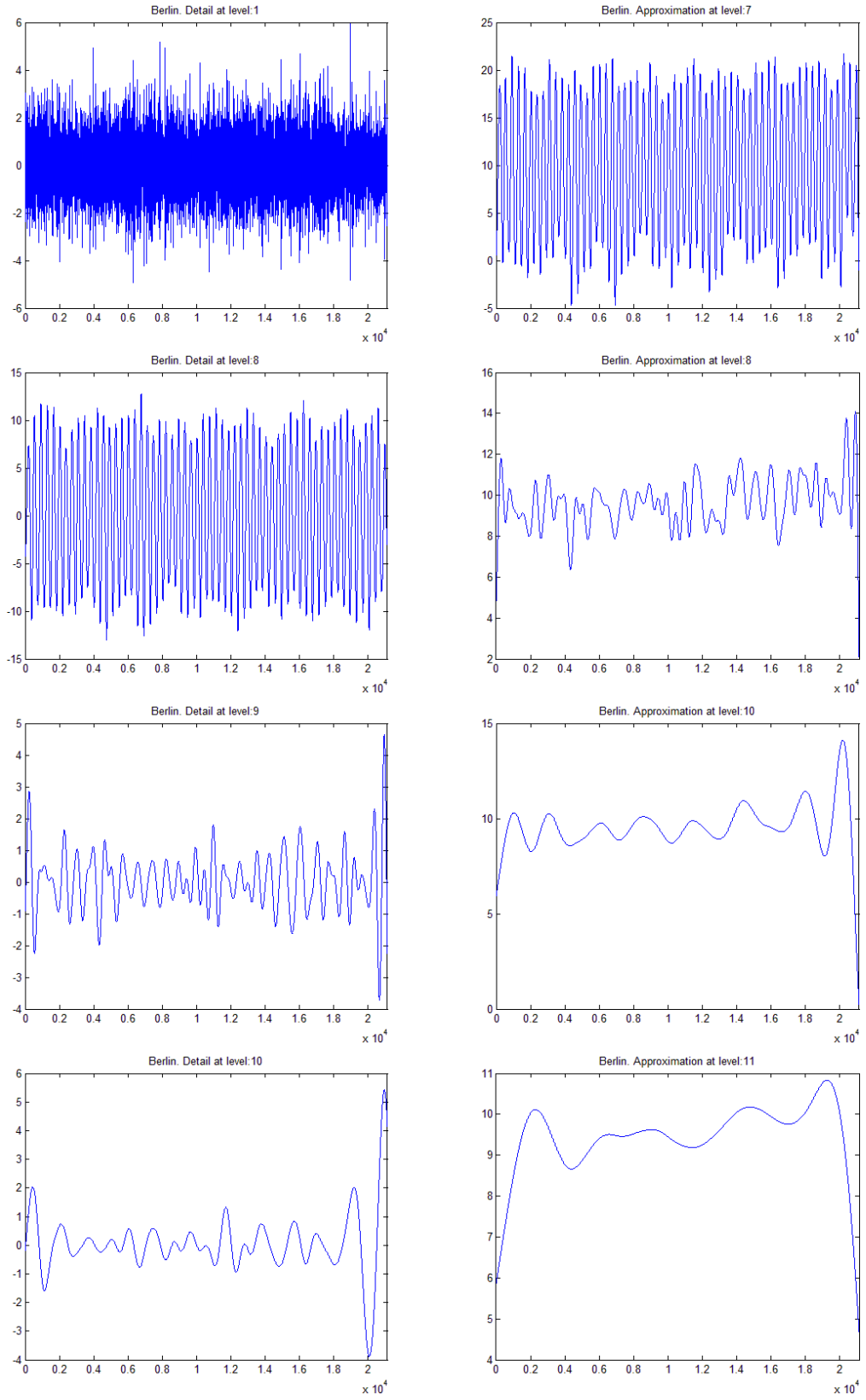


Figure 3. Selected parts of the discrete wavelet decomposition in Berlin; approximations (a_j) and details (d_j). The Daubechies 11 at level 11 wavelet was applied

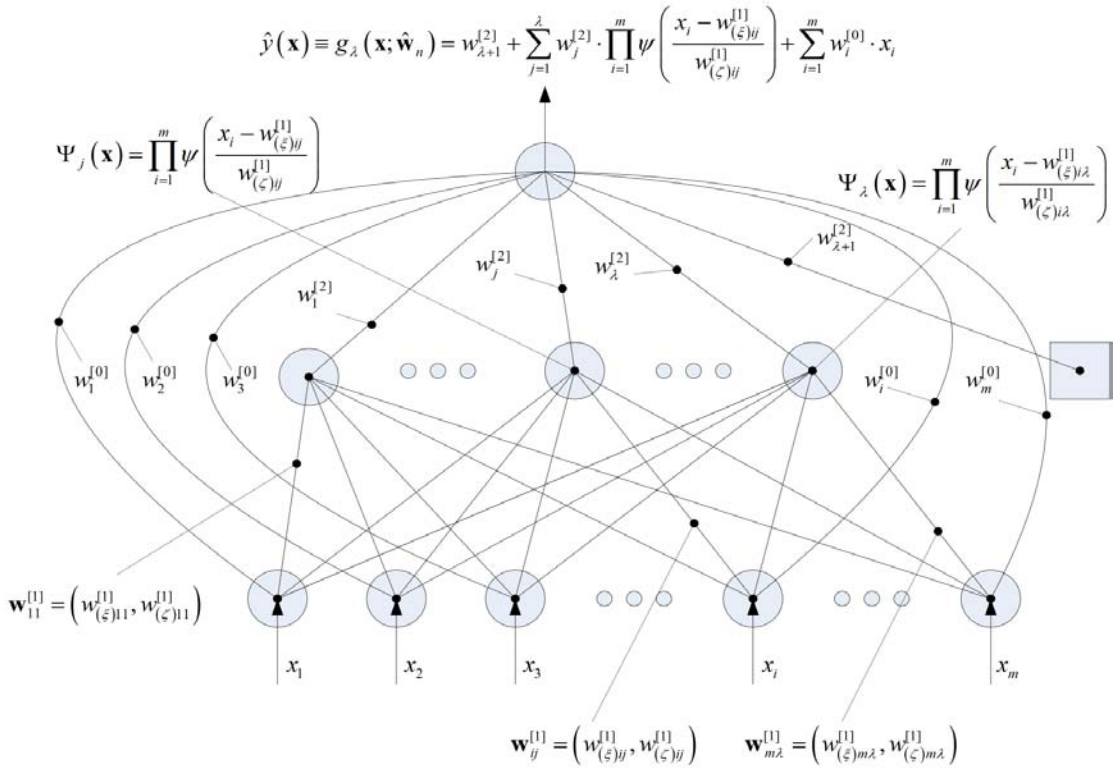


Figure 4. Structure of a Wavelet Network

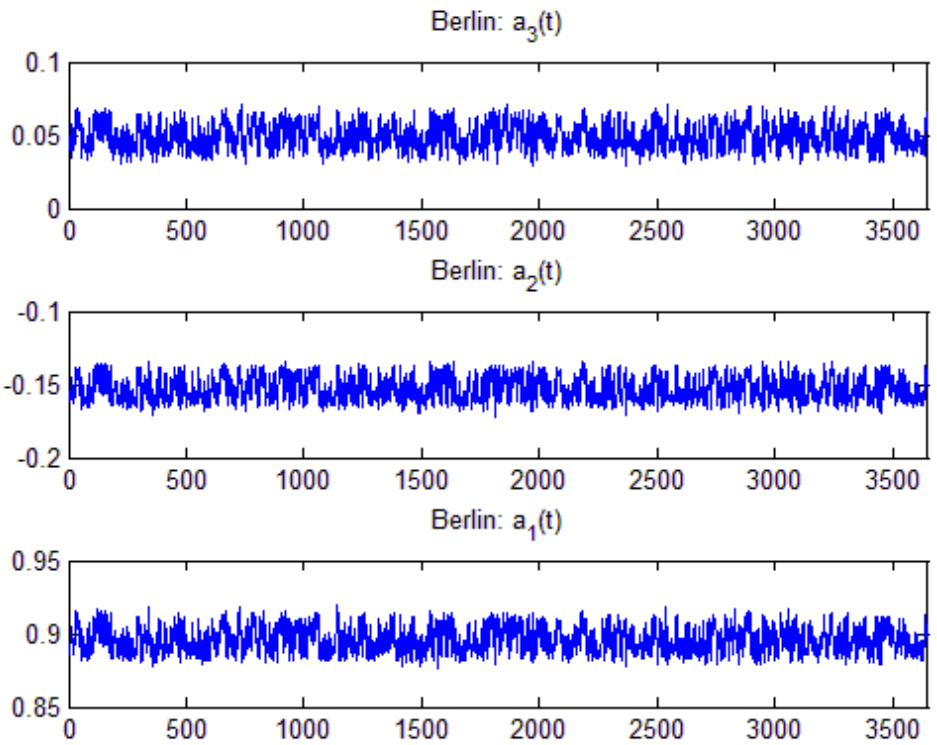


Figure 5. Daily variation of the speed of mean reversion functions a_i in Berlin

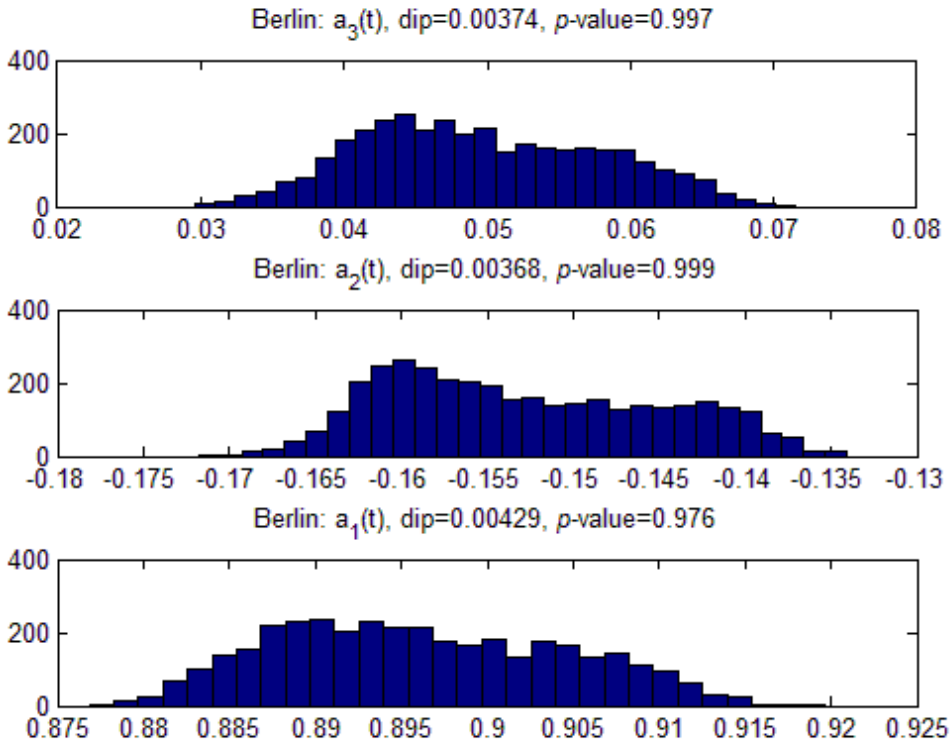


Figure 6. Frequency distribution of the speed of mean reversion function a_i in Berlin

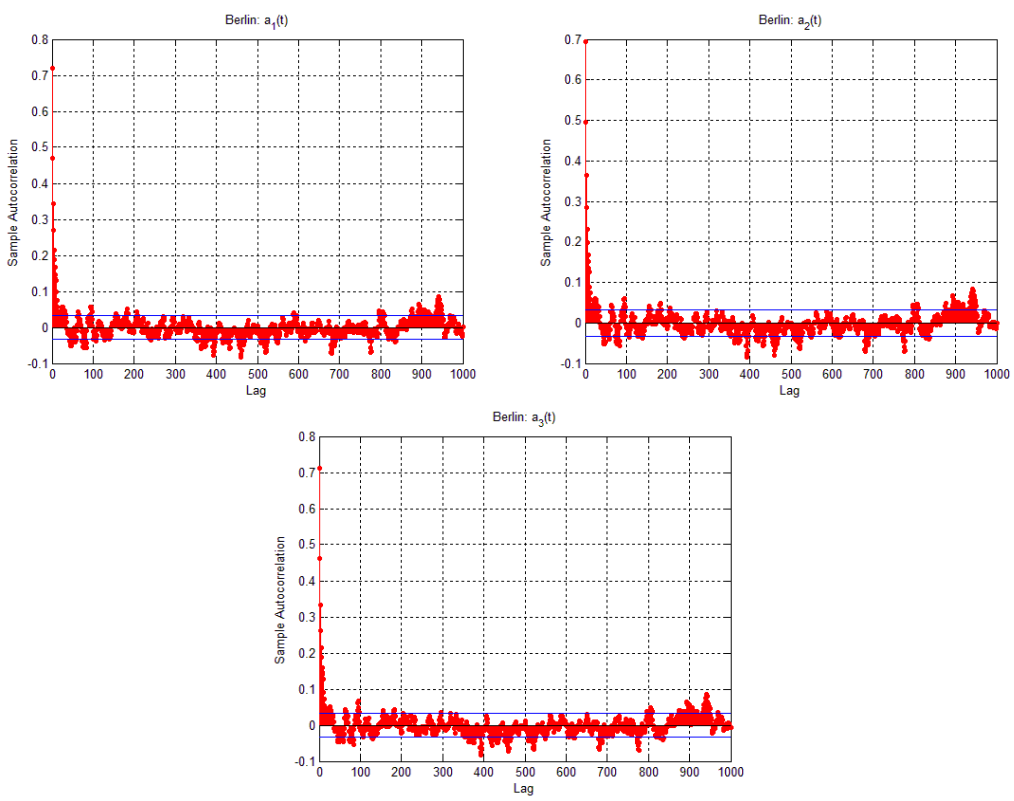


Figure 7. Autocorrelation function of the speed of mean reversion functions of the nonlinear AR model, $a_i(t)$, in Berlin

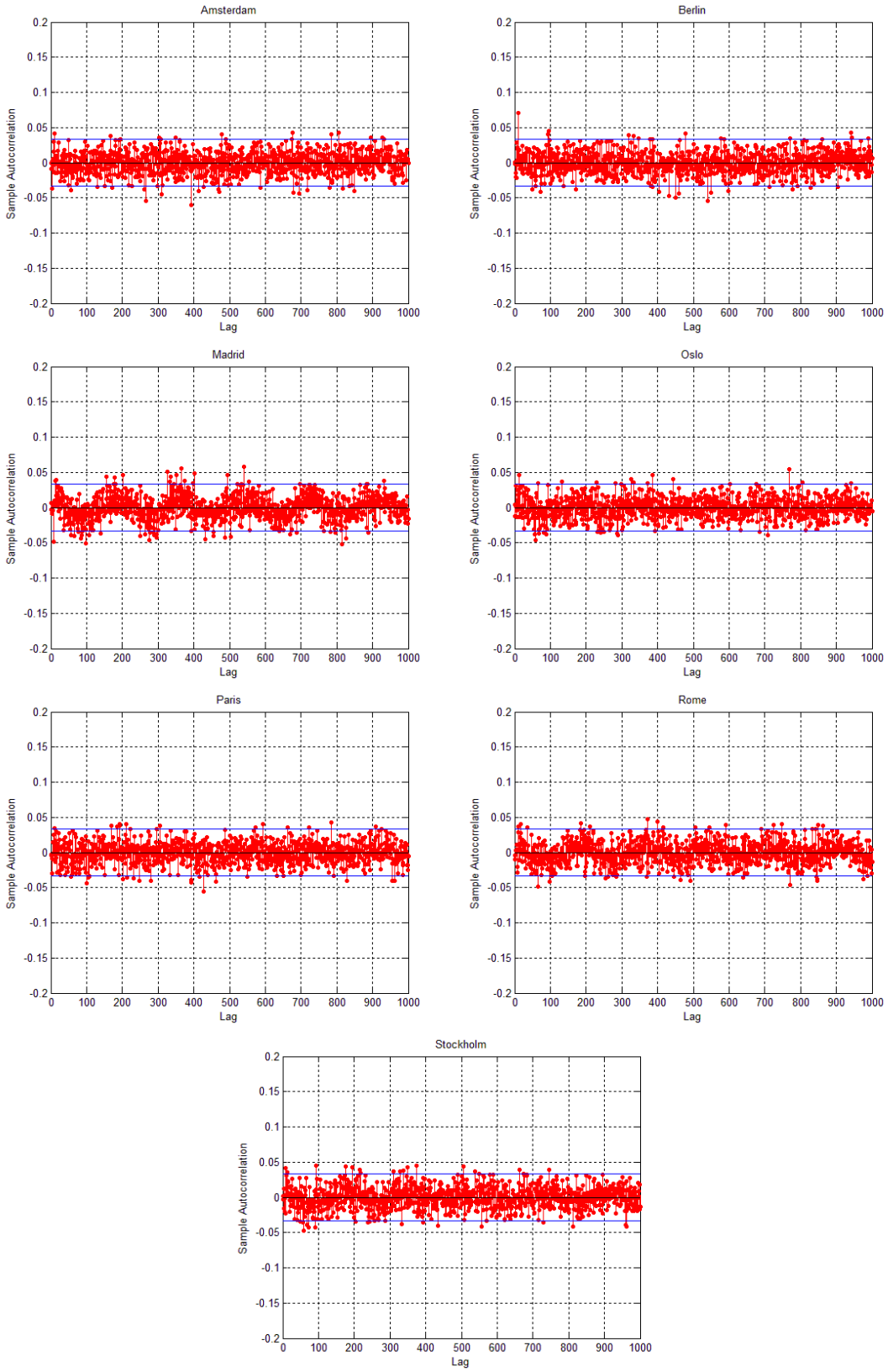


Figure 8. Autocorrelation function of the residuals of the wavelet network of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

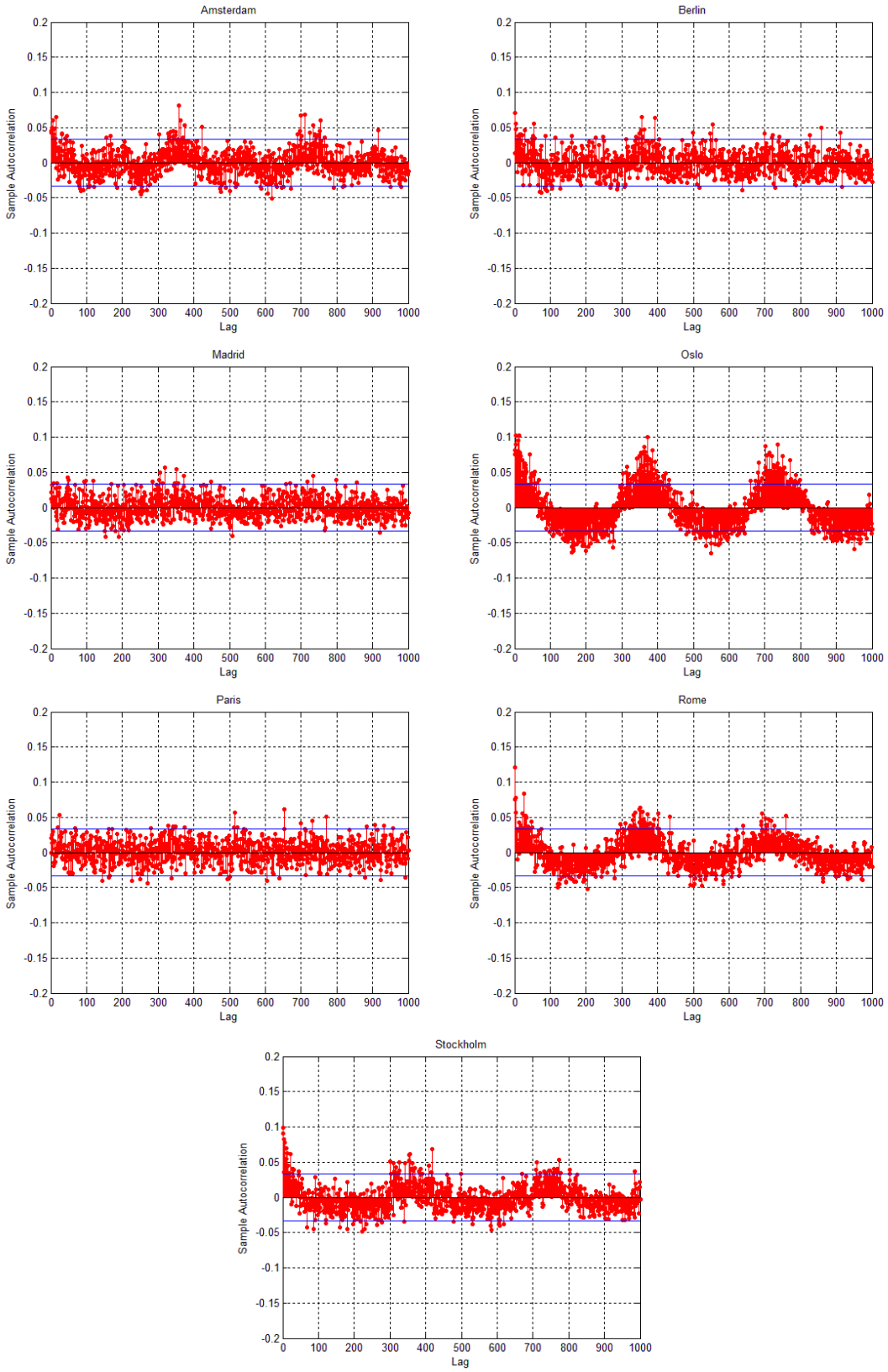


Figure 9. Autocorrelation function of the squared residuals of the wavelet network of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

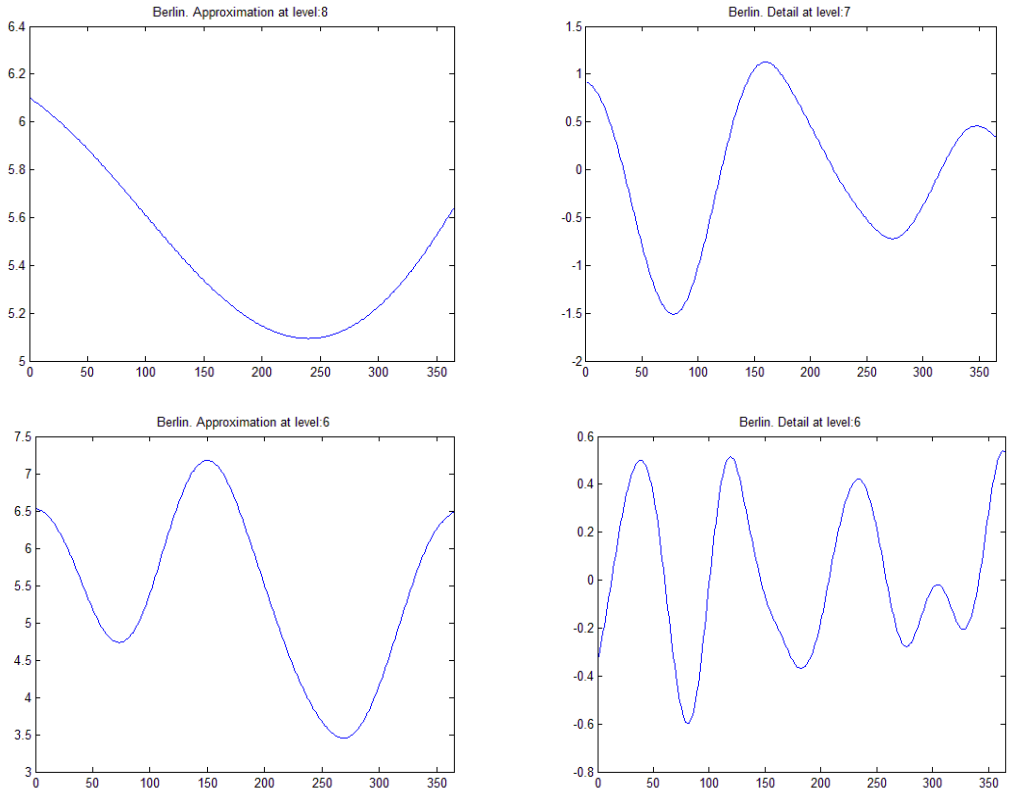


Figure 10. Selected parts of the discrete wavelet decomposition of the seasonal variance in Berlin; approximations (a_j) and details (d_j). The Daubechies 8 wavelet at level 8 was used

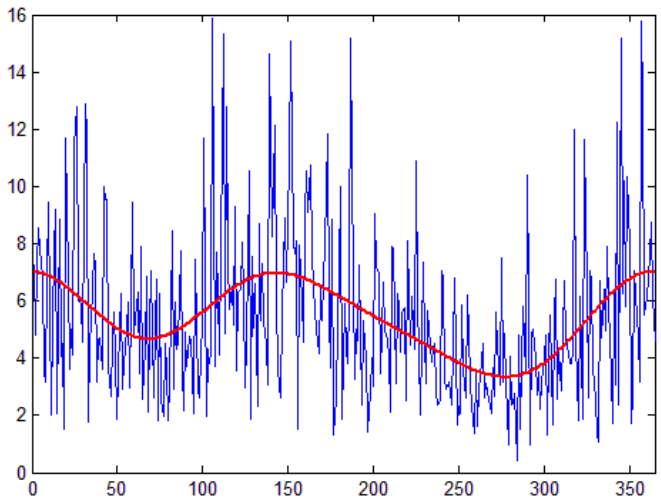


Figure 11. Empirical and fitted variance in Berlin

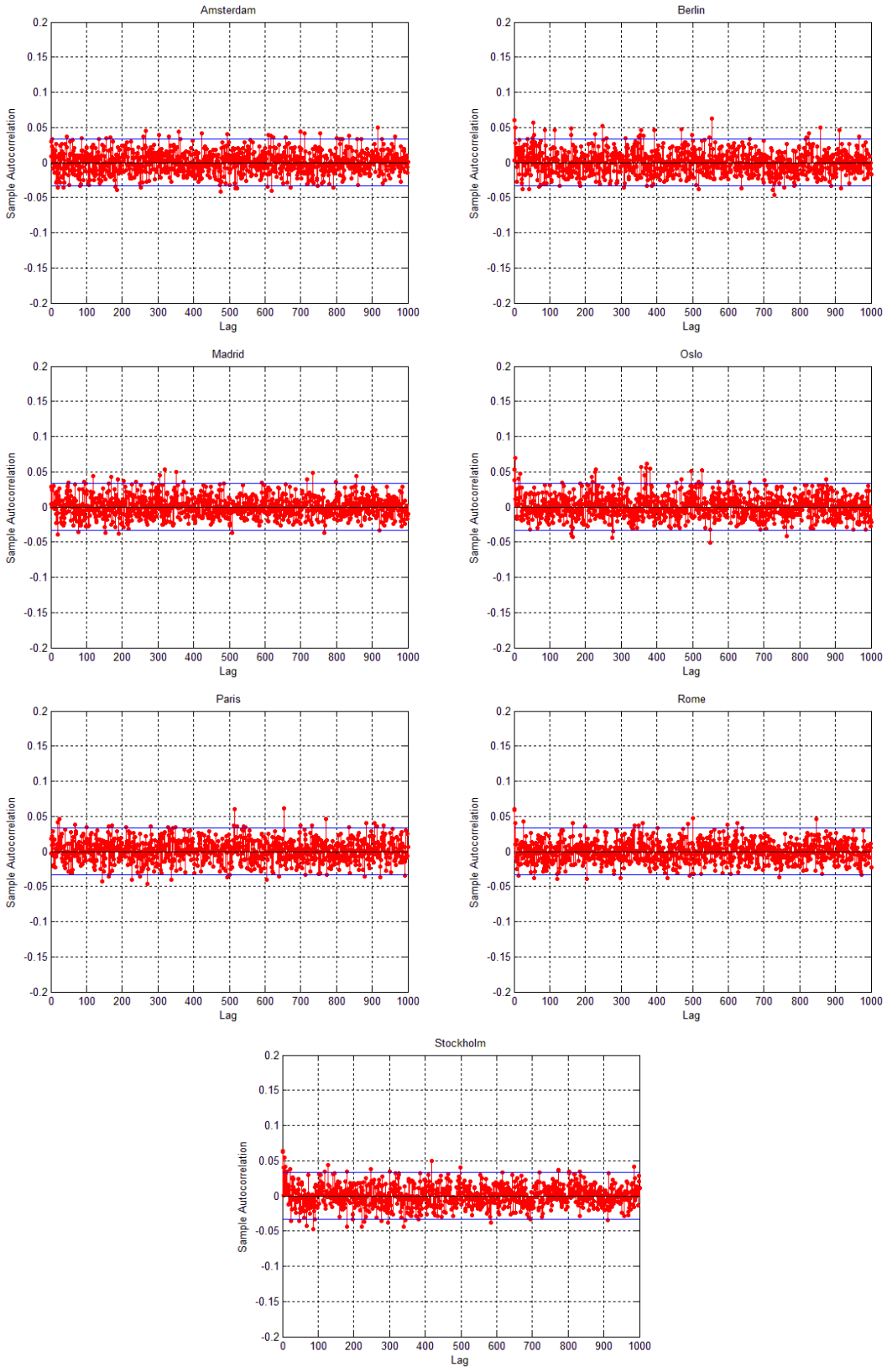


Figure 12. Autocorrelation function of the squared residuals after dividing out the volatility function of the seven cities: Amsterdam, Berlin, Madrid, Oslo, Paris, Rome, Stockholm

Table 1. 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 intercept the and *b* is the slope

Table 2. 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
..	-	-	-	-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.

Table 3. 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	$\overline{R}^2$
	-	-	7	5 (83)	43.9	1.5928	3.2004	63.98%
1	X_6	-	6	2 (33)	110.4	1.5922	3.1812	64.40%
2	X_7	-	5	1 (17)	214.3	1.5927	3.1902	64.59%
3	X_5	-	4	1 (14)	260.2	1.6004	3.2056	64.61%
4	X_4	-	3	1 (11)	331.2	1.5969	3.1914	64.61%

The algorithm concluded in 4 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

Table 4. 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 are 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

Table 5. Descriptive statistics of the mean reverting functions

Amsterdam	Mean	St.Dev	Max	Median	Min	Skewness	Kurtosis	KS	*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
$a_3(t)$	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 6. 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.

Table 7. 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 8. 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 9. 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 Alton mode. κ is the speed of mean reversion, A is the intercept and 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.

Table 10. 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 11. 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.

Table 12. 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

Table 13. 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.

Table 14. Out-of sample performance of the proposed, the Alaton and Benth models

	1-day-ahead	Out-of-sample	Total
Proposed	44 (66%)	35 (50%)	81 (58%)
Alaton	14 (20%)	21 (30%)	35 (25%)
Benth	10 (14%)	14 (20%)	24 (17%)

References

Aczel, A. D. (1993). *Complete business statistics* (2nd ed.). Homewood, IL: Irwin.

Alaton, P., Djehince, B., & Stillberg, D. (2002). On Modelling and Pricing Weather Derivatives. *Applied Mathematical Finance, 9*, 1-20.

Alexandridis, A. (2010). *Modelling and Pricing Temperature Derivatives Using Wavelet Networks and Wavelet Analysis.* Unpublished Ph.D. Thesis, University of Macedonia, Thessaloniki, Greece.

Alexandridis, A., & Zapranis, A. (2011). Wavelet Neural Networks: A Practical Guide. SSRN.

Bellini, F. (2005). *The weather derivatives market: Modelling and pricing temperature.* Unpublished Ph.D Thesis, University of Lugano, Lugano, Switzerland.

Benth, F. E. (2003). On arbitrage-free pricing of weather derivatives based on fractional Brownian motion. *Applied Mathematical Finance, 10*, 303-324.

Benth, F. E., Hardle, W. K., & Lopez Cabrera, B. (2009). Pricing of Asian temperature risk, *SFB649 Working Paper*. Berlin: Humboldt-Universitat zu Berlin.

Benth, F. E., & Saltyte-Benth, J. (2004). The normal inverse Gaussian distribution and spot price modeling in energy markets. *International Journal of Theoretical and Applied Finance, 7*(2), 177-192.

Benth, F. E., & Saltyte-Benth, J. (2005). Stochastic Modelling of Temperature Variations With a View Towards Weather Derivatives. *Applied Mathematical Finance, 12*(1), 53-85.
Benth, F. E., & Saltyte-Benth, J. (2007). The volatility of temperature and pricing of weather derivatives. *Quantitative Finance, 7*(5), 553-561.
Benth, F. E., Saltyte-Benth, J., & Koekebakker, S. (2007). Putting a price on temperature. *Scandinavian Journal of Statistics,, 34*, 746-767.
Benth, F. E., Saltyte-Benth, J., & Koekebakker, S. (2008). *Stochastic Modelling of Electricity and Related Markets*. Singapore: World Scientific.
Bhowan, A. (2003). *Temperature Derivatives.* University of Wiwatersand.
Brockwell, P. J., & Marquardt, T. (2005). Levy-driven and fractionality integrated ARMA process with continuous time parameter. *Statistica Sinica, 15*, 477-494.
Brody, C. D., Syroka, J., & Zervos, M. (2002). Dynamical Pricing of Weather Derivatives. *Quantitave 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. *Working Paper* Retrieved July, 2002, from http://ssrn.com/abstract=405800
Caballero, R., Jewson, S., & Brix, A. (2002). Long Memory in Surface Air Temperature: Detection Modelling and Application to Weather Derivative Valuation. *Climate Research, 21*, 127-140.
Campbell, S., D., & Diebold, F., X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association, 100*, 6-16.
Cao, M., Li, A., & Wei, J. (2004). Watching the Weather Report. *Canadian Investment Review, Summer*, 27-33.
Cao, M., & Wei, J. (1999). Pricing weather derivatives: An equilibrium approach. *Working Paper*, from 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 And Power Risk Management*, 67-70.
Cao, M., & Wei, J. (2003). Weather derivatives: A new class of financial instruments. *Working Paper*, 2003, from http://www.rotman.utoronto.ca/~wei/research/JAI.pdf
Cao, M., & Wei, J. (2004). Weather Derivatives valuation and market price of weather risk. *Journal of Future Markets, 24*(1), 1065-1089.
Carmona, R. (1999). *Calibrating Degree Day Options.* Paper presented at the 3rd Seminar on Stochastic Analysis, Random Field and Applications, Ecole Polytechnique de Lausanne.
Carr, M., & Madan, D., B. (1999). Option Valuation Using The Fast Fourier Transform. *Journal of Computational Finance, 2*(4), 61-73.
Challis, S. (1999). Bright Forecast for Profits. *Reactions, June edition*.
CME. (2005). An Introduction to CME Weather Products. Retrieved January, 2007, from http://www.cme.com/edu/res/bro/cmeweather
Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia, USA: SIAM.
Davis, M. (2001). Pricing weather derivatives by marginal value. *Quantitative Finance, 1*, 1-4.
Dischel, B. (1998a). At least: A model for weather risk. *Weather risk special report, Energy And Power Risk Management*, 30-32.
Dischel, B. (1998b). Black-Scholes won't do. *Weather risk special report, Energy And Power Risk Management*, 8-9.
Dischel, B. (1999). Shaping history for weather risk management. *Energy And Power Risk, 12*(8), 13-15.
Dorfleitner, G., & Wimmer, M. (2010). The Pricing of Temperature Futures at The Chicago Mercantile Exchange. *Journal of Banking & Finance*.
Dornier, F., & Queruel, M. (2000). Caution to the wind. *Weather risk special report, Energy Power Risk Management*, 30-32.
Dunis, C. L., & Karalis, V. (2003). Weather derivative pricing and filling analysis for missing temperature data. *Derivative Use, Trading & Regulation, 9*(1), 61-83.
Franses, P. H., Neele, J., & van Dijk, D. (2001). Modeling Asymetric Volatility in Weekly Dutch Temperature Data. *Enivromental Modelling & Software, 16*, 37-46.
Garman, M., Blanco, C., & Erickson, R. (2000). Weather derivatives: Instruments and pricing issues. *Environmental Finance*.
Geman, H. (1999). *Insurance and Weather Derivatives*: RISK Books.
Geman, H., & Leonardi, M.-P. (2005). Alternative approaches to weather derivatives pricing. *Managerial Finance, 31*(6), 46-72.
Hamisultane, H. (2006a). Extracting Information from the Market to Price the Weather Derivatives. *Working Paper* Retrieved November, 2006, from http://halshs.archives-ouvertes.fr/docs/00/17/91/89/PDF/weathderiv_extraction.pdf
Hamisultane, H. (2006b). Pricing the weather derivatives in the presence of long memory in temperatures. *Working Paper* Retrieved May, 2006, from http://halshs.archives-ouvertes.fr/docs/00/08/87/00/PDF/weathderiv_longmemory.pdf
Hamisultane, H. (2007). Utility-based Pricing of the Weather Derivatives. *Working Paper* Retrieved September, 2006, from http://halshs.archives-ouvertes.fr/docs/00/17/91/88/PDF/weathderiv_utility.pdf

Hamisultane, H. (2008). Which Method for Pricing Weather Derivatives ? *Working Paper* Retrieved July, 2008, from http://halshs.archives-ouvertes.fr/docs/00/35/58/56/PDF/wpaper0801.pdf

Hanley, M. (1999). Hedging the Force of Nature. *Risk Professional, 1*, 21-25.

Hardle, W. K., & Lopez Cabrera, B. (2009). Infering the market price of weather risk, *SFB649 Working Paper*. Berlin: Humboldt-Universitat zu Berlin.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics, 13*, 70-84.

Hull, C. J. (2003). *Option, Futures and Other Derivatives* (Fifth ed.). New Jersey: Prentice Hall.

Ikeda, N., & Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. Tokyo: North-Holland and Kodansha.

Jewson, S., Brix, A., & Ziehmann, C. (2005). *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations*. Cambridge, UK: Campbridge University Press.

Jewson, S., & Caballero, R. (2003a). Seasonality in the dynamics of surface air temperature and the pricing of weather derivatives. *Meteorological Applications, 10*(4), 377-389.

Jewson, S., & Caballero, R. (2003b). Seasonality in the Statistics of Surface Air Temperature and the Pricing of Weather Derivatives. *Meteorological Applications, 10*(4), 367-376.

Koutsouyiannis, D. (2003). Climate change, the Hurst phenomenon, and hydrological statistics / Changement climatique, phénomène de Hurst et statistiques hydrologiques. *Hydrological Sciences Journal, 48*(1), 3 - 24.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? *Journal of Econometrics, 54*(1-3), 159-178.

Lau, K. M., & Weng, H. Y. (1995). Climate signal detecting using wavelet signal transform. *Bulletin of American Meteorological Society, 76*, 2391-2401.

Mallat, S. G. (1999). *A Wavelet Tour of Signal Processing*. San Diego: Academic Press.

McIntyre, R., & Doherty, S. (1999). An example from the UK. *Energy And Power Risk Management*.

Moody, J. E., & Utans, J. (1992). Principled Architecture Selection for Neural Networks: Applications to Corporate Bond Rating Prediction. In A. P. Refenes (Ed.), *Neural Networks in the Capital Markets*: John Wiley & Sons.

Moreno, M. (2000). Riding the temp. *Weather Derivatives, FOW Special Support*.

Oetomo, T., & Stevenson, M. (2005). Hot or cold? A comparison of different approaches to the pricing of weather derivatives. *Journal of Emerging Market Finance, 4*(2), 101-133.

Platen, E., & West, J. (2005). A fair pricing approach to weather derivatives. *Asia-Pacific Financial Markets, 11*, 23-53.

Richards, T., J., Manfredo, M., R., & Sanders, D., R. (2004). Pricing Weather Derivatives. *American Journal of Agricultural Economics, 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.

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.

Schiller, F., Seidler, G., & Wimmer, M. (2008). Temperature models for pricing weather derivatives. *SSRN*, 2008, from http://ssrn.com/abstract=1280826

Svec, J., & Stevenson, M. (2007). Modelling and forecasting temperature based weather derivatives. *Global Finance Journal, 18*(2), 185-204.

Taylor, J., W., & Buizza, R. (2002). Neural network load forecasting with weather ensemble predictions. *IEEE Transactions on Power Systems, 17*(3), 626-632.

Taylor, J., W., & Buizza, R. (2004). A Comparison of Temperature Density Forecasts from GARCH and Atmospheric Models. *Journal of Forecasting, 23*, 337-355.

Tol, R. S. J. (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. *Journal of Risk Finance, 4*(4), 6-17.

Turvey, C., G. (2005). The Pricing of Degree-Day Weather Options. *Agricultural Finance Review, Spring 2005*, 59-85.

Xu, W., Odening, M., & Musshof, O. (2008). Indifference Pricing of Weather Derivatives. *American Journal of Agricultural Economics, 90*(4), 979-993.

Yoo, S. (2003). Weather derivatives and seasonal forecast. *Working paper* Retrieved January, 2003, from http://www.card.iastate.edu/faculty/profiles/bruce_babcock/Shiyong_Yoo_papers/WD_2.pdf

Zapranis, A., & Alexandridis, A. (2006). *Wavelet analysis and weather derivatives pricing*. Paper presented at the 5th Hellenic Finance and Accounting Association (HFAA). Retrieved from http://www.hfaa.gr/Docs/HFAA%202005/Session%202%20Options%20and%20Futures/Wavelet%20Analysis%20&%20Weather%20Derivatives%20Pricing.pdf

Zapranis, A., & Alexandridis, A. (2007, 29-31 August). *Weather Derivatives Pricing: Modelling the Seasonal Residuals Variance of an Ornstein-Uhlenbeck Temperature Process With Neural Networks*. Paper presented at the EANN 2007, Thessaloniki, Greece.

Zapranis, A., & Alexandridis, A. (2008). Modelling Temperature Time Dependent Speed of Mean Reversion in the Context of Weather Derivetive Pricing. *Applied Mathematical Finance, 15*(4), 355 - 386.

Zapranis, A., & Alexandridis, A. (2009a). Model Identification in Wavelet Neural Networks Framework. In L. Iliadis, I. Vlahavas & M. Bramer (Eds.), *Artificial Intelligence Applications and Innovations III* (Vol. IFIP 296, pp. 267-277). New York, USA: Springer.

Zapranis, A., & Alexandridis, A. (2009b, 27-29 August). *Modeling and Forecasting CAT and HDD Indices for Weather Derivative Pricing*. Paper presented at the EANN 2009, London,UK.

Zapranis, A., & Alexandridis, A. (2009c). 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 Computing & Applications, 20*(6), 787-801.

Zapranis, A., & Refenes, A. P. (1999). *Principles of Neural Model Indentification, Selection and Adequacy: With Applications to Financial Econometrics*: Springer-Verlag.

Zhang, Q. (1997). Using Wavelet Network in Nonparametric Estimation. *IEEE Trans. Neural Networks, 8*(2), 227-236.