8

Modeling Financial Temperature Derivatives

In this chapter a real data set is used to demonstrate the application of our proposed framework. We focus on a financial application in forecasting the prices of weather derivatives. A weather derivative is a financial instrument that has a payoff derived from variables such as temperature, snowfall, humidity, and rainfall. Since their inception in 1996, weather derivatives have shown substantial growth. The first parties to arrange for, and issue, weather derivatives in 1996 were energy companies, which, after the deregulation of energy markets, were exposed to weather risk. In September 1999, the Chicago Mercantile Exchange (CME) launched the first exchange-traded weather derivatives. Today, weather derivatives are being used for hedging purposes by companies and industries whose profits can be affected adversely by unseasonal weather or, for speculative purposes, by hedge funds and others interested in capitalizing on those volatile markets.

Weather risk is unique in that it is highly localized, and despite great advances in meteorological science, still cannot be predicted precisely and consistently. Weather derivatives are also different from other financial derivatives in that the underlying weather index (i.e., HDD, CDD, CAT, etc.) cannot be traded. Furthermore, the corresponding market is relatively illiquid. Consequently, since weather derivatives cannot be replicated cost-efficiently by other weather derivatives, arbitrage pricing cannot apply to them directly. The weather derivatives market is a classically incomplete market, because the underlying weather variables are not tradable. When the market is incomplete, prices cannot be derived from the no-arbitrage condition, since it is

Wavelet Neural Networks: With Applications in Financial Engineering, Chaos, and Classification, First Edition. Antonios K. Alexandridis and Achilleas D. Zapranis.

not possible to replicate the payoff of a given contingent claim by a controlled portfolio of basic securities. Consequently, the classical Black–Scholes–Merton pricing approach, which is based on no-arbitrage arguments, cannot be applied directly. In addition, market incompleteness is not the only reason; weather indices do not follow random walks (as the Black and Scholes approach assumes), and the payoffs of weather derivatives are determined by indices that are average quantities, whereas the Black–Scholes payoff is determined by the underlying value at exactly the maturity date of the contract (European options).

In the remainder of the chapter we introduce the weather derivative market and then focus on the modeling of the temperature process and the forecasting of future outcomes of weather derivatives. More precisely, using data from detrended and deseasonalized daily average temperatures (DATs), a wavelet network is constructed, initialized, and trained. At the same time, the significant variables are selected, in this case the correct number of lags. Finally, the trained wavelet network will be used to predict the future evolution of the weather derivative and to construct confidence and prediction intervals.

WEATHER DERIVATIVES

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 on the price of some fundamental underlying, 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 cannot be stored or traded; 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. However, it is estimated that over 90% of the weather derivatives now traded are based on temperature.

Today, weather derivatives are being used for hedging purposes by companies and industries whose profits can be affected adversely 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 U.S. economy is directly exposed to weather risk. It is estimated that nearly 30% of the U.S. economy and 70% of U.S. companies are affected by weather (CME, 2005). The electricity sector is especially sensitive to temperature. According to Li and Sailor (1995) and Sailor and Munoz (1997), temperature is the most significant weather factor explaining electricity and gas demand in the United States. The impact of temperature on both electricity demand and price has been considered in many papers, including those of Engle et al. (1992), Henley and Peirson (1998), and Peirson and Henley (1994). Unlike insurance- and catastrophe-linked instruments, which cover high-risk and low-probability events, weather derivatives shield revenues against low-risk and high-probability events (e.g., mild or cold winters).

Today, the weather market is one of the fastest-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.

Pricing and Modeling Methods

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 previous years is considered to be the payoff of the derivative (Jewson et al., 2005).

Alternatively, one can model the corresponding index directly, a method known as *index modeling*: 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 the literature, a few papers suggest that temperature index modeling (HDD or CDD Index) might be more appropriate (Davis, 2001; Dorfleitner and Wimmer, 2010; Geman and Leonardi, 2005; Jewson et al., 2005).

Another approach to estimating the temperature-driving process is to use models based on daily temperatures. Daily modeling can, in principle, lead to more accurate pricing than that from 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. Temperatures observed show seasonality in the mean, variance, distribution, and autocorrelation, and there is evidence of long memory in the autocorrelation. The risk with daily modeling is that small misspecifications in the models can lead to large mispricing of the temperature contracts (Jewson et al., 2005).

In the literature, two methods have been proposed for the modeling of the DAT, the use of a discrete or a continuous process. Moreno (2000) argues against the use of continuous processes in temperature modeling based on the fact that the values of temperature are in discrete form; hence, a discrete process should be used directly. Caballero and Jewson (2002), Caballero et al. (2002), Campbell and Diebold (2005), Cao et al. (2004), Cao and Wei (1999, 2000, 2003, 2004), Carmona (1999), Franses et al. (2001), Jewson and Caballero (2003a,b), Moreno (2000), Roustant et al. (2003a,b), Svec and Stevenson (2007), Taylor and Buizza (2002, 2004), and Tol (1996) make use of a general ARMA framework.

On the other hand, a wide range of studies suggest a temperature diffusion stochastic differential equation: Alaton et al. (2002), Bellini (2005), Benth (2003), Benth and Saltyte-Benth (2005, 2007), Benth et al. (2007, 2008), Bhowan (2003), Brody et al. (2002), Dischel (1998a,b, 1999), Dornier and Queruel (2000), Geman and Leonardi (2005), Hamisultane (2006a,b, 2007, 2008), McIntyre and Doherty (1999),

Oetomo and Stevenson (2005), Richards et al. (2004), Schiller et al. (2008), Torro et al. (2003), Yoo (2003), Zapranis and Alexandridis (2006, 2007, 2008, 2009a,b), and Zapranis and Alexandridis (2011). The continuous processes used for modeling daily temperatures usually take a mean-reverting form, which has to be discretized to estimate its various parameters. Once the parameters of the process are estimated, one can then value any contingent claim by taking the 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, MonteCarlo (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.

DATA DESCRIPTION AND PREPROCESSING

For accurate pricing and efficient weather risk management, the weather data must be of both adequate amount and highly quality (Dunis and Karalis, 2003). The data corresponding to the DATs for Berlin were provided by the European Climate Assessment and Dataset (ECAD; http://eca.knmi.nl/). The weather index we are interested in is the DAT. In the ECAD the DAT is measured as the average of the daily maximum and minimum temperature and is measured in Celsius degrees (°C). European weather contracts traded on the CME use the same measurement for the temperature. The precision with which temperature is measured in the ECAD is 0.1°C.

The data set consists of 3650 values, corresponding to 10 years' (1991–2000) of detrended and deseasonalized DATs in Berlin. So that each year would have equal observations, February 29 was removed from the data.

DATA EXAMINATION

In this section the data of the DATs of Berlin are examined to build a mathematical model that describes the dynamics of the temperature. In Figure 8.1a the DATs for Berlin for the period 1/1/1991 to 12/31/2000 are presented. Closer inspection of the figure reveals a seasonal cycle of one year, as expected. Moreover, extreme values in summer and winter are evident. To obtain a better insight of the temperature dynamics, the descriptive statistics of the DATs are examined. The mean temperature is 10.1, while variation of the DAT is quite large, with a standard deviation of 7.91. The difference between the maximum and minimum temperatures is around 44°C. These results indicate that temperature is very volatile and it is expected to be difficult to model and predict it accurately.

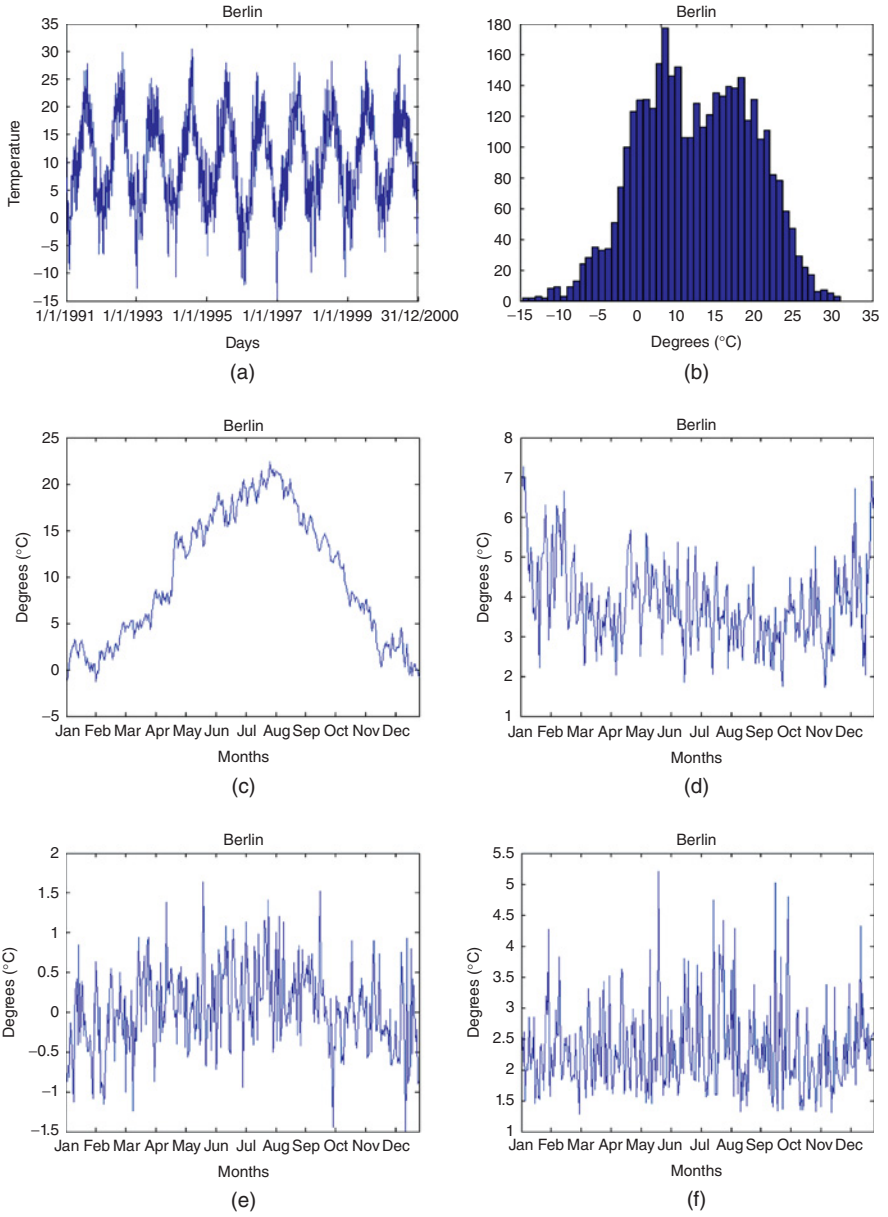

Figure 8.1 *(a) Daily average temperature, (b) empirical distribution, (c) mean, (d) standard deviation, (e) skewness, and (e) kurtosis of the DAT in Berlin.*

Negative skewness is evident, while the temperature in Berlin exhibits excess negative kurtosis. More precisely, the skewness is −0.08 and the kurtosis is 2.38. The results above indicate that the distribution of the DAT in Berlin is platykurtic, with a lower and wider peak where the mass of the distribution is concentrated on the left tail.

Finally, a normality test is performed. The normality is strongly rejected by a Jarque–Bera (JB) test. JB tests the null hypothesis that the sample data come from a normal distribution with unknown mean and variance, against the alternative that it does not come from a normal distribution:

$$\begin{aligned} H_0 &= \text{sample comes from a normal distribution} \\ H_1 &= \text{sample does not come from a normal distribution} \end{aligned} \tag{8.1}$$

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

$$\text{JB} = \frac{n}{6}\left[s^2 + \frac{(k-3)^2}{4}\right] \tag{8.2}$$

where n is the sample size, s is the sample skewness, and k is the sample kurtosis. The JB statistic is over 62, and the p-value is zero, indicating the rejection of the null hypothesis that the temperature at the seven European cities follows a normal distribution. Figure 8.1b presents the empirical distribution of the DAT in Berlin. A closer inspection of the figure reveals a bimodal distribution, with the two peaks corresponding to the summer and winter temperatures.

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, $T_{\text{avg},t}$, was estimated using only observations for each particular day t. In Figure 8.1c the seasonal pattern is clear. 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 Berlin fluctuates from −1.2 to 22.4°C.

Next, the standard deviation of the DAT is estimated. In Figure 8.1d the standard deviation for Berlin is presented. The standard deviation is higher in the winter period, while it is smaller in summer. Our results confirm the studies of Bellini (2005), Benth and Saltyte-Benth (2005, 2007), Benth et al. (2007), and Zapranis and Alexandridis (2008, 2009b).

Figure 8.1e presents the estimated skewness for each day t for Berlin. The figure reveals that the skewness tends to increase during the summer months while it decreases during the winter months. In general, the skewness is negative in winter months and positive in 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 Figure 8.1f. The figure does not reveal any seasonal pattern of the kurtosis. On the other hand, it is clear that the kurtosis has small deviations around 2, with many large upward spikes.

Two unit root tests were performed in the DAT. The temperature time series is tested for a unit root using an augmented Dickey–Fuller (ADF) test. The ADF was 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 smaller than the critical value at the 5% significance level. More precisely, the ADF statistic is -5.1116, with p-value zero, while the critical value at the 5% significance level is -2.8621.

When a more powerful test is required, the Kwiatkowski et al. (1992) (KPSS) unit root test is employed. In contrast to the ADF test, the KPSS tests the null hypothesis that the time series is stationary versus the alternative that the time series is nonstationary (a unit root exists). The optimal bandwidth number was estimated using the Newery–West method. The KPSS statistic is 0.0505. The KPSS statistic has a value smaller than the critical value 0.463; hence, the null hypothesis that the time series is stationary cannot be rejected.

MODEL FOR THE DAILY AVERAGE TEMPERATURE: GAUSSIAN ORNSTEIN–UHLENBECK PROCESS WITH LAGS AND TIME-VARYING MEAN REVERSION

Many different models have been proposed to describe the dynamics of a temperature process. In this section a model for the seven cities studied in the preceding section is derived. Studying temperature data, Cao et al. (2004) and Cao and Wei (1999, 2000, 2003) built 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 winter than in summer.

As shown in the rest of the section, our results confirm the foregoing assumptions. It is known that temperature follows a predicted cycle. As expected and as shown in Figure 8.1, 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 8.1a,c, and d. In other words, it is not possible to observe temperatures of 20°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 in an increment in the observed measurements of temperature. Finally, observing Figure 8.1d it is clear that the temperature volatility is higher in winter than in summer.

Following Zapranis and Alexandridis (2008, 2011) and Alexandridis (2010), a model that describes the temperature dynamics is given by a Gaussian mean-reverting Ornstein–Uhlenbeck process and is defined as follows:

$$dT(t) = dS(t) + \kappa(t)\,[T(t) - S(t)]\;dt + \sigma(t)\,dB(t) \tag{8.3}$$

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 shown by Dornier and Queruel (2000), the term $dS(t)$ should be added for a proper mean-reversion toward 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), it is expected that the speed of mean reversion is high (i.e., the difference of today's and tomorrow's temperature is expected to be high). In contrast, if the temperature today is close to the seasonal variance, we expect the temperature to revert slowly to its seasonal average. To capture this feature, the speed of mean reversion is modeled by a time-varying function $\kappa(t)$.

It is clear that strong seasonalities and trends exist in the data. Before we present the time series to the wavelet network, both the trend and the seasonality must be removed. The usual approach is to model the seasonality by a sinusoid given by

$$S(t) = \text{Trend}_t + A\,\sin\frac{2\pi(t - f_i)}{365} \tag{8.4}$$

and the trend by a first-order polynomial given by

$$\text{Trend}_t = a + bt \tag{8.5}$$

Alexandridis and Zapranis (2013) used a more complex method where the seasonal mean and variance were extracted using wavelet analysis:

$$\begin{aligned} S(t) = \text{Trend}_t &+ \sum_{i=1}^{I_1} a_i \sin\frac{2\pi(t - f_i)}{p_i \cdot 365} \\ &+ a_{I_1+1}\left[1 + \sin\frac{2\pi(t - f_{I_1+1})}{p_{I_1+1} \cdot 365}\right]\sin\frac{2\pi t}{365} \end{aligned} \tag{8.6}$$

$$\sigma^2(t) = c + \sum_{i=1}^{I_2} c_i \sin\frac{2p_i'\pi t}{365} + \sum_{j=1}^{J_2} d_j \cos\frac{2p_j'\pi t}{365} \tag{8.7}$$

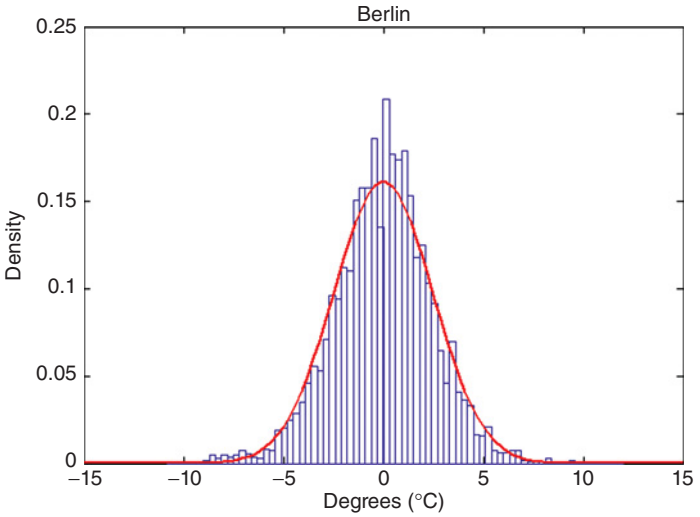

Figure 8.2 *Empirical and normal distribution (solid line) of the first difference of the daily average temperature in Berlin.*

To identify the terms I_1 and p_i in (8.6) and I_2, J_2, and p_i' in (8.7), we decompose the temperature series using a wavelet transform.

Finally, the driving noise process $B(t)$ is modeled by a standard Brownian motion (BM). The histogram of the first difference of the DAT and the normal distribution (solid line) is presented in Figure 8.2. A closer inspection of the figure 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 is tested further later.

Following Benth and Saltyte-Benth (2007) and Zapranis and Alexandridis (2008), a discrete approximation of (8.3) is obtained and is given by

$$\Delta T(t) \approx \Delta S(t) + \kappa \left[T(t-1) - S(t-1)\right] + \sigma(t)\, \Delta B(t) \tag{8.8}$$

Next, by setting

$$\tilde{T}(t) = T(t) - S(t) \tag{8.9}$$

we have that

$$\tilde{T}(t) = a\tilde{T}(t-1) + \sigma(t)\varepsilon(t) \tag{8.10}$$

with

$$a = 1 + \kappa \tag{8.11}$$

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

$$\tilde{T}(t) = a(t-1)\tilde{T}(t-1) + \sigma(t)\varepsilon(t) \tag{8.12}$$

$$a(t) = 1 + \kappa(t) \tag{8.13}$$

The detrended and deseasonalized temperature series, $\tilde{T}(t)$, can be modeled with an AR(1) process with a zero constant term, as shown in (8.12). In the context of such a model, the mean reversion parameter a is typically assumed to be constant over time. Brody et al. (2002) mentioned that in general a should be a function of time, but no evidence was presented. On the other hand, Benth and Saltyte-Benth (2005), using a data set comprising 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 reported that none was found. However, 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 (8.12). On the other hand, averaging techniques, on a yearly or monthly basis, run the danger of filtering out too much variation and consequently of 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 and Alexandridis (2008) estimated daily values of the variable $a(t)$ for the city of Paris using a nonparametric nonlinear NN. Their results indicate strong time dependence in the daily values of $a(t)$.

In this section we address that issue by using a wavelet network to estimate relationship (8.12) nonparametrically and then estimate a as a function of time. In addition, we apply the variable selection framework to select the number of lags in the temperature process. 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 the network input, we obtain a series of daily values for $a_i(t)$. This gives us a much better insight into temperature dynamics and temperature derivative pricing. As shown by Alexandridis and Zapranis (2013), the daily variation of $a_i(t)$ is quite significant after all. In addition it is expected that the waveform of the wavelet network will provide a better fit to the DATs that are governed by seasonalities and periodicities.

In previous studies (Alaton et al., 2002; Bellini, 2005; Benth and Saltyte-Benth, 2005, 2007); Zapranis and Alexandridis, 2008, 2009b), it has been shown that an AR(1) model is not complex enough to completely remove the autocorrelation in the residuals. Alternatively, more complex models were suggested (Carmona, 1999; Geman and Leonardi, 2005). Using wavelet networks, the generalized version of (8.12) is estimated nonlinearly and nonparametrically; that is,

$$\tilde{T}(t+1) = \phi\left(\tilde{T}(t), \tilde{T}(t-1), ...\right) + e(t) \tag{8.14}$$

where

$$e(t) = \sigma(t)\varepsilon(t) \tag{8.15}$$

where $\tilde{T}$ is the detrended and deseasonalized DAT and the $e(t)$ are the residuals of the wavelet network.

After estimation of the wavelet network, the seasonal variance must be removed from the residuals. The exact modeling of the temperature process is beyond our scope in this chapter. To compare the out-of-sample forecast of a proposed model, the complete system, consisting of equations (8.5), (8.6), (8.7), (8.14), and (8.15), must be estimated. Here we focus on the construction of an optimal wavelet network for the system (8.14) that represents the detrended and deseasonalized DATs. Although the remaining steps are of no interest in this book, an out-of-sample comparison between the wavelet network and two linear models used widely in the literature is presented at the end of the chapter.

ESTIMATION USING WAVELET NETWORKS

Model (8.14) 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 those of Alaton et al. (2002), Benth and Saltyte-Benth (2007), and Zapranis and Alexandridis (2008). However, the length of the lag series must be selected. Since the wavelet network is a nonparametrical 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 utilized. Hence, the variable significance algorithm presented in Chapter 5 is employed to determine the number of significant lags.

Variable Selection

The target values of the wavelet network are the detrended and deseasonalized DATs. The explanatory variables are lagged versions of the target variable. Choosing the length of a lag distribution in linear models can be done by minimizing an information criterion such as the Akaike or Schwarz criterion. Alternatively, the ACF and the PACF can be studied. The ACF suggests that the first 35 lags are significant. On the other hand, the PACF suggests that the first six lags, as well as lags 8 and 11, must be included in the model. However, results from these methods are not necessarily true in nonlinear nonparametric models.

Alternatively, to select only the significant lags, the variable selection algorithm presented in the preceding section will be employed. Initially, the training set contains the dependent variable and seven lags. Hence, the training set consists of 7 inputs, 1 output, and 3643 training pairs. The relevance of a variable to the model is quantified by the SBP criterion introduced earlier, which outperformed similar criteria.

Table 8.1 summarizes the results of the model identification algorithm for Berlin. Both the model selection and variable selection algorithms are included in the table. The algorithm concluded in four steps and the final model contains only three variables. The prediction risk for the reduced model is 3.1914, whereas for the original model it was 3.2004. On the other hand, the empirical loss increased slightly, from

TABLE 8.1 Variable Selection with Backward Elimination in Berlin[a]

Step	Variable to Remove (Lag)	Variable to Enter (Lag)	Variables in Model	Hidden Units (Parameters)	n/p Ratio	Empirical Loss	Prediction Risk
—	—	—	7	5 (83)	43.9	1.5928	3.2004
1	X_6	—	6	2 (33)	110.4	1.5922	3.1812
2	X_7	—	5	1 (17)	214.3	1.5927	3.1902
3	X_5	—	4	1 (14)	260.2	1.6004	3.2056
4	X_4	—	3	1 (11)	331.2	1.5969	3.1914

[a]The algorithm concluded in four steps. In each step the following are presented: which variable is removed, the number of hidden units for the particular set of input variables and the parameters used in the wavelet network, the empirical loss, and the prediction risk.

1.5928 for the initial model to 1.5969 for the reduced model, indicating that the explained variability (unadjusted) decreased slightly. However, the explained variability (adjusted for degrees of freedom) was increased for the reduced model to 64.61%, whereas it was 63.98% initially. Note that the noise term in $\tilde{T}$ is large, since the seasonal and trend parts were removed from the data; hence, we expect low values of R^2 and $\bar{R}^2$.

Finally, the number of parameters is reduced significantly in the final model. The initial model needed 5 hidden units and 7 inputs. Hence, 83 parameters were adjusted during the training phase, so the ratio of the number of training pairs n to the number of parameters p was 43.9. In the final model, only 1 hidden unit 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.

The statistics for the wavelet network model at each step are given in Table 8.2. More precisely, the first part of the table reports the value of the SBP and its p-value. In the second part of the table, various fitting criteria are reported: the mean absolute error, the maximum absolute error (MaxAE), the normalized mean squared error (NMSE), the mean absolute percentage error (MAPE), $\bar{R}^2$, the empirical loss, and the prediction risk.

In the full model it is clear that the value of the SBP for the last three variables is very small, in contrast to the first two variables. Observing the p-values, we conclude that the last four variables have p-value greater than 0.1, while the sixth lag has a p-value of 0.8826, strongly indicating a "not significant" variable. The wavelet network was converged after 43 iterations. In general, a very good fit was obtained. The empirical loss is 1.5928, and the prediction risk is 3.2004. MaxAE is 11.1823, MAE is 1.8080, and NMSE is 0.3521. MAPE is 3.7336. Finally, the $\bar{R}^2 = 63.98\%$.

The statistics for the wavelet network at step 1 are also presented in Table 8.2. The network had 6 inputs, two wavelets were used to construct the wavelet network, and 33 weights were adjusted during the training phase. The wavelet network converged after 17 iterations. By removing X_6 from the model, we observe from the table that 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. The empirical loss was decreased slightly to 1.5922. However, MAE and NMSE were increased slightly, to 1.8085 and 0.3529, respectively. On the other hand, MaxAE and the MAPE were decreased to 11.1446 and 3.7127, respectively.

TABLE 8.2 Step-by-Step Variable Selection in Berlin[a]

	Full Model		Step 1		Step 2		Step 3		Step 4	
Variable	SBP	*p*-Value	SBP	*p*-Value	SBP	*p*-Value	SBP	*p*-Value	SBP	*p*-Value
7	0.0026	0.7796	0.0031	**0.5700**						
6	0.0032	**0.8826**								
5	0.0053	0.6757	0.0131	0.0000	0.0206	**0.1907**				
4	0.0161	0.3500	0.0149	0.1403	0.0216	0.1493	−0.0052	**0.4701**		
3	0.2094	0.0000	0.2368	0.0000	0.2285	0.0000	0.1991	0.0000	0.2244	0.0000
2	1.1123	0.0000	1.0318	0.0000	1.0619	0.0000	0.9961	0.0000	0.9363	0.0000
1	9.8862	0.0000	10.0160	0.0000	9.9858	0.0000	10.0537	0.0000	10.1933	0.0000
MAE	1.8080		1.8085		1.8083		1.8093		1.8095	
MaxAE	11.1823		11.1446		11.1949		11.0800		11.0925	
NMSE	0.3521		0.3529		0.3525		0.3526		0.3530	
MAPE	3.7336		3.7127		3.7755		3.7348		3.7171	
$\bar{R}^2$	63.98%		64.40%		64.59%		64.61%		64.61%	
Empirical loss	1.5928		1.5922		1.5927		1.6004		1.5969	
Prediction risk	3.2004		3.1812		3.1902		3.2056		3.1914	
Iterations	43		17		19		4		19	

[a]SBP was the average for each variable of 50 bootstrapped samples, the standard deviation, and the *p*-value; SBP, sensitivity-based pruning; MAE, mean absolute error; MaxAE, maximum absolute error; NMSE, normalized mean squared error; MSE, mean squared error; MAPE, mean absolute percentage error.

Next, the decision to remove X_6 is tested. The new prediction risk was reduced to 3.1812, while the explained variability adjusted for degrees of freedom increased to 64.40%. Hence, the removal of X_6 reduced the complexity of the model, while its predictive power was increased.

At step 2, X_7, which had the largest p-value, 0.5700, at the previous step, was removed from the model. Table 8.2 shows the statistics for the wavelet network at step 2. The new wavelet network had 5 inputs, 1 hidden unit was used, and 17 weights were adjusted during the training phase. The WN converged after 19 iterations. A closer inspection of Table 8.2 reveals that the removal of X_7 resulted in an increase in the error measures and a worse fit was obtained. The new $\bar{R}^2$ is 64.59%. The new prediction risk increased to 3.1902, which is smaller than the threshold. In other words, by removing X_7 the total predictive power of our model was slightly decreased; however, adding the variable X_7 to the model, only 0.28% additional variability of our model was explained, while the computational burden was increased significantly.

Examining the values of the SBP in Table 8.2 it is observed that the first two variables still have significantly larger values than the remaining variables. The p-values reveal that in the third step the X_5 must be removed from the model since its p-value is 0.1907.

At step 3 the network had 4 inputs, 1 hidden unit was used, and 14 weights were adjusted during the training phase. The wavelet network converged after four iterations. When removing X_5 from the model, we observe from Table 8.2 that only X_4 has a p-value greater than 0.1. Again the empirical loss and the prediction risk were increased. More precisely, the empirical loss is 1.6004 and the prediction risk increased from 0.48% to 3.2056%. The new prediction risk is greater than the estimated prediction risk of the initial model of about 0.16%. Again the increase in the prediction risk was significantly smaller than the threshold. On the other hand, $\bar{R}^2$ was increased to 64.61%, indicating an improved fit. Hence, the decision to remove X_5 was accepted.

In the final step, the variable X_4 had a p-value of 0.4701 and it was removed from the model. The network had 3 inputs, one wavelet was used in the construction of the wavelet network, and only 11 weights were adjusted during the training phase. The wavelet network converged after only 19 iterations. After the removal of X_4, a new wavelet network was trained with only one wavelet. The new empirical loss was decreased to 1.5969. The MAE and NMSE are 1.8095 and 0.3530, respectively, while MaxAE and MAPE are 11.0925 and 3.7171, respectively. Next the decision to remove X_4 was tested. The new prediction risk was reduced to 3.1914, while the explained variability adjusted for degrees of freedom was 64.61%. Hence, the removal of X_4 reduced the complexity of the model while its performance was increased. The p-values of the remaining variables are zero, indicating that the remaining variables are characterized as very significant variables. Hence, the algorithm stops. Our proposed algorithm indicates that only the three most recent lags should be used, while the PACF suggested the first six lags as well as lags 8 and 11.

In the final model only three of the seven variables were used. The complexity of the model was reduced significantly, since from 83 parameters in the initial model only 11 parameters have to be trained in the final model. In addition, in the reduced

TABLE 8.3 Prediction Risk at Each Step of the Variable Selection Algorithm for the First 5 Hidden Units for Berlin

	Hidden Units				
Step	1	2	3	4	5
0	3.2009	3.2026	3.2023	3.2019	**3.2004**
1	3.1817	**3.1812**	3.1828	3.1861	3.1860
2	**3.1902**	3.1915	3.1927	3.1972	3.1974
3	**3.2056**	3.2077	3.2082	3.2168	3.2190
4	**3.1914**	3.2020	3.2182	3.2158	3.2169

model the prediction risk was minimized when only 1 hidden unit was used, while 5 hidden units were needed initially. Our results indicate that the in-sample fit was slightly decreased in the reduced model. However, when an adjustment is made for the degrees of freedom, we observe that $\bar{R}^2$ 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.

Model Selection

In each step the appropriate number of hidden units is determined by applying the model selection algorithm presented in Chapter 4. Table 8.3 shows the prediction risk for the first 5 hidden units at each step of the variable selection algorithm for Berlin. Ideally, the prediction risk will decrease (almost) monotonically until a minimum is reached and will then start to increase (almost) monotonically. The number of hidden units 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 1 hidden unit is only 3.2009. When 1 additional hidden unit is added to the model, the prediction risk increases. Then, as more hidden units are added to the model, the prediction risk decreases monotonically. The minimum is reached when 5 hidden units are used and is 3.2004. When additional hidden units are added in the topology of the model, the prediction risk increases. Hence, the architecture of the wavelet network contains 5 hidden units. In other words, to construct the wavelet network, the five higher-ranking wavelets should be selected to form the wavelet basis. Observing Table 8.3, it is clear that the prediction risk in the initial model with only 1 hidden unit is almost the same as in the model with 5 hidden units. This is due to the small number of parameters that were adjusted during the training phase when only 1 hidden unit is used, not due to a better fit.

At the second step, when variable X_6 was removed, the prediction risk is minimized when 2 hidden units are used. Similarly, at steps 2, 3, and 4, the prediction risk is minimized when only 1 hidden unit is used. Additional hidden units do not improve the fitting or the predictive power of the model.

Initialization and Training

After the training set and the correct topology of the wavelet network are selected, the wavelet network can be constructed and trained. The BE method is used to initialize

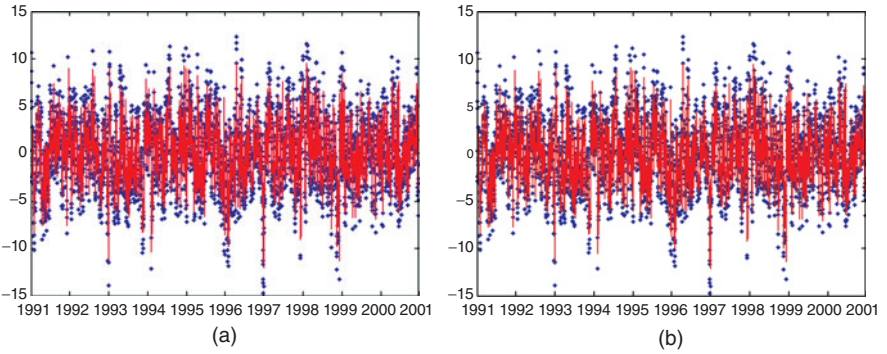

Figure 8.3 *Initialization of the final model for the temperature data in Berlin using the BE method (a) and the fit of the trained network with 1 hidden unit (b). The wavelet network converged after 19 iterations.*

the wavelet network. A wavelet basis is constructed by scanning the first four levels of the wavelet decomposition of the DAT in Berlin.

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 fewer than five sample points of the training data in their support are removed. Seventy-six wavelets that do not contribute significantly to the approximation of the original time series were identified. The truncated basis contains 92 wavelet candidates. Applying the BE method, the wavelets are ranked in order of significance. The wavelets in the wavelet library are ranked as follows: The BE starts the regression by selecting all the available wavelets from the wavelet library. Then the wavelet that contributes the least in the fitting of the training data is repeatedly eliminated. Since only 1 hidden unit is used on the architecture of the model, only the wavelet with the highest ranking is used to initialize the wavelet network. Figure 8.3a presents the initialization of the final model using only 1 hidden unit. The initialization is very good and the wavelet network converged after only 19 iterations. The training stopped when the minimum velocity, 10^{-5}, of the training algorithm was reached. The fitting of the trained wavelet network can be found in Figure 8.3b.

Next, various fitness criteria of the wavelet network corresponding to the $\tilde{T}_t$ in Berlin are estimated. Our results reveal that the wavelet networks fit the detrended and deseasonalized DATs reasonably well. The overall fit for Berlin is $\bar{R}^2 = 64.61\%$, while the MSE is 5.4196 and the MAE is only 1.8090. As mentioned earlier, the noise term is very large compared to the detrended and deseasonalized data, and as a result, a relatively small $\bar{R}^2$ is expected.

Next, 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 estimated. These three criteria examine the ability of the network to predict changes, independent of the size of the change, and they are reported as percentages. The POS measures the ability of the network to predict the sign of the target values, positive or

negative. The POS for the detrended and deseasonalized DATs is very high, 81.49%. The POCID is 60.15%, and the IPOCID is 52.30%.

For our temperature model we assumed that the noise term follows a normal distribution. Examining the residuals of the wavelet network, $e(t)$, given by equation (8.14), we find that the residuals have a mean of 0.01 and a standard deviation of 2.33. The skewness is −0.03 and the kurtosis is 3.73. Finally, the KS normality test has a p-value of 0 and rejects the normality of the residuals, while the Ljung–Box Q-statistic is 31.469 with a p-value of 0.0493, barely accepting the hypothesis of uncorrelated residuals.

However, after removal of the seasonal variance, the final residuals of the model, $\varepsilon(t)$, given by (8.15) have a mean of 0 and a standard deviation of 1. The skewness is –0.02 and the kurtosis is 3.53. Finally, the KS normality test has a p-value of 0.3081 and accepts the normality of the residuals, while the Ljung–Box Q-statistic is 29.616, with a p-value of 0.0763 accepting the hypothesis of uncorrelated residuals. Hence, our model does not violate our initial assumption and can be used for predictions.

Confidence and Prediction Intervals

After the wavelet network is constructed and trained, it can be used for prediction. Hence, confidence and prediction intervals can be constructed. In this section both confidence and prediction intervals are constructed using the balancing method. Using the BS method, 200 training samples are created and then divided into eight groups. In each group the average output of the wavelet networks is estimated. Next, new 1000 bootstrapped samples are created for the 8 average outputs to estimate the model variance given by (7.25). Then the confidence intervals is estimated with a level of significance a of 5%.

Figure 8.4 presents the confidence intervals for the detrended and deseasonalized DAT in Berlin as well as the average wavelet network output obtained from 200 bootstrapped samples. As the intervals are very narrow, to obtain a clear figure only the five first values are presented.

Next, the prediction intervals are constructed for the out-of-sample data set. The out-of-sample data consist of 365 values of detrended and deseasonalized DATs in Berlin for the period 2000–2001. The out-of-sample performance criteria are presented in Table 8.4. The overall fit adjusted for degrees of freedom is $\bar{R}^2 = 59.27\%$. The NMSE is 0.3961, while the MAPE is only 2.4108. Figure 8.5 illustrates the prediction intervals together with the real data and the average forecast of the wavelet network for the 200 bootstrapped samples. PICP = 93.46%.

Out-of-Sample Comparison

As mentioned earlier, to compare the out-of-sample forecast of a proposed model, the complete system consisting of equations (8.5), (8.6), (8.7), (8.14), and (8.15) must be estimated. An analytical discussion of temperature modeling in the context of weather derivatives was presented by Alexandridis and Zapranis (2013).

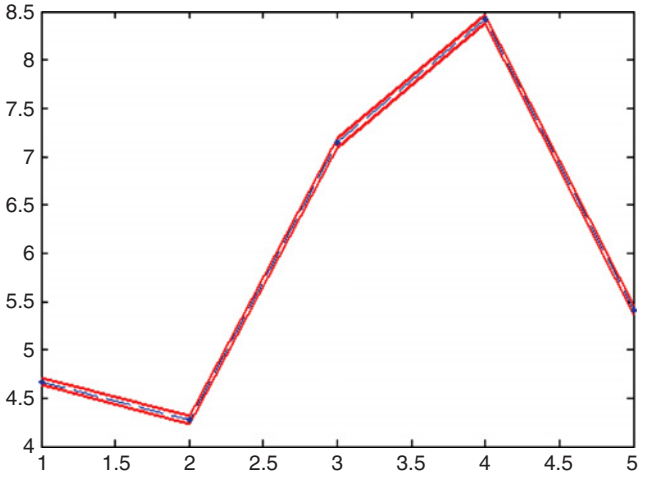

Figure 8.4 *Confidence intervals and the average WN output using the balancing method and 200 bootstrapped samples. The five presents only the first five values for simplicity.*

TABLE 8.4 Out-of-Sample Performance Criteria for Berlin

MAE	1.7340
MaxAE	9.3330
NMSE	0.3961
MAPE	2.4108
$\bar{R}^2$	59.27%

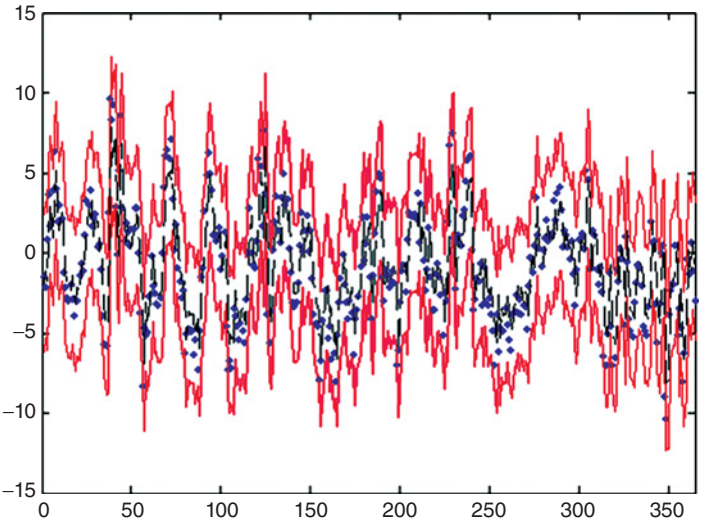

Figure 8.5 *Prediction intervals (red solid lines), the real data (dotted), and the average wavelet network output (black solid line) using the balancing method and 200 bootstrapped samples of the detrended and deseasonalized DATs in Berlin for the period 2000–2001. PICP = 93.46%.*

We presented only the part that is modeled by a wavelet network—in other words, the estimation of model (8.14)—but it would also be interesting to examine the forecasting ability of our model. To do so we examine the predictive power of the wavelet network against two linear models proposed previously and used widely in the literature: the first proposed by Alaton et al. (2002) and the second proposed by Benth and Saltyte-Benth (2007). The three models are similar, but in the two linear models, the speed of mean reversion, κ, is assumed to be constant and is computed through an AR(1) model given by (8.10).

The three models will be compared in the prediction of two temperature indices in seven European cities; the forecasting window will be 1, 2, 3, 6, and 12 months. Moreover, two forecasting methods will be used: one-day-ahead forecasting and out-of-sample forecasting. Hence, the three methods will be compared in 140 samples. The results indicate that the wavelet network outperforms the other two methods in 81 case (58%). On the other hand, the model proposed by Alaton et al. (2002) gave the best results in 35 cases (25%), while the model proposed by Benth and Saltyte-Benth (2007) gave the best results in only 24 cases (17%). For a complete analysis of temperature modeling and presentation of the forecasting comparison, we refer the reader to the book by Alexandridis and Zapranis (2013).

CONCLUSIONS

In this chapter a temperature time series was studied to develop a model that describes the temperature evolution in the context of weather derivative pricing. A mean reverting Ornstein–Uhlenbeck with seasonal mean and variance and time-varying speed of mean reversion was proposed.

In the context of an Ornstein–Uhlenbeck temperature process, the time dependence of the speed of the mean reversion $\kappa(t)$ was examined using a wavelet network. First, the variable selection framework proposed in Chapter 7 was utilized to estimate the number of lags of the nonlinear nonparametric AR model. At each step of the variable selection algorithm, the optimal number of wavelons was obtained by applying the minimum prediction risk criterion using bootstrapping.

The initial model had seven variables, whereas the final had one only three. Our results indicated that the reduced model has the same fitting to the data but increased predicting power. Moreover, a significantly smaller network and fewer parameters were needed during the training phase of the reduced wavelet network. Similarly, the training time was reduced significantly.

We compared the fit of the residuals with the normal distributions with two types of models. The first type was the nonlinear nonparametric model proposed, where κ is a function of time. The second category of models consists of two linear models proposed previously and often cited in the 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.

Testing our model for adequacy, we find that the wavelet network does not violate any of the initial conditions. The residuals are uncorrelated and follow a normal distribution. On the other hand, in the two linear models, as presented by Alexandridis and Zapranis (2013), the normality hypothesis was rejected while strong autocorrelation in the residuals was evident.

The model proposed was evaluated out-of-sample. The predictive power of the model proposed was evaluated using two out-of-sample forecasting methods: out-of-sample forecasts over a period and one-day-ahead forecasts over a period. Modeling the DAT using wavelet networks significantly enhanced the fitting and the predictive accuracy of the temperature process. Modeling the DAT assuming a time-varying speed of mean reversion resulted in 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.

Concluding, wavelet networks can model very well the temperature process in the context of financial weather derivatives pricing. Consequently, they constitute an accurate and efficient tool for weather derivative pricing.

REFERENCES

Alaton, P., Djehince, B., and 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," Ph.D. thesis, University of Macedonia, Thessaloniki, Greece.

Alexandridis, A. K., and Zapranis, A. D. (2013). *Weather Derivatives: Modeling and Pricing Weather-Related Risk*. Springer-Verlag, New York.

Bellini, F. (2005). "The Weather Derivatives Market: Modelling and Pricing Temperature," 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., and 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., and 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., and Koekebakker, S. (2007). "Putting a price on temperature." *Scandinavian Journal of Statistics*, 34, 746–767.

Benth, F. E., Saltyte-Benth, J., and Koekebakker, S. (2008). *Stochastic Modelling of Electricity and Related Markets*. World Scientific, Singapore.

Bhowan, A. (2003). *Temperature Derivatives*. University of Witwatersrand, Johannesburg, South Africa.

Brody, C. D., Syroka, J., and Zervos, M. (2002). "Dynamical pricing of weather derivatives." *Quantitative Finance*, 2, 189–198.

Caballero, R., and 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., and 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., and Diebold, F. X. (2005). "Weather forecasting for weather derivatives." *Journal of the American Statistical Association*, 100, 6–16.

Cao, M., and Wei, J. (1999). "Pricing Weather Derivatives: An Equilibrium Approach." Working Paper, Rotman Graduate School of Management, University of Toronto, Toronto.

Cao, M., and Wei, J. (2000). "Pricing the weather." Weather Risk Special Report, *Energy and Power Risk Management*, May, 67–70.

Cao, M., and Wei, J. (2003). "Weather Derivatives: A New Class of Financial Instruments." Working Paper, University of Toronto, Toronto, Canada.

Cao, M., and Wei, J. (2004). "Weather derivatives valuation and market price of weather risk." *Journal of Future Markets*, 24(11), 1065–1089.

Cao, M., Li, A., and Wei, J. (2004). "Watching the weather report." *Canadian Investment Review*, Summer, 27–33.

Carmona, R. (1999). "Calibrating degree day options." *3rd Seminar on Stochastic Analysis, Random Field and Applications*, Ecole Polytechnique de Lausanne, Ascona, Switzerland.

Challis, S. (1999). "Bright forecast for profits." *Reactions*, June.

CME. (2005). "An introduction to CME weather products." Retrieved January, 2007, from http://www.cme.com/edu/res/bro/cmeweather

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*, March, 30–32.

Dischel, B. (1998b). "Black–Scholes won't do." Weather Risk Special Report, *Energy and Power Risk Management*, October, 8–9.

Dischel, B. (1999). "Shaping history for weather risk management." *Energy and Power Risk Management*, 12(8), 13–15.

Dorfleitner, G., and Wimmer, M. (2010). "The pricing of temperature futures at the Chicago Mercantile Exchange." *Journal of Banking and Finance*.

Dornier, F., and Queruel, M. (2000). "Caution to the wind." Weather Risk Special Report, *Energy and Power Risk Management*, August, 30–32.

Dunis, C. L., and Karalis, V. (2003). "Weather derivative pricing and filling analysis for missing temperature data." *Derivative Use, Trading and Regulation*, 9(1), 61–83.

Engle, R. F., Mustafa, C., and Rice, J. (1992). "Modelling peak electricity demand." *Journal of Forecasting*, 11, 241–251.

Franses, P. H., Neele, J., and van Dijk, D. (2001). "Modeling asymetric volatility in weekly dutch temperature data." *Enivromental Modelling and Software*, 16, 37–46.

Geman, H., and 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.

Henley, A., and Peirson, J. (1998). "Residential energy demand and the interaction of price and temperature: British experimental evidence." *Energy Economics*, 20, 157–171.

Jewson, S., and 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., and Caballero, R. (2003b). "Seasonality in the statistics of surface air temperature and the pricing of weather derivatives." *Meteorological Applications*, 10(4), 367–376.

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

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and 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.

Li, X., and Sailor, D. J. (1995). "Electricity use sensitivity and climate and climate change." *World Resource Review*, 3, 334–346.

McIntyre, R., and Doherty, S. (1999). "Weather risk: an example from the UK." *Energy and Power Risk Management*, June.

Moreno, M. (2000). "Riding the temp." *Weather Derivatives, FOW Special Supplement*, December.

Oetomo, T., and 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.

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

Richards, T. J., Manfredo, M. R., and Sanders, D. R. (2004). "Pricing weather derivatives." *American Journal of Agricultural Economics*, 4(86), 1005–1017.

Roustant, O., Laurent, J.-P., Bay, X., and Carraro, L. (2003a). "A Bootstrap Approach to Price Uncertainty of Weather Derivatives." Ecole des Mines, Saint-Etienne and Ecole ISFA, Villeurbanne.

Roustant, O., Laurent, J.-P., Bay, X., and Carraro, L. (2003b). "Model Risk in the Pricing of Weather Derivatives." Ecole des Mines, Saint-Etienne and Ecole ISFA, Villeurbanne.

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

Schiller, F., Seidler, G., and Wimmer, M. (2008). "Temperature models for pricing weather derivatives." *Quantitative Finance*, 12(3), 489–500.

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

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

Taylor, J. W., and 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., and Valor, E. (2003). "Single factor stochastic models with seasonality applied to underlying weather derivatives variables." *Journal of Risk Finance*, 4(4), 6–17.

Yoo, S. (2003). "Weather Derivatives and Seasonal Forecast." Working Paper, Department of Applied Economics and Management, Cornell University, Ithaca, NY.

Zapranis, A., and Alexandridis, A. (2006). "Wavelet Analysis and Weather Derivatives Pricing." 5th Hellenic Finance and Accounting Association, Thessaloniki, Greece.

Zapranis, A., and Alexandridis, A. (2007). "Weather Derivatives Pricing: Modelling the Seasonal Residuals Variance of an Ornstein–Uhlenbeck Temperature Process with Neural Networks." *EANN 2007*, Thessaloniki, Greece.

Zapranis, A., and 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., and Alexandridis, A. (2009a). "Modeling and Forecasting CAT and HDD Indices for Weather Derivative Pricing." *EANN*, London,UK.

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

Zapranis, A., and Alexandridis, A. (2011). "Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing." *Neural Computing and Applications*, 20(6), 787–801.

Zhang, Q. (1997). "Using wavelet network in nonparametric estimation." *IEEE Transactions on Neural Networks*, 8(2), 227–236.