EANN 2009

Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing

A. Zapranis · A. Alexandridis

Abstract In this paper, we use wavelet neural networks in order to model a mean-reverting Ornstein–Uhlenbeck temperature process, with seasonality in the level and volatility and time-varying speed of mean reversion. We forecast up to 2 months ahead out of sample daily temperatures, and we simulate the corresponding Cumulative Average Temperature and Heating Degree Day indices. The proposed model is validated in 8 European and 5 USA cities all traded in the Chicago Mercantile Exchange. Our results suggest that the proposed method outperforms alternative pricing methods, proposed in prior studies, in most cases. We find that wavelet networks can model the temperature process very well and consequently they constitute an accurate and efficient tool for weather derivatives pricing. Finally, we provide the pricing equations for temperature futures on Cooling and Heating Degree Day indices.

Keywords Weather derivatives · Pricing · Forecasting · Wavelet networks

1 Introduction

Neural networks (NNs) have been used with success in a broad range of applications. NNs have the ability to approximate any deterministic nonlinear process, with little

A. Zapranis · A. Alexandridis (✉)
Department of Accounting and Finance,
University of Macedonia of Economics and Social Studies,
156 Egnatia St., P.O. 54006, Thessaloniki, Greece
e-mail: aalex@uom.gr; alex@uom.gr

A. Zapranis
e-mail: zapranis@uom.gr

knowledge and no assumptions regarding the nature of the process. Typically, the initial values of the NNs' weights are randomly chosen. However, random weights initialization is generally accompanied with extended training times. In addition, when the transfer function is of a sigmoidal type, there is always a significant chance that the training algorithm will converge to local minima. Finally, there is no theoretical link between the specific parameterization of a sigmoidal activation function and the optimal network architecture, i.e. model complexity (the opposite holds true for wavelet neural networks).

Wavelet analysis (WA) is often regarded as a "microscope" in mathematics [1], and it is a powerful tool for representing nonlinearities [2]. WA has proved to be a valuable tool for analyzing a wide range of time-series and has already been used with success in image processing, signal denoising, density estimation, signal and image compression and time-scale decomposition, [3–7]. However, WA is limited to applications of small input dimensions, since the construction of a wavelet basis, when the dimensionality of the input vector is relatively high, is computationally expensive [8].

In [9], it has been demonstrated that it is possible to construct a theoretical formulation of a feedforward NN in terms of wavelet decompositions. Wavelet Networks (WNs) were proposed by [10] as an alternative to feedforward NNs, which would alleviate the aforementioned weaknesses associated with each method. The WNs are a generalization of radial basis function networks (RBF). WNs are one hidden-layer networks that use a wavelet as an activation function, instead of the classic sigmoidal family. It is important to mention here that multidimensional wavelets preserve the "universal approximation" property that characterizes NNs. The nodes (or wavelons) of WNs are the wavelet coefficients of the function

expansion that have a significant value. In [11], various reasons were presented in why wavelets should be used instead of other transfer functions. In particular, first, wavelets have high compression abilities, and secondly, computing the value at a single point or updating the function estimate from a new local measure involves only a small subset of coefficients.

WNs have been used in a variety of applications so far, i.e., in short term load forecasting [12–16], in time-series prediction [1, 17, 18], signal classification and compression [19–21], signal denoising [22], static, dynamic [9, 10, 23–26] and nonlinear modeling [27], nonlinear static function approximation [28–30], to mention the most important. In [31] WNs were even proposed as a multivariate calibration method for simultaneous determination of test samples of copper, iron, and aluminum.

In contrast to classical "sigmoid NNs", WNs allow for constructive procedures that efficiently initialize the parameters of the network. Using wavelet decomposition, a "wavelet library" can be constructed. In turn, each wavelon can be constructed using the best wavelet of the wavelet library. The main characteristics of these procedures are (1) convergence to the global minimum of the cost function, (2) initial weight vector into close proximity of the global minimum, and as a consequence drastically reduced training times [8, 10, 25]. Finally, WNs provide information on the relative participation of each wavelon to the function approximation and the estimated dynamics of the generating process.

As it was already mentioned, WNs are a generalization of RBF networks. Since support vector machines (SVMs) are theoretically better than RBFs, it would be extremely interesting to compare the two approaches. In contrast to NNs, SVMs have very good properties and advantages against the classical NNs [32]. First, SVMs do not suffer from local minima since the solution to an SVM is global and unique. In addition, SVMs have a simple geometric interpretation and give a sparse solution [33]. The reason that SVMs often outperform ANNs in practice is that are less prone to overfitting [33].

SVMs perform very well in classification problems and usually outperform RBF [34, 35]. SVMs can also be applied in regression problems. In [36], SVMs are compared to classical NNs and RBF networks in predicting financial time-series. Their results indicate that SVMs perform significantly than classical networks. On the other hand, the performance between SVMs and RBFs is similar.

On the other hand, SVMs have a series of drawbacks. In [33], a series of limitations are presented. First, there is little theory about choosing the Kernel functions and its parameters. Secondly, SVMs encounter problems with discrete data. Thirdly, in SVMs very large training times are needed and extensive memory for solving the quadratic

programming is required. When the number of data points is large (say over 2,000), the quadratic programming problem becomes extremely difficult to solve [37]. In this study, 11 years of detrended and deseasonalized daily average temperature resulting to 4,015 training patterns are used. Hence, our large data set restricts the use of SVMs. Hence, in this study the WNs will be compared against two methods widely used by market participants and often cited in the literature.

In this paper, we use a WN in the context of temperature modeling and weather derivative pricing. Relatively, recently a new class of financial instruments, known as "weather derivatives", has been introduced. 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. However, in the majority of the weather derivatives, the underlying asset is a temperature index.

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. Hence, a model that describes accurate the temperature dynamics, the evolution of temperature, and which can be used to derive closed form solutions for the pricing of temperature derivatives is essential.

According to [38, 39], nearly $1 trillion of the US economy is directly exposed to weather risk. 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. Weather derivatives are used to hedge volume risk, rather than price risk.

According to the annual survey by the Weather Risk Management Association (WRMA), the estimated notional value of weather derivatives—OTC and exchange-traded—traded in 2008/2009 was $15 billion, compared to $32 billion the previous year, itself down from the all-time record year of 2005–2006 with $45 billion. However, it was significantly up from 2005s $8.4 billion, and 2004s $2.2 billion [40]. According to Chicago Mercantile Exchange (CME), the recent decline reflected a shift from

seasonal to monthly contracts. However, it is anticipated that the weather market will continue to develop, broadening its scope in terms of geography, client base, and inter-relationship with other financial and insurance markets. In order to fully exploit all the advantages that this market offers, an adequate pricing approach is required [41].

The list of traded contracts in the weather derivatives market is extensive and constantly evolving. However, over 90% of the contracts are written on temperature Heating Degree Days (HDD), Cooling Degree Days (CDD) and Cumulative Average Temperature (CAT) indices. In Europe, CME weather contracts for the summer months are based on an index of CAT. The CAT index is the sum of the daily average temperatures over the contract period. The average temperature is measured as the simple average of the minimum and maximum temperature over 1 day. The value of a CAT index for the time interval $[\tau_1, \tau_2]$ is given by the following expression:

$$\int_{\tau_1}^{\tau_2} T(s)\mathrm{d}s \tag{1}$$

where the temperature is measured in degrees of Celsius. In USA, CME weather derivatives are based on HDD or CDD index. A HDD is the number of degrees by which daily temperature is below a base temperature, while a CDD is the number of degrees by which the daily temperature is above the base temperature,

i.e., $\mathrm{Daily\,HDD} = \max(0, \mathrm{base\,temperature} - \mathrm{daily\,average})$,
$\mathrm{Daily\,CDD} = \max(0, \mathrm{base\,temperature} - \mathrm{base\,temperature})$.

The base temperature is 65° Fahrenheit in the US and 18° Celsius in Europe. HDDs and CDDs are usually accumulated over a month or over a season. At the end of 2008, at CME were traded weather derivatives for 24 US cities,[1] 10 European cities,[2] 2 Japanese cities[3] and 6 Canadian cities.[4]

Weather risk is unique in that it is highly localized, and despite great advances in meteorological science, it still cannot be predicted precisely and consistently. Weather derivatives are also different than other financial derivatives in that the underlying weather indexes, like HDD, CDD, CAT, etc., cannot be traded. Furthermore, the corresponding market is relatively illiquid. Consequently, since weather derivatives cannot be cost-efficiently replicated with other weather derivatives, arbitrage pricing cannot be directly applied to them. Since the underlying weather variables are not tradable, the weather derivatives market is a classic incomplete market.

The first and simplest method that has been used in weather derivative pricing is historical Burn analysis (HBA). HBA is just a simple calculation of how a weather derivative would perform in the past years. By taking the average of these values, an estimate of the price of the derivative is obtained. HBA is very easy in calculation since there is no need to fit the distribution of the temperature or to solve any stochastic differential equation. Moreover, HBA is based on very few assumptions. First, we have to assume that the temperature time-series is stationary. Next, we have to assume that the data for different years are independent and identically distributed. For a detailed explanation of HBA, the reader can refer to [42].

A closer inspection of a temperature time-series shows that none of these assumptions are correct. It is clear that the temperature time-series is not stationary since it contains seasonalities, jumps and trends [43, 44]. Also, the independence of the temperature data for different years is under question. In [42], it is shown that these assumptions can be used if the data can be cleaned and detrended. However, their results show that pricing still remains inaccurate. Other methods as index and daily modeling are more accurate but still HBA is usually a good first approximation of the derivative's price.

In contrast to the previous methods, a dynamic model that directly simulates the future behavior of temperature can be used. Using models for daily temperatures can, in principle, lead to more accurate pricing than modeling temperature indices. Using models for daily temperatures can, in principle, lead to more accurate pricing than modeling temperature indices. Daily models very often show greater potential accuracy than the HBA [42], since daily modeling makes a complete use of the available historical data. In the contrary, calculating the temperature index, such as HDDs, as a normal or lognormal process, a lot of information both in common and in extreme events is lost (e.g., HDD is bounded by zero). It is clear that using index modeling a different model must be estimated for each index. On the other hand, using daily modeling only one model is fitted to the data and can be used for all available contracts on the market on the same location. Also, using a daily model an accurate representation of all indices and their distribution can be obtained. Finally, in contrast to index modeling and HBA, it is easy to incorporate meteorological forecasts.

On the other hand, deriving an accurate model for the daily temperature is not a straightforward process. Observed

---

[1] Atlanta, Detroit, New York, Baltimore, Houston, Philadelphia, Boston, Jacksonville, Portland, Chicago, Kansas City, Raleigh, Cincinnati, Las Vegas, Sacramento, Colorado Spring, Little Rock, Salt Lake City, Dallas, Los Angeles, Tucson, Des Moines, Minneapolis-St. Paul, Washington, D.C.

[2] Amsterdam, Barcelona, Berlin, Essen, London, Madrid, Paris, Rome, Stockholm, Oslo.

[3] Tokyo, Osaka.

[4] Calgary, Montreal, Vancouver, Edmonton, Toronto, Winnipeg.

temperatures show seasonality in all of the mean, variance, distribution, and autocorrelations and long memory in the autocorrelations. The risk with daily modeling is that small misspecifications in the models can lead to large mispricing in the contracts.

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. The most common approach is to model the temperature dynamics with a mean-reverting Ornstein–Uhlenbeck process where the noise part is driven by a Brownian motion [43–49]. The Ornstein–Uhlenbeck process can capture the following characteristics of temperature. Temperature 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; and its volatility is higher in the winter than in summer [50–53]. Alternatively, instead of the classical Brownian Motion, the use of a fractional Brownian Motion is proposed in [49, 54]; however, [55] suggests that fractionality is not justified, if all seasonal components are removed from temperature. In [46], a Levy process as the driving noise process is suggested since the normality hypothesis is often rejected [47, 48]. On the other hand, a Levy process does not allow for closed form derivation of the pricing formula.

In order to rectify the rejection of the normality hypothesis, in more recent papers, [56] and [44] replaced the simple AR(1) model by more complex ones. They used Autoregressive Moving Average, ARMA(3,1), Autoregressive Fractionally Integrated Moving Average, ARFIMA, and ARFIMA-FIGARCH (Fractionally Integrated Autoregressive Conditional Heteroskedasticity) models. Their results from the DAT in Paris indicate that as the model gets more complex, the noise part draws away from the normal distribution. They conclude that although the AR(1) model probably is not the best model for describing temperature anomalies, increasing the model complexity and thus the complexity of theoretical derivations in the context of weather derivative pricing does not seem to be justified. Next, [44] model nonparametrically the seasonal residual variance with NNs. The improvement regarding the distributional properties of the original model is significant. The examination of the corresponding Q-Q plot reveals that the distribution is quite close to Gaussian, while the Jarque–Bera statistic of the original model is almost halved. The NN approach gives a good fit for the autocorrelation function and an improved and reasonable fit for the residuals.

In [43], three different decades of daily average temperatures in Paris are examined using the mean-reverting O-U process proposed by Benth and Saltyte-Benth [47]. The seasonality and the seasonal variance were modeled using WA. Previous studies assume that the parameter of

the speed of mean reversion, κ, is constant. However, the findings of [43] indicate some degree of time dependency in $\kappa(t)$. Since $\kappa(t)$ is important for the correct and accurate pricing of temperature derivatives a significant degree of time dependency in $\kappa(t)$ can be quite important, [45]. A novel approach to estimate nonparametrically a nonlinear time depended $\kappa(t)$ with a NN was presented. Daily values of the speed of the mean reversion were computed. In contrast to averaging techniques, in a yearly or monthly basis, which run the danger of filtering out too much variation, it is expected that daily values will provide more information about the driving dynamics of the temperature process. Results from [43] indicate that the daily variation of the value of the speed of mean reversion is quite high. Intuitively, it is expected $\kappa(t)$ not to be constant. If the temperature today is far from the seasonal average (a cold day in summer), then it is expected that the mean reversion speed will be high, i.e. the difference between today's temperature and tomorrow's temperature is expected to be high. In contrast, if the temperature today is close to the seasonal variance, we expect the temperature to revert to its seasonal average slowly. In [43], $\kappa(t)$ is studied. Their data from Paris indicate that $\kappa(t)$ has a bimodal distribution with an upper threshold, which is rarely exceeded. Also it was examined whether $\kappa(t)$ is a stochastic process itself. Both an Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests were used. Both tests conclude that $\kappa(t)$ is stationary. Finally, using a constant speed of mean reversion parameter, the normality hypothesis was rejected in all three cases while in the case of the NN the normality hypothesis was accepted in all three different samples.

This study extends in various ways the framework presented in [43]. In [43], WA was used in order to identify the trend and the seasonal part of the temperature signal and then a NN was used for modeling the detrended and deseasonalized series. Deducting the form of the seasonal mean and variance was based on observing the wavelet decomposition. In this paper, first, we combine these two steps using WNs. It is expected that the waveform of the activation function and the wavelet decomposition that is performed in the hidden layer of the WN will provide a better fit to the temperature. In particular, a WN is constructed in order to fit the daily average temperature in 13 cities and to forecast the daily average temperature up to 2 months. Second, we compare our model with a similar linear model and the improvement using a nonconstant speed of mean reversion is measured. Third, our approach is evaluated out-of-sample with two methods widely used by researchers and market participants. More precisely, the proposed methodology is compared against HBA and the Benth and Benth's model [47], in forecasting CAT and HDD indices. Finally, we provide the pricing equations on

temperature futures written on CDDs and HDDs indices. The pricing equations for the future CAT contracts when the speed of mean reversion is not constant can be found in [43]. For a concise treatment of wavelet analysis, the reader can refer to [3, 4, 57], while for wavelet networks the reader can refer to [10, 24, 58].

The rest of the paper is organized as follows. In Sect. 2, the wavelet network used to model the detrended and deseasonalized daily average temperature is presented. In Sect. 3, we describe our data and the process used to model the daily average temperature. Next, our model is used to forecast out-of-sample daily average temperatures, and our results are compared against other models previously proposed in literature. In Sect. 4, we discuss pricing of temperature derivatives written on CAT, CDDs, and HDDs indices. Finally, in Sect. 5 we conclude.

2 Wavelet neural networks for multivariate process modeling

In this section, the WN used for modeling the detrended and deseasonalized series is described, with the emphasis being on the theory and mathematics of wavelet neural networks. Until today, various structures of a WN have been proposed [10, 24, 25]. In this study, we implement a multidimensional WN with a linear connection between the wavelons and the output. Moreover, in order for the model to perform well in the presence of linearity, we use direct connections from the input layer to the output layer. Hence,

a network with zero hidden units (HUs) is reduced to the linear model.

The structure of a single hidden-layer feedforward wavelet network is given in Fig. 1. WNs usually have the form of a three layer network. The lower layer represents the input layer, the middle layer is the hidden layer, and the upper layer is the output layer. In the input layer, the explanatory variables are introduced to the WN. The hidden layer consists of the HUs. The HUs are often referred as wavelons, similar to neurons in the classical sigmoid NNs. In the hidden layer, the input variables are transformed to dilated and translated version of the mother wavelet. Finally, in the output layer the approximation of the target values is estimated. The network output is given by the following expression:

$$\hat{y}(\mathbf{x}) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \Psi_j(\mathbf{x}) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i. \quad (2)$$

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 hidden units, and w stands for a network weight. Following [8] we use as a "mother wavelet" the second derivative of the Gaussian function, the so-called "Mexican Hat". The multidimensional wavelets are computed as follows:

$$\Psi_j(\mathbf{x}) = \prod_{i=1}^{m} \psi(z_{ij}) \quad (3)$$

where ψ is the Mexican Hat mother wavelet given by

Fig. 1 A feedforward wavelet network
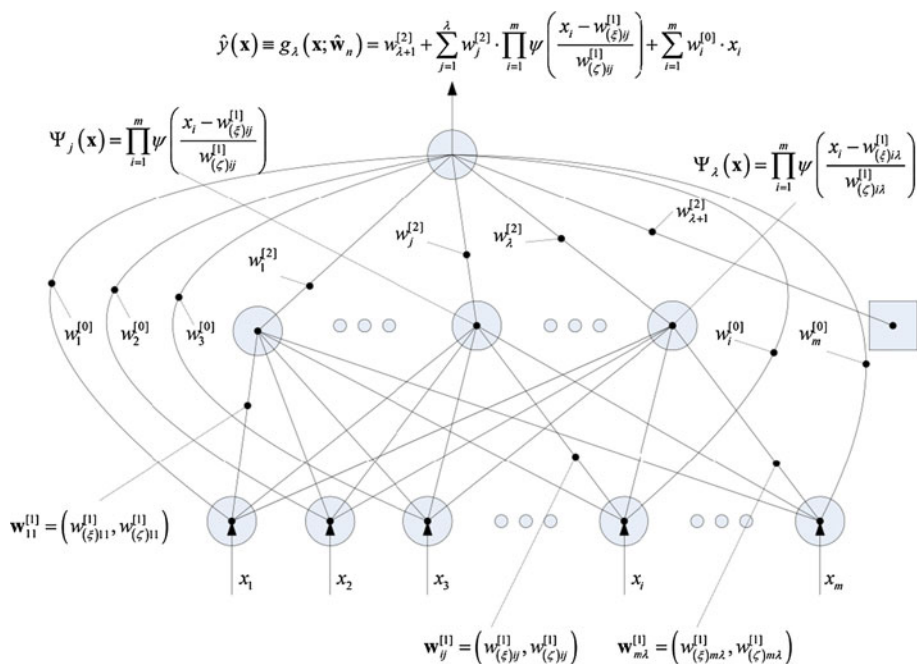
$$\psi(z_{ij}) = \left(1 - z_{ij}^2\right) e^{-\frac{1}{2}z_{ij}^2} \tag{4}$$

and

$$z_{ij} = \frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}}. \tag{5}$$

In the above expression, $i = 1,\ldots, m, j = 1, \ldots, \lambda + 1$, and the weights w correspond to the "translation" $\left(w_{(\xi)ij}^{[1]}\right)$ and the "dilation" $\left(w_{(\zeta)ij}^{[1]}\right)$ factors of the wavelet decomposition. The complete vector of the network parameters comprises:

$$w = \left(w_i^{[0]}, w_{(\xi)ij}^{[1]}, w_{(\zeta)ij}^{[1]}, w_j^{[2]}, w_{\lambda+1}^{[2]}\right). \tag{6}$$

In WNs, in contrast to NNs that use sigmoid functions, selecting initial values of the dilation and translation parameters randomly may not be suitable [25]. A wavelet is a waveform of effectively limited duration that has an average value of zero and localized properties; hence, a random initialization may lead to wavelons with a value of zero. Also, random initialization affects the speed of training and may lead to a local minimum [26]. In literature, more complex initialization methods have been proposed [8, 24, 59]. All methods can be summed in the following three steps.

1. Construct a library W of wavelets.
2. Remove the wavelets that their support does not contain any sample points of the training data.
3. Rank the remaining wavelets and select the best regressors.

The wavelet library can be constructed either by an orthogonal wavelet or a wavelet frame. However, orthogonal wavelets cannot be expressed in closed form. It can be shown that a family of compactly supported nonorthogonal wavelets is more appropriate for function approximation [14]. However, constructing a WN using wavelet frames is not a straightforward process. The wavelet library may contain a large number of wavelets since only the input data were considered in the construction of the wavelet frame. In order to construct a WN, the "best" wavelets must be selected. However, arbitrary truncations may lead to large errors [60]. In the second step, [61] proposes to remove the wavelets that have very few training patterns in their support. Alternatively, in [62], magnitude-based methods were used to eliminate wavelets with small coefficients. In this study, we follow the first approach.

In the third step, the remaining wavelets are ranked and the wavelets with the highest rank are used for the construction of the WN. In [8], three alternative methods were proposed in order to reduce and rank the wavelet in the

wavelet library: Residual-Based Selection (RBS), Stepwise Selection by Orthogonalization (SSO), and Backward Elimination (BE). In this study, we use the BE initialization method, since results from previous studies indicates that it outperforms the other two [8, 58]. The BE algorithm starts by building a WN with all the wavelets in the wavelet library W. Then, the wavelet that contributes the least in the fitting of the training data is repeatedly eliminated. The BE is used only for the initialization of the dilation and translation parameters.

It is clear that additional computational burden is added in order to initialize efficiently the parameters of the WN. However, the efficient initialization significantly reduces the training phase; hence, the total amount of computations is significantly smaller than in a network with random initialization.

After the initialization stage, the weights of the WN are further adjusted. The WN is further trained in order to obtain the vector of the parameters $w = w_0$, which minimizes the cost function. There are several approaches to train a WN. In our implementation, the ordinary back-propagation (BP) was used. BP is probably the most popular algorithm used for training WNs [2, 8, 10, 22, 24–26, 28]. Ordinary BP is less fast but also less prone to sensitivity to initial conditions than higher order alternatives [63]. The basic idea of BP is to find the percentage of contribution of each weight to the error.

The architecture of the WN is selected by minimizing the prediction risk [63]. The prediction risk is estimated by applying the *v-fold* cross-validation. The algorithm of the *v-fold* cross-validation is analytically described in the next section.

The weights $w_i^{[0]}$, $w_j^{[2]}$ and parameters $w_{(\xi)ij}^{[1]}$ and $w_{(\zeta)ij}^{[1]}$ are trained for approximating the target function. Under the assumption that the architecture of the WN and the number of wavelets were selected by minimizing the prediction risk the training is stopped when one of the following criteria is met—the cost function reaches a fixed lower bound or the variations of the gradient or the variations of the parameters reaches a lower bound or the number of iterations reaches a fixed maximum, whichever is satisfied first. In our implementation, the fixed lower bound of the cost function, of the variations of the gradient, and of the variations of the parameters was set to 10^{-5}.

3 Modeling the temperature process and forecasting CAT and HDD indices

Many different models have been proposed in order to describe the dynamics of a temperature process. Early models were using AR(1) processes or continuous

equivalents [45, 50, 64]. Others like [65] and [66] have suggested versions of a more general ARMA(p,q) model. In [67], it has been shown, however, that all these models fail to capture the slow time decay of the autocorrelations of temperature and hence lead to significant underpricing of weather options. Thus, more complex models were proposed. The most common approach is to model the temperature dynamics with a mean-reverting Ornstein–Uhlenbeck process where the noise is driven by a Brownian motion, [43–49]. An Ornstein–Uhlenbeck process is given by:

$$dT(t) = dS(t) - \kappa(T(t) - S(t))dt + \sigma(t)dB(t) \quad (7)$$

where, $T(t)$ is the daily average temperature, $B(t)$ is a standard Brownian motion, $S(t)$ is a deterministic function modeling the trend and seasonality of the average temperature, while $\sigma(t)$ is the daily volatility of temperature variations and κ is the speed of mean reversion. In [47], both $S(t)$ and $\sigma^2(t)$ were modeled as truncated Fourier series:

$$S(t) = a + bt + \sum_{i=1}^{I_1} a_i \sin(2i\pi(t - f_i)/365) $$
$$+ \sum_{j=1}^{J_1} b_j \cos(2j\pi(t - g_j)/365) \quad (8)$$

$$\sigma^2(t) = c + \sum_{i=1}^{I_2} c_i \sin(2i\pi t/365) + \sum_{j=1}^{J_2} d_j \cos(2j\pi t/365) \quad (9)$$

while in [43] and [44], the form of (8) and (9) were determined by WA.

From the Ito formula, an explicit solution for (7) can be derived:

$$T(t) = s(t) + (T(t - 1) - s(t - 1))e^{-\kappa t} $$
$$+ \int_{t-1}^{t} \sigma(u)e^{-\kappa(t-u)}dB(u). \quad (10)$$

According to this representation, $T(t)$ is normally distributed at t and it is reverting to a mean defined by $S(t)$. A discrete approximation to the Ito formula, (10), which is the solution to the mean-reverting Ornstein–Uhlenbeck process (7), is

$$T(t + 1) - T(t) = S(t + 1) - S(t) - (1 $$
$$- e^{-\kappa})(T(t) - S(t)) $$
$$+ \sigma(t)(B(t + 1) - B(t)) \quad (11)$$

which can be written as

$$\tilde{T}(t + 1) = a\tilde{T}(t) + \tilde{\sigma}(t)\varepsilon(t) \quad (12)$$

where $\varepsilon(t) \sim$ i.i.d. and follow the normal $N(0,1)$ distribution and

$$\tilde{T}(t) = T(t) - S(t) \quad (13)$$
$$a = e^{-\kappa} \quad (14)$$

and $e^{-\kappa}$ is the Euler's number.

In order to estimate model (12), we need first to remove the trend and seasonality components from the daily average temperature series. The trend and the seasonality of daily average temperatures is modeled and removed as in [47]. Next, a WN is used to model and forecast daily detrended and deseasonalized temperatures. Hence, (12) reduces to:

$$T(t) = \varphi(T(t - 1)) + e_t \quad (15)$$

where $\varphi(\bullet)$ is estimated nonparametrically by a WN and e_t are the residuals of the network. As it is shown later, strong autocorrelation is observed in e_t. Hence, α is not constant but a time-varying function. Once we have the estimator of the underlying function φ, then we can compute the daily values of α as follows:

$$a = d\tilde{T}(t + 1)/d\tilde{T}(t) = d\varphi/d\tilde{T}. \quad (16)$$

The analytic expression for the WN derivative $d\varphi/d\tilde{T}$ can be found in [58]. For analytic details on the estimation of parameters in (8), (9), (12), and (14), we refer to [43, 47].

In this section, real weather data will be used in order to validate our model and compare it against models proposed in previous studies. Our model is validated in data consisting of 2 months, January and February, of daily average temperatures (2005–2006) corresponding to 59 values. Note that meteorological forecasts over 10 days are not considered accurate. The data set consists of 4,015 values, corresponding to the average daily temperatures of 11 years (1995–2005) in Paris, Stockholm, Rome, Madrid, Barcelona, Amsterdam, London and Oslo in Europe and New York, Atlanta, Chicago, Portland and Philadelphia in USA. The data were collected by the University of Dayton.[5] Temperature derivatives on the above cities are traded in CME. In order for each year to have equal observations, the 29th of February was removed from the data.

Table 1 shows the descriptive statistics of the daily average temperature in each city for the past 11 years, 1995–2005. The mean CAT and mean HDD represent the mean of the HDD and CAT index for the past 11 years for a period of 2 months, January and February. For consistency, all values are presented in degrees Fahrenheit. It is clear that the HDD index exhibits large variability. Similar the difference between the maximum and minimum is close to 70° Fahrenheit in average, for all cities, while the standard deviation of temperature is close to 15° Fahrenheit. Also, for all cities there is kurtosis significant smaller

---

[5] http://www.engr.udayton.edu/weather/.

Table 1 Descriptive statistics of temperature in each city

	Mean	SD	Max	Min	Skewness	Kurtosis	Mean HDD	Mean CAT
Paris	54.38	12.10	89.90	13.80	−0.04	2.50	1368.40	2466.60
Rome	60.20	11.37	85.80	31.10	−0.04	1.96	1075.00	2759.99
Stockholm	45.51	14.96	79.20	−5.00	−0.09	2.33	2114.36	1720.64
Amsterdam	51.00	11.00	79.90	12.20	−0.18	2.54	1512.07	2322.93
Barcelona	61.56	10.59	85.70	32.60	0.09	2.03	899.23	2935.79
Madrid	58.61	13.84	89.80	24.90	0.17	1.94	1262.56	2572.44
New York	55.61	16.93	93.70	8.50	−0.15	2.08	1783.44	2051.56
London	52.87	10.03	83.00	26.70	0.02	2.36	1307.11	2527.89
Oslo	41.47	15.65	74.60	−8.70	−0.31	2.50	2404.53	1430.47
Atlanta	62.18	14.52	89.60	13.70	−0.45	2.25	1130.75	2704.78
Chicago	50.61	19.40	91.40	−12.90	−0.25	2.17	2221.33	1613.67
Portland	46.80	17.36	83.20	−3.70	−0.22	2.22	2382.04	1452.96
Philadelphia	56.02	17.13	90.50	9.50	−0.19	2.03	1766.98	2068.02

Descriptive statistics of the daily average temperature for the period 1995–2005

Mean HDD mean of the HDD index for the past 11 years for a period of January–February, *Mean CAT* mean of the CAT index for the past 11 years for a period of January–February, *SD* standard deviation

than 3 and, with the exceptions of Barcelona, Madrid, and London, there is negative skewness.

First, the linear trend and the mean seasonal part in the daily average temperature in each city are quantified. We simplify (8) and (9) by setting $I_1 = 1$, $J_1 = 0$, $I_2 = 1$ and $J_2 = 1$ and as in [47]. The estimated parameters of the seasonal part $S(t)$ can be found in Table 2. Parameter b indicates that Rome, Stockholm, Amsterdam, Barcelona, London, Oslo, Chicago, Portland, and Philadelphia have an upward trend while a downward trend is clear in the remaining cities. Parameter b ranges from −0.000569 to 0.00064. This means that the in the last 11 years there is a decrease in temperature of −2.1°F in Madrid and an increase in temperature of 2.1°F in Amsterdam. The amplitude a_1 indicates that the difference between the daily winter and daily summer temperature is around 24°F in London and 49°F in Chicago. All parameters are statistically significant with p values smaller than 0.05. In Fig. 2, the seasonal fit of the daily average temperature in Barcelona can be found. For simplicity, we refer only to Barcelona; the results from the remaining cities are similar.

Then, the function $\varphi(\bullet)$ was estimated nonparametrically by a WN. Table 3 shows the necessary number of HUs needed for each network and the estimated prediction risk. The correct topology of each network is selected using the *v-fold* cross-validation criterion, where $v = 20$. The data set is split in 20 equal sets. Each data set contained 5% of the original data, randomly selected without replacement from the initial time-series. Starting with zero HUs, one data set, D_i, where $i = 1,…, 20$, is left out of the training sample Then, the trained network is validated on D_i and the Mean Square Error (MSE) is estimated. Then, D_i is put

Table 2 Estimated parameters of the seasonal part $S(t)$

	a	b	a_1	f_1
Paris	54.85	−0.000234	−14.56	18.54
Rome	60.15	0.000023	−14.83	26.77
Stockholm	44.64	0.000437	−18.94	20.69
Amsterdam	49.94	0.000527	−13.26	21.57
Barcelona	61.46	0.000046	−13.74	27.02
Madrid	59.75	−0.000569	−17.90	19.40
New York	55.72	−0.000052	−21.82	−339.68
London	52.56	0.000154	−12.09	23.34
Oslo	40.18	0.000640	−19.67	18.16
Atlanta	62.45	−0.000136	−17.91	16.41
Chicago	49.90	0.000353	24.56	−526.63
Portland	46.75	0.000027	−22.46	−340.57
Philadelphia	55.87	0.000074	22.00	−160.75

The parameter a is the coefficient of the linear trend while b is the slope. The parameter a_1 is the amplitude of the sinusoid of the seasonal variance and f_1 is the angle that referrers to the maximum and minimum of the temperature in the year. All parameters are statistical significant with p values < 0.05

back into the training sample while D_{i+1} is left out. A new network is trained and a new MSE is calculated. When $i = v$, the average MSE, which represents the prediction risk, is calculated. Then, one more HU is added to the network and the whole procedure is repeated. The algorithm stops when the number of HUs reaches the maximum allowed number (we chose a relative large maximum number of HUs, that is 10). Finally, the WN with the number of HUs that correspond to the smallest prediction risk is selected. As expected, only a few HUs were needed
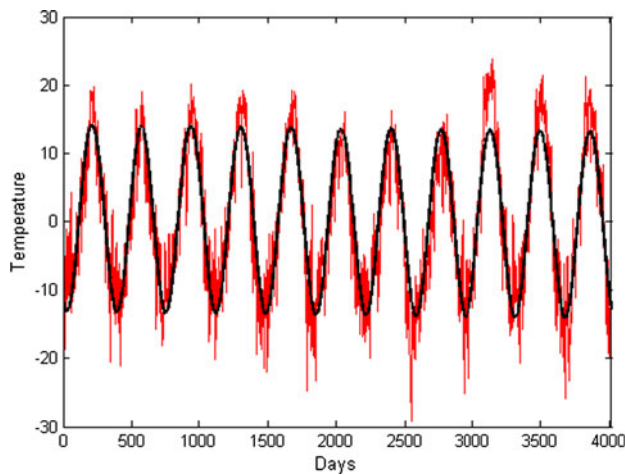
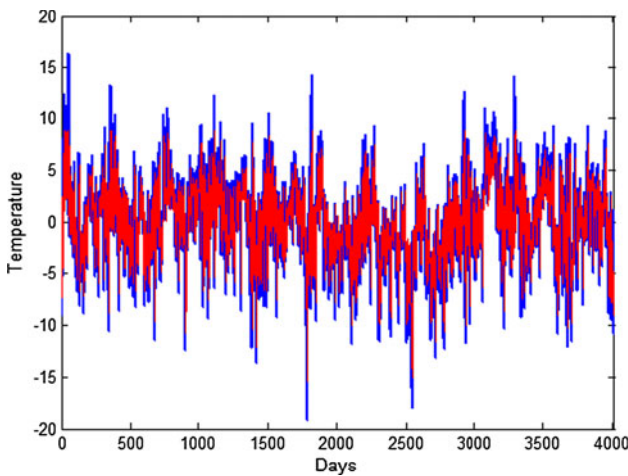Fig. 2 Seasonal fit of the daily average temperature in Barcelona

Fig. 3 The WN fit of the deseasonalized and detrended data

Table 3 Network topology and prediction risk, i.e., out-of-sample absolute relative (percentage) error

	HU	Prediction risk (%)
Paris	1	2.03
Rome	1	2.76
Stockholm	1	2.02
Amsterdam	1	1.84
Barcelona	4	1.42
Madrid	1	2.73
New York	3	2.77
London	1	3.31
Oslo	1	2.20
Atlanta	1	2.60
Chicago	1	2.25
Portland	2	3.25
Philadelphia	3	2.64

to fit the detrended and deseasonalized daily average temperature in model (15). In Barcelona, the hidden layer of the WN consists of 4 HUs while in New York and Philadelphia of 3 HUs. Similarly, in Portland only 2 HUs were needed while for the remaining cities a WN with only one HU was used.

The fitting of the WN to the data is very good and can be found on Fig. 3. The adjusted R^2 in Barcelona is 68.88% and the Mean Square Error (MSE) is 5.5 while the Mean Absolute Error (MAE) is 1.73. The results from the remaining cities are similar.

The residuals of model (15) exhibit strong seasonal variance given by (9), [43, 44, 47, 48]. In Table 4, the estimated parameters of the seasonal variance $\sigma^2(t)$ are presented. Again, all parameters are statistical significant with p values smaller than 0.1. Comparing the autocorrelation function of the squared residuals in Figs. 4 and 5, it is clear that the seasonal variance were successfully removed from the residuals.

Next, the trained WNs were used to forecast 2 months ahead, 59 days, out-of-sample forecasts for the CAT and cumulative HDD indices. Our method is validated and compared against two methods proposed in prior studies that are widely used by market participants. The historical burn analysis (HBA) and the Benth's and Saltyte-Benth's (B–B) model, presented in [47], which is the starting point for our methodology for the detrended and deseasonalized Barcelona daily average temperature.
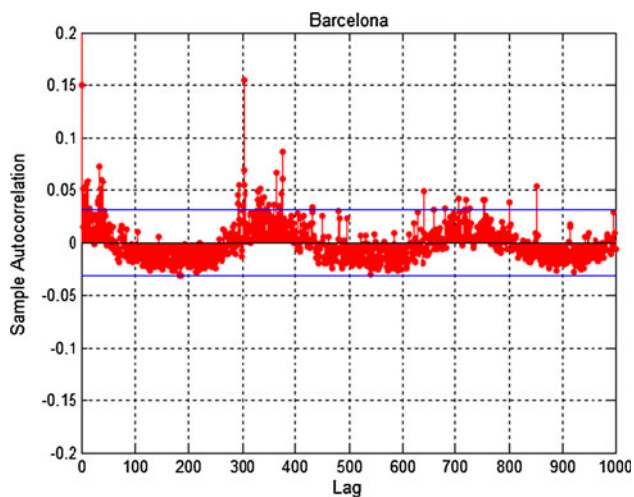
In Table 5, the absolute relative (percentage) errors for the CAT index of each approach are presented while in Table 6 the relative (percentage) errors for the HDD index are shown. It is clear that the proposed WN approach outperforms both HBA and B–B. More precisely, The WN approach is characterized by smaller out-of-sample errors (in 9 out of 13 times) and clearly outperforms B–B (in 11 out of 13 times). Our findings indicate that the WN approach can provide better accuracy in temperature forecasts for European cities, and it is associated with significant smaller errors in comparison with the alternative approaches. Only for Oslo and Amsterdam the WN approach underperforms HBA, but still produces more accurate forecasts when compared with B–B. For the cities in USA, the WN approach gives the smallest out-of-sample error in three cases, while HBA and B–B in one and two cases, Atlanta and Chicago, respectively.

However, by examining the fit of the WNs in Atlanta and Chicago, we observe that $\bar{R}^2$ is 52.39 and 55%, respectively (the lowest values for all locations). This is most probably the reason why the B–B model outperforms WNs in these two cases. Furthermore, when the temperature indices are very close to their historical mean, it is expected HBA to be most accurate. Hence, the out-of-sample values of the indices are compared against their

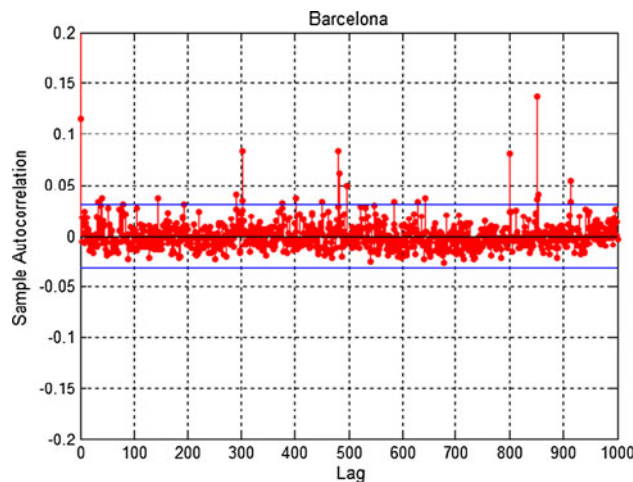**Table 4** Estimated parameters seasonal variance σ_t^2

	c	c_1	c_2	c_3	c_4	d_1	d_2	d_3	d_4
Paris	23.01	1.77	−1.14	−0.07	0.62	4.24	5.43	−0.10	−0.04
Rome	13.06	0.43	−0.88	−0.33	−0.48	8.83	2.16	0.06	0.24
Stockholm	27.05	4.40	−1.14	3.01	1.34	14.07	7.62	2.63	0.09
Amsterdam	16.87	2.13	−2.43	0.13	0.21	4.88	3.91	0.41	−0.81
Barcelona	8.08	0.02	−0.59	−0.16	0.29	4.60	1.68	0.33	−0.13
Madrid	15.61	1.28	−1.02	−0.19	0.60	2.41	3.75	0.90	0.26
New York	58.99	15.83	−1.35	3.94	−0.61	26.70	5.48	1.46	3.49
London	19.01	0.97	−2.02	−0.53	0.06	7.25	3.63	0.35	−0.60
Oslo	30.46	4.28	2.07	4.84	0.36	22.98	10.74	2.86	−1.05
Atlanta	41.26	15.08	2.21	−0.50	1.62	32.22	1.45	−1.70	−1.14
Chicago	67.95	18.23	−6.52	0.80	3.28	36.44	1.25	1.74	1.15
Portland	79.14	10.61	6.80	14.31	3.38	47.00	13.37	−3.00	−0.23
Philadelphia	53.60	16.89	−2.51	−0.99	−0.77	28.30	1.61	0.59	3.24

The estimated parameters of the seasonal variance. The seasonal variance consists of 4 sinusoids and 4 cosines. c_0 is the coefficient, c_1–c_4 are the amplitudes of the sinusoids and d_1–d_4 are the coefficients of the cosines. All parameters are statistical significant with p values < 0.1

Fig. 4 Autocorrelation function of squared residuals of the WN model

Fig. 5 Autocorrelation function of squared residuals of the WN model after removing the seasonal variance

mean. In order to do so, a two-sided t test for small samples is performed.[6] In Table 7, the real and the mean values of the CAT indices are presented as well as the t values and the p values of the t test. Observing the p values, we conclude that the out-of-sample values of the CAT index are not statistically different than the mean value with level of significance $\alpha = 5\%$ for the following three cities: Barcelona, Oslo, and Atlanta. As a result, HBA performs better in these cases with the exception of Barcelona. The p-values for the HDDs index are the same.

On the other hand, for Barcelona and Madrid, where the $\bar{R}^2$ is over 68%, the forecasting error of the proposed method is quite small especially when compared with HBA and B–B (0.03 and 0.74% respectively). Furthermore, by comparing the mean CAT (form Table 1) and the real CAT observed in the out-of-sample period, we can conclude that when the corresponding index deviates from its average historical value, then HBA produces large estimation errors which subsequently lead to large pricing errors. On the other hand, the WN approach gives significant smaller errors even in cases where the temperature deviates significantly from its historical mean. In Table 6, we can see the absolute relative (percentage) errors for the HDD index of each method. The results are similar.

Finally, we examine the fitted residuals in model (12). Note that the B–B model is based on the hypothesis that the

---

[6] The t test assumes that the population is normally distributed. This assumption for the two indices is justified in many papers, in [68] for example.

Table 5 Absolute relative errors for the three forecasting models. CAT index

Errors	Real CAT	HBA (%)	B–B (%)	WN (%)
Paris	2229.50	10.63	8.34	7.12
Rome	2641.40	4.49	4.39	3.93
Stockholm	1559.40	10.34	9.47	9.29
Amsterdam	2162.90	7.40	8.60	8.55
Barcelona	2893.40	1.46	0.19	0.03
Madrid	2400.20	7.18	2.10	0.74
New York	2295.70	10.63	9.02	8.76
London	2340.80	7.99	6.07	5.75
Oslo	1390.50	2.87	5.62	4.53
Atlanta	2765.30	2.21	1.83	2.58
Chicago	1910.90	15.55	10.22	10.90
Portland	1689.90	14.02	8.87	8.32
Philadelphia	2258.90	8.45	5.95	5.92

Table 6 Absolute relative errors for the three forecasting models (HDD index)

	Real CAT	HBA (%)	B–B (%)	WNN (%)
Paris	1605.50	14.77	11.58	9.88
Rome	1193.60	9.94	9.71	8.70
Stockholm	2275.60	7.09	6.49	6.37
Amsterdam	1672.10	9.57	11.13	11.06
Barcelona	941.60	4.50	0.57	0.10
Madrid	1434.80	12.00	3.51	1.24
New York	1539.30	15.86	13.45	13.07
London	1494.20	12.52	9.51	9.01
Oslo	2444.50	1.64	3.20	2.58
Atlanta	1069.70	5.71	4.74	6.67
Chicago	1924.10	15.45	10.15	10.82
Portland	2145.10	11.05	6.99	6.56
Philadelphia	1576.10	12.11	8.53	8.49

Table 7 Test for equality of mean of the CAT index

	Mean CAT	Real CAT	t Value	p Value
Paris	2466.60	2229.50	5.83	0.0002
Rome	2759.99	2641.40	3.24	0.0089
Stockholm	1720.64	1559.40	2.70	0.0223
Amsterdam	2322.94	2162.90	2.81	0.0186
Barcelona	2935.79	2893.40	1.37	0.2014
Madrid	2572.44	2400.20	4.44	0.0012
New York	2051.56	2295.70	−3.91	0.0029
London	2527.89	2340.80	6.28	0.0001
Oslo	1430.46	1390.50	0.56	0.5870
Atlanta	2704.78	2765.30	−1.66	0.1287
Chicago	1613.67	1910.90	−4.65	0.0009
Portland	1452.95	1689.90	−4.08	0.0022
Philadelphia	2068.02	2258.90	−2.99	0.0135

Table 8 Normality test for the B–B residuals

	Jarque–Bera	p Value
Paris	5.78	0.0549
Rome	170.12	0.0000
Stockholm	60.36	0.0000
Amsterdam	44.64	0.0000
Barcelona	685.84	0.0000
Madrid	69.52	0.0000
New York	53.92	0.0000
London	11.67	0.0039
Oslo	37.27	0.0000
Atlanta	403.06	0.0000
Chicago	44.33	0.0000
Portland	21.92	0.0000
Philadelphia	89.55	0.0000

remaining residuals follow the normal distribution. It is clear from Table 8 that only for Paris the normality hypothesis can be marginally accepted. The Jarque–Bera statistic is slightly higher than 0.05. In every other case, the normality hypothesis is rejected. More precisely, the Jarque–Bera statistics are very large and the p values are close to zero, indicating that both the kurtosis and the skewness of the residuals are significantly different than 0 and 3, respectively.

The previous extensive analysis indicates that our results are very promising. Modeling the DAT using WA and WNs enhanced the fitting and the predictive accuracy of the temperature process. Modeling the DAT assuming a time-varying speed of mean reversion resulted 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.

4 Temperature derivative pricing

So far, we modeled the temperature using an Ornstein–Uhlenbeck process. We have shown in [43] that the mean reversion parameter α in model (12) is characterized by significant daily variation. Recall that parameter α is connected to our initial model as $a = e^{-\kappa}$, where κ is the speed of mean reversion. It follows that the assumption of a constant mean reversion parameter introduces significant error in the pricing of weather derivatives. In this section, the pricing formulae for a future contract written on the HDD or the CDD index that incorporate the time dependency of the

speed of the mean reversion parameter is derived. The corresponding equations for the CAT index have already been presented in [43].

The CDD, HDD indices over a period $[\tau_1, \tau_2]$ are given by

$$\text{HDD} = \int_{\tau_1}^{\tau_2} \max(c - T(s), 0) \mathrm{d}s \tag{17}$$

$$\text{CDD} = \int_{\tau_1}^{\tau_2} \max(T(s) - c, 0) \mathrm{d}s. \tag{18}$$

Hence, the pricing equations are similar for both indices. The CDD, HDD, and CAT futures prices are linked by the following relation:

$$F_{\text{HDD}}(t, \tau_1, \tau_2) = c(\tau_2 - \tau_1) - F_{\text{CAT}}(t, \tau_1, \tau_2) + F_{\text{CDD}}(t, \tau_1, \tau_2). \tag{19}$$

Hence, if we derive the futures price of the CDDs price, then the price of a futures contract written on the HDDs can also be easily estimated. First, we rewrite (7) where parameter κ, now is a function of time t, $\kappa(t)$.

$$\mathrm{d}T(t) = \mathrm{d}S(t) + \kappa(t)(T(t) - S(t)) + \sigma(t)\mathrm{d}B(t). \tag{20}$$

From the Ito formula, an explicit solution can be derived:

$$T(t) = S(t) + e^{\int_0^t \kappa(u)\mathrm{d}u}(T(0) - S(0)) + e^{\int_0^t \kappa(u)\mathrm{d}u} \int_0^t \sigma(s) e^{-\int_0^s \kappa(u)\mathrm{d}u} \mathrm{d}B(s). \tag{21}$$

Note that $\kappa(t)$ is bounded away from zero [43].

Our aim is to give a mathematical expression for the CDD future price. The weather derivatives market is an incomplete market, since cumulative average temperature contracts are written on a temperature index, which is not a tradable or storable asset. In order to derive the pricing formula, first we must find a risk-neutral probability measure $Q \sim P$, where all assets are martingales after discounting. In the case of weather derivatives, any equivalent measure Q is a risk-neutral probability. If Q is the risk-neutral probability and r is the constant compounding interest rate, then the arbitrage-free future price of a CDD contract at time $t \le \tau_1 \le \tau_2$ is given by:

$$e^{-r(\tau_2-t)} E_Q \left[\int_{\tau_1}^{\tau_2} \max(0, T(\tau) - c)d\tau - F_{\text{CDD}}(t, \tau_1, \tau_2) | F_t \right]$$

$$= 0 \tag{22}$$

and since F_{CDD} is F_i adapted, we derive the price of a CDD futures to be

$$F_{\text{CDD}}(t, \tau_1, \tau_2) = E_Q \left[\int_{\tau_1}^{\tau_2} \max(0, T(\tau) - c)\mathrm{d}\tau | F_t \right]. \tag{23}$$

Using the Girsanov's Theorem, under the equivalent measure Q, we have that

$$\mathrm{d}W(t) = \mathrm{d}B(t) - \theta(t)\mathrm{d}t \tag{24}$$

and note that $\sigma(t)$ is bounded away from zero. Hence, by combining (20) and (24) the stochastic process of the temperature in the risk-neutral probability Q is

$$\mathrm{d}T(t) = \mathrm{d}S(t) + (\kappa(t)(T(t) - s(t)) + \sigma(t)\theta(t))\mathrm{d}t + \sigma(t)\mathrm{d}W(t) \tag{25}$$

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 by historical data. More specifically $\theta(t)$ can be calculated by looking the market price of contracts. The value that makes the price of the model fits the market price is the market price of risk. Using Ito formula, the solution of (25) is

$$T(t) = S(t) + e^{\int_0^t \kappa(u)\mathrm{d}u}(T(0) - S(0))$$
$$+ e^{\int_0^t \kappa(u)\mathrm{d}u} \int_0^t \sigma(s)\theta(s) e^{-\int_0^s \kappa(u)\mathrm{d}u} \mathrm{d}s.$$
$$+ e^{\int_0^t \kappa(u)\mathrm{d}u} \int_0^t \sigma(s) e^{-\int_0^s \kappa(u)\mathrm{d}u} \mathrm{d}B(s) \tag{26}$$

By replacing this expression to (23), we find the price of future contract on CDD index at time t where $0 \le t \le \tau_1 \le \tau_2$. Following the notation of [48], we have the following proposition.

Proposition 1 *The CDD future price for $0 \le t \le \tau_1 \le \tau_2$ is given by*

$$F_{\text{CDD}}(t, \tau_1, \tau_2) = E_Q \left[\int_{\tau_1}^{\tau_2} \max(T(s) - c)\mathrm{d}s | F_t \right]$$
$$= \int_{\tau_1}^{\tau_2} v(t, s)\Psi \left(\frac{m(t, s)}{v(t, s)} \right) \mathrm{d}s \tag{27}$$

where,

$$m(t, s) = S(s) + e^{\int_t^s \kappa(z)\mathrm{d}z} \widetilde{T}(t)$$
$$+ e^{\int_t^s \kappa(z)\mathrm{d}z} \int_t^t \sigma(u)\theta(u) e^{-\int_t^u \kappa(z)dz} \mathrm{d}u - c \tag{28}$$

$$v^2(t, s, x) = e^{2\int_t^s \kappa(z)\mathrm{d}z} \int_t^t \sigma^2(u)\theta(u) e^{-2\int_t^u \kappa(z)\mathrm{d}z} \mathrm{d}u \tag{29}$$

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

Proof From (23) and (26) we have that

$$F_{\text{CDD}}(t, \tau_1, \tau_2) = E_Q\left[\int_{\tau_1}^{\tau_2} \max(T(s) - c)\mathrm{d}s | F_t\right]$$

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

$$E_Q\left[\int_{\tau_1}^{\tau_2} \max(T(s) - c)|F_t\right] = \int_{\tau_1}^{\tau_2} E_Q[\max(T(s) - c)|F_t]\mathrm{d}s$$

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

$$E_Q[T(s)|F_t] = S(s) + e^{\int_t^s \kappa(z)\mathrm{d}z}\widetilde{T}(t)$$
$$+ e^{\int_t^s \kappa(z)\mathrm{d}z} \int_t^s \sigma(u)\theta(u)e^{-\int_t^u \kappa(z)\mathrm{d}z}\mathrm{d}u$$

$$\mathrm{Var}_Q[T(s)|F_t] = e^{2\int_t^s \kappa(z)\mathrm{d}z} \int_t^s \sigma^2(u)\theta(u)e^{-2\int_t^u \kappa(z)\mathrm{d}z}\mathrm{d}u.$$

Hence, $T(S) - c$ is normally distributed with mean given by $m(t, s)$ and variance given by $v^2(t, s)$ and the proposition follows by standard calculations using the properties of the normal distribution.

5 Conclusions

This paper proposes and implements a modeling and forecasting approach for temperature based weather derivatives, which is an extension of [43] and [44]. Here, the speed of mean reversion parameter is considered time varying and it is modeled by a WN. WNs combine WA and NNs in one step. The basic contributions of this paper can be summarized as follows:

First, our results show that the waveform of the activation function and the wavelet decomposition that is performed in the hidden layer of the WN provide a better fit to the temperature data. The WNs were constructed and applied in order to fit the daily average temperature in 13 cities. In [44], a linear model was used to model the temperature in Paris while in [43] a NN were applied to fit the temperature in the same location.

Secondly, we compared our model with a similar linear model, and the improvement using a nonconstant speed of mean reversion was measured. In [44], a NN was used in order to model the seasonal mean and variance while in [43] WA was also used in order to capture the seasonalities in the mean and variance.

Thirdly, our approach, in contrast to [43, 44], was validated in a 2-month (ahead) out-of-sample forecast period. The proposed method was compared against two methods, often cited in the literature and widely used by market practitioners, in forecasting CAT and cumulative HDDs indices. The absolute relative errors produced by the WN are compared against the original B–B model and HBA. Our results indicate that the WN approach significantly outperforms the other methods. More precisely, the WN forecasting ability is better than B–B and HBA in 11 times out of 13. Our results indicate that HBA is accurate only when the value of the index is close to the historical mean while when the value of the index deviates from its average historical value, then HBA produces large estimation errors that subsequently lead to large pricing errors. On the other hand, the WN approach gives significant smaller errors even in cases where the temperature deviates significantly from its historical mean. Moreover, testing the fitted residuals of B–B we observe that the normality hypothesis can be (almost always) rejected. Hence, B–B may induce large errors in both forecasts and pricing.

Finally, we provide the pricing equations for temperature futures on cumulative CDDs and HDDs indices, when the speed of mean reversion is time depended while in [43] the pricing equations of the CAT index were presented.

In our model, the number of sinusoids in (8) and (9) (representing the seasonal part of the temperature and the variance of residuals) are chosen according to [47]. Further research in alternative approaches may improve the fitting of the original data and enhance forecasting accuracy.

Another important aspect of all approaches is the length of the forecasting horizon. Currently, meteorological forecasts more than 10 days ahead are considered inaccurate. Hence, it is quite important to develop models than can accurately predict daily average temperatures for larger horizons. Concluding, it would be extremely interesting to compare our approach that utilizes WNs with SMVs.

References

1. Cao L, Hong Y, Fang H, He G (1995) Predicting chaotic time series with wavelet networks. Phys D 85:225–238
2. Fang Y, Chow TWS (2006) Wavelets based neural network for function approximation. Lecture notes in computer science, vol 3971, pp 80–85
3. Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia
4. Mallat SG (1999) A wavelet tour of signal processing. Academic Press, San Diego
5. Kobayashi M (1998) Wavelets and their applications. SIAM, Philadelphia

6. Donoho DL, Johnstone IM (1998) Minimax estimation via wavelet shrinkage. Ann Stat 26:879–921
7. Donoho DL, Johnstone IM (1994) Ideal spatial adaption by wavelet shrinkage. Biometrika 81:425–455
8. Zhang Q (1997) Using wavelet network in nonparametric estimation. IEEE Trans Neural Netw 8:227–236
9. Pati YC, Krishnaprasad PS (1993) Analysis and synthesis of feedforward neural networks using discrete affine wavelet transforms. IEEE Trans Neural Netw 4:73–85
10. Zhang Q, Benveniste A (1992) Wavelet networks. IEEE Trans Neural Netw 3:889–898
11. Bernard C, Mallat S, Slotine J-J (1998) Wavelet interpolation networks. In: Proceedings of ESANN '98, Bruges, Belgium, pp 47–52
12. Bashir Z, El-Hawary ME (2000) Short term load forecasting by using wavelet neural networks. In: Proceedings of Canadian conference on electrical and computer engineering, pp 163–166
13. Benaouda D, Murtagh G, Starck J-L, Renaud O (2006) Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting. Neurocomputing 70:139–154
14. Gao R, Tsoukalas HI (2001) Neural-wavelet methodology for load forecasting. J Intell Robot Syst 31:149–157
15. Ulugammai M, Venkatesh P, Kannan PS, Padhy NP (2007) Application of bacterial foraging technique trained artificial and wavelet neural networks in load forecasting. Neurocomputing 70:2659–2667
16. Yao SJ, Song YH, Zhang LZ, Cheng XY (2000) Wavelet transform and neural networks for short-term electrical load forecasting. Energy Convers Manag 41:1975–1988
17. Chen Y, Yang B, Dong J (2006) Time-series prediction using a local linear wavelet neural wavelet. Neurocomputing 69:449–465
18. Cristea P, Tuduce R, Cristea A (2000) Time series prediction with wavelet neural networks. In: Proceedings of 5th seminar on neural network applications in electrical engineering, Belgrade, Yugoslavia, pp 5–10
19. Kadambe S, Srinivasan P (2006) Adaptive wavelets for signal classification and compression. Int J Electron Commun 60:45–55
20. Pittner S, Kamarthi SV, Gao Q (1998) Wavelet networks for sensor signal classification in flank wear assessment. J Intell Manuf 9:315–322
21. Subasi A, Alkan A, Koklukaya E, Kiymik MK (2005) Wavelet neural network classification of eeg signals by using AR model with MLE pre-processing. Neural Netw 18:985–997
22. Zhang Z (2007) Learning algorithm of wavelet network based on sampling theory. Neurocomputing 71:224–269
23. Allingham D, West M, Mees AI (1998) Wavelet reconstruction of nonlinear dynamics. Int J Bifurcat Chaos 8:2191–2201
24. Oussar Y, Dreyfus G (2000) Initialization by selection for wavelet network training. Neurocomputing 34:131–143
25. Oussar Y, Rivals I, Presonnaz L, Dreyfus G (1998) Training wavelet networks for nonlinear dynamic input output modelling. Neurocomputing 20:173–188
26. Postalcioglu S, Becerikli Y (2007) Wavelet networks for nonlinear system modelling. Neural Comput Appl 16:434–441
27. Billings SA, Wei H-L (2005) A new class of wavelet networks for nonlinear system identification. IEEE Trans Neural Netw 16:862–874
28. Jiao L, Pan J, Fang Y (2001) Multiwavelet neural network and its approximation properties. IEEE Trans Neural Netw 12:1060–1066
29. Szu H, Telfer B, Kadambe S (1992) Neural network adaptive wavelets for signal representation and classification. Opt Eng 31:1907–1916
30. Wong K-W, Leung AC-S (1998) On-line successive synthesis of wavelet networks. Neural Process Lett 7:91–100
31. Khayamian T, Ensafi AA, Tabaraki R, Esteki M (2005) Principal component-wavelet networks as a new multivariate calibration model. Anal Lett 38:1447–1489
32. Vapnik V, Golowich SE, Smola A (1996) Support vector method for function approximation. Regression estimation, and signal processing. Adv Neural Inf Process Syst 9:281–287
33. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Knowl Discov Data Min 121–167
34. Scholkopf B, Sung K, Burges C, Girosi F, Niyogi P, Poggio T, Vapnik V (1997) Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Trans Signal Process 45:1–8
35. Morris CW, Autret A, Boddy L (2001) Support vector machines for identifying organisms —a comparison with strongly partitioned radial basis function networks. Ecol Model 146:57–67
36. Cao LJ, Tay EHF (2003) Support vector machine with adaptive parameters in financial time series forecasting. IEEE Trans Neural Netw 14:1506–1518
37. Kecman V (2001) Learning and soft computing : support vector machines, neural networks, and fuzzy logic models. MIT Press, Cambridge
38. Challis S (1999) Bright forecast for profits. Reactions, June edition
39. Hanley M (1999) Hedging the force of nature. Risk Prof 1:21–25
40. Ceniceros R (2006) Weather derivatives running hot. Bus Insur 40(32):24–35
41. WRMA (2009) Celebrating 10 years of weather risk industry growth, vol 2009. http://www.wrma.org/pdf/WRMA_Booklet_%20FINAL.pdf. Accessed Aug 2009
42. Jewson S, Brix A, Ziehmann C (2005) Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations. Cambridge University Press, Cambridge
43. Zapranis A, Alexandridis A (2008) Modelling temperature time dependent speed of mean reversion in the context of weather derivative pricing. Appl Math Finance 15:355–386
44. Zapranis A, Alexandridis A (2009) Weather derivatives pricing: modelling the seasonal residuals variance of an Ornstein-Uhlenbeck temperature process with neural networks. Neurocomputing 73:37–48
45. Alaton P, Djehince B, Stillberg D (2002) On modelling and pricing weather derivatives. Appl Math Finance 9:1–20
46. Benth FE, Saltyte-Benth J (2005) Stochastic modelling of temperature variations with a view towards weather derivatives. Appl Math Finance 12:53–85
47. Benth FE, Saltyte-Benth J (2007) The volatility of temperature and pricing of weather derivatives. Quant Finance 7:553–561
48. Benth FE, Saltyte-Benth J, Koekebakker S (2007) Putting a price on temperature. Scand J Stat 34:746–767
49. Brody CD, Syroka J, Zervos M (2002) Dynamical pricing of weather derivatives. Quant Finance 2:189–198
50. Cao M, Wei J (2000) Pricing the weather. Risk weather risk special report, energy and power risk management, pp 67–70
51. Cao M, Wei J (1999) Pricing weather derivatives: an equilibrium approach. Working Paper. The Rotman Graduate School of Management, The University of Toronto, Toronto
52. Cao M, Wei J (2003) Weather derivatives: a new class of financial instruments. Working Paper, vol 2003. University of Toronto, Toronto
53. Cao M, Li A, Wei J (2004) Watching the weather report. Can Invest Rev Summer 27–33
54. Benth FE (2003) On arbitrage-free pricing of weather derivatives based on fractional Brownian motion. Appl Math Finance 10:303–324
55. Bellini F (2005) The weather derivatives market: Modelling and pricing temperature. Faculty of Economics, vol. Ph.D. University of Lugano, Lugano, Switzerland
56. Zapranis A, Alexandridis A (2007) Weather derivatives pricing: modelling the seasonal residuals variance of an Ornstein-

Uhlenbeck temperature process with neural networks. In: EANN 2007, Thessaloniki, Greece

57. Zapranis A, Alexandridis A (2006) Wavelet analysis and weather derivatives pricing. Paper presented at 5th Hellenic Finance and Accounting Association (HFAA), Thessaloniki, Greece

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

59. Xu J, Ho DWC (2002) A basis selection algorithm for wavelet neural networks. Neurocomputing 48:681–689

60. Xu J, Ho DWC (2005) A constructive algorithm for wavelet neural networks. Lecture notes in computer science, pp 730–739

61. Zhang Q (1993) Regressor selection and wavelet network construction. Technical report, INRIA

62. Cannon M, Slotine J-JE (1995) Space-frequency localized basis function networks for nonlinear system estimation and control. Neurocomputing 9:293–342

63. Zapranis A, Refenes AP (1999) Principles of neural model identification, selection and adequacy: with applications to financial econometrics. Springer, London

64. Davis M (2001) Pricing weather derivatives by marginal value. Quant Finance 1:1–4

65. Dornier F, Queruel M (2000) Caution to the wind. Weather risk special report, energy power risk management, pp 30–32

66. Moreno M (2000) Riding the temp. Weather derivatives, FOW special support

67. Caballero R, Jewson S, Brix A (2002) Long memory in surface air temperature: detection modelling and application to weather derivative valuation. Clim Res 21:127–140

68. Geman H, Leonardi M-P (2005) Alternative approaches to weather derivatives pricing. Manag Finance 31:46–72