Preprint - Published in the Applied Mathematical Journal, 14 (4) pp. 355-386, 2008

Modeling Temperature Time-Dependent Speed of Mean Reversion in the Context of Weather Derivatives Pricing

A. Zapranis[1], A. Alexandridis[2]

Department of Accounting and Finance, University of Macedonia of Economics and Social Studies, 156 Egnatia St., P.O. 54006, Thessaloniki, Greece.

Abstract—In this paper, in the context of an Ornstein-Uhlenbeck temperature process we use neural networks to examine the time dependence of the speed of the mean reversion parameter α of the process. We estimate non-parametrically with a neural network a model of the temperature process and then we compute the derivative of the network output w.r.t. the network input, in order to obtain a series of daily values for α. To our knowledge, this is done for the first time, and it gives us a much better insight in temperature dynamics and in temperature derivative pricing. Our results indicate strong time dependence in the daily values of α but no seasonal patterns. This is important, since in all relevant studies so far, α was assumed to be constant. Furthermore, the residuals of the neural network provide a better fit to the normal distribution, when compared with the residuals of the classic linear models which are being used in the context of temperature modeling (where α is constant). It follows, that by setting the mean reversion parameter to be a function of time we improve the accuracy of the pricing of the temperature derivatives. Finally, we provide the pricing equations for temperature futures, when α is time dependent.

Keywords—Neural networks, Weather derivatives pricing.

[1] Corresponding Author. Phone: +30 2310 891690, Fax: +30 2310 891689. e-mail: zapranis@uom.gr.
[2] Phone: +30 2310 891631, e-mail: aalex@uom.gr

1. Introduction

Since their inception in 1996, weather derivatives have known a substantial growth. The first parties to arrange for, and issue weather derivatives in 1996, were energy companies, which after the deregulation of energy markets were exposed to weather risk. In September 1999, the Chicago Mercantile Exchange (CME) launched the first exchange traded weather derivatives. In 2004, the notional value of CME weather derivatives was $2.2 billion and grew nine-fold 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 is still more active than the exchange, so the bid-ask spreads are quite large. 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.

A weather derivative is a financial instrument that has a payoff derived from variables such as temperature, snowfall, humidity and rainfall. However, it is estimated that 98-99% of the weather derivatives now traded are based on temperature. This is not surprising since, it is estimated that 30% of the US economy is affected by temperature (CME, 2005). The electricity sector is especially sensitive to the temperature. According to Li and Sailor (1995) and Sailor and Munoz (1997), temperature is the most significant weather factor explaining electricity and gas demand in the United States. The impact of temperature in both electricity demand and price has been considered in many papers, including Henley and Peirson (1998), Peirson and Henley (1994) and Engle et al (1992). 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).

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 than other financial derivatives in that the underlying weather index (HDD, CDD, CAT, etc.) cannot be traded. Furthermore, the corresponding market is relatively illiquid. Consequently, since weather derivatives cannot be cost-efficiently replicated with other weather derivatives, arbitrage pricing cannot directly apply to them. The weather derivatives market is a classic incomplete market, because the underlying weather variables are not tradable. When the market is incomplete, prices cannot be derived from the no-arbitrage condition, since it is not possible to replicate the payoff of a given contingent claim by a controlled portfolio of the basic securities. Consequently, the classical Black-Scholes-Merton pricing approach, which is based on no-arbitrage arguments, cannot be directly applied. And market incompleteness is not the only reason for that; weather indices do not follow random walks (as the Black & Scholes approach assumes) and the payoffs of weather derivatives are determined by indices, which are average quantities, whilst the Black-Scholes payoff is determined by the value of the underlying exactly at the maturity date of the contract (European options).

There are several approaches for dealing with incomplete markets. One of them is to introduce the ‘market price of risk’ for the particular type of the incomplete market, namely a ‘factor model’, where there are some non-traded underlying objects. Since, weather derivatives are path depended they are very similar to the average Asian option and similar analytical pricing approaches can be used in this case too. A characteristic example is the approach of Geman and Yor (1993), which used Bessel processes to obtain an exact analytical expression of the Laplace transformation in time of the option price.

A pricing methodology for weather derivatives that is widely used in insurance is the actuarial (or insurance) method. It is based on statistical analysis and it is less applicable in contracts with underlying variables that follow recurrent, predictable patterns. Since, this is the case for most of the weather derivatives contracts, actuarial analysis is not considered the most appropriate pricing approach unless the contract is written on rare weather events such as extreme cold or heat.

Another approach for weather derivatives pricing, is performing simulations based on historical data, known as historical Burn analysis. That is, computing the average payoff of the weather derivatives in the past n years. The central assumption of this method is that the historical record of weather contracts payoffs gives a precise illustration of the distribution of the potential payoffs (Dischel, 1999). If weather risk is calculated as the payoffs standard deviation, then the price of the contract will be $P(t) = D(t, T) \times (\mu \pm \alpha \times \sigma)$, where $D(t, T)$ is the discount factor from contract maturity T to the pricing time t, μ is the historical average payoff, σ is the historical standard deviation of payoffs and a is a positive number denoting risk tolerance. However, since the weather processes are not stationary and this approach does not incorporate forecasts, it is bound to be biased and inaccurate. In fact, the historical Burn analysis is considered as the simplest pricing method in terms of implementation, and the most probable to cause large pricing errors.

In contrast to the previous methods, a dynamic model can be used which directly simulates the future behavior of temperature. Using models for daily temperatures can, in principle, lead to more accurate pricing than modeling temperature indices. In the process of calculating the temperature index, such as HDD, as a normal or lognormal process, a lot of information is lost (e.g., HDD is bounded by zero). On the other hand, deriving an accurate model for the daily temperature is not a straightforward process.

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

In this paper, we address the problem of pricing the European CAT options. For this purpose we extent the mean-reverting process with seasonality in the level and volatility proposed by Benth and Saltyte-Benth (2007a) - a generalization of (Dornier and Querel, 2000) which is descretized in the form of an AR(1) model. We estimate non-parametrically a non-linear AR(1) model with a neural network. This removes the constraint of a constant mean reverting parameter. By computing the derivative of the network output w.r.t. the network input, we take a series of daily values for the mean reversion parameter for a period of 30 years for the city of Paris. Analytical expressions for the various network derivatives are given by Zapranis and Refenes (1999).

It is important to mention here, that up to date the mean reversion parameter was assumed constant in all relevant studies. However, our findings indicate exactly the opposite. The daily variation of the value of the mean reversion parameter is quite high. The non-linear neural model which encapsulated this time dependency provides a much better fit to the temperature data than the classic linear alternative. The implications in the accuracy of the pricing process of this type of derivatives are obvious. Furthermore, the complexity of the pricing equations is not being increased significantly by using a time dependent mean reversion parameter. Below, first we describe the basic steps of our analysis and then the organization of the rest of the paper.

Given the temperature model, the first step is to identify and remove from the temperature series the (possible) trend and the non-stationary seasonal cycle, hoping that what is left will be stationary. This is usually done by modeling the seasonal variations as deterministic and the same every year (seasonally stationary). The stochastic variability of the temperature is then moved entirely from the seasonal cycle into the residuals.

In modeling the seasonal cycle deterministically, there are several approaches. The discrete Fourier transform (DTF) is considered to be the most accurate, since, in principle at least, removes the seasonal cycle both in the mean and in the variance. For a detailed discussion on this subject see Jewson and Brix (2005). However, recently Zapranis and Alexandridis (2006, 2007) proposed a novel approach in modeling the seasonal cycle which is an extension of the DFT approach. Since small misspecifications in a dynamical model can lead to large pricing errors, we incorporate wavelet analysis in the modeling process in order to calibrate our model. The fundamental idea behind wavelets is to analyze according to scale. Wavelet analysis is an extension of the Fourier transform, which superposes sines and cosines to represent

other functions. Wavelet analysis decomposes a general function or signal into a series of (orthogonal) basis functions, called wavelets, with different frequency and time locations. The wavelet analysis procedure adopts a particular wavelet function, called a mother wavelet. Temporal analysis is performed with a contracted high-frequency version of the mother wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same mother wavelet. Because the original signal can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients. A particular feature of the analyzed signal can be identified with the positions of the wavelets into which it is decomposed. Results of the wavelet transform can be presented as a contour map in frequency-time plane (spectrogram), allowing the changing spectral composition of non-stationary signals to be measured and compared. As illustrated in Donoho et al (1995) the wavelet approach is very flexible in handling very irregular data series. Wavelet analysis has the ability to represent highly complex structures without knowing the underlying functional form, which is of great benefit in economic and financial research. In order to capture the seasonality of the volatility of the temperature we use a truncated Fourier series. The specific terms of the Fourier series are being selected on the basis of the results of a wavelet analysis of the temperature. As we demonstrate here, wavelet analysis is very useful in offering guidance as to which terms of the Fourier series to select. The wavelet decomposition brings out the structure of the underlying temperature series as well as trends, periodicities, singularities or jumps that could not be observed originally (Alaton *et al*., 2000 and Davis, 2001). Our approach was tested in 40 years of temperature data collected from Paris (from 1960 to 2000), and the improvement in

terms of distributional properties was found to be significant, Zapranis & Alexandridis (2007). The same approach is used in this paper.

Once the trend and the seasonal cycle in the mean and the variance have being removed, one has to investigate the distributional properties of the residuals (anomalies) of the temperature process. To the extent that this part of the modeling approach and the initial temperature process are accurate, the residuals must follow a normal distribution with mean zero and standard deviation of one at all times of the year. However, often the hypothesis of normality is rejected, Benth & Saltyte-Benth (2005).

As it is shown in the next section, the temperature process can be written as an AR(1) model after removing the linear trend and the seasonal component. Or, as we propose here as a non-linear AR(1) fitted non-parametrically with a neural network, which allows us to examine the time structure of the speed of the mean reversion of the temperature process. We show that temperature is a mean reverting process where the speed of mean reversion depends on time. Our findings were compared against a linear AR(1) process with a constant parameter.

Since, there is time dependency in the variance of the residuals we have to extract that variance. In doing so, we group the residuals in 365 groups, each group corresponding to a particular day of the year. Each group comprises 30 observations. Each observation corresponds to a different year. Then we take the average for each group. Using those 365 values we model the residual variance with a neural network having as inputs the harmonics corresponding to the seasonal cycles of the residuals, identified by a second wavelet analysis.

The rest of the paper is organized as follows. In section 2, we give an introduction to wavelet analysis. In section 3, describe the process used to model the average daily temperature in Paris. The calibration of the temperature model is done based on the

results of wavelet analysis. We also estimate and then remove the linear trend and the seasonality component. In section 4, we estimate a linear AR(1) model with a constant speed of mean reversion parameter. Then we model the seasonal residual variance, again using the wavelet analysis approach. The analysis is repeated in section 5 where we estimate the AR(1) model non parametrically using a neural network. We extract daily values of the speed of mean reversion of our process. Then we estimate the seasonal residual variance, again using wavelet analysis. In section 6, we give the analytical expressions for pricing temperature futures with a time dependent speed of mean reversion parameter. Finally, in section 7, we conclude.

2. Introduction to Wavelet Analysis: Examples of Its Application to Simulated Time-Series

2.1. Fourier Transform and Wavelet Analysis

Wavelet analysis is a mathematical tool used in various areas of research. Especially, during the last years wavelets are frequently used in order to analyse time-series, data and images. Time-series are represented by local information such as frequency, duration, intensity and time-position and by global information such as the mean states over different time periods. Both global and local information is needed for a correct analysis of a signal. The Wavelet transform (WT) is a generalization of Fourier and windowed Fourier transforms (FT and WFT).

FT breaks down a signal into a linear combination of constituent sinusoids of different frequencies; hence the FT is decomposition on a frequency by frequency basis. However, in transforming to the frequency domain, time information is lost. When looking at a FT of a signal, it is impossible to tell when a particular event took place. This is a serious drawback if the signal properties change a lot over time, i.e., if they contain non-stationary or transitory characteristics: drift, trends, abrupt changes, and beginnings and ends of events. These characteristics are often the most important part of the signal, and FT is not suited to detecting them.

In order to achieve a sort of compromise between frequency and time, FT was expanded in Windowed Fourier Transform. WFT uses a window across the time series and then uses the FT of the windowed series. This is a decomposition of two parameters, time and frequency. However, since the window size is fixed with respect to frequency, WFT cannot capture events that appear outside the width of the window. Many signals require a more flexible approach that is one where we can vary the window size to determine more accurately either time or frequency.

Wavelet Transform, on the other hand is localized in both time and frequency and overcomes the fixed time-frequency partitioning. The new time-frequency partition is long in time in low- frequencies and long in frequency in high-frequencies. This means that the WT has good frequency resolution for low-frequency events and good time resolution for high-frequency events. Also, the WT adapts itself to capture features across a wide range of frequencies. Consequently the assumption of stationarity can be avoided.

In addition, wavelets have the ability to decompose a signal or a time-series in different levels. As a result, this decomposition brings out the structure of the underlying signal as well as trends, periodicities, singularities or jumps that cannot be

observed originally. Wavelets can prove to be a valuable tool for analyzing a wide range of time-series and they have already been used with success in image processing, signal de-noising, density estimation, signal and image compression and time-scale decomposition. Wavelet techniques are being used in finance, for detecting the properties of quick variation of values.

2.2. Wavelets

A wavelet is a waveform of effectively limited duration that has an average value of zero. A *wavelet family* is a set of orthogonal basis functions generated by dilation and translation of a compactly supported *scaling function*, ϕ (or *father wavelet*), and a *wavelet function*, ψ (or *mother wavelet*). The wavelet family consists of *wavelet children* which are dilated and translated forms of a mother wavelet:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2.1}$$

where, *a* is the *scale* or *dilation* parameter and *b* is the *shift* or *translation* parameter. The value of the scale parameter determines the level of stretch or compression of the wavelet. The term $1/\sqrt{a}$ normalizes $\left\|\psi_{a,b}(t)\right\| = 1$. In most cases, we will limit our choice of *a* and *b* values by using a discrete set, because calculating wavelet coefficients at every possible scale is computationally intensive. However, if we choose only a subset of scales and translations based on powers of two (the *dyadic* lattice) then our analysis will be much more efficient and just as accurate. We obtain such an analysis from the Discrete Wavelet Transform (DWT). The wavelet family is taken from a double indexed regular lattice:

$$\left\{\left(a_j, b_k\right) = \left(p^j, kqp^j\right) : j,k \in Z\right\} \tag{2.2}$$

where the parameters p and q denote the step sizes of the dilation and the translation parameters. For $p = 2$ and $q = 1$ we have the standard dyadic lattice:

$$\left\{\left(a_j, b_k\right) = \left(2^j, k2^j\right) : j, k \in Z\right\} \tag{2.3}$$

The scaling function ϕ generates for each $j \in \mathbf{Z}$ the sets $V_j = \text{span}\left\{\phi_{j,k}, k \in \mathbf{Z}\right\}$, where $\mathbf{Z}$ denotes the set of integers and

$$\phi_{j,k}(t) = 2^{-j/2}\phi\left(2^{-j}t - k\right), \quad j, k \in \mathbf{Z} \tag{2.4}$$

The basis wavelet functions are usually of the form:

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right), \quad j, k \in \mathbf{Z} \tag{2.5}$$

It follows from above that there is a sequence $\{h_k\}$ (where h_k is a *scaling filter* associated with the wavelet) such that $\sum |h_k|^2 = 1$ and

$$f(t) = \sqrt{2}\sum_{k=0}^{\infty} h_k f\left(2t - k\right) \tag{2.6}$$

where ϕ is normalized so that $\int_{-\infty}^{\infty} \phi(t)dt = 1$.

When $\{h_k\}$ is finite, a compactly supported scaling function is the solution to the above *dilation equation*. The wavelet function is defined in terms of the scaling function as:

$$\psi(t) = \sqrt{2}\sum_{k=0}^{\infty} g_k \phi\left(2t - k\right) \tag{2.7}$$

where $\int_{-\infty}^{\infty} \psi(t)dt = 0$ and $g_k = (-1)^{k+1} h^{1-k}$ is a *wavelet filter*.

Then $W_j = \text{span}\left\{\psi_{j,k}, k \in \mathbf{Z}\right\}$ is the orthogonal complement of V_j in V_{j+1}, $\forall j \in \mathbf{Z}$.

Over the years a substantial number of wavelet functions have been proposed in the literature, (Mallat, 1999 and Daubechies, 1992). In this study we use the Daubechies wavelet family.

2.3. Signal Reconstruction

Representing a signal as a function $T(t)$, the Continuous Wavelet Transform (CWT) of this function comprises the *wavelet coefficients* $C(a,b)$, which are produced through the convulsion of a mother wavelet function $\psi(t)$ with the analyzed signal $T(t)$:

$$C(a,b) = \int_{-\infty}^{+\infty} T(t)\psi\left(\frac{t-b}{a}\right)dt \tag{2.8}$$

The wavelet coefficients are localized in time and frequency. We term *approximations* the high scale, low frequency components and *details* the low scale, high frequency components. Given the wavelet coefficients we can perform continuous synthesis of the original signal:

$$T(t) = \frac{1}{K_\psi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{a^2}\left[C(a,b)\right]\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dadb \tag{2.9}$$

The DWT of the signal function comprises the *wavelet coefficients* $C(j,k)$, which are produced through the convulsion of a mother wavelet function $\psi_{j,k}(t)$ with the analyzed signal $T(t)$:

$$C(j,k) = \int_{-\infty}^{+\infty} T(t)\psi_{j,k}(t)\,dt \tag{2.10}$$

Thus, the discrete synthesis of the original signal is:

$$T(t) = \sum_{j\in Z}\sum_{k\in Z} C(j,k)\psi_{j,k}(t) \tag{2.11}$$

At each level j, we build the j-level approximation a_j, or approximation at level j, and a deviation signal called the j-level detail d_j, or detail at level j. We can consider the original signal as the approximation at level 0, denoted by a_0. The words approximation and detail are justified by the fact that a_1 is an approximation of a_0 taking into account the low frequencies of a_0, whereas the detail d_1 corresponds to the high frequency correction. For detailed expositions on the mathematical aspects of wavelets we refer to, for example Mallat (1999), Wojtaszczyk (1997) and Daubechies (1992).

3. Dynamic Modeling Of The Temperature Process.

Many different models have been proposed in order to describe the dynamics of a temperature process. The common assumptions in all these models concerning the temperature are the following: it follows a predictable cycle, it moves around a seasonal mean, it is affected by global warming, it appears to have autoregressive changes and its volatility is higher in the winter than in summer.

Early models were using AR(1) processes or continuous equivalents (see for example Alaton *et al.* (2000), Cao and Wei (2000), Davis (2000)). Other researchers (e.g., Dornier and Querel, 2000, Moreno, 2000) have suggested versions of a more general ARMA(p,q) model. However, it has been shown 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, Caballero *et al.* (2002). In order to deal

with this problem, more complex models were proposed, with a characteristic example being the model of Brody *et al.* (2002) , which is an Ornstein-Uhlenbeck process. This model was further extended, at first by replacing the noise part of the process (Brownian) by a fractional Brownian noise and then by a Levy process (Benth and Saltyte-Benth, 2005).

Our analysis is based on the model of Benth & Saltyte-Benth, where the temperature is expressed as a mean reverting Ornstein-Uhlenbeck process, i.e.

$$dT(t) = dS(t) + \kappa\big(T(t) - S(t)\big) + \sigma(t)dB(t) \tag{3.1}$$

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 *σ(t)* is the daily volatility of temperature variations.

In Benth & Saltyte-Benth (2007a) both *S(t)* and $\sigma^2(t)$ are being modeled as a truncated Fourier series, i.e.:

$$S(t) = a + bt + a_0 + \sum_{i=1}^{I_1} a_i \sin(2i\pi(t - f_i)/365) + \sum_{j=1}^{J_1} b_j \cos(2j\pi(t - g_j)/365) \tag{3.2}$$

$$\sigma^2(t) = c + \sum_{i=1}^{I_2} c_i \sin(2i\pi t/365) + \sum_{j=1}^{J_2} d_j \cos(2j\pi t/365) \tag{3.3}$$

From the Ito formula an explicit solution for (3.1) can be derived:

$$T(t) = S(t) + e^{\kappa}\left(T(0) - S(0)\right) + \int_0^t \sigma(s)e^{\kappa(t-s)}dB(t) \tag{3.4}$$

According to this representation $T(t)$ is normally distributed at t and it is reverting to a mean defined by $S(t)$. Previous works (Benth and Saltyte-Benth 2007a, Alaton *et al.* 2000) suggest that $S(t)$ is modeled by a sinusoid with a period of one year. In this paper, the exact specification of models (3.2) and (3.3) is decided based on the results of wavelet analysis of the temperature series.

In this section we derive the characteristics and dynamics of the daily temperature of the city of Paris, France. The data consists average daily temperatures of 30 years (1971-2000). The distribution of the data is not normal, indicating a temperature process that is generally hard to model.

In order to identify the number of terms I_1, J_1 in (3.2) and I_2, J_2 in (3.3) we decompose the temperature series using a wavelet transform. Lau *et al.* (1995) confirmed seasonalities in the temperature series with a period greater than one year. This conclusion was also reached in Zapranis and Alexandridis (2006) when the Daubechies 11 wavelet at level 11 was used for the decomposition of 100 years of the average daily temperature time-series of Paris. Specifically, in these articles, wavelet analysis captured the following dynamics of temperature in Paris: first, an upward trend exists in the temperature reflecting the global and urban warming. This is clearly shown in figure 1, in approximations 8 to 11. Also a series of cycles affects the dynamics of temperature. An one year cycle exists in the first seven approximations, as expected. Moreover, cycles of 2, 4, 8 and 13 years also exist and affect the temperature dynamics (details 9,10,11 and approximation 11 respectively). Also, wavelet analysis captures a product of two sinusoids, with a period of 1 and 7 years respectively (detail 8). Finally, the lower details reflect the noise part of the time-series. A closer inspection of the noise part reveals seasonalities, which will be extracted later on.

A discrete approximation to (3.4), which is the solution to the mean reverting Ornstein-Uhlenbeck process (3.1), is:

$$T(t+1)-T(t)=S(t+1)-S(t)+\kappa\big(T(t)-S(t)\big)+\sigma(t)\big(B(t+1)-B(t)\big) \tag{3.5}$$

which can be written as:

$$T_{t+1}=\alpha T_t+\sigma(t)\varepsilon_t \tag{3.6}$$

where

$$T(t)=T(t)-S(t) \tag{3.7}$$

$$\alpha=1+\kappa \tag{3.8}$$

In order to estimate (3.6) we need first to remove the trend and seasonality components from the average temperature series.

Firstly, we quantify the upward trend indicated by the results of the wavelet analysis by fitting a linear regression to the temperature data. The regression is statistically significant with intercept a =11.171 and slope b=0.00010562. The corresponding standard errors are 0.095717 and 0.00000908 and the t-statistics are 116.4089 and 8.641278 while both p-values are zero. Subtracting the trend form the original data we obtain the de-trended temperature series.

After removing the linear trend from the data we use wavelet analysis to indentify the seasonal part. The results of the wavelet analysis indicate that the seasonal part of the temperature takes the following form:

$$\begin{aligned} S(t) = a + bt + a_0 + b_1 \sin(2\pi(t - f_1)/365) + b_2 \sin(2\pi(t - f_2)/(2\cdot 365)) \\ + b_3 \sin(2\pi(t - f_3)/(13\cdot 365)) + b_4(1 + \sin(2\pi(t - f_4)/(7\cdot 365)))\sin(2\pi t/365) \\ + b_5 \sin(2\pi(t - f_5)/(8\cdot 365)) + b_6 \sin(2\pi(t - f_6/(4\cdot 365))\cdot 365)) \end{aligned} \quad (3.9)$$

The estimated parameters of the above model are as follows: α_0 = -0.0008, b_1 = -7.6994, b_2 = 0.1317, b_3 = 0.0469, b_4 = -0.2743, b_5 = -0.3445, b_6 = 0.0796, f_1 = -73.2644, f_2 = 95.0642, f_3 = -640.2319, f_4 = 183.1090, f_5 = -13.1151 and f_6 = -134.5803. The mean of the residuals 5.9091e-009 and the standard deviation is 3.3708. Next the temperature series is de-seasonalized by removing *S(t)*. The approximation of the trend and the remaining seasonal part of the temperature is done in two different steps. Thus the inclusion of a_0 represents part of the trend that the simple linear fitting did not capture. The detailed wavelet analysis of the seasonal part *S(t)* can be found in Zapranis & Alexandridis (2007).

4. The Linear Regression Approach.

Our temperature data consisted of 30 years up to 2000 of de-trended and de-seasonalized daily average temperatures, $T(t)$, from the city of Paris. We separate the data in 3 groups, each group corresponding to one decade. First we remove the 29th of February to from 3 equal groups of 3650 data points each. Then we use the linear AR(1)

model, proposed by Benth & Saltyte-Benth (2007a). We estimate the parameter α for the AR(1) model for each decade. For all three decades the constant was found to be zero, as it was expected, while the reversion parameter α takes the values: $\alpha_1 = 0.797$, $\alpha_2 = 0.7989$, $\alpha_3 = 0.8005$ (the subscript indicates the decade); these values are also statistically significant ($t = 79.35, 79.88, 80.33$). For all three decades the adjusted R^2 is over 0.63 and F is over 6297. For simplicity we will refer only to the last decade. The results for all decades can be found in Table 1.

For the 3rd (last) decade the distributional statistics of the residuals of the AR(1) model (3.6), indicate a significant deviation from the normal distribution (Fig. 2). There is a negative skewness (-0.174117), positive kurtosis (3.021718) and the value of the Jarque-Bera statistic is 18.50932. The p-value is less than 0.05, so that the hypothesis of normal distribution has to be rejected. The results for the 1st and 2nd decade are very similar.

Previous works (Benth & Saltyte-Benth 2007a, 2005, Zapranis and Alexandridis, 2006, 2007) suggest seasonal variance in the residuals of the AR(1) model. Figures 3 and 4 show the autocorrelation of the residuals and squared residuals respectively. The autocorrelation of the residuals is significant in the first three lags with lag 1 to be positive while lag 2 and 3 are negative (Fig. 3). As Benth *et al.* (2007b) suggest the autocorrelation may be modeled by GARCH process. The autocorrelation of the squared residuals indicates time dependency in the variance of the residuals (Fig. 4). Especially in fig. 4, we can observe the seasonal variation although is not so clear as in Benth *et al.* (2007b) due to the fact that part of it was removed by the wavelet analysis.

Since, for the residuals $e(t)$ of AR(1) model it is true that

$$e(t) = \sigma(t)\varepsilon(t) \tag{4.1}$$

where $\varepsilon(t)$ are i.i.d. N(0,1), we can extract the seasonal variance of the residuals as follows: Firstly, we group the residuals in 365 groups, comprising 10 observations each (each group corresponds to a single day of the year). Then, by taking the average of the 10 squared values we obtain the variance for that day. That is, we assume that

$$\sigma^2(365+t) = \sigma^2(t) \tag{4.2}$$

where t = 1, …, 3,650 (for each decade).

In deciding which terms of a truncated Fourier series to use in order to model the variance $\sigma^2(t)$, we perform again a wavelet analysis, which indicates the presence of five cycles within $\sigma^2(t)$. The wavelet decomposition of the seasonal variance is shown in Fig.5. Approximation a_7 and details d_7, d_6, d_5 suggest an one-year cycle, a half-year cycle, a 1/4 of a year cycle, a 1/9 of a year cycle and a 1/18 of a year cycle, respectively. We model accordingly the variance $\sigma^2(t)$, as follows:

$$\begin{aligned}\sigma^2(t) = {} & c_0 + c_1 \sin(2\pi t/365) + c_2 \sin(4\pi t/365) + c_3 \sin(8\pi t/365) + c_4 \sin(18\pi t/365) \\ & + c_5 \sin(36\pi t/365) + d_1 \cos(2\pi t/365) + d_2 \cos(4\pi t/365) + d_3 \cos(8\pi t/365) \\ & + d_4 \cos(18\pi t/365) + d_5 \cos(36\pi t/365)\end{aligned} \tag{4.3}$$

The values of the estimated parameters of (4.3) for the 3rd decade are: c_0 = 4. 0968, c_1 = 0.4412, c_2 = -0.1352, c_3 = 0.1913, c_4 = 0.0898, c_5 =0.1733, d_1 = 0.0527, d_2 = 0.5410, d_3 = -0.056 d_4 = 0.1229 and d_5 = -0.0059.

The estimated parameters for all three decades can be shown in Table 2 (bottom). Figure 6 shows the empirical and the fitted seasonal variance of the AR(1)

model while in figure 7 it is shown the autocorrelation of the squared residuals after dividing out the seasonal variance. It is clear that the seasonal variance was removed successfully. Note that although the seasonal variance has the same number of estimated parameters as in Benth and Saltyte-Benth (2007a) the selected cycles are different and resulted from wavelet analysis. Correct modeling of the seasonal variance results to correct pricing of weather derivates. As we show in the last section the price of weather derivatives is sensitive to the variance σ^2, hence, correct modeling is essential.

After dividing out the seasonal variance of the residuals we left with the noise part of the temperature signal. According to our theoretical model we expect to have a normally distributed noise. By examining the statistics of the noise part (Fig. 8) we have to reject the hypothesis of normal distribution. For the last decade α_3 = 0.8005. The standard deviation of the remaining noise part is 1.0007 and the mean is 0.0253. The Jarque-Bera is 14.22022 and its p-value is 0.000817. We conclude to the same result by observing the QQ-plot (Fig. 9) of the residuals after dividing out the seasonal variance. In Table 1, we can see the distributional statistics for all three decades. From the table we conclude that only in first decade we marginally accept the hypothesis of normality.

The findings of Benth and Saltyte-Benth (2007a) for the Stockholm temperature series are very similar. Although, they did not use wavelet analysis to calibrate their model, they had managed to remove seasonality from the residuals, but their distribution proved to be non-normal. They suggested that a more refined model would probably rectify this problem, but they did not proceed in estimating one. In an earlier paper regarding Norwegian temperature data, Benth and Saltyte-Benth (2005) suggested to model the residuals by a generalized hyperbolic distribution. However, as

the same authors comment the inclusion of a non-normal model leads to a complicated Levy process dynamics. Recently Benth *et al.* (2007b) proposed a continuous-time autoregressive process with lag *p* (CAR(p)-process) but their distribution proved again to be non-normal.

Zapranis and Alexandridis (2006) eastimated a number of alternatives to the original AR(1) model. In particular they estimated an ARMA(3,1) model, a long-memory homoscedastic ARFIMA model and a long-memory heteroscedastic ARFIMA-FIGARCH model. Their findings suggest that, increasing the model complexity and thus the complexity of theoretical derivations in the context of weather derivative pricing does not seem to be justified.

5. The Neural Network Approach.

The de-trended and de-seasonalized temperature series, T , can be modeled with an AR(1) process with a zero constant term, as shown in (3.6). In the context of such a model the mean reversion parameter α is typically assumed to be constant over time. In Brody *et al.* (2002) it is mentioned that in general α should be a function of time, but no evidence was presented. On the other hand, Benth and Saltyte-Benth (2005), using a dataset comprising 10 years of Norwegian temperature data, calculated mean annual values of α. They reported that their variation from year to year was not significant. They also investigated the seasonal structures in monthly averages of α and they reported that none was found. However, since to date, no one has computed daily values of the mean reversion parameter, since there is no obvious way to do this in the context of model (3.6). On the other hand, averaging techniques, in a yearly or monthly basis,

run the danger of filtering out too much variation and consequently presenting a distorted picture regarding the true nature of α. The impact of a false specification of α, on the accuracy of the pricing of temperature derivatives is significant, Alaton *et al.* (2000) .

In this section, we address that issue, by using a neural network (N.N.) to estimate non-parametrically relationship (3.6) and then estimate α as a function of time. By computing the derivative of the network output w.r.t. the network input we obtain a series of daily values for α. This is done for the first time, and it gives us a much better insight in temperature dynamics and in temperature derivative pricing. As we will see the daily variation of α is quite significant after all.

5.1 The Neural Networks Approach: Time Dependent Mean Reversion Parameter

Using neural networks we estimated non-parametrically the generalized version of (3.6), that is:

$$T(t+1) = \phi\Big(T(t)\Big) + e(t) \tag{5.1}$$

Once we have the estimator of the underlying function ϕ, then we can compute the daily values of α as follows:

$$\alpha(t) = dT(t+1)/dT(t) = d\phi/dT \tag{5.2}$$

The analytic expression for the neural network derivative $d\phi/dT$ can be found in Zapranis & Refenes (1999).

We estimate $\phi(\bullet)$ non-parametrically with the neural network $g(\bullet)$. Given an input vector x (the harmonics) and a set of weights w (parameters), the network response (output) $g_\lambda(x;w)$ is:

$$g_\lambda(\mathbf{x};\mathbf{w}) = \gamma\left(\sum_{j=1}^{\lambda} w_j^{[2]}\gamma\left(\sum_{i=1}^{m} w_{ij}^{[1]}x_i + w_{m+1,j}^{[1]}\right) + w_{\lambda+1}^{[2]}\right) \tag{5.3}$$

where, $w_{i,j}^{[1]}$ is a weight corresponding to the connection between the i^{th} input and the j^{th} hidden unit, $w_{m+1,j}^{[1]}$ is a bias term corresponding to the j^{th} hidden unit, $w_j^{[2]}$ is the weight of the connection between the j^{th} hidden unit and the output unit, and $w_{\lambda+1}^{[2]}$ is the bias term of the output unit, and the function $\gamma(\bullet)$ is a sigmoidal function.

For the 3rd decade the daily values of α (3,650 values) are depicted in Fig. 10. The corresponding frequency histogram is given in Fig. 11. The graphs for the 1st and 2nd decades are very similar. The relevant statistics for all three decades are given in Table 1.

It is clear, that the mean reversion parameter is not constant. On the contrary, its daily variation is quite significant; this fact naturally has an impact on the accuracy of the pricing equations and it has to be taken into account. Intuitively, it was expected $\alpha(t)$ not to be constant. If the temperature today is away from the seasonal average (a cold day in summer) then it is expected that the speed of mean reversion to be high; i.e. the difference of today and tomorrows temperature it is expected to be high. In contrast

if the temperature today is close to the seasonal variance we expect the temperature to revert to its seasonal average slowly.

Referring now to Fig. 10, we observe that the spread between the maximum and minimum value is quite high (0.8586 and 0.2303 correspondingly). The standard deviation is 0.0587 and the mean is 0.7573. We also observe, that there is an upper threshold in the values of $\alpha(t)$ (0.8376) which is rarely exceeded. This can also be seen in the frequency distribution of $\alpha(t)$ in Fig. 11. As expected, the average values of $\alpha(t)$ which were derived from the neural network models are actually very close to the values of α which were derived in the previous section. First we examine if $\alpha(t)$ is a stochastic process by itself. Both an Augmented Dickey-Fuller (A.D.F.) and Kwiatkowski-Phillips-Schmidt-Shin (K.P.S.S.) tests are used. The A.D.F. test statistic is -10.73455 with p-value=0 that leads to the rejection of the null hypothesis that $\alpha(t)$ has unit root. The K.P.S.S test statistic is 0.157341 and less that the critical values in 1%, 5% and 10% suggesting the acceptance of the null hypothesis that $\alpha(t)$ is stationary. Table 3 shows the A.D.F. test and table 4 the K.P.S.S test for all three decades. The histogram in figure 11 may suggest that the distribution of $\alpha(t)$ is bimodal. Hartigan's DIP statistic is a measure of departure from unimodality. If a distribution is unimodal then the DIP converges to zero otherwise converges to a positive constant (Hartigan & Hartigan, 1985). We found that the DIP is 0.0226 and the *p-value=0*. Hence, we reject the hypothesis that $\alpha(t)$ follows a unimodal distribution. Moreover figure 10 suggest seasonalities in the structure of $\alpha(t)$. The autocorrelation function of $\alpha(t)$ is shown in figure 12. A seasonality of a half year is clear in the autocorrelation function. Also the first 23 lags are statistically important. For a closer inspection we use wavelet analysis to decompose the signal of $\alpha(t)$. The decomposition is shown in figure 13. Wavelet

analysis confirms the half year cycle while reveals a major cycle of two years and a seasonality of one year.

The distributional statistics of the residuals of the neural network (Fig. 14), do not indicate a significant deviation from the normal distribution in contrast to the AR(1) model with constant parameter. There is a small negative skewness (-0.094027), positive kurtosis (3.031307) and the value of the Jarque-Bera statistic is 5.525856. The probability is 0.063107 (>0.05), indicating that we have to accept the normality hypothesis in contrast to the linear regression approach. Next we examine the autocorrelation function of the residuals. The autocorrelation of the residuals is significant in the first lag (Fig. 15), while the autocorrelation of the squared residuals indicates, although not clear, time dependency in the variance of the residuals (Fig. 16). In order to verify the existence of seasonality in the variance we use wavelet analysis.

In this section we will remove the seasonality in variance using the same approach as in previous section.

In deciding which terms of a truncated Fourier series to use in order to model the variance $\sigma^2(t)$, we perform again a wavelet analysis. As it was expected, our results are similar to the ones from section 4. Hence, we model the seasonal variance as in (4.3). The wavelet decomposition can be found in figure 17.

The values of the estimated parameters of for the 3rd decade when the speed of mean reversion is a function of time are: c_0 = 4. 3390, c_1 = 0.5095, c_2 = -0.0721, c_3 = 0.1883, c_4 = 0.1533, c_5 =0.1379, d_1 = 0.1260, d_2 = 0.6230, d_3 = -0.2897 d_4 = 0.0637 and d_5 = -0.0431. Again, the estimated parameters for all three decades can be found in Table 2 (top).

The empirical values of the variance of the residuals (365 values) together with the fitted variance can be seen in Fig. 18. We observe that the variance takes its highest

values during the winter months, while it takes its lowest values during early autumn. This is consistent to our initial hypothesis in section 3.

The standard deviation of the residuals is 2.0697, while the standard deviation of the remaining noise part is 0.9962 and its mean is 0.1165.

In Fig. 19, we can see the autocorrelation function of the squared residuals of the process, after dividing out the volatility from the residuals.

We observe that the seasonality has been removed, but there is still autocorrelation in the first lag while using the linear regression method we had three significant lags. Moreover, the Jarque-Bera statistic is reduced to 2.568741 with a p-value of 0.276825 leading to the acceptance of the hypothesis of normal distribution. Figure 20 presents the distribution statistics of the residuals after dividing out the volatility function of the residuals.

As we have seen, the hypothesis of normality was accepted only in the case of the neural models. In Fig. 21 and Fig. 9 we can see the normality plots for the residuals (after dividing out the seasonal variance) of the neural network and the AR(1) model for the 3rd decade. Clearly, in the case of the neural network the residuals provide a better fit to the normal distribution.

In Table 1, we can see the distributional statistics for all three decades. The neural networks approach always gives a smaller Jarque-Bera and higher p-value. Moreover, the skewness of the distributions corresponding to the N.N.s is always significantly lower, although, the kurtosis is lower only for the first decade. Finally, in the last decade the normality hypothesis using the linear regression is rejected (p=0.000817) while it is accepted using the neural network approach (p=0.276825).

Finally, the estimated parameters from the seasonal variance are presented in Table 2, for all three decades. The top table refers to the N.N. model, while the bottom table refers to the AR(1) model.

6. Temperature Derivative Pricing.

The list of traded contracts in weather derivatives market is extensive and constantly evolving. In Europe, CME weather contracts for the summer months are based on an index of Cumulative Average Temperature (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 one day. The value of a CAT index for the time interval $[\tau_1,\tau_2]$ is given by the following expression:

$$\int_{\tau_1}^{\tau_2} T(s)ds \tag{6.1}$$

where the temperature is measured in degrees of Celsius. In USA, CME weather derivatives are based on Heating Degree Days (HDD) or Cooling Degree Days (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., Daily HDD = max (0, base temperature – daily average temperature),

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

The base temperature is usually 65 degrees Fahrenheit in the US and 18 degrees Celsius in Europe and Japan. HDDs and CDDs are usually accumulated over a month or over a season. For the two Japanese cities, weather derivatives are based on the Pacific Rim index. The Pacific Rim index is simply the average of the CAT index over the specific time period. At the end of 2006, at CME were traded weather derivatives for 18 US cities[3], 9 European cities[4], 2 Japanese cities[5], as well as seasonal strip and frost contracts.

So far, we modeled the temperature using an Ornstein-Uhlenbeck process (as in Benth & Saltyte-Benth, 2007a) and we also used wavelet analysis to identify and filter out the seasonal component. Moreover, we have shown that the mean reversion parameter α in model (3.6) is characterized by significant daily variation. Recall that parameter α is connected to our initial model (3.1) with $\alpha=1+\kappa$ where κ is the speed of mean reversion. It follows that, the assumption of a constant mean reversion parameter introduces significant error in the pricing of weather derivatives. In this section, we give the pricing formula for a future contract written on the CAT index that incorporate the time dependency of the speed of the mean reversion parameter. First, we re-write (3.1) where parameter κ, now is a function of time t, $\kappa(t)$.

$$dT(t) = dS(t) + \kappa(t)\big(T(t) - S(t)\big) + \sigma(t)dB(t) \tag{6.2}$$

[3] Atlanta, Baltimore, Boston, Chicago, Cincinnati, Dallas, Des Moines, Detroit, Houston, Kansas City, Las Vegas, Minneapolis-St. Paul, New York, Philadelphia, Portland, Sacramento, Salt Lake City, Tucson.
[4] Amsterdam, Barcelona, Berlin, Essen, London, Madrid, Paris, Rome, Stockholm.
[5] Tokyo, Osaka.

From the Ito formula an explicit solution can be derived:

$$T(t) = S(t) + e^{\int_0^t \kappa(u)du}\left(T(0) - S(0)\right) + e^{\int_0^t \kappa(u)du}\int_0^t \sigma(s)e^{-\int_0^s \kappa(u)du} dB(s) \tag{6.3}$$

Note that $\kappa(t)$ is bounded away from zero and that the derivatives of $\kappa'(t)$ have been explicitly calculated in section 5.1.

Our aim is to give a mathematical expression for the CAT future price. It is clear that the weather derivative market is an incomplete market. 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 CAT contract at time $t \le \tau_1 \le \tau_2$ is given by:

$$e^{-r(\tau_2 - t)}\mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \mathrm{T}(\tau)d\tau - F_{CAT}(t,\tau_1,\tau_2) \mid \mathcal{F}_t\right] = 0 \tag{6.4}$$

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

$$F_{CAT}(t,\tau_1,\tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} \mathrm{T}(\tau)d\tau \mid \mathcal{F}_t\right] \tag{6.5}$$

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

$$dW(t) = dB(t) - \theta(t)dt \tag{6.6}$$

and note that $\sigma(t)$ is bounded away from zero. Hence, by combining equations (6.2) and (6.6) the stochastic process of the temperature in the risk neutral probability Q is:

$$dT(t) = dS(t) + \left(\kappa(t)\left(T(t) - s(t)\right) + \sigma(t)\theta(t)\right)dt + \sigma(t)dW(t) \tag{6.7}$$

where *θ(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 *θ(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 equation (6.7) is:

$$\begin{aligned} T(t) = S(t) + e^{\int_0^t \kappa(u)du}\left(T(0) - S(0)\right) + e^{\int_0^t \kappa(u)du}\int_0^t \sigma(s)\theta(s)e^{-\int_0^s \kappa(u)du}ds \\ + e^{\int_0^t \kappa(u)du}\int_0^t \sigma(s)e^{-\int_0^s \kappa(u)du}dB(s) \end{aligned} \tag{6.8}$$

By replacing this expression to (6.5) we find the price of future contract on CAT index at time t where $0 \le t \le \tau_1 \le \tau_2$.

Proposition 6.1 *The CAT future price for* $0 \le t \le \tau_1 \le \tau_2$ *is given by*

$$F_{CAT}(t, \tau_1, \tau_2) = \mathrm{E}_Q\left[\int_{\tau_1}^{\tau_2} T(s)ds \,|\mathcal{F}_t\right] = \int_{\tau_1}^{\tau_2} S(s)ds + I_1 + I_2 \tag{6.9}$$

where,

$$I_1 = \int_{\tau_1}^{\tau_2} e^{\int_t^s \kappa(z)dz} T(t)ds \tag{6.10}$$

$$I_2 = \int_t^{\tau_1}\int_{\tau_1}^{\tau_2} e^{\int_0^s \kappa(z)dz}\sigma(u)\theta(u)e^{\int_u^0 \kappa(z)dz}dsdu + \int_{\tau_1}^{\tau_2}\int_u^{\tau_2} e^{\int_0^s \kappa(z)dz}\sigma(u)\theta(u)e^{\int_u^0 \kappa(z)dz}dsdu \tag{6.11}$$

Proof. From equation (6.5) and (6.8) we have that:

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

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

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

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

$$= I_1 + I_2$$

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

$$I_2 = \int_{\tau_1}^{\tau_2}\int_t^s \sigma(u)\theta(u)e^{\int_u^s \kappa(z)dz} duds = \int_{\tau_1}^{\tau_2}\int_t^{\tau_2} 1_{[t,s]}(u)\sigma(u)\theta(u)e^{\int_u^s \kappa(z)dz} duds$$

where $1_{[t,s]}$ is zero outside the interval [t,s]. Then we can change the order of the integrals.

$$= \int_t^{\tau_2}\int_{\tau_1}^{\tau_2} 1_{[t,s]}(u)\sigma(u)\theta(u)e^{\int_u^s \kappa(z)dz} dsdu$$

Next we split the outer integral in two parts.

$$= \int_t^{\tau_1}\int_{\tau_1}^{\tau_2} 1_{[t,s]}(u)\sigma(u)\theta(u)e^{\int_u^s \kappa(z)dz} dsdu + \int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2} 1_{[t,s]}(u)\sigma(u)\theta(u)e^{\int_u^s \kappa(z)dz} dsdu$$

The second part is zero when s>u. Hence we can change the limits of the inner integral

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

or equivalently .

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

Practitioners often prefer easy to implement models than realistic ones. A classic example is the Black-Scholes equation. The above solution of the price of a CAT future is not easy to solve although is not much more complex than the pricing formulas presented by Benth & Saltyte-Benth (2007a). To calculate equation (6.9) is not a straightforward process. Alternatively we can calculate the price of a future contract on a CAT index using numerical methods. According to our analysis in section 5.1, $\alpha(t)$ and hence according to (3.8) $\kappa(t)$ can be modeled as a truncated Fourier series where:

$$\kappa(t) = e + \sum_{i=1}^{I_3} e_i \sin\left(2i\pi t / 365\right) + \sum_{j=1}^{J_3} g_j \sin\left(2j\pi t / 365\right) \tag{6.12}$$

hence, the integral $e^{\int_0^s \kappa(z)dz}$ can be easily calculated using numerical procedures.

7. Conclusions

In this paper, in the context of an Ornstein-Uhlenbeck temperature process we have used neural networks to examine the time dependence of the speed of the mean reversion parameter α of the process. By computing the derivative $dT(t+1)/dT(t)$ of the fitted neural model, we obtained daily values for α. To our knowledge, we are the first to have done so. Our results, indicate strong time dependence in the daily values of α but no seasonal patterns. We compared the fit of the residuals to the normal

distribution of two types of models. Neural networks, were α is a function of time, and AR(1) models, were α is constant. Generally, in the case of neural networks we have a better fit. It follows, that by setting the speed of mean reversion parameter to be a function of time we improve the accuracy of the pricing of the temperature derivatives. Also, since small misspecifications in dynamic models lead to large mispricing errors, we presented an approach to estimate and calibrate the seasonal component in both mean reversion and variance using wavelet analysis. Finally, we provided the pricing equations for temperature futures of a CAT index derivative, when α is time dependent.

Acknowledgements

We would like to thank the anonymous referee for the constructive comments that improved substantially the final version of this paper.

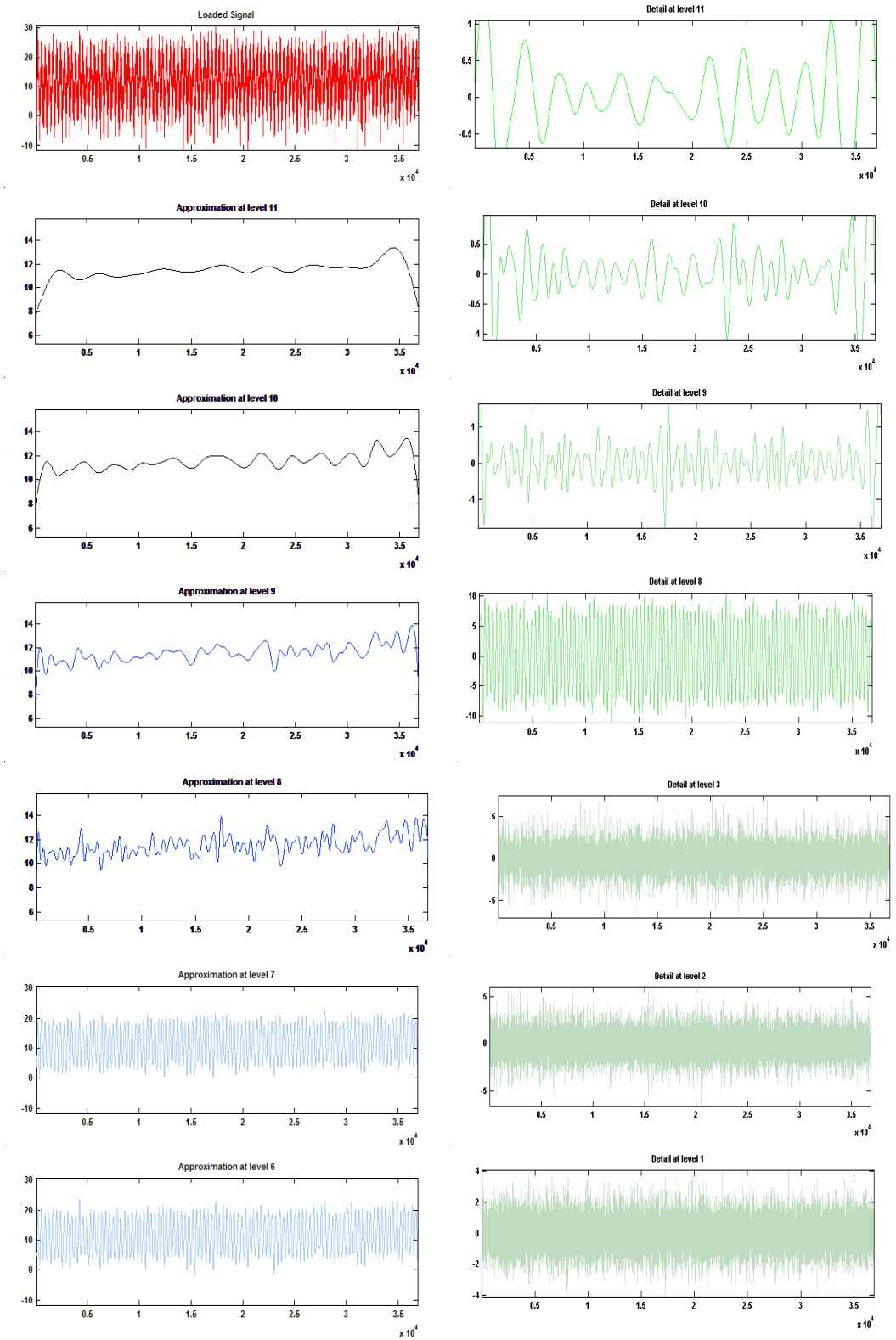

Figure 1. Daily temperature time-series (s) for Paris, France, approximations (a_j) and details (d_j) produced by the wavelet decomposition**.**

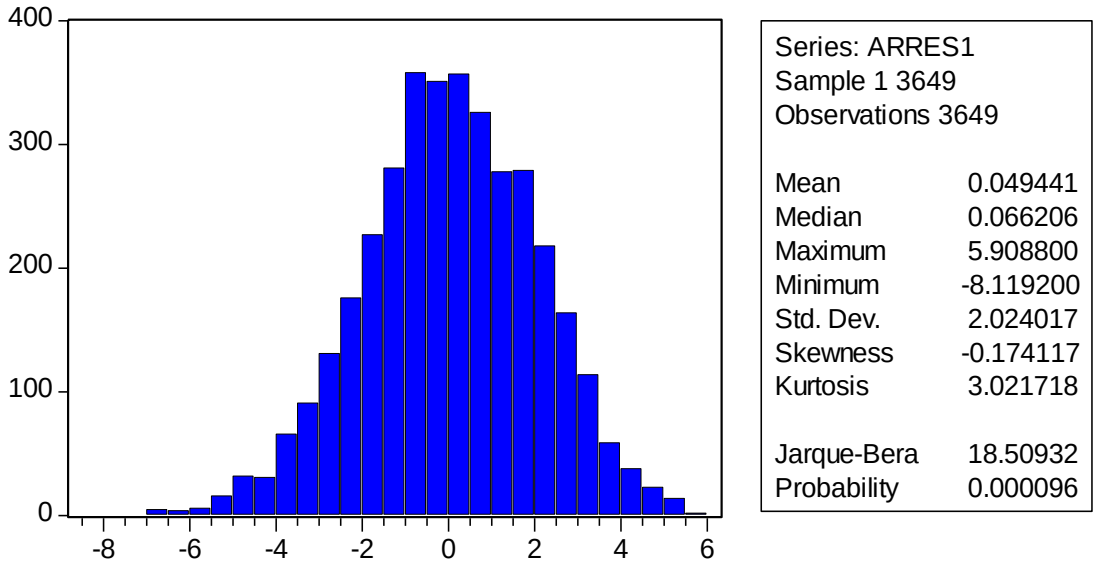

Fig.2. Distribution statistics of the residuals of the AR(1) model

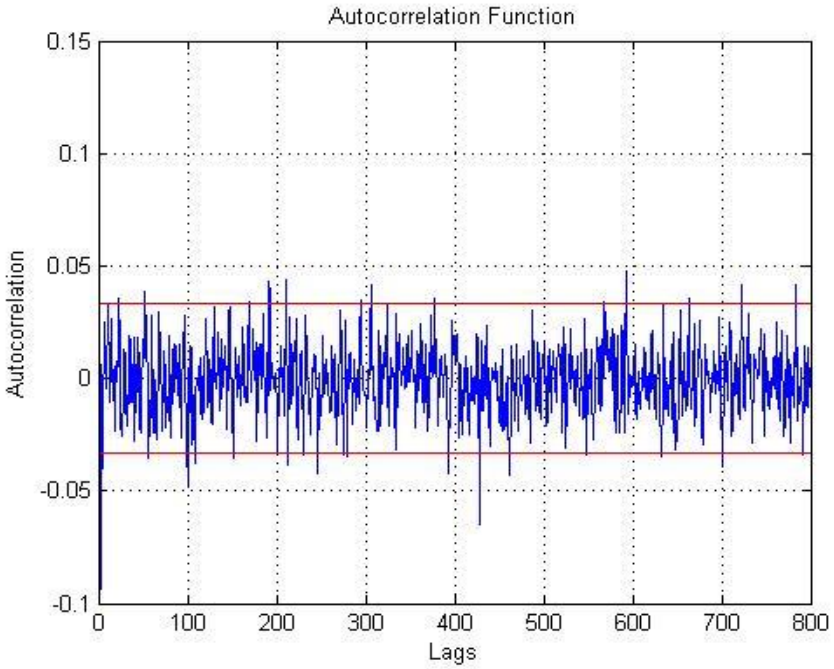

Fig.3. ACF of the residuals of the AR(1) model for the de-trended and de-seasonalized Paris average daily data

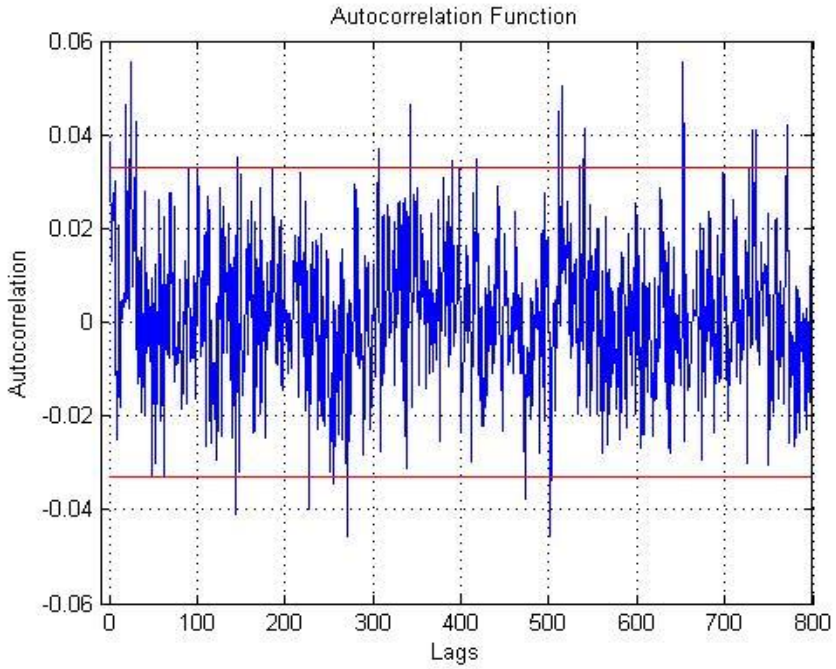

Fig.4. ACF of the squared residuals of the AR(1) model for the de-trended and de-seasonalized Paris average daily data.

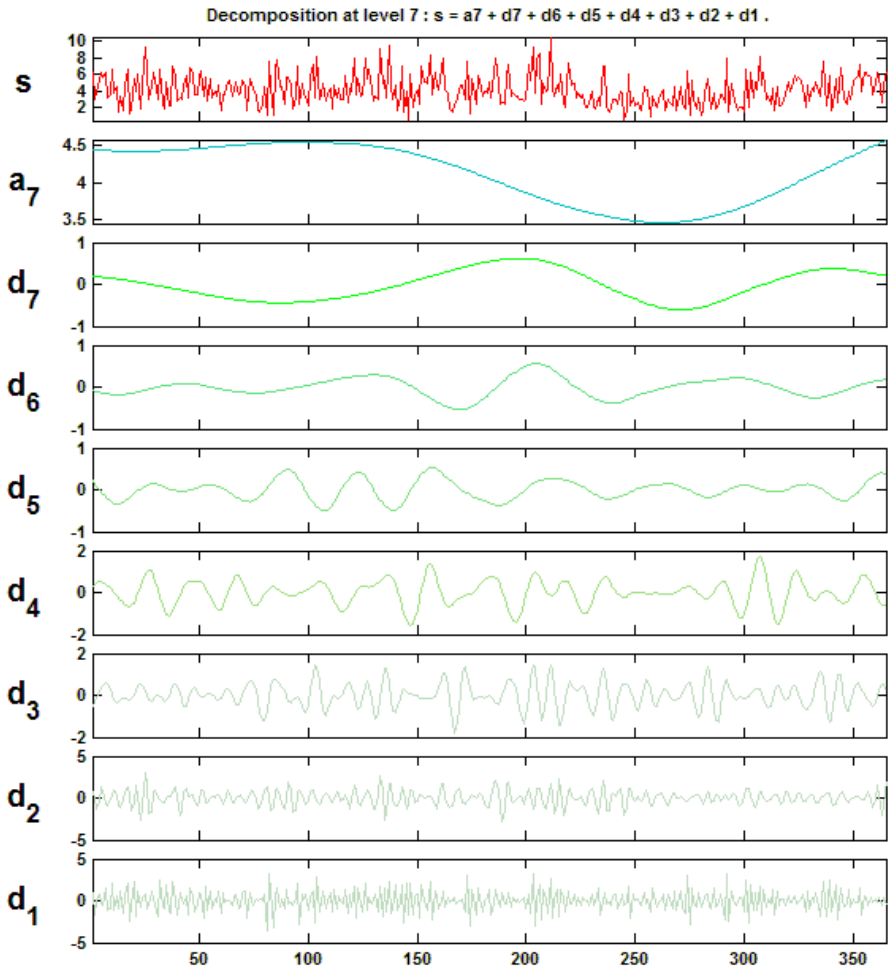

Fig.5. Wavelet decomposition of the averaged squared variance.

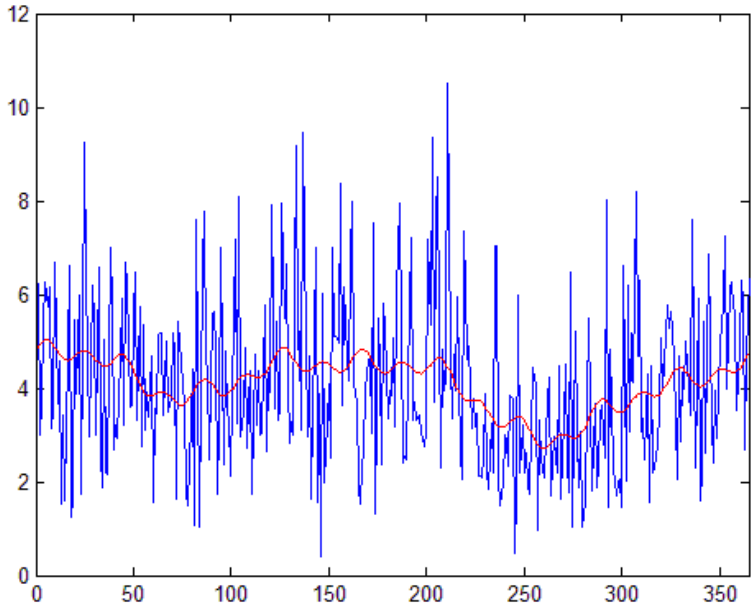

Fig.6. Empirical variance and fitted variance for the AR(1) model.

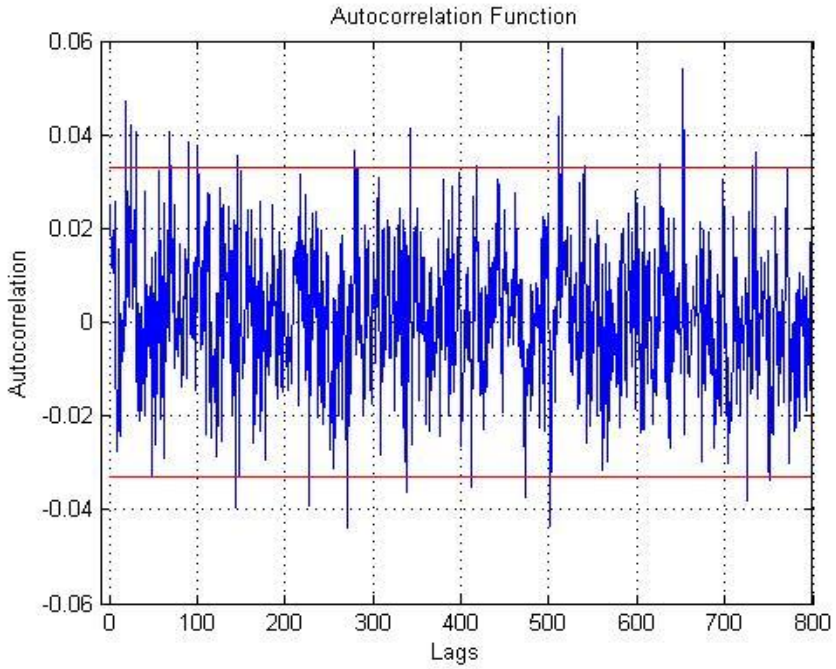

Fig.7. ACF of the squared residuals of the AR(1) model after dividing out the volatility function from the residuals.

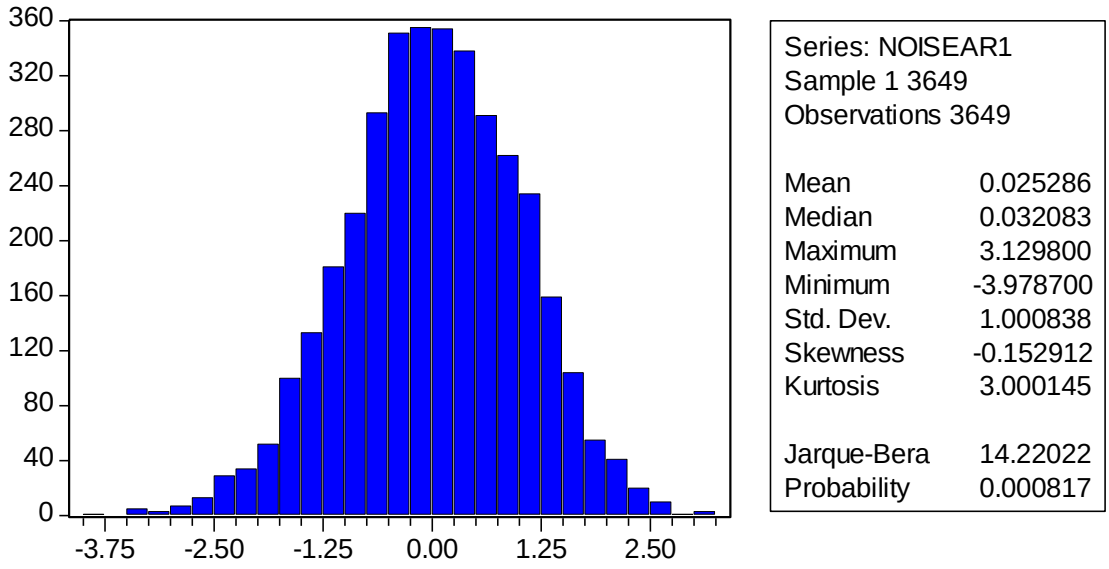

Fig.8. Distribution statistics of the residuals of the AR(1) model after dividing out the volatility function from the regression residuals

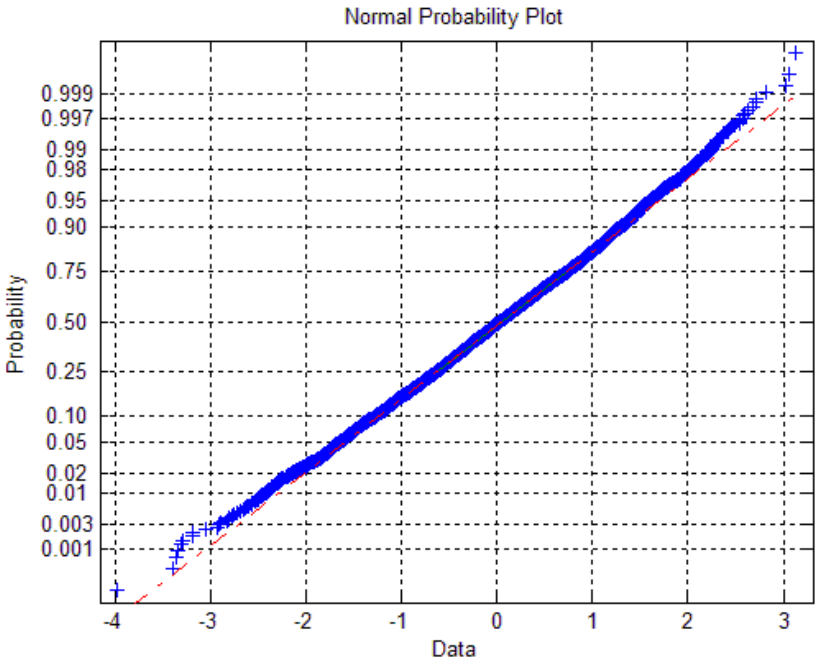

Fig.9. Normal probability plot of the of the residuals of the AR(1) model after dividing out the volatility function from the residuals

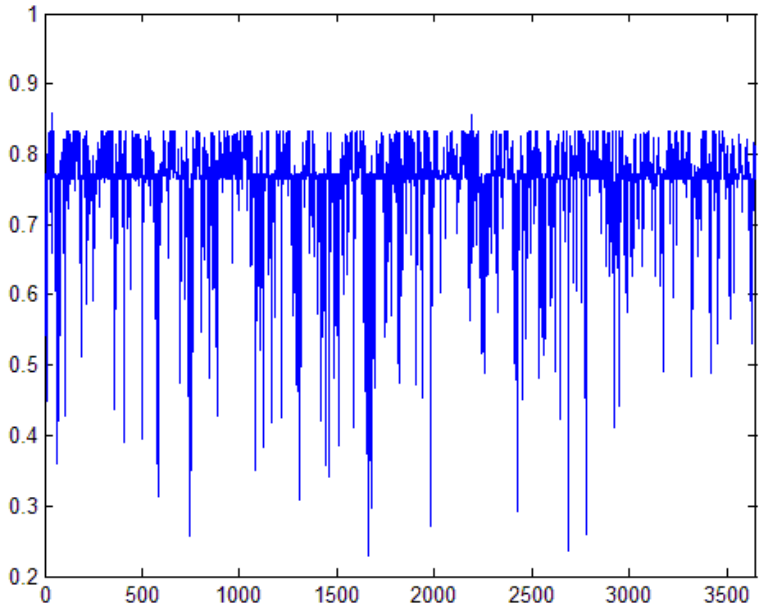

Fig.10. Daily variation of the mean reversion parameter α.

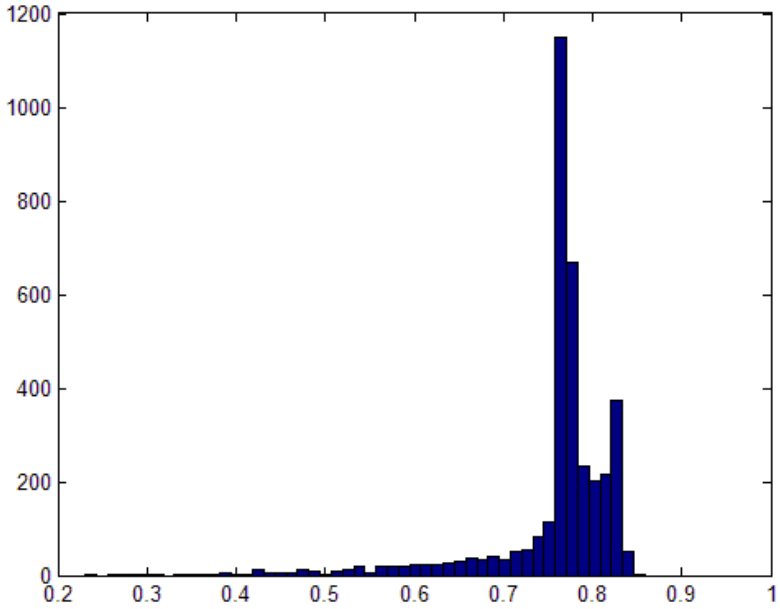

Fig.11. Frequency distribution of the mean reversion parameter α.

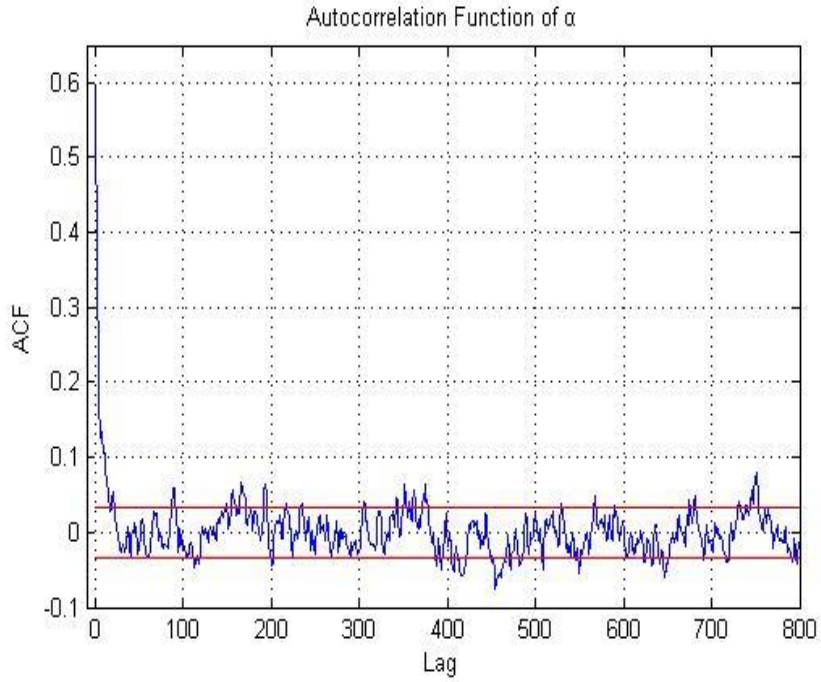

Fig.12. ACF of the speed of mean reversion *α(t)*

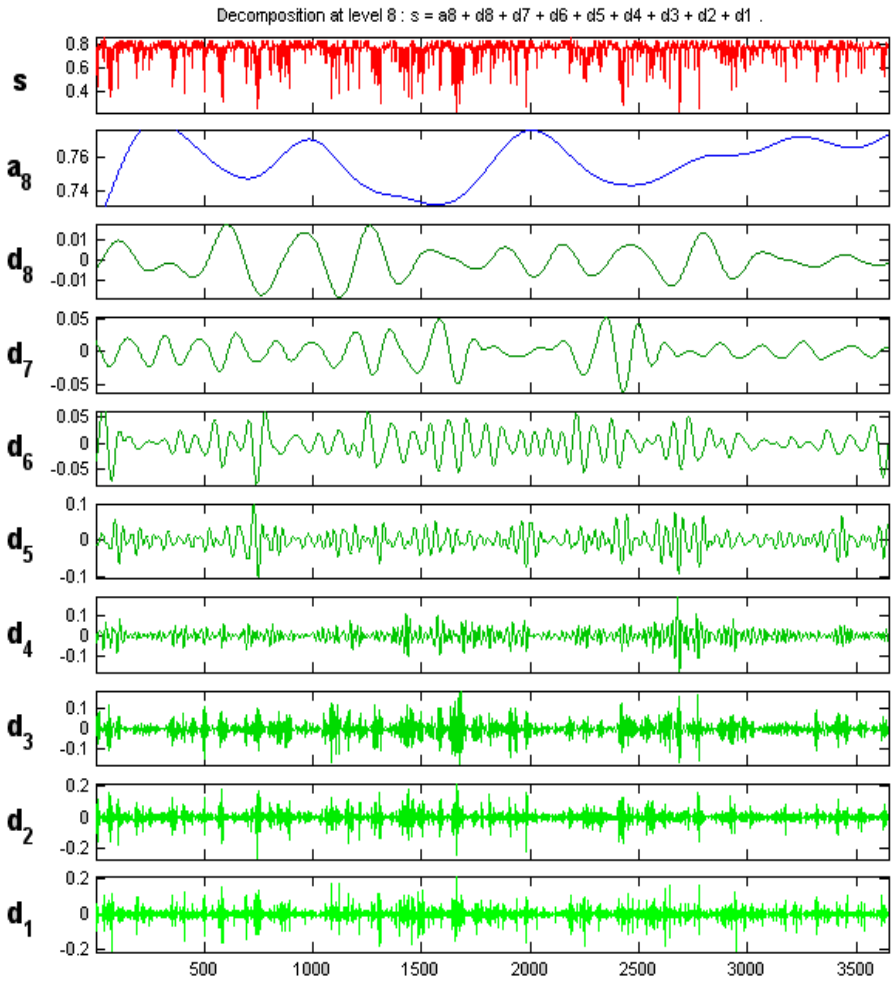

Fig.13. Wavelet decomposition of *α(t)*.

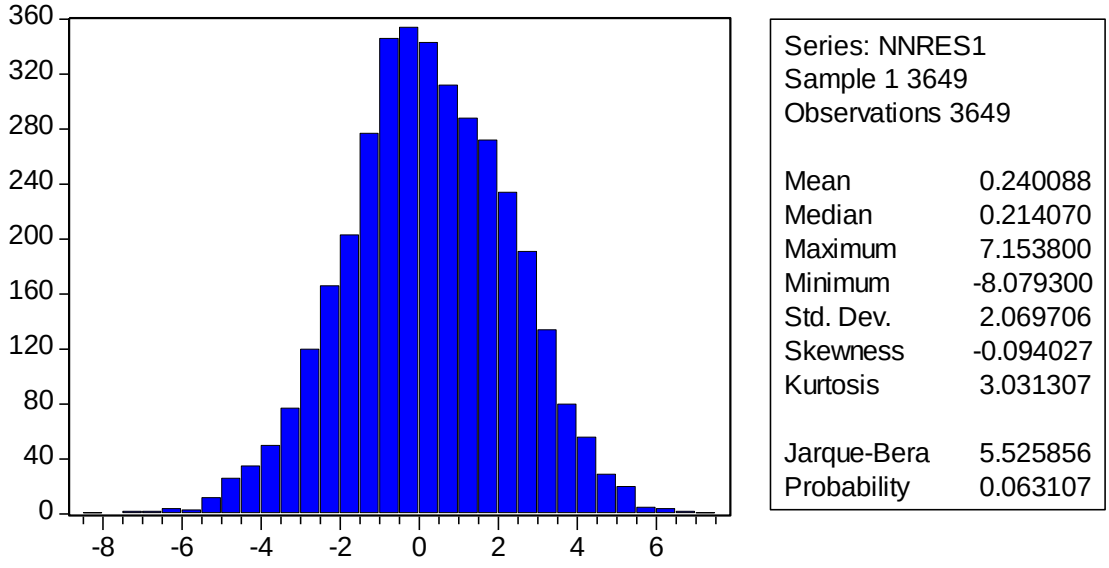

Fig.14. Distribution statistics of the residuals of the N.N. model

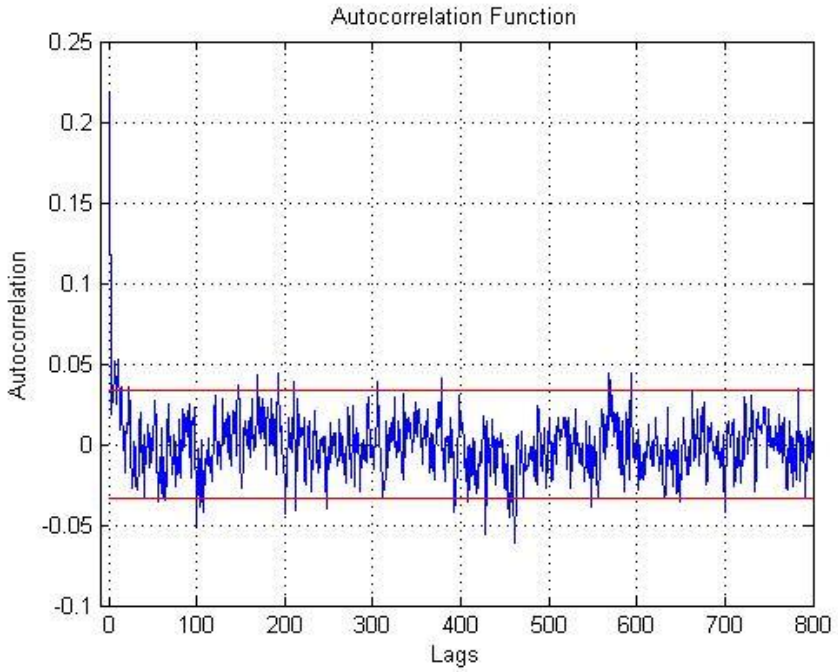

Fig.15. ACF of the residuals of the N.N. model for the de-trended and de-seasonalized Paris average daily data

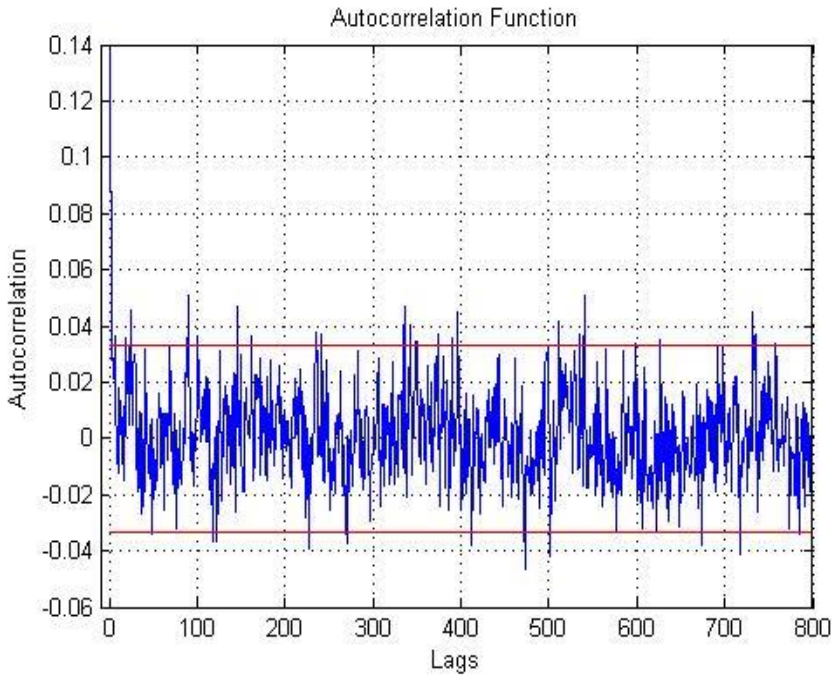

Fig.16. ACF of the squared residuals of the N.N. model for the de-trended and de-seasonalized Paris average daily data

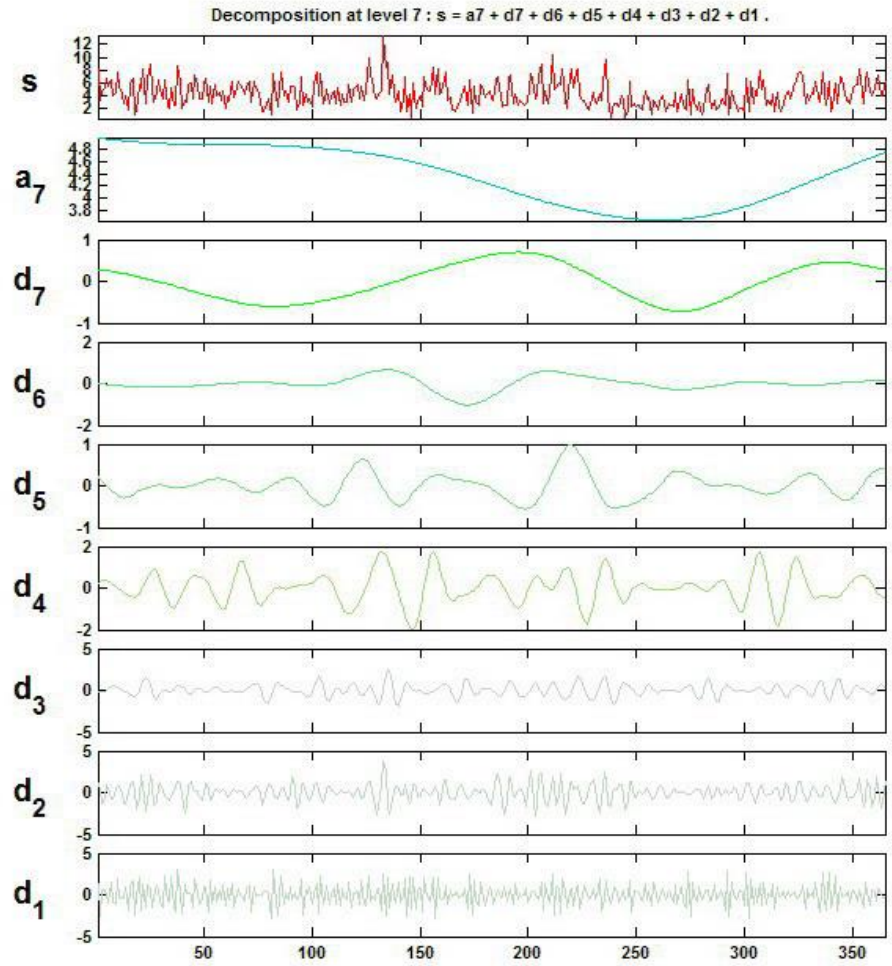

Fig.17. Wavelet decomposition of the averaged squared variance.

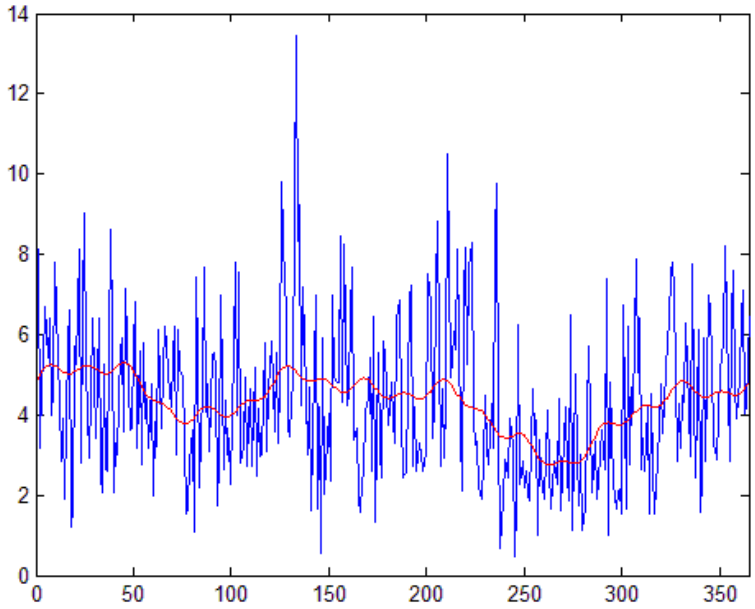

Fig.18. Empirical variance and fitted variance for the NN model.

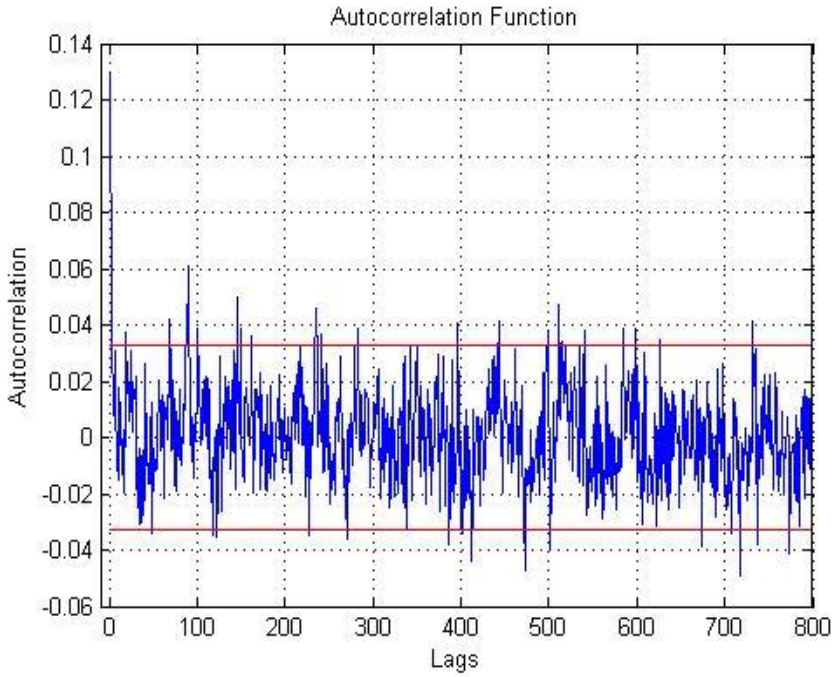

Fig.19. ACF of the squared residuals of the NN model after dividing out the volatility function from the residuals.

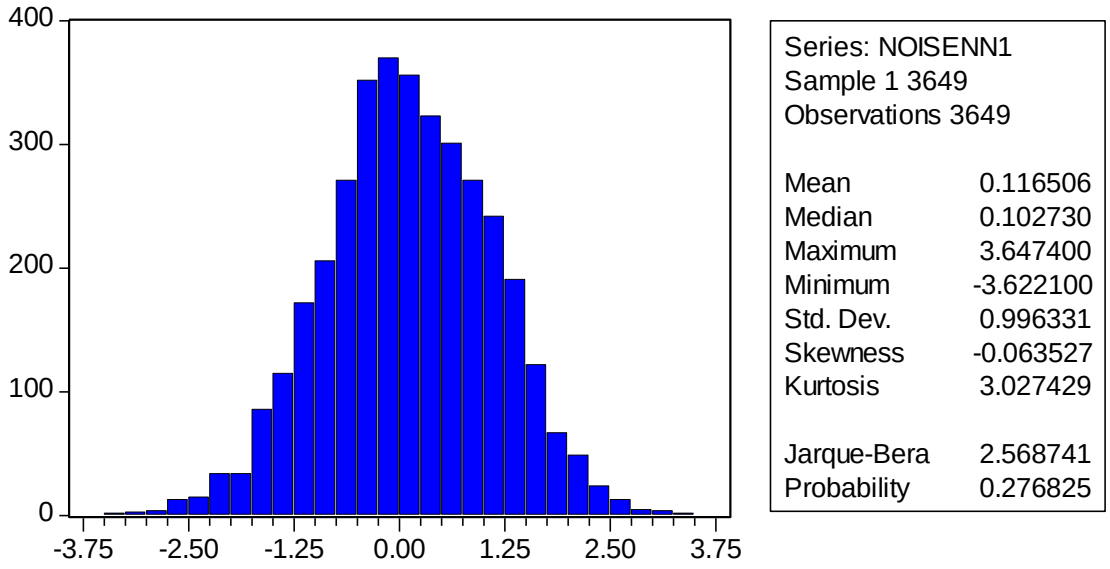

Fig.20. Distribution statistics of the residuals of the NN model after dividing out the volatility function from the residuals.

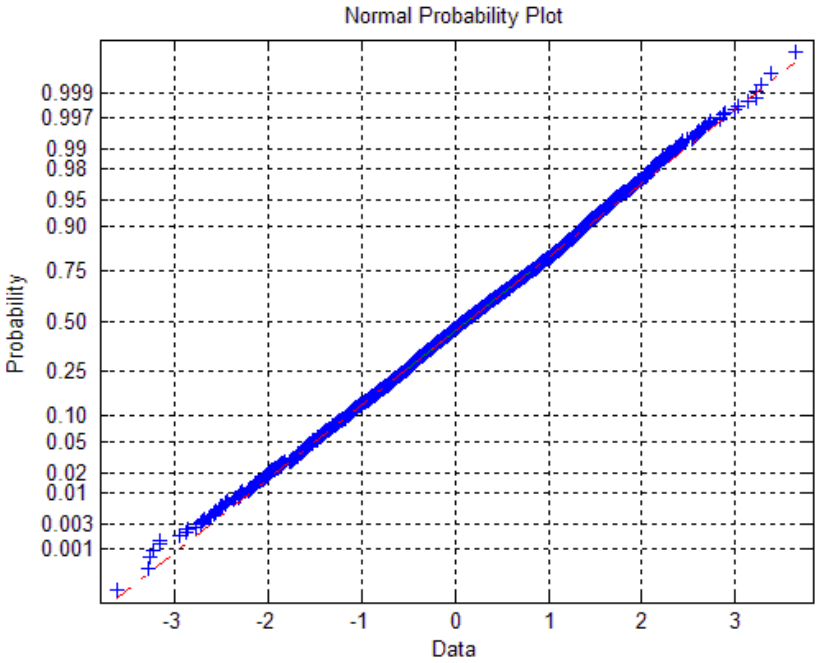

Fig.21. Normal probability plot of the of the residuals of the NN model after dividing out the volatility function from the residuals

TABLE 1

DISTRIBUTIONAL STATISTICS

	Decade1	Decade2	Decade3
Mean	-0.02237	-0.00374	0.025286
	0.027346	0.067015	0.116506
Median	-0.001483	-0.00012	0.032083
	0.039475	0.065229	0.10273
Maximum	3.6872	3.4283	3.1298
	3.6395	3.3923	3.6474
Minimum	-4.0096	-3.4861	-3.9787
	-3.7338	-3.3591	-3.6221
Std. Dev	1.000689	1.000795	1.000838
	1.000445	0.998902	0.996331
Skewness	-0.07411	-0.13531	-0.15291
	-0.05289	-0.11002	-0.06353
Kurtosis	3.10874	2.984192	3.000145
	3.05332	2.916664	3.027429
Jarque-Bera	5.139474	11.1759	14.22022
	2.134051	8.420029	2.568741
Probability	0.076556	0.003743	0.000817
	0.34403	0.014846	0.276825

Distributional statistic for each decade after dividing out the seasonal variance. The first row of each statistic corresponds to the AR(1) model. The second row corresponds to the N.N. model.

TABLE 2
PARAMETER ESTIMATION FOR THE SEASONAL VARIANCE

	Decade1	Decade2	Decade3

c0	4.3078	4.2618	4.339
c1	0.16928	0.35246	0.5095
c2	-0.33575	-0.22978	-0.0721
c3	0.079308	0.11756	0.1883
c4	0.006756	0.23951	0.1533
c5	-0.23018	0.21061	0.1379
d1	0.72976	0.34437	0.126
d2	0.72429	0.35277	0.0623
d3	0.11016	0.1796	-0.2897
d4	-0.20968	0.027802	0.0637
d5	-0.2206	0.068786	-0.0431

	Decade1	Decade2	Decade3
c0	4.2707	4.1713	4.0968
c1	0.12959	0.28903	0.44127
c2	-0.37756	-0.24132	-0.13527
c3	0.16133	0.16338	0.19132
c4	0.065612	0.26635	0.089875
c5	-0.23215	0.22791	0.17336
d1	0.57271	0.33498	0.052762
d2	0.71875	0.41667	0.54105
d3	0.12225	0.13236	-0.05608
d4	-0.19856	0.076904	0.12299
d5	-0.21465	0.082671	0.005984

Parameter estimation of the seasonal variance using N.N. (top) and an AR(1)model (bottom)

TABLE 3
AUGMENTED DICKEY-FULLER TEST

Null Hypothesis: $\alpha(t)$ has a unit root

Exogenous: Constant

First Decade		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-8.823206	0.0000
Test critical values:	1% level	-3.431969	
	5% level	-2.862141	
	10% level	-2.567133	
Second Decade		**t-Statistic**	**Prob.***
Augmented Dickey-Fuller test statistic		-10.35569	0.0000
Test critical values:	1% level	-3.431968	
	5% level	-2.862140	
	10% level	-2.567133	
Third Decade		**t-Statistic**	**Prob.***
Augmented Dickey-Fuller test statistic		-10.73455	0.0000
Test critical values:	1% level	-3.431965	
	5% level	-2.862139	
	10% level	-2.567132	

*MacKinnon (1996) one-sided p-values.

TABLE 4
K.P.S.S. TEST

Null Hypothesis: $\alpha(t)$ is stationary

Exogenous: Constant

First Decade		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.121631
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000
Second Decade		**LM-Stat.**
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.575217
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000
Third Decade		**LM-Stat.**
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.157341
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

REFERENCES

[1] Alaton, P. & Djehince, B., Stillberg, D., 2000. On Modeling and Pricing Weather Derivatives. *Applied Mathematical Finance*, 9, pp.1-20.

[2] Benth, F.E. & Saltyte-Benth, J., 2005. Stochastic Modelling of Temperature Variations With a View Towards

Weather Derivatives. *Applied Mathematical Finance*, 12 (1), pp.53-85.
[3] Benth, F.E. & Saltyte-Benth, J., 2007a. The Volatility Of Temperature And Pricing Of Weather Derivatives. *Quantitative Finance.* 7 (5), pp. 553-561
[4] Benth, F.E., Benth, J.S., Koekebakker S. 2007b. Putting A Price On Temperature. *Scandinavian Journal of Statistics*, 34 (4), pp. 746-767.
[5] Brix, A., Jewson, S. & Ziehmann, C., 2002. *Weather Derivative Modelling and Valuation*, Risk Books.
[6] Brody, C.D., Syroka, J. & M. Zervos, M. 2002. Dynamical Pricing of Weather Derivatives" *Quantitative Finance*, 2, pp.189-198.
[7] Caballero, R., Jewson, S. & Brix A. 2002.Long Memory in Surface Air Temperature: Detection, Modelling and Application to Weather Derivative Valuation. *Climate Research*, 21, pp.127-140.
[8] Cao, M. & Wei, J. 2000. Pricing The Weather. *Risk*, 13 (5).
[9] Davis, M., 2001. Pricing Weather Derivatives by Marginal Value. *Quantitative Finance*, 1, pp.305-308.
[10] Daubechies, I., 1992. *Ten Lectures on Wavelets (CBMS-NSF Regional Conference Series in Applied Mathematics*.
[11] Dischel, B., 1999. Shaping history. *Weather Risk*, 8-9, 1999.
[12] Donoho, D.L. & Johnstone, I.M., 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, pp.425-455.
[13] Dornier, F. & Querel, M., 2000. Caution to the Wind. *Energy Power Risk Management*, p. 30-32.
[14] Engle, R.F., Mustafa, C. & Rice, J., 1992. Modelling peak electricity demand. *Journal of forecasting*, 11, pp. 241-251.
[15] Geman, H. & Yor, M. 1993. Bessel processes, Asian options, and perpetualities. *Mathematical Finance*, 3, pp.349-375.
[16] Hartigan, J.A. & Hartigan P.M., 1985. The Dip Test of Unimodality. *The Annals of Statistics.*13 (1), pp. 70-84.
[17] Henley, A. & Peirson, J, 1998. Residential energy demand and the interaction of price and temperature: British experimental evidence. *Energy Economics*, 20, pp.157-171.
[18] Jewson, S. & Brix, A. 2005. *Weather Derivative Valuation*, Cambridge University Press.
[19] Lau, K. M. & Weng, H.Y., 1995. Climate Signal Detecting Using Signal Transform. *Bulletin of American Meteorological Society*, 76, pp. 2391-2402.
[20] Li, X. & Sailor, D.J., 1995. Electricity use sensitivity and climate and climate change. *World Resource Review*, 3, pp.334-346.
[21] Mallat, S.G., 1999. *A Wavelet Tour of Signal Processing.* San Diego: Academic Press.
[22] Moreno, M., 2000. Riding the Temp. *Futures and Options World*, 11.
[23] Peirson, J. & Henley, A., 1994. Electricity load and temperature issues in dynamic specification. *Energy Economics*, 16, pp.235-243.
[24] Sailor, D.J. & 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, pp.987-998.
[25] Wojtaszczyk, P., 1997. *A Mathematical Introduction to Wavelets.* Cambridge: Cambridge University Press.
[26] Zapranis, A. & Alexandridis, A., 2006. Weather Analysis & Weather Derivative Pricing in proc. HFAA, Thessaloniki, 15-16 December.
[27] Zapranis, A. & Alexandridis, A., 2007. Weather Derivatives Pricing: Modelling The Seasonal Residual Variance of an Ornstein-Uhlenbeck Temperature Process With Neural Networks in proc. EANN, Thessaloniki, 29-31 August.
[28] Zapranis, A. & Refenes, A.-P., 1999. *Principles of Neural Model Identification, Selection and Adequacy: With Applications to Financial Econometrics*, Springer-Verlag, 1999.