LÉVY PROCESS BASED ORNSTEIN-UHLENBECK TEMPERATURE MODEL WITH TIME VARYING SPEED OF MEAN REVERSION

Article *in* Advances and Applications in Statistics · September 2018

DOI: 10.17654/AS053030199

Nelson Dzupire
University of Malawi

Philip Ngare
University of Nairobi

Advances and Applications in Statistics

http://www.pphmj.com
http://dx.doi.org/10.17654/AS053030199
Volume 53, Number 3, 2018, Pages 199-224 ISSN: 0972-3617

LÉVY PROCESS BASED ORNSTEIN-UHLENBECK TEMPERATURE MODEL WITH TIME VARYING SPEED OF MEAN REVERSION

Nelson Christopher Dzupire[1], **Philip Ngare**[1,2] **and Leo Odongo**[1,3]

[1]Pan African University
Institute of Basic Sciences, Technology and Innovation
Juja, Kenya

[2]University of Nairobi
Kenya

[3]Kenyatta University
Kenya

Abstract

In this study we develop a Lévy process driven Ornstein-Uhlenbeck daily temperature model. The model takes into account a time-dependent speed of mean reversion. It is statistically demonstrated that historical data and temperature differences are not normally distributed and hence we have argued against modeling temperature residuals as a Wiener process rather we have used the normal inverse Gaussian distribution which can ably describe skewed and heavy tailed data. Neural networks have been applied to estimate parameters of the detrended and deseasonalized temperature data because there is no prior knowledge on the nature of the function that describes the speed of mean reversion in the model.

Received: December 15, 2017; Accepted: March 19, 2018
2010 Mathematics Subject Classification: 62Pxx.
Keywords and phrases: Lévy process, Ornstein-Uhlenbeck, mean reversion, Wiener process, normal inverse Gaussian, neural networks, deseasonalized, detrended, temperature, residuals.

1. Introduction

Climate change remains a serious environment threat to the fight against hunger, malnutrition, diseases and poverty in Africa as its impact is manifested mainly through serious reduction in agriculture productivity (Iboteye and Shaibu [12]). Water shortages and heat stress are the main important environmental factors limiting crop growth, development and yield especially in maize which is rainfed (Harrison et al. [10]). High temperatures coupled with drought can significantly affect the pollination process especially if it occurs during and within 10 days of pollination period (Harrison et al. [10]) resulting in low yield. Temperature can also have a great impact on business like tourism, recreation, energy producers and consumers.

In this study we develop a Lévy based Ornstein-Uhlenbeck (O-U) daily temperature process with a varying speed mean of reversion. We assume the model is driven by the Lévy process as the available temperature data is found not to be normally distributed hence it is modeled using the normal inverse Gaussian distribution. Besides the speed at which the temperature process reverts to its mean is considered a time varying function since it depends on how far the temperature is from its mean on a particular day. However the form of the function is not known therefore we use neural networks to estimate the temperature process.

The model presented can be used in forecasting future temperatures of a location which is significant in drought and floods risk management. It is also important in agriculture which is the backbone of developing countries economy as crop yields is greatly affected by temperature of the location. In addition temperature is used as underlying variable in pricing weather derivatives. As such an accurate daily temperature process helps to correctly to price weather derivatives which mitigate the risks arising from the effects of climate change in business and agriculture. Daily temperature models are preferred in pricing weather derivatives as they can easily incorporate metrological forecasts and can used for all available contracts unlike index modeling and burn analysis (Alexandridis and Zapranis [2]).

The study is organized as follows: Section 2 presents the overview of different temperature models sampled from literature. We specify and formulate the Lévy process based mean reverting O-U process with a time varying speed of mean reversion in Section 3. The discrete form of the model is presented in Section 4 since the available temperature data is in discrete form. This is followed by parameter estimation and simulation in Section 5. Finally we conclude the study in Section 6.

2. Review of Temperature Process Models

Temperature process is characterized by four components namely seasonality, existence of long term trends, randomness and mean reversion (Alaton et al. [1]). The temperature process shows seasonality patterns and normally reverts to a mean whose value depends on the time of the year. In addition temperature does not grow or fall indefinitely. Therefore it is recommended that a well defined model should incorporate all these temperature characteristics.

In literature different forms of stochastic mean reversion models are suggested for the time dynamics of temperature with seasonal mean and volatility. Most of such models are extension of Dornier and Queruel [9]. We hereby analyze some of these models.

Dornier and Queruel [9] modeled temperature fluctuations as a regression between daily deseasonalized temperatures. The proposed model separates the daily average temperature evolution into two parts namely seasonal trend and random walk. The seasonality is formulated as a sine function with both seasonal change and global warming.

Alaton et al. [1] modified model by Dornier and Queruel [9] by expressing volatility as a piece-wise constant function representing monthly variation in volatility. It was observed that the quadratic variation of the volatility were almost constant over each month in the data set hence validating their choice of volatility. Though no statistical test for normality is provided the justification of using Wiener process as the driving noise in the model because the observed temperature differences were close to being normally distributed. In both studies, there is no mention of possible time dependencies in the residuals observed in the regression models.

In (Brody et al. [8]) the dynamics of temperature is modeled by means of a stochastic process known as fractional Brownian motion as follows: The change in temperature is regressed on the previous day's deseasonalized temperature. Clear signs of fractional behavior in the temperature fluctuations were discovered after removing the seasonal mean based on the data series of daily average temperature from Central England. As observed by Benth and Šaltytė-Benth [6], the authors did not perform the same fractional analysis for the residuals in the model, hence it is not clear if the time dependence of the residuals will follow characteristics of fractional noise. In accordance with Dornier and Queruel [9] argument one should add the changes of seasonal variation $ds(t)$ in the model to have a consistent mean reversion model.

Based on the O-U model by Dornier and Queruel [9], Benth and Šaltytė-Benth [6] proposed a mean reverting model driven by a Lévy process. This was as result of the rejection of the normality test by empirical Norwegian data. In addition the variance of the model is empirically based function estimated from observed variances. A generalized hyperbolic distribution is suggested as one which is flexible for capturing semi-heavy tails and skewness observed in the data. However the model assumes constant speed of mean reversion, and the inclusion of the Lévy process complicates the model. In addition the use of AR(1) fails to capture the slow time decay of the auto correlations of temperature which may lead to significant under pricing of weather derivatives.

In a related work (Benth and Benth [5]) the authors worked on an O-U mean reverting model with Brownian motion as the driving noise whereas the seasonal mean and volatility were modeled as truncated Fourier series. The order of both series was arbitrarily chosen and no statistical tests were presented for the significance of each parameter. The model was found to be capable to describe temperature dynamics and allowed for a closed form solutions of pricing weather derivatives. Another model different from the O-U is proposed in (Benth and Šaltytė-Benth [7]) as a continuous time autoregressive model for the temperature dynamics with the volatility function as the product of seasonal function and a stochastic process using Barndoff-Nielsen and Shephard model for stochastic volatility.

Zapranis and Alexandridis [16] extended the O-U mean reverting model developed by Benth and Šaltytė-Benth [6] with seasonality in the level and volatility, validated by more than 100 years of temperature data collected in Paris. Unlike Benth and Šaltytė-Benth [6] here wavelet analysis is used to identify the seasonality component in the temperature process as well as the volatility of the temperature residuals. It was observed that the distribution statistics of the residuals of AR(1) showed presence of negative skewness and positive kurtosis (> 3) indicating a significant deviation from the normal distribution. In addition the effect of replacing the AR(1) process with ARMA, ARFIMA, and ARFIMA-FIGARCH were also explored. However all these processes failed to capture the slow time decay of the autocorrelations of temperature.

Zapranis and Alexandridis [17] developed an O-U stochastic temperature model driven by the Wiener process and used neural networks to examine the time dependence of the speed of mean reversion parameter $k(t)$ on time. The model is an extension of Benth and Benth [5] which is a generalization of the Dornier and Queruel [9] and it is discretized as an AR(1) model.

Estimating non-parametrically with neural networks the temperature process and computing the derivative of the network output with respect to the network input, a series of daily values of $k(t)$ are obtained. This removes the constraint of a constant mean reverting speed which has been assumed in several models. The results show daily variation in the speed of mean reversion is quiet high hence expressing the speed of mean reversion as a function of time improves the accuracy of the model and significantly improves pricing of weather derivatives.

Wang et al. [15] developed a feasible model for daily average temperature for Zhenzhou area which is applied in weather derivative pricing. This is an O-U mean reverting model like Alaton et al. [1], Benth and Šaltytė-Benth [6] where volatility is expressed as cyclic function using the truncated Fourier series for simplicity.

3. Model Specification and Formulation

Definition 1. Given a probability space $(\Omega, \mathfrak{F}, \mathbf{P})$ a *Lévy process* $L = \{L_t, t \geq 0\}$ is defined as infinitely divisible continuous time stochastic process $L_t : \Omega \to \mathbb{R}$ with stationery and independent increments.

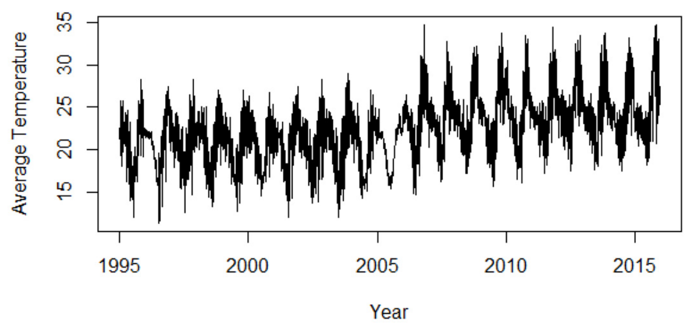


Figure 1. Daily average temperature for Balaka district in Malawi.

Lévy process structures allow the representation of jumps, skewness and excess kurtosis. In this study we aim to develop a mean reverting O-U stochastic model that accurately describes the dynamics of daily average temperature. The model is basically an extension of Benth and Šaltytė-Benth [6] which is a generalization of Dornier and Queruel [9]. Based on observation by several researchers (Alaton et al. [1], Benth and Šaltytė-Benth [6], Benth and Benth [5], Zapranis and Alexandridis [16]) a feasible temperature model is supposed to take into account the following properties. Firstly it should take into account that temperature process follows a predicted cycle, and moves around the seasonal mean. It is affected by global warming and urban effects and appears to have auto-regressive changes. Finally the volatility of temperature is high in the winter than is in the summer. Looking at Figure 1 which is an empirical plot for daily average temperature of Balaka district in Malawi, we can as well observe the temperature process properties: The mean reverting O-U stochastic model is as follows:

$$dT(t) = ds(t) - k(t)[T(t) - s(t)]dt + \sigma(t)dL(t), \tag{1}$$

where

$$T(t) = \frac{T_{\max} + T_{\min}}{2}$$

is the daily average temperature, $s(t)$ is the cyclic seasonal mean, $k(t)$ is a speed of mean reverting dependent on time t, $\sigma(t)$ is the seasonal volatility of daily average temperature which we assume to be strictly positive valued measurable and bounded function. The term $ds(t)$ adjusts the mean $s(t)$ so that in the long term the temperature process $T(t)$ reverts to the seasonal mean $s(t)$, since the mean is time dependent function.

In Benth and Šaltytė-Benth [6], Benth and Benth [5], Alaton et al. [1], Dornier and Queruel [9] it was assumed that the speed of mean reversion is constant. In Benth and Šaltytė-Benth [6], the yearly variation of the speed of mean reversion were found to be very small over the years, however the variations were bigger when computed monthly for the same study. Brody et al. [8] observed that in general the speed of mean reversion should be a function of time though the authors provided no evidence as no studies had been undertaken to compute the daily mean reversion as the process is very complex.

So unlike the generalized model of (Benth and Šaltytė-Benth [6]) where it was assumed the speed of mean reversion is constant, here we develop a model that considers a varying speed of mean reversion. As observed by Alexandridis and Zapranis [2] if the temperature today is far from the seasonal mean, then it is expected that the speed of mean reversion is higher as compared to if the temperature is close to the seasonal mean where then the speed of reversion should be slow. Hence this can only be captured if the speed of reversion is considered as a function of time.

In literature (Alaton et al. [1], Zapranis and Alexandridis [16], Benth and Benth [5], Alexandridis and Zapranis [2]) and many more, the driving noise of the model is the Wiener process because the temperature differences histogram looked normal however no tests were provided. We tested whether the temperature differences for Balaka district is normal using the Anderson Darling test and the hypothesis was rejected with $A = 85.889$ and p-value $< 2.2e - 16$. We also measured both the skewness and kurtosis of the data and found that they were respectively 0.2500 and 3.2083 such that the daily average temperature is positively skewed to the right and heavy tailed. Figure 2 is the histogram of the temperature differences for Balaka district.

Therefore the Lévy process $L(t)$ are used as a suitable noise driving the temperature dynamics. The generalized Lévy distributions are flexible family of distributions which can model skewed and heavy tailed data. The process $L(T)$ is a *cádlág*, adapted real value general Lévy process with independent stationery increments and stochastically continuous. In this study we use the normal inverse Gaussian distribution to model the driving noise of the temperature dynamics.

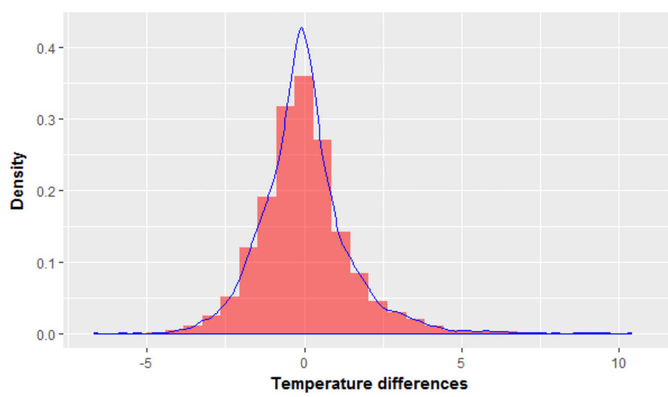


Figure 2. Histogram for temperature differences of Balaka district, Malawi.

They follow the infinitely divisible distribution with the density function (Benth and Šaltytė-Benth [6]),

$$f_{gh}(x; \lambda, \mu, \alpha, \beta, \sigma)$$

$$= c(\sigma^2 + (x-\mu)^2)^{\frac{\lambda - \frac{1}{2}}{2}} \exp(\beta(x-\mu)) K_{\lambda - \frac{1}{2}}(\alpha\sqrt{\sigma^2 + (x-\mu)^2}),$$

where K_s is the modified Bessel function of the third kind with index s and c as the normalizing constant given by

$$c = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda} - \frac{1}{2}\sigma^{\lambda} K_{\lambda}(\sigma\sqrt{\alpha^2 - \beta^2})}.$$

The parameter α controls the fatness of the tail which is the steepness of the distribution, μ is the location of the distribution, β determines the skewness and σ is the scaling parameter. In this study we adopt the Normal Inverse Gaussian (NIG) distribution $\left(\lambda = -\frac{1}{2}\right)$ to model the noise process with density function

$$f_{nig}(x; \alpha, \beta, \sigma, \mu)$$

$$= \frac{\alpha\sigma}{\pi} \exp(\sigma(\sqrt{\alpha^2 - \beta^2}) + \beta(x-\mu)) \frac{K_1(\alpha\sqrt{\sigma^2 + (x-\mu)^2})}{\sqrt{\sigma^2 + (x-\mu)^2}},$$

where

$$K_1(x) = \frac{1}{2}\int_0^{\infty} \exp\left(-\frac{1}{2}x(t + t^{-1})\right)dt$$

is the modified Bessel function of third kind. The normal inverse Gaussian distribution is suitable for data that is hevy or semi tailed, skewed and non-normal since the distribution can model heavy tailed, excess kurtosis and jumps in addition it has closed form solution for its parameters. It is the only subclass of the generalized hyperbolic distributions which is closed under the convolution and it is a normal variance mixture where the mixing distribution is a generalized inverse gamma distribution which generalizes the Gamma distribution (Necula [14]).

The seasonal mean is modeled as sinusoidal function as it can be observed from Figure 1, the dynamic of temperatures mean is cyclic. We also observe that there is a small trend in the temperature data which can be due to global warming and urban heating effects which can be expressed as a linear function. Therefore the trend and the seasonal mean can be expressed as follows:

$$s(t) = a + bt + c\sin(\omega t + \theta), \quad \omega = \frac{2\pi}{365},$$

where c is the amplitude, θ is phase shift as the maximum and minimum temperature do not occur at the beginning and middle of the year. Similarly we express the function σ^2 as a cyclic function based on empirical results from (Benth and Benth [5], Zapranis and Alexandridis [16])

$$\sigma^2 = c + \sum_{i=1}^{I} c_i \sin\left(\frac{2\pi i}{365}\right) + \sum_{j=1}^{J} c_i \cos\left(\frac{2\pi j}{365}\right). \tag{2}$$

4. Discrete Model Formulation

Before discretizing the model to estimate the parameters we show that the proposed model reverts to the seasonal mean by the addition of the term $ds(t)$ and try to find its analytic solution in closed form:

Proposition 2. *If the mean of the temperature process* $s(t)$ *is not a constant, then the temperature process* $T(t)$ *such that*:

$$dT(t) = ds(t) + k(T(t) - s(t))dt + \sigma(t)dL(t)$$

reverts to the mean $s(t)$, *i.e.*, $\mathbf{E}[T(t)] = s(t)$ *by the addition of* $ds(t)$ *to the model*.

Proof. Let $Z(t) = e^{\int_0^t kds}[T(t) - s(t)]$. By Itó lemma we have

$$dZ(t) = -e^{\int_0^t kds} s'(t) + ke^{\int_0^t kds}[T(t) - s(t)]dt + e^{\int_0^t kds} dT$$

$$= e^{\int_0^t kds}[-s'(t) - k(T(t) - s(t))dt + s'(t)$$

$$+ k(T(t) - s(t))dt + \sigma(t)dL(t)]$$

$$= e^{\int_0^t kds}\sigma(t)dL(t)$$

$$Z(t) = Z(0) + \int_0^t e^{\int_0^t kds}\sigma(u)dL(u).$$

Setting $T(0) = s(0) = c$ we have the following

$$e^{\int_0^t kds}[T(t) - s(t)] = \int_0^t e^{\int_0^t kds}\sigma(u)dL(u)$$

$$T(t) = s(t) + e^{-\int_0^t kds}\int_0^t e^{\int_0^t kds}\sigma(u)dL(u).$$

Taking expectation,

$$\mathbf{E}[T(t)] = \mathbf{E}\left[s(t) + e^{-\int_0^t kds}\int_0^t e^{\int_0^t kds}\sigma(u)dL(u)\right]$$

$$= s(t).$$

By direct application Ito Lemma coupled with the fact that Lévy processes are semi-martingales, the explicit solution of the model (1) is as follows:

$$T(t) = s(t) + \exp(-kt)[T(0) - s(0)] + \int_0^t \sigma(u)\exp[k(t - u)]dL(u). \tag{3}$$

The explicit solution $T(t)$ shows that daily average temperature follows a Lévy distribution at each instant time and also reverts to the seasonal mean $s(t)$ while the variance moves along the volatility $\sigma^2(t)$.

The stochastic integral $\int_0^t \sigma(u)\exp[k(t-u)]dL(u)$ is the Wiener-Lévy integral which is well defined as L^2 limiting approximating Riemann-Stieltjes sums (Applebaum [4]).

The stochastic model (3) is time continuous mean reverting process, however the available daily average temperature data is in discrete form, measured in days. Hence there is a need to reformulate the time continuous model into a discrete-time model for purposes of fitting the data and estimating its various parameters obtained by subtracting $T(t)$ from $T(t+1)$:

$$\Delta T(t) = \Delta s(t) - (1-e^k)[T(t) - s(t)] + e^k \int_t^{t+1} \sigma(u) e^{k(t-u)} dL(u),$$

where $\Delta Y = Y(t+1) - Y(t)$.

The stochastic integral can be approximated by

$$\Delta T(t) = \Delta s(t) - (1-e^k)[T(t) - s(t)] + e^k \sigma(t) \Delta L(t).$$

This model can be reorganized as follows:

$$T(t+1) - T(t) = s(t+1) - s(t) - (1-e^k)[T(t) - s(t)] + e^k \sigma(t) \Delta L(t)$$

$$T(t+1) - s(t+1) = T(t) - s(t) - (1-e^k)[T(t) - s(t)] + e^k \sigma(t) \Delta L(t).$$

Letting $\widetilde{T}(t+1) = T(t+1) - s(t+1)$ we have

$$\widetilde{T}(t+1) = e^k \widetilde{T}(t) + \widetilde{\xi}(t), \quad \text{where } |e^k| \le 1, \quad \widetilde{\xi}(t) = e^k \sigma(t) \Delta L(t). \tag{4}$$

The time-discretized model (4) represents the temperature dynamics that has been deseasonalized and detrended where $\widetilde{\xi}(t)$ is the randomness in the model. This will be our basic model when analysing the daily average temperature data to estimate various parameters.

5. Parameter Estimation

5.1. Data description

In this study, daily average temperature data measured in degrees Celsius, observed in Balaka district, Malawi is analyzed. The data covers a period from 1995-2015, resulting in 7300 data series. For uniformity across the years we have removed temperature observations on 29th February of each leap year. Balaka district is located in the southern region of Malawi.

The 20 year length period is considered a better sample to study temperature dynamics compared to a very large sample which may run the danger of estimated parameters being affected by dynamics that do not represent future behavior of temperature anymore like urban effects whereas if the period is very small there is a possibility that important dynamics may not be revealed which may lead to an incorrect model (Alexandridis and Zapranis [2]).

In order to better understand the dynamics of temperature, we compute the descriptive statistics of the data namely maximum, minimum, mean, median, mode, standard deviation, skewness, and kurtosis as shown in Table 1.

It can be observed that the data is slightly skewed and non-normal since its kurtosis is in the excess of 0.2083, it is heavily tailed to the right.

Table 1. Descriptive statistics of temperature for Balaka

Maximum	34.7
Minimum	11.4
Mean	22.307
Median	22.3
Mode	22.3
Variance	12.8455
Standard Deviation	3.5841
Skewness	0.2500
Kurtosis	3.2083

5.2. Trend and seasonality

From Figure 1 it can be observed that the temperature process follows a seasonal pattern. Therefore we model the seasonal mean which also takes into account the trend as

$$s(t) = a + bt + c\sin(\omega t + \theta), \quad \omega = \frac{2\pi}{365},$$

where $a + bt$ is the trend due to urban effects and global warming, c is the amplitude and θ controls when we have maximum or minimum temperature since such occurrences do not occur at exactly the beginning and middle of the year. The $s(t)$ takes into account a yearly cyclic for the temperature process.

Using trigonometric formula we have

$$\begin{aligned} s(t) &= a + bt + c\sin(\omega t + \theta) \\ &= a + bt + c_1 \sin(\omega t) + c_2 \cos(\omega t), \end{aligned}$$

where

$$c = \sqrt{c_1^2 + c_2^2}$$

$$\theta = \arctan\left(\frac{c_2}{c_1}\right) - \pi.$$

Therefore we fit $s(t)$ using method of least squares so that we find parameters $A = \{\hat{a}, \hat{b}, \hat{c}_1, \hat{c}_2\}$ that solves the optimization problem

$$\min_A \| s(t) - X(t) \|,$$

where $X(t)$ is the data vector. In Table 2 we report the estimated values of the parameters which indicated that all of them are significant. Inserting the estimated values into $s(t)$ we have

$$s(t) = 1.946 + 0.0074t + 2.9968\sin\left(\frac{2\pi}{365} - 4.33\right). \tag{5}$$

It can be observed in equation (5) that the constant mean level of the temperature process is 1.946 and the amplitude of the seasonal mean is 2.9968 with a phase shift of 4.33 hence the difference between a typical winter day and a summer day is about 13°C. The trend is minimal probably due to the length of period under study though it is still significant in the model.

Table 2. Estimated parameter values for the seasonal mean

Parameter	Estimated	Std Error	t value	$Pr(>\|t\|)$
$\hat{a}$	1.946	$5.391e-02$	361.03	$< 2e-16$ ***
$\hat{b}$	0.00742	$1.218e-05$	60.88	$< 2e-16$ ***
$\hat{c}_1$	–1.124	$3.812e-02$	–29.48	$< 2e-16$ ***
$\hat{c}_2$	2.778	$3.810e-02$	72.93	$< 2e-16$ ***

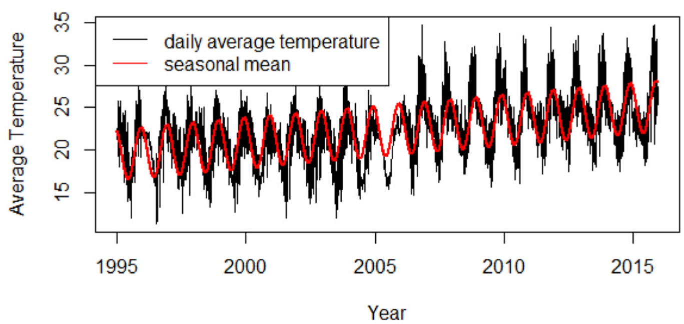


Figure 3. Predicted season mean and the daily average temperature.

Figure 3 shows a plot of the seasonal mean and the daily average temperature. Clearly the mean fits the data fairly well.

5.3. Estimating the cyclic component with the varying speed of mean reversion $k(t)$

In the model (4) k has been assumed to be a constant in several studies. However such studies have shown that such a model is not complex enough to completely remove the autocorrelations of the residuals (Alaton et al. [1], Benth and Šaltytė-Benth [6], Benth and Benth [5], Dornier and Queruel [9]). In addition there has been suggestions that the speed of mean reversion cannot be a constant function (Brody et al. [8], Alexandridis and Zapranis [2]) though no one has taken the task of calculating the daily values of the speed of mean reversion until (Alexandridis and Zapranis [2]).

Therefore we generalize the model (4) and estimate it nonparametrically as

$$\widetilde{T}(t) = \Phi(\widetilde{T}(t-1), \widetilde{T}(t-2), \ldots) + \xi(t). \tag{6}$$

The model (6) uses past detrended and deseasonalized temperature like the AR(1) model of (4).

In this model, today's detrended and deseasonalized daily average temperature data is regressed against those of previous day. This process was observed to perform better than fitting an autoregressive lags (Benth and Šaltytė-Benth [6]). However there is no prior knowledge and assumptions about $\Phi(t)$ which makes approximating its estimator by parametric regression methods impossible. Therefore in order to estimate the function $\Phi(t)$ requires procedures that can model a nonlinear function whose nature of form is not known and does not require any assumptions about the function.

In this study we estimate the speed of mean reversion using neural networks modeling since little is known about the form of the function of the speed of mean reversion. These are nonparametric data driven approaches which can capture nonlinear data structures without prior assumptions about the underlying relationship in the particular problem (Zhang et al. [18]). It has been proved that neural networks models are capable of approximating any deterministic nonlinear function with little knowledge and no prior assumptions hence making this applicable in our case (Hornik et al. [11], Lantz [13]). In addition neural networks are attractive due to their flexibility as well as capable of dealing with noisy and seasonal data compared to the auto-regressive models.

Neural networks mimic the structure of animal brain to model an arbitrary function and despite their complexity, can easily be applied to real world problems as they do not require any prior knowledge of the process. They comprises of an activation function which transforms a neuron's combined input into an output. Each neural network model has a typical topology which tells the number of layers and neurons in the model in addition there is a training algorithm which specifies how the connections weights are set in order to inhibit the neurons (Lantz [13]).

In this study the neural network model is based on the multilayer perceptron (MLP) feed forward neural network which comprises of three layers namely the input, hidden and output layer. The input layer receives the data whereas the output layer sends data out of the network. In between the hidden layer transforms the input variables based on the activation function for use in the output layer (Alexandridis and Zapranis [3]). The MLP overcomes the limitation of the single layer perceptron by adding several layers which enables neural networks to solve complex problems.

We define the neural model as follows:

$$\hat{\Phi}(t) = f_0\left[b_k^0 + \sum_{j=1}^{p} w_{jk}^0 f_h\left(\sum_{i=1}^{n} w_{ij}^h x_i + b_j^h\right)\right], \tag{7}$$

where w_{ij}^h are the weights for the connections between ith input and jth hidden layer, w_{jk}^0 are weights between the jth hidden unit and the kth output layer. b_j^h is the bias for the hidden unit j and similarly b_k^0 is the bias for the kth output layer.

The function f_0 and f_h are the activation function at the output and hidden layers respectively. This is a nonlinear function applied to the net input to produce output. The most commonly used activation is the sigmoid activation function also called a logistic function defined as $f : \mathbb{R} \to (0, 1)$ such that

$$f(x) = \frac{1}{1 + e^{-x}}.$$

Clearly $f(x)$ is differentiable, hence it is possible to calculate derivatives across the entire range of inputs.

After initializing the weights of the neural network model, it is trained further to find the weights that minimize the error function:

$$E = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{8}$$

where $E \geq 0$, y_i is the target value and $\hat{y}_i$ is the output value. The error function describes the deviation of the predicted outcomes from the observed values such that large deviation suggests a poorly fitted model and the weights have to be adjusted.

In this study the model is trained by backpropagation whereby the weights and biases are adjusted so that they minimize the mean sum of square error by propagating the error backwards at each step. Backpropagation is a popular algorithm used in training neural models since it is insensitive to initial conditions. In backpropagation the weights are modified and adjusted until weights that minimizes the error function are found, by calculating the gradient of the function with respect to the weights

$$\frac{\partial E}{\partial w} = \frac{1}{2n}\sum_{i=1}^{n}\frac{\partial E_i}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial w_i}, \text{ where } E_i = \frac{1}{2}(y_i - \hat{y}_i)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} - (y_i - \hat{y}_i)\frac{\partial \hat{y}_i}{\partial w_i}.$$

The weights w_i are iteratively calculated as follows:

$$w_{i+1} = w_i - \eta\frac{\partial E}{\partial w_t},$$

where η is called a learning rate that is normally fixed in the algorithm.

A multilayer perceptron feed forward neural model is fitted to four lags of the detrended and deseasonalised data with one hidden layer since it is capable of estimating the nonlinear function. The number of neurons were varied from 2 to 6 so that a better model could be selected based on the error, number of iterations and AIC. In selecting a good model we base our choice depending on small error, few number of iterations and a small AIC. The error which is the sum of squares of the errors defined by (8) helps to estimate the model performance on the training data whereas the number of iterations helps to check whether the model is overfitted or not.

It was found that the model with 4 neurons performed better in terms of error and number of iterations but not AIC as shown in Table 3. However neural network is a nonparametric method, the main objective is to minimize the error with minimal number of iterations so as to avoid overfitting data, a model of 4 neurons was selected as best model for the study.

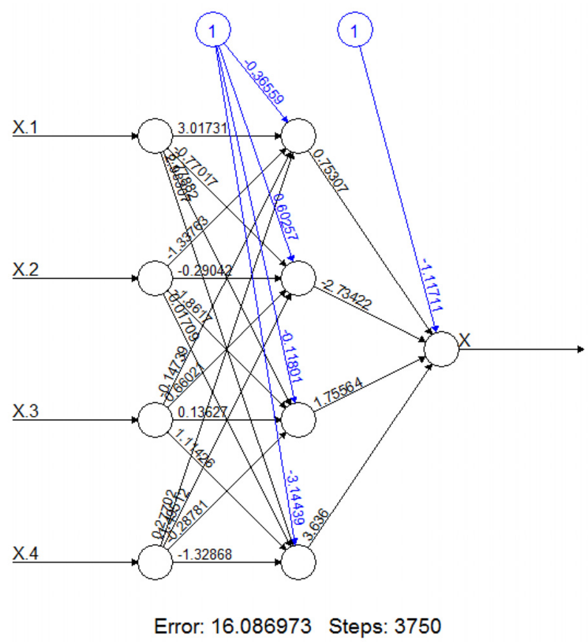


Figure 4. Neural network model.

Figure 4 shows the neural network model with the weights and biases indicated by the blue line. The $\{X.1, ..., X.4\}$ are the input variables representing daily average temperature lags, with corresponding weights to each neuron. The number in blue at each neuron is a bias term.

The neural network model was fitted using 75% of the available data, and the 25% was used for testing and predicting so as to compare if the model really fits the data. Figure 5 compares the observed data and the network output values. It can be shown that the fitted model fits the data very well as there are not many variations in pattern and values. The correlation of the predicted values and the test values was found to be 0.4714 which indicates positive relationship between the target values and the predicted values confirming a well fitted model.

Table 3. Different neural network models based on number of neurons

Number of Neurons	Error	AIC	Number of Iterations
2	16.0943	58.1687	13556
3	16.0961	70.1922	6379
4	16.0870	82.174	3750
5	16.9354	83.0907	13497
6	16.9466	82.3162	10455

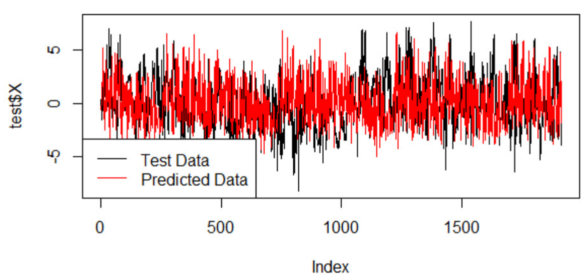


Figure 5. Test data vs predicted values.

We compared the results from the neural network model with and those of Benth and Šaltytė-Benth [6], where the speed of mean reversion is assumed constant on the same data from Balaka district in Malawi. From Figure 6, it was observed that the model underestimates in most instances unlike the one where neural networks were used.

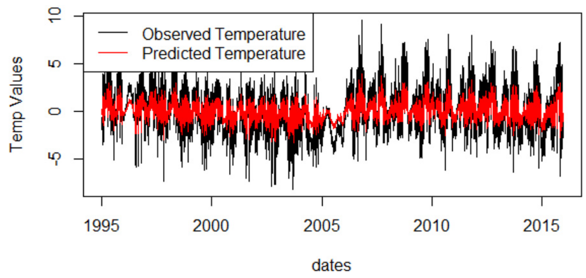


Figure 6. Observed temperature and predicted temperature.

5.4. Estimating the volatility of the temperature process $\sigma^2(t)$

The residuals of the neural network model are shown in Figure 7. From the plot it can be concluded that the residuals are not normally distributed as many points have deviated from the normal line.

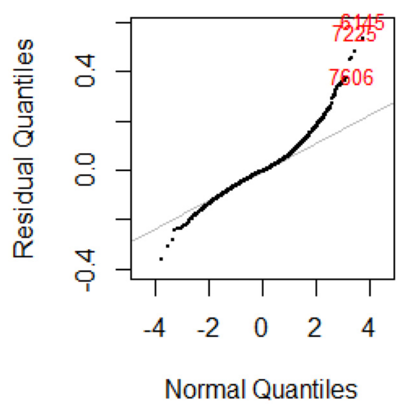
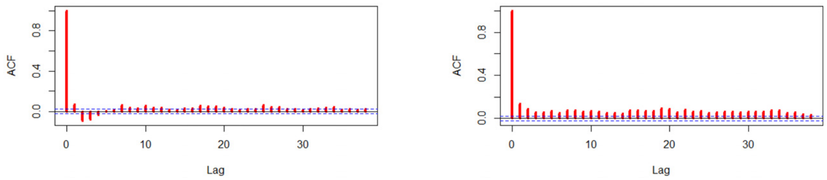


Figure 7. Plot of the residuals.

The autocorrelation functions (ACFs) for the both the residuals and squared residuals are shown in Figures 8a and 8b. There are high values of autocorrelation for several lags in both ACFs which may prompt the use of higher order autoregressive models. In addition the ACFs reveal seasonality presence in the residuals and time dependency in the variance of residuals. Therefore we need to extract the $\sigma^2(t)$ from the residuals before analyzing and modeling the random noise process.



(a) ACF of residuals (b) ACF of squared residuals

Figure 8

The volatility is extracted as described in (Benth and Benth [5]). We group the residuals into 365 groups each corresponding to a day of the year. Then we find the mean of the squared residuals in each group being the estimate of the expected daily residuals

$$\hat{\sigma}^2(t) = \mathbf{E}[(\sigma(t)\tilde{\xi}(t))^2].$$

These values are taken as empirical values of the daily variance based on the observations over the years for that particular day as shown in Figure 9. It can be seen that temperature volatility is higher during dry season as compared to the rainy season. However contrary to the perception that temperature volatility is higher in winter as compared to summer, but this result was also observed on Stockholm data by the work of Benth and Benth [5].

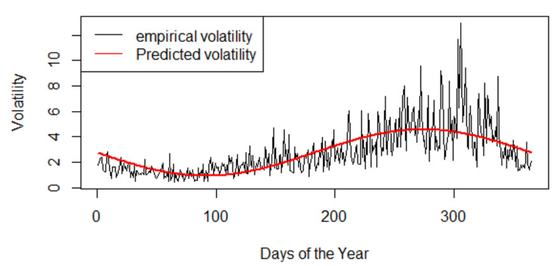


Figure 9. Empirical volatility.

Then based on volatility model (9) with $I = J = 1$ we fit the data to obtain the volatility model

$$\hat{\sigma}^2(t) = 2.7541 - 1.8219 \sin\left(\frac{2\pi t}{365}\right).$$

We compared the empirical volatility and the estimated model as shown in Figure 9 which shows that the estimated model fits the empirical data very well. After removing temperature volatility the distribution of the residuals deviates from normality as observed by the Jarque-Bera test where X-squared is 1329.4 and the P-value of $< 2.2e-16$. With the skewness of -0.4911 and kurtosis of 1.7865 we can conclude that the residuals are not normally distributed.

Therefore it is not effective to describe the random noise as Brownian motion as some researchers (Alexandridis and Zapranis [2], Alaton et al. [1], Benth and Benth [5]) have done. As such we try to model the random process by using generalized hyperbolic distributions.

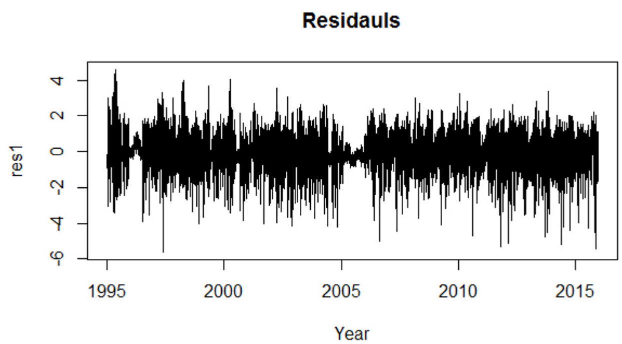


Figure 10. Residuals.

Table 4. Parameters for NIG and HY

	μ	δ	α	β
NIG	0.2962	1.4108	1.4536	–0.2882
GHY	0.2948	0.2948	1.8876	–0.2872

5.5. Modeling residuals using a Lévy distribution

We modeled the randomness of the temperature process by two different distributions namely the normal inverse Gaussian (NIG) and hyperbolic (HY) distribution. Based on the results of estimation and goodness-of-fit looking at the quantile plots, it was found out that the NIG fitted the residuals better than the hyperbolic distribution. The likelihood of NIG was found to be

–10781.97 compared to –10786.23 for generalized hyperbolic distribution. Table 4 shows the estimated parameters for both distributions: The quantile plot for the two distributions are as shown in Figures 11 and 12.

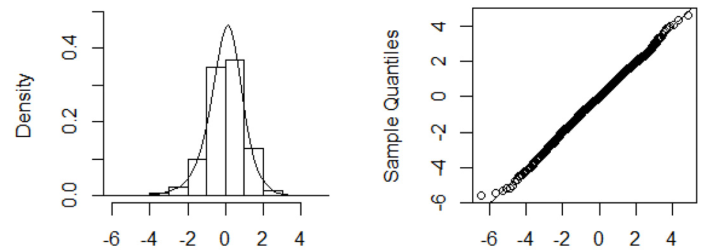


Figure 11. Histogram and Q-Q plot of residuals for NIG.

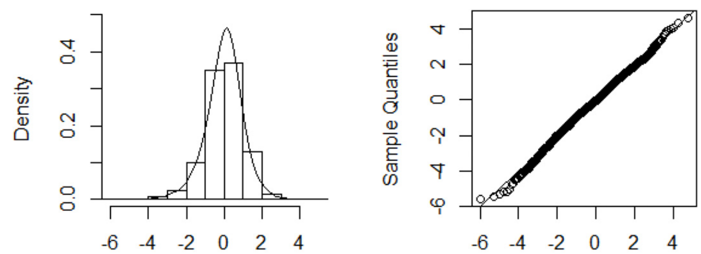


Figure 12. Histogram and Q-Q plot of residuals for HY.

6. Conclusion

In this study we developed a normal inverse Gaussian mean reverting Orstein-Uhlenbeck temperature model. The model is driven by Lévy process unlike several models in literature which are driven by the Wiener process by assuming that temperature differences are nearly normally distributed. We showed that both the historical average temperature data and temperature differences are not normally distributed and hence we argued against modeling the residuals by Wiener process preferring a Lévy distribution, the normal inverse Gaussian, which is able to capture the skewness and heaviness of the tails of the data. Both the seasonal mean and volatility are modeled as cyclic functions.

Another unique feature of this model is that the speed at which temperature reverts to its mean is modeled as nonlinear function. In literature several authors have assumed that the speed is a constant (Alaton et al. [1], Benth and Šaltytė-Benth [6], Benth and Benth [5], Dornier and Queruel [9]), however in this study we argue that the speed at which temperature reverts to its mean depends on how far the temperature of that particular day is from the mean and therefore cannot be modeled as a constant. However little is known about the nature of the nonlinear function and hence it is difficult to calculate daily values of the speed of mean reversion. Therefore we estimate the parameters of the model using neural networks as they do not require prior knowledge of the function. Employing the neural networks in the estimation led to a significant improvement regarding the cyclic component of the deseasonalized and detrended temperature.

References

[1] P. Alaton, B. Djehiche and D. Stillberger, On modelling and pricing weather derivatives, Applied Mathematical Finance 9(1) (2002), 1-20.

[2] A. K. Alexandridis and A. D. Zapranis, Weather Derivatives: Modeling and Pricing Weather-related Risk, Springer Science & Business Media, 2012.

[3] A. K. Alexandridis and A. D. Zapranis, Wavelet Neural Networks: With Applications in Financial Engineering, Chaos, and Classification, John Wiley & Sons, 2014, pp. 61-80.

[4] D. Applebaum, Lévy Processes and Stochastic Calculus, Cambridge University Press, Cambridge, 2009.

[5] F. E. Benth and J. Š. Benth, The volatility of temperature and pricing of weather derivatives, Quant. Finance 7(5) (2007), 553-561.

[6] F. E. Benth and J. Šaltytė-Benth, Stochastic modelling of temperature variations with a view towards weather derivatives, Applied Mathematical Finance 12(1) (2005), 53-85.

[7] F. E. Benth and J. Šaltytė-Benth, Weather derivatives and stochastic modelling of temperature, Int. J. Stoch. Anal. 2011, Art. ID 576791, 21 pp.

[8] D. C. Brody, J. Syroka and M. Zervos, Dynamical pricing of weather derivatives, Quant. Finance 2(3) (2002), 189-198.

[9] F. Dornier and M. Queruel, Caution to the wind, Energy & Power Risk Management 13(8) (2000), 30-32.

[10] L. Harrison, J. Michaelsen, C. Funk and G. Husak, Effects of temperature changes on maize production in Mozambique, Climate Research 46(3) (2011), 211-222.

[11] K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2(5) (1989), 359-366.

[12] S. Iboteye and U. Shaibu, The effect of rainfall and temperature on maize yield in Kogi state, Nigeria, Asian Journal of Basic and Applied Sciences 1(2) (2014).

[13] B. Lantz, Machine Learning with R, Packt Publishing Ltd., 2013.

[14] C. Necula, Modeling heavy-tailed stock index returns using the generalized hyperbolic distribution, Romanian Journal of Economic Forecasting 10(2) (2009), 118-131.

[15] Z. Wang, P. Li, L. Li, C. Huang and M. Liu, Modeling and forecasting average temperature for weather derivative pricing, Advances in Meteorology 2015 (2015), Art. ID 837293, 8 pp.

[16] A. Zapranis and A. Alexandridis, Wavelet analysis and weather derivatives pricing, 5th Hellenic Finance and Accounting Association (HFAA), Thessaloniki, 2006, pp. 15-16.

[17] A. Zapranis and A. Alexandridis, Modelling the temperature time-dependent speed of mean reversion in the context of weather derivatives pricing, Appl. Math. Finance 15(4) (2008), 355-386.

[18] G. P. Zhang, B. E. Patuwo and M. Y. Hu, A simulation study of artificial neural networks for nonlinear time-series forecasting, Computers & Operations Research 28(4) (2001), 381-396.

Nelson Christopher Dzupire: ndzupire@cc.ac.mw