Temperature Forecasting in the Concept of Weather Derivatives: A Comparison between Wavelet Networks and Genetic Programing

Antonios K. Alexandiris[1], Michael Kampouridis[2]

[1]School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK.
A.Alexandridis@kent.ac.uk
[2]School of Computing, University of Kent, Medway, UK.
M.Kampouridis@kent.ac.uk

Abstract. The purpose of this study is to develop a model that accurately describes the dynamics of the daily average temperature in the context of weather derivatives pricing. More precisely we compare two state of the art algorithms, namely wavelet networks and genetic programming against the classic linear approaches widely using in the contexts of temperature derivative pricing. The accuracy of the valuation process depends on the accuracy of the temperature forecasts. Our proposed models were evaluated and compared in-sample and out-of-sample in various locations. Our findings suggest that the proposed non-linear methods significantly outperform the alternative linear models and can be used for accurate weather derivative pricing.

Keywords: weather derivatives, wavelet networks, temperature derivatives, genetic programing

1 Introduction

In this paper, we use a Wavelet Neural Networks (WN) and Genetic Programming (GP) in the context of temperature modeling and weather derivative pricing. Relatively, recently a new class of financial instruments, known as ‘‘weather derivatives’’, has been introduced. Weather derivatives are financial instruments that can be used by organizations or individuals as part of a risk management strategy to reduce risk associated with adverse or unexpected weather conditions, [1]. Just as traditional contingent claims, whose payoffs depend upon the price of some fundamental, a weather derivative has an underlying measure such as: rainfall, temperature, humidity, or snowfall. The difference from other derivatives is that the underlying asset has no value and it cannot be stored or traded while at the same time the weather should be quantified in order to be introduced in the weather derivative. To do so, temperature, rainfall, precipitation, or snowfall indices are introduced as underlying assets. How-

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

ever, in the majority of the weather derivatives, the underlying asset is a temperature index.

According to [2, 3] nearly $1 trillion of the US economy is directly exposed to weather risk. 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 capitalising on those volatile markets. Weather derivatives are used to hedge volume risk, rather than price risk. Hence, a model that describes accurate the temperature dynamics, the evolution of temperature, and which can be used to derive closed form solutions for the pricing of temperature derivatives is essential.

In this study two state of the art algorithms are used, namely WN and GP, in order to model the temperature dynamics. WNs were proposed by [4] as an alternative to Neural Networks, which would alleviate the weaknesses associated with Neural Networks and Wavelet Analysis. In [5], various reasons were presented in why wavelets should be used instead of other transfer functions. In particular, first, wavelets have high compression abilities, and secondly, computing the value at a single point or updating the function estimate from a new local measure involves only a small subset of coefficients. WNs have been used in a variety of applications so far, i.e., in short term load forecasting, in time-series prediction, signal classification and compression, signal denoising, static, dynamic and nonlinear modeling, nonlinear static function approximation, [5], to mention the most important and as it was presented in [1], they can constitute an accurate forecasting method in the context of weather derivatives pricing.

On the other hand, GP is a nature-inspired algorithm, which uses the principles of evolution to find computer programs that perform well in a given task, [6-8]. One of the main advantages of GP is its ability to perform well in high-dimensional combinatorial problems, such as the one of weather derivatives pricing. An additional advantage of GP is that it is a white-box technique, which thus allows the traders to visualize the trees and thus the temperature models. To our knowledge GP was applied to weather derivatives only in [9, 10]. In addition the proposed GP in [9, 10] was used for seasonal forecasting. In contrast in this study a GP is used in order to forecast daily average temperatures (DAT) in 3 European cities in which weather derivatives are actively traded.

Using models for daily temperatures can, in principle, lead to more accurate pricing than modelling temperature indices. Daily models very often show greater potential accuracy than the Historical Burn Analysis or seasonal forecasts, [1, 11], since daily modelling makes a complete use of the available historical data. The results produced by the GP and WN are compared to two traditional linear temperature modelling methods proposed by [12] and [13]. Our results are compared in 1-day-ahead forecast and to out-of-sample forecasts.

The rest of the paper is organized as follows. In Section 2 the various methods for forecasting DAT are presented. More precisely in Section 2.1 the linear models are presented while in Sections 2.2 and 2.3 the WN and the GP are discussed respectively. The data set is described in Section 3 while in Section 4 our results are presented. Finally, in Section 5 we conclude.

2 Methodology

According to [1, 14] temperature shows the following characteristics: it follows a predicted cycle, it moves around a seasonal mean, it is affected by global warming and urban effects, it appears to have autoregressive changes, its volatility is higher in winter than in summer. Following [13] a model that describes the temperature dynamics is given by a Gaussian mean-reverting Ornstein-Uhlenbeck (O-U) process defined as follows:

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

where $T(t)$ is the average daily temperature, κ is the speed of mean reversion, $S(t)$ is a deterministic function modelling the trend and seasonality, $\sigma(t)$ is the daily volatility of temperature variations and $B(t)$ is the driving noise process. As it was shown in [15] the term $dS(t)$ should be added for a proper mean-reversion towards the historical mean, $S(t)$. For more details on temperature modelling we refer the reader to [1].

2.1 Linear Models

Alaton. In [12] the model given by (1) is used where the seasonality in the mean is incorporated by a sinusoid function

$$S(t) = A + Bt + C\sin(\omega t + \varphi) \tag{2}$$

where φ is the phase parameter that defines the day of the yearly minimum and maximum temperature. Since it is known that the DAT has a strong seasonality of an one year period, the parameter ω was set to $\omega = 2\pi / 365$. The linear trend caused by urbanization or climate changes is represented by $A + Bt$. The time, measured in days, is denoted by *t*. The parameter C defines the amplitude of the difference between the yearly minimum and maximum DAT. Another innovative characteristic of the framework presented in [12] is the introduction of seasonalities in the standard deviation modelled by a piecewise function.

Benth. In [13] a mean reverting O-U process where the noise process is modelled by a simple BM as in (1) was suggested. Both seasonal mean and (square of) daily volatility of temperature variations are modelled by truncated Fourier series:

$$S(t) = a + bt + \sum_{i=1}^{I_1} a_i \sin\left(2\pi i\left(t - f_i\right)/365\right) + \sum_{j=1}^{J_1} b_j \cos\left(2\pi j\left(t - g_j\right)/365\right) \tag{3}$$

$$\sigma^2(t) = c + \sum_{i=1}^{I_2} c_i \sin\left(2\pi i t / 365\right) + \sum_{j=1}^{J_2} d_j \cos\left(2\pi j t / 365\right) \tag{4}$$

Using truncated Fourier series a good fit for both the seasonality and the variance component can be obtained while keeping the number of parameters relative low. The above representation simplifies the needed calculations for the estimation of the parameters and for the derivation of the pricing formulas. Equations (3) and (4) allow both larger and smaller periodicities than the classical one year temperature cycle.

2.2 Wavelet Networks

In [1] a more complex model was used by applying WNs. As it was shown in [1] the solution of model (1) can be written as an AR(1) model:

$$\tilde{T}(t+1) = a\tilde{T}(t) + \tilde{\sigma}(t)\varepsilon(t) \tag{5}$$

where $T(t)$ is given by $T(t) = T(t) - S(t)$, $a = e^{-\kappa}$ and $\tilde{\sigma}(t) = a\sigma(t)$.

Intuitively, it is expected that the speed of mean reversion is not constant. If the temperature today is away from the seasonal average (a cold day in summer) then it is expected that the speed of mean reversion is high; i.e. the difference of today and tomorrows temperature is expected to be high. In contrast if the temperature today is close to the seasonal variance we expect the temperature to revert to its seasonal average slowly. To capture this feature the speed of mean reversion is modelled by a time-varying function $\kappa(t)$. Hence the structure to model the dynamics of the temperature evolution becomes:

$$dT(t) = dS(t) + \kappa(t)\big(T(t) - S(t)\big)dt + \sigma(t)dB(t) \tag{6}$$

Model (5) is a lineal AR(1) model with a zero constant. Since in our analysis the speed of mean reversion is not considered constant but a time-varying function, equation, (5) can be written as follows:

$$\tilde{T}(t) = a(t-1)\tilde{T}(t-1) + \sigma(t)\varepsilon(t) \tag{7}$$

where

$$a(t) = 1 + \kappa(t) \tag{8}$$

The impact of a false specification of a, on the accuracy of the pricing of temperature derivatives is significant, [12]. In this section, we address that issue, by using a WN to estimate non-parametrically relationship (7) and then estimate a as a function of time. Moreover, previous studies [12, 13, 16-19] show that an AR(1) model is not complex enough to completely remove the autocorrelation in the residuals. Alternatively more complex models were suggested, [20, 21].

Using WNs the generalized version of (7) is estimated nonlinearly and non-parametrically, that is:

$$\tilde{T}(t+1) = \phi\big(\tilde{T}(t), \tilde{T}(t-1), ...\big) + e(t) \tag{9}$$

Model (9) uses past temperatures (detrended and deseasonalized) over one period. Using more lags we expect to overcome the strong correlation found in the residuals in models such as in [12], [13] and [18]. However, the length of the lag series must be selected. For additional details on modelling the temperature using WN we refer to [1, 5, 22].

2.3 Genetic Programing

While the previous methods are directly using a functional form for their predictions (e.g., linear), the GP operates in a different manner. It can evolve different arithmetic expressions that can take the form of regression models. This has the advantage of flexibility, since different temperature models can be derived for each city that we are interested in.

In this work, a simple GP was used to evolve trees that predict the temperatures of a given city over a future period. The function set of the GP contained standard arithmetic operators (ADD, SUB, MUL, DIV (protected division)), along with MOD (modulo), LOG(x), SQRT(x) and the trigonometric functions of sine and cosine. The terminal set was composed of the index t representing the current day, $1 \le t \le$ size of training and testing set the temperatures of the last three days $\tilde{T}(t-1)$, $\tilde{T}(t-2)$ and $\tilde{T}(t-3)$, the constant π, and 10 random numbers in the range of (-10, 10). In this study the GP is based on DAT of the three previous days. Similar, structures were proposed in previous studies, [23]. Nevertheless, our future work will be focused on selecting this window dynamically. The details of the GP is summarized in Table 1[1]

Table 1. GP Experimental Parameters

Parameter	Value
Max initial depth	2
Max depth	4
Generations	50
Population size	500
Tournament size	4
Subtree crossover	30%
Subtree mutation	40%
Point mutation	30%
Fitness function	Mean Square Error (MSE)
Function set	ADD, SUB, MUL, DIV, MOD, LOG, SQRT, SIN, COS
Terminal set	Index t corresponding to the current day $Temp_{t-1}$, $Temp_{t-2}$, $Temp_{t-3}$ Constant π 10 random constants in (-10, 10)

[1] These parameters were selected after careful experimental tuning.

Finally, we should note that traditionally in the GP literature the algorithm is run many times and then statistical results are reported, e.g., the average fitness over the multiple runs, standard deviation, and the best result. This is done in order to get an overall picture of the algorithm's performance. However, because of the fact that the other algorithms tested in this paper are producing a single model only, it is not meaningful for our comparative analysis in Section 4 to use average results. Thus, we obtain the best tree in terms of training fitness (per algorithm), and compare it to the models produced by the two linear methods and the WN.

3 Data Description

For this study DATs for Amsterdam, Berlin and Paris were obtained. Temperature derivatives are actively traded in these cities through the Chicago Mercantile Exchange (CME). The data were provided by the ECAD[2].

The dataset consists of 4,015 values, corresponding to the DAT of 11 years, (1991-2001). In order for each year to have equal observations the 29th of February was removed from the data. Next the seasonal mean and trend were removed from the data. In order to do so, equation (2) was used in Alaton’s method and (3) was used in Benth’s and GP methods. In the case of WNs the seasonal mean was captured using wavelet analysis, [1].

In our analysis, the four methods will be used in order to model and then forecast detrended, deseasonalized DATs. This procedure is followed in order to avoid possible over-fitting problems of the WN and the GP in the presence of seasonalities and periodicities. Then, the forecasts are transformed back to the original temperature time-series in order to compare the performance of each algorithm.

The objective is to accurate forecast two temperature indices, namely Heating Degree Day (HDD) and Cumulative Average Temperature (CAT). Temperature derivatives are commonly written on these two temperature indices.

4 Results

In this section our proposed models will be validated out of sample. Our methods are validated and compared against two forecasting methods proposed in prior studies, the Alaton’s and Benth’s models. The four models will be used for forecasting out-of-sample DATs for different periods. Usually, temperature derivatives are written for a period of a month or a season and sometimes even for a year. Hence, DATs for 1, 2, 3, 6 and 12 months will be forecasted. The out-of-sample period corresponds to the period of 1st January – 31st December 2001 and every time interval starts at 1st January of 2001. Note that the DATs from 2001 were not used for the estimation of the parameters of the four models. Next the corresponding HDDs and CAT indices will be constructed.

---

2 European Climate Assessment & Dataset project: http://eca.knmi.nl

The predictive power of the four models will be evaluated using two out-of-sample forecasting methods. First, we will estimate out-of-sample forecasts over a period and then 1-day-ahead forecasts over a period. The first case, in the out-of-sample forecasts, today (time step 0) temperature is known and is used to forecast the temperature tomorrow (time step 1). However, tomorrow's temperature is unknown and cannot be used to forecast the temperature 2 days ahead. Hence, we use the forecasted temperature at time step 1 to forecast the temperature at time step 2 and so on. We call this method the out-of-sample over a period forecast. The second case, the 1-day-ahead forecast, the procedure is as follows. Today (time step 0) temperature is known and is used to forecast the temperature tomorrow (time step 1). Then tomorrow's real temperature is used to forecast the temperature at time step 2 and so on. We will refer to this method as the 1-day-ahead over a period forecast. The first method can be used for out-of-period valuation of a temperature derivative, while the second one for in-period valuation. Naturally, it is expected the first method to cause larger errors.

In the USA, Canada and Australia, CME weather derivatives are based on the HDD index. A HDD is the number of degrees by which the daily temperature is below a base temperature, i.e.

$$Daily\ HDD\ =\ max\left(0,\ base\ temperature\ -\ daily\ average\ temperature\right)$$

The base temperature is usually 65 degrees Fahrenheit in the U.S. and 18 degrees Celsius in Europe and Japan. HDDs are usually accumulated over a month or over a season. The accumulated HDD index over a period $[\tau_1, \tau_2]$ is given by

$$HDD = \int_{\tau_1}^{\tau_2} \max\left(c - T(s), 0\right) ds \tag{10}$$

Similarly, the CAT index indicates the cumulative average temperature over a specified period. Hence, over a specified period $[\tau_1, \tau_2]$ the CAT index is given by

$$CAT = \int_{\tau_1}^{\tau_2} T(s) ds \tag{11}$$

Since we are studying 3 cities and 2 indices for 5 different time periods using two forecasting schemes, the four models are compared in 60 datasets. Our results are very promising. In the 1-day ahead forecasts the WN outperformed the alternative methods in 18 cases out of the 30. The Benth methods gave the best results 8 times while the GP in only 4. On the other hand in out-of-sample forecasts the GP outperformed the other methods in 12 cases out of 30 while the WN was best model in only 4 cases. Due to space limitations the results of the 1-day ahead forecasts for one month (1-31 January 2001) for the HDD index are presented in Table 2. The results for the remaining datasets are similar and are available from the authors upon request. In total the WN had the best predictive performance in 36.67% of the samples while the GP and Benth's method both in 26.67% and Alaton's model in only 10%. A summary of the results is presented in Table 3. More precisely, Table 3 shows the number of samples in which each method outperforms the others, i.e. has the best predictive accuracy. Percentages are reported in parentheses.

Furthermore, we were interested in statistically ranking the 4 algorithms. We thus run the non-parametric Friedman test, with the Holm's post-hoc test [24, 25]. For the out-of-sample tests the WN ranked first with an average ranking of 2.13, then the GP and Alaton rank with 2.33, and lastly Benth had a ranking of 3.19. Holm's test found that WN was significantly better than the remaining 3 algorithms, and also that the GP was significantly better than Benth (at 5% level, where the *p*-value of the algorithm is compared and found lower than the critical value of the Holm's test). Similarly, the ranks for 1-day-ahead tests, the rankings are as follows: 1. WN (1.46), 2. GP (2.50), 3. Alaton (2.83), 4. Benth (3.19). Holm's post-hoc test showed again that the WN is significantly better than all other 3 algorithms, at 5% significance level. Lastly, we were interested in ranking the 4 algorithms under all 60 datasets tested in this paper (we thus merged the out-of-sample and 1-day-ahead results into a single table). The best overall rank was obtained by WN (1.80), with the GP ranked second with an average rank of 2.41. Alaton and Benth were ranked third and fourth, respectively, with average ranks of 2.58 and 3.20. Holm's post-hoc test also showed that the WN's ranking is significantly better than all other 3 algorithms. In addition, the test showed that the GP's ranking is significantly better than Benth's.

Table 2. Day ahead comparison for a period of 1 month using the HDD index and the relative percentage errors.

HDD/1month	*Real*	*Historical*	*Alaton*	*Benth*	*WN*	*GP*
Amsterdam	463.6	449.5	460.4	458.3	463.8	464.3
Berlin	522.4	517.9	524.8	523.0	523.8	524.7
Paris	378.6	394.7	381.3	379.9	380.2	384.8
Relative Percentage Errors						
Amsterdam			0.69%	1.14%	**0.04%**	0.14%
Berlin			0.46%	**0.11%**	0.27%	0.43%
Paris			0.72%	**0.35%**	0.41%	1.63%

Real and historical HDDs for the period 1 January – 31 January 2001 and estimated HDDs using the Alaton's, Benth's and the two proposed (WN and GP) methods. The second panel corresponds to the relative absolute percentage errors.

5 Conclusions

The previous analysis indicates that our results are very promising. Modelling the DAT using WNs enhanced the predictive accuracy of the temperature process. The additional accuracy of the proposed model will have an impact on the accurate pricing of temperature derivatives. In addtion the GP performed very well in the out-of-sample forecasting method which is very useful for pricing weather contracts before the temperature measuring period.

Our results are preliminary and additional analysis must be contacted. First, the proposed methodologies must be tested in more locations. Second, an extensive analysis of the residuals must be contacted in both in-sample and out-of-sample sets. An understanding of the dynamics that govern the residuals will provide additional information of the validity of the proposed models. The space limitation of this paper prevents us from doing so. Other potential future work could be to further improve the GP models. At the moment, a simple GP was used. However, such GPs are open to criticisms of effective model generalization. A way of tackling this can be by using ensemble learning techniques. We aim to do this next. Also as it was mentioned earlier, the GP is currently based on DAT of the three previous days. Our goal is to allow this window to be changed dynamically through GP operators. We believe that this could lead to even more effective models.

Nevertheless, our preliminary results indicate that the proposed methods can model the dynamics of the temperature very well and they can constitute an accurate method for temperature derivatives pricing.

Table 3. Predictive performance of the four methods

	1-day-ahead	Out-of-sample	Total
WN	18 (60%)	4 (13%)	22 (36.6%)
GP	4 (13%)	12 (40%)	16 (26.7%)
Alaton	0 (0%)	6 (20%)	6 (10.0%)
Benth	8 (27%)	8 (27%)	16 (26.7%)

The number of datasets that each method has the best predictive accuracy. Percentages are reported in parentheses.

References

1. Alexandridis, A., Zapranis, A.: Weather Derivatives: Modeling and Pricing Weather-Related Risk. Springer (2013)
2. Challis, S.: Bright Forecast for Profits. Reactions June edition, (1999)
3. Hanley, M.: Hedging the Force of Nature. Risk Professional 1, 21-25 (1999)
4. Pati, Y.C., Krishnaprasad, P.S.: Analysis and Synthesis of Feedforward Neural Networks Using Discrete Affine Wavelet Transforms. IEEE Trans. on Neural Networks 4, 73-85 (1993)
5. Alexandridis, A.K., Zapranis, A.D.: Wavelet neural networks: A practical guide. Neural Networks 42, 1-27 (2013)
6. Banzhaf, W., Nordin, P., Keller, R.E., Francone, F.D.: Genetic Programming—An Introduction: On the Automatic Evolution of Computer Programs and Its Applications, dpunkt. verlag and Morgan Kaufmann Publishers. Inc., San Francisco, California (1998)
7. Koza, J.R.: Genetic programming as a means for programming computers by natural selection. Statistics and Computing 4, 87-112 (1994)

8. Poli, R., Langdon, W.W.B., McPhee, N.F.: Field Guide to Genetic Programming. Lulu Enterprises Uk Limited (2008)
9. Agapitos, A., O'Neill, M., Brabazon, A.: Evolving seasonal forecasting models with genetic programming in the context of pricing weather-derivatives. Applications of Evolutionary Computation, pp. 135-144. Springer (2012)
10. Agapitos, A., O'Neill, M., Brabazon, A.: Genetic Programming for the Induction of Seasonal Forecasts: A Study on Weather Derivatives. Financial Decision Making Using Computational Intelligence, pp. 159-188. Springer (2012)
11. Jewson, S., Brix, A., Ziehmann, C.: Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations. Campbridge University Press, Cambridge, UK (2005)
12. Alaton, P., Djehince, B., Stillberg, D.: On Modelling and Pricing Weather Derivatives. Applied Mathematical Finance 9, 1-20 (2002)
13. Benth, F.E., Saltyte-Benth, J.: The volatility of temperature and pricing of weather derivatives. Quantitative Finance 7, 553-561 (2007)
14. Cao, M., Wei, J.: Weather Derivatives valuation and market price of weather risk. Journal of Future Markets 24, 1065-1089 (2004)
15. Dornier, F., Queruel, M.: Caution to the wind. Weather risk special report, Energy Power Risk Management August, 30-32 (2000)
16. Bellini, F.: The weather derivatives market: Modelling and pricing temperature. Faculty of Economics, vol. Ph.D. University of Lugano, Lugano, Switzerland (2005)
17. Benth, F.E., Saltyte-Benth, J.: Stochastic Modelling of Temperature Variations With a View Towards Weather Derivatives. Applied Mathematical Finance 12, 53-85 (2005)
18. Zapranis, A., Alexandridis, A.: Modelling Temperature Time Dependent Speed of Mean Reversion in the Context of Weather Derivetive Pricing. Applied Mathematical Finance 15, 355 - 386 (2008)
19. Zapranis, A., Alexandridis, A.: Weather Derivatives Pricing: Modelling the Seasonal Residuals Variance of an Ornstein-Uhlenbeck Temperature Process With Neural Networks. Neurocomputing 73, 37-48 (2009)
20. Carmona, R.: Calibrating Degree Day Options. In: 3rd Seminar on Stochastic Analysis, Random Field and Applications. (Year)
21. Geman, H., Leonardi, M.-P.: Alternative approaches to weather derivatives pricing. Managerial Finance 31, 46-72 (2005)
22. Zapranis, A., Alexandridis, A.: Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing. Neural Computing & Applications 20, 787-801 (2011)
23. Šaltytė Benth, J., Benth, F.E., Jalinskas, P.: A Spatial-temporal Model for Temperature with Seasonal Variance. Journal of Applied Statistics 34, 823 - 841 (2007)
24. Demšar, J.: Statistical Comparisons of Classifiers over Multiple Data Sets. J. Mach. Learn. Res. 7, 1-30 (2006)
25. Garcıa, S., Herrera, F.: An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons. Journal of Machine Learning Research 9, 66 (2008)