Evolving Seasonal Forecasting Models with Genetic Programming in the Context of Pricing Weather-Derivatives

Alexandros Agapitos, Michael O'Neill, and Anthony Brabazon

Financial Mathematics and Computation Research Cluster
Natural Computing Research and Applications Group
Complex and Adaptive Systems Laboratory
University College Dublin, Ireland
{alexandros.agapitos,m.oneill,anthony.brabazon}@ucd.ie

Abstract. In this study we evolve seasonal forecasting temperature models, using Genetic Programming (GP), in order to provide an accurate, localised, long-term forecast of a temperature profile as part of the broader process of determining appropriate pricing model for weather-derivatives, financial instruments that allow organisations to protect themselves against the commercial risks posed by weather fluctuations. Two different approaches for time-series modelling are adopted. The first is based on a simple system identification approach whereby the temporal index of the time-series is used as the sole regressor of the evolved model. The second is based on iterated single-step prediction that resembles autoregressive and moving average models in statistical time-series modelling. Empirical results suggest that GP is able to successfully induce seasonal forecasting models, and that autoregressive models compose a more stable unit of evolution in terms of generalisation performance for the three datasets investigated.

1 Introduction

Weather conditions affect the cash flows and profits of businesses in a multitude of ways. For example, energy companies (gas or electric) may sell fewer supplies if a winter is warmer than usual, leisure industry firms such as ski resorts, theme parks, hotels are affected by weather metrics such as temperature, snowfall or rainfall, construction firms can be affected by rainfall, temperatures and wind levels, and agricultural firms can be impacted by weather conditions during the growing or harvesting seasons [3]. Firms in the retail, manufacturing, insurance, transport, and brewing sectors will also have weather "exposure". Less obvious weather exposures include the correlation of events such as the occurrence of plant disease with certain weather conditions (i.e. blight in potatoes and in wheat) [9]. Another interesting example of weather risk is provided by the use of "Frost Day" cover by some of the UK town/county councils whereby a payout is obtained by them if a certain number of frost days (when roads would require gritting - with an associated cost) are exceeded. Putting the above into context,

C. Di Chio et al. (Eds.): EvoApplications 2012, LNCS 7248, pp. 135–144, 2012.

it is estimated that in excess of $1 trillion of activity in the US economy is weather-sensitive [5]. In response to the existence of weather risk, a series of financial products have been developed in order to help organisations manage these risks. *Weather derivatives* are financial products that provide a payout which is related to the occurrence of pre-defined weather events [7].

A key component of the accurate pricing of a weather derivative are forecasts of the expected value of the underlying weather variable and its associated volatility. The goal of this study is to produce predictive models by the means of Genetic Programming [4] (GP) of the stochastic process that describes temperature. Section 2 introduces weather derivatives, motivates the need for seasonal temperature forecasting, and reviews the major statistical and heuristic time-series modelling methods. Section 3 describes the experiment design, Section 4 discusses the empirical findings, and finally Section 5 draws our conclusions.

2 Background

2.1 OTC Weather Derivatives

The earliest weather derivatives were traded over-the-counter (OTC) as individually negotiated contracts. In OTC contracts, one party usually wishes to hedge a weather exposure in order to reduce cash flow volatility. The payout of the contract may be linked to the value of a weather index on the Chicago Mercantile Exchange (CME) or may be custom-designed. The contract will specify the weather metric chosen, the period (a month, a season) over which it will be measured, where it will be measured (often a major weather station at a large airport), the scale of payoffs depending on the actual value of the weather metric and the cost of the contract. The contract may be a simple "swap" where one party agrees to pay the other if the metric exceeds a pre-determined level while the other party agrees to pay if the metric falls below that level.

In the US, many OTC (and all exchange-traded) contracts are based on the concept of a *'degree-day'*. A degree-day is the deviation of a day's average temperature from a reference temperature. Degree days are usually defined as either *'Heating Degree Days'* (HDDs) or *'Cooling Degree Days'* (CDDs). The origin of these terms lies in the energy sector which historically (in the US) used 65 degrees Fahrenheit as a baseline, as this was considered to be the temperature below which heating furnaces would be switched on (a heating day) and above which air-conditioners would be switched on (a cooling day). As a result HDDs and CDDs are defined as

$$HDD = \text{Max (0, 65}^o\text{F - average daily temperature)} \tag{1}$$

$$CDD = \text{Max (0, average daily temperature - 65}^o\text{F)} \tag{2}$$

For example, if the average daily temperature for December 20th is 36^oF, then this corresponds to 29 HDDs (65 - 36 = 29). The payoff of a weather future is usually linked to the aggregate number of these in a chosen time period (one HDD or CDD is typically worth $20 per contract). Hence, the payoff to a December contract for HDDs which (for example) trades at 1025 HDDs on 1st December - assuming that there was a total of 1080 HDDs during December - would be $1,100 ($20 * (1080-1025). A comprehensive introduction to weather derivatives is provided by [7].

2.2 Seasonal Forecasting for Pricing a Weather Derivative

A substantial literature exists concerning the pricing of financial derivatives. However, models from this literature cannot be simply applied for pricing of weather derivatives as there are a number of important differences between the two domains. The *underlying* (variable) in a weather derivative (a weather metric) is non-traded and has no intrinsic value in itself (unlike the underlying in a traditional derivative which is typically a traded financial asset such as a share or a bond). It is also notable that changes in weather metrics do not follow a pure random walk as values will typically be quite bounded at specific locations. Standard (arbitrage-free) approaches to derivatives pricing (such as the Black-Scholes option pricing model) are inappropriate as there is no easy way to construct a portfolio of financial assets which replicates the payoff to a weather derivative [6].

One method that is used to price weather risk is *index modelling*. This approach attempts to build a model of the distribution of the underlying weather metric (for example, the number of seasonal cumulative heating degree days), typically using historical data. A wide variety of forecasting approaches such as time-series models, of differing granularity and accuracy, can be employed.

In considering the use of weather forecast information for derivatives pricing, we can distinguish between a number of possible scenarios. In this paper we are focusing on weather derivatives that are traded long before the start of the relevant weather period. In this case we can only use *seasonal forecasting* methods as current short run weather forecasts have no useful information content in predicting the weather than will arise during the weather period. *Seasonal forecasts* are long-term forecasts having a time horizon beyond one month [10]. There are a plethora of methods for producing these forecasts ranging from the use of statistical time-series models based on historic data to the use of complex, course-grained, simulation models which incorporate ocean and atmospheric data. The following sections briefly review some of the major techniques that fall into the two families of *statistical* and *heuristic* approaches to time-series forecasting.

2.3 Statistical Time-Series Forecasting Methods

Statistical time-series forecasting methods fall into the following five categories: (a) *exponential smoothing methods*; (b) *regression methods*; (c) *autoregressive integrated moving average methods* (ARIMA); (d) *threshold methods*; (e) *generalised autoregressive conditionally heteroskedastic methods* (GARCH). The first three categories can be considered as linear, whereas the last two as non-linear methods.

In *exponential smoothing*, a forecast is given as a weighted moving average of recent time-series observations. The weights assigned decrease exponentially as the observations get older. In *regression*, a forecast is given as a linear combination of one or more explanatory variables. ARIMA models give a forecast as a linear function of past observations and error values between the time-series itself and past observations of explanatory variables. These models are essentially based on a composition of *autoregressive models* (linear prediction formulas that attempt to predict an output of a system based on the previous outputs), and moving average models (linear prediction model based on a *white noise* stationary time-series). For a discussion on smoothing, regression and ARIMA methods see [8]. Linear models cannot capture some featured that commonly occur in real-world data such as asymmetric cycles and outliers.

Threshold methods [8] assume that extant asymmetric cycles are cause by distinct underlying phases of the time-series, and that there is a transition period between these phases. Commonly, the individual phases are given a linear functional form, and the transition period is modelled as an exponential or logistic function. GARCH methods [2] are used to deal with time-series that display non-constant variance of residuals (error values). In these methods, the variance of error values is modelled as a quadratic function of past variance values and past error values.

2.4 Genetic Programming for Time-Series Modelling

In GP-based time-series prediction [1] the task is to induce a model that consists of the best possible approximation of the stochastic process that could have generated an observed time-series. Given *delayed vectors* v, the aim is to induce a model f that maps the vector v to the value x_{t+1}. That is,

$$x_{t+1} = f(v) = f(x_{t-(m-1)\tau}, x_{t-(m-2)\tau}, \dots, x_t) \tag{3}$$

where m is embedding dimension and τ is delay time. The embedding specifies on which historical data in the series the current time value depends. These models are known as *single-step predictors*, and are used to predict to predict one value x_{t+1} of the time series when all inputs $x_{t-m}, \dots, x_{t-2}, x_{t-1}, x_t$ are given. For long-term forecasts, *iterated single-step prediction models* are employed to forecast further than one step in the future. Each predicted output is fed back as input for the next prediction while all other inputs are shifted back one place. The input consists partially of predicted values as opposed to observables from the original time-series. That is,

$$\begin{aligned}
x'_{t+1} &= f(x_{t-m}, \dots, x_{t-1}, x_t); m < t \\
x'_{t+2} &= f(x_{t-m+1}, \dots, x_t, x'_{t+1}); m < t \\
&\vdots \\
x'_{t+k} &= f(x_{t-m+k-1}, \dots, x'_{t+k-2}, x'_{t+k-1}); m < t, k \geq 1
\end{aligned} \tag{4}$$

where k is the prediction step.

2.5 Scope of Research

The goal of this study is to produce predictive models of the stochastic process that describes temperature. More specifically, we are interested in modelling aggregate monthly HDDs. The incorporation of this model into a complete pricing model for weather derivatives is left for future work. We also restrict attention to the case where the contract period for the derivative has not yet commenced. Hence, we ignore short-run weather forecasts, and concentrate on seasonal forecasting.

We investigate two families of program representations for time-series modelling. The first is the standard GP technique, genetic symbolic regression (GSR), applied to the forecasting problem in the same way that is applied to symbolic regression problems. The task is to approximate a periodic function, where *temperature* (HDDs) is the dependent variable, and *time* is the sole regressor variable. The second representation allows the induction of iterated single-step predictors that can resemble autoregressive (GP-AR) and autoregressive moving average (GP-ARMA) time-series models as described in Section 2.3.

3 Experiment Design

3.1 Model Data

Three US weather stations were selected: (a) Atlanta (ATL); (b) Dallas, Fort Worth (DEN); (c) La Guardia, New York (DFW). All the weather stations were based at major domestic airports and the information collected included date, maximum daily temperature, minimum daily temperature, and the associated HDDs and CDDs for the day. This data was preprocessed to create new time-series of *monthly* aggregate HDDs and CDDs for each weather station respectively.

There is generally no agreement on the appropriate length of the time-series which should be used in attempts to predict future temperatures. Prior studies have used lengths of twenty to fifty years, and as a compromise this study uses data for each location for the period 01/01/1979 - 31/12/2002. The monthly HDDs data for each location is divided into a *training set* (15 years) that measures the performance during the learning phase, and a *test set* (9 years) that quantifies model generalisation.

3.2 Forecasting Model Representations and Run Parameters

This study investigates the use of two families of seasonal forecast model representations, where the forecasting horizon is set to 6 months. The first is based on standard GP-based symbolic regression (GSR), where *time* serves as the regressor variable (corresponding to a month of a year), and *monthly HDDs* is the regressand variable. Assuming that time t is the start of the forecast, we can obtain a 6-month forecast by executing the program with inputs $\{t+1, \ldots, t+6\}$.

Table 1. Learning algorithm parameters

EA	panmictic, generational, elitist GP with an expression-tree representation
No. of generations	51
Population size	1,000
Tournament size	4
Tree creation	ramped half-and-half (depths of 2 to 6)
Max. tree depth	17
Subtree crossover	30%
Subtree mutation	40%
Point mutation	30%
Fitness function	Root Mean Squared Error (RMSE)

The second representation for evolving seasonal forecasting models is based on the iterated single-step prediction that can emulate autoregressive models, as described in Section 2.3. This method requires that delayed vectors from the monthly HDDs time-series are given as input to the model, with each consecutive model output being added at the end of the delayed input vector, while all other inputs are shifted back one place.

Table 2. Forecasting model representation languages

Forecasting model	Function set	Terminal set
GSR	`add`, `sub`, `mul`, `div`, `exp`, `log`, `sqrt`, `sin`, `cos`	index t corresponding to a month 10 rand. constants in -1.0, ..., 1.0 10 rand. constants in -10.0, ..., 10.0
GP-AR(12)	`add`, `sub`, `mul`, `div`, `exp`, `log`, `sqrt`	10 rand. constants in -1.0, ..., 1.0 10 rand. constants in -10.0, ..., 10.0 $HDD_{t-1}, \ldots, HDD_{t-12}$
GP-AR(24)	`add`, `sub`, `mul`, `div`, `exp`, `log`, `sqrt`	10 rand. constants in -1.0, ..., 1.0 10 rand. constants in -10.0, ..., 10.0 $HDD_{t-1}, \ldots, HDD_{t-24}$
GP-AR(36)	`add`, `sub`, `mul`, `div`, `exp`, `log`, `sqrt`	10 rand. constants in -1.0, ..., 1.0 10 rand. constants in -10.0, ..., 10.0 $HDD_{t-1}, \ldots, HDD_{t-36}$
GP-ARMA(36)	`add`, `sub`, `mul`, `div`, `exp`, `log`, `sqrt`	10 rand. constants in -1.0, ..., 1.0 10 rand. constants in -10.0, ..., 10.0 $HDD_{t-1}, \ldots, HDD_{t-36}$ $\mathrm{M}(HDD_{t-1},\ldots, HDD_{t-6})$, $\mathrm{SD}(HDD_{t-1},\ldots, HDD_{t-6})$ $\mathrm{M}(HDD_{t-1},\ldots, HDD_{t-12})$, $\mathrm{SD}(HDD_{t-1},\ldots, HDD_{t-12})$ $\mathrm{M}(HDD_{t-1},\ldots, HDD_{t-18})$, $\mathrm{SD}(HDD_{t-1},\ldots, HDD_{t-18})$ $\mathrm{M}(HDD_{t-1},\ldots, HDD_{t-24})$, $\mathrm{SD}(HDD_{t-1},\ldots, HDD_{t-24})$ $\mathrm{M}(HDD_{t-1},\ldots, HDD_{t-30})$, $\mathrm{SD}(HDD_{t-1},\ldots, HDD_{t-30})$ $\mathrm{M}(HDD_{t-1},\ldots, HDD_{t-36})$, $\mathrm{SD}(HDD_{t-1},\ldots, HDD_{t-36})$

Table 2 shows the primitive single-type language elements that are being used for forecasting model representation in different experiment configurations. For GSR, the function set contains standard arithmetic operators (protected division) along with e^x, $log(x)$, $\sqrt{x}$, and finally the trigonometric functions of sine and cosine. The terminal set is composed of the index t representing a month, and random constants within specified ranges. GP-AR(12), GP-AR(24), GP-AR(36), all correspond to standard autoregressive models that are implemented as iterated single-step prediction models. The argument in the parentheses specifies the number of past time-series values that are available as input to the model. The function set in this case is similar to that of GSR excluding the trigonometric functions, whereas the terminal set is augmented with historical monthly HDD values. For the final model configuration, GP-ARMA(36), the function set is identical to the one used in the other autoregressive models configurations, however the terminal set contains moving averages, denoted by $M(HDD_{t-1}, \ldots, HDD_{t-\lambda})$, where λ is the time-lag and HDD_{t-1} and $HDD_{t-\lambda}$ represent the bounds of the moving average period. For every moving average, the associated standard deviation for that period is also given as model input, and is denoted by $SD(HDD_{t-1}, \ldots, HDD_{t-\lambda})$. Finally, Table 1 presents the parameters of our learning algorithm.

4 Results

Table 3. Comparison of training and testing RMSE obtained by different forecasting configurations, each experiment was ran for 50 times. Standard error for mean in parentheses. Bold face indicates best performance on test data.

Dataset	Forecasting configuration	Mean Training RMSE	Best Training RMSE	Mean Testing RMSE	Best Testing RMSE
ATL	GSR	140.52 (9.55)	68.82	149.53 (8.53)	**72.73**
	GP-AR(12)	92.44 (0.54)	81.78	111.87 (0.41)	103.60
	GP-AR(24)	91.33 (0.68)	83.33	96.15 (0.51)	91.26
	GP-AR(36)	88.96 (0.81)	77.30	90.38 (0.81)	79.44
	GP-ARMA	85.20 (0.86)	75.84	85.71 (0.82)	74.31
DEN	GSR	165.76 (11.46)	103.09	180.46 (11.74)	**95.23**
	GP-AR(12)	133.18 (0.43)	121.38	126.78 (0.25)	117.19
	GP-AR(24)	130.41 (0.73)	111.48	124.36 (0.66)	110.31
	GP-AR(36)	131.13 (1.08)	114.86	111.41 (0.57)	103.73
	GP-ARMA	126.46 (1.29)	106.18	108.90 (0.64)	101.57
DFW	GSR	118.96 (8.02)	66.49	118.69 (7.20)	66.12
	GP-AR(12)	88.75 (0.66)	80.64	90.37 (0.26)	86.57
	GP-AR(24)	96.14 (0.95)	83.55	85.36 (0.42)	78.24
	GP-AR(36)	89.52 (0.69)	81.12	62.11 (0.43)	55.84
	GP-ARMA	87.09 (0.82)	75.41	60.92 (0.52)	**55.10**

We performed 50 independent evolutionary runs for each forecasting model configuration presented in Table 2. A summary of average and best training and test results obtained by different models is presented in Table 3. The distributions of training and test errors obtained at the end of the evolutionary runs are depicted in Figure 1 for the DFW time-series. Graphs for the other time-series exhibited a similar trend and were omitted due to lack of space. Results suggest that the family of autoregressive moving average models perform better on average than those obtained with standard symbolic regression. A statistical significance difference (unpaired t-test, two-tailed, $p < 0.0001$, degrees of freedom $df = 98$) was found between the average test RMSE for GSR and GP-ARMA in all three datasets. Despite the fact that the ARMA representation space offers a more stable unit for evolution than the essentially free-of-domain-knowledge GSR space, best testing RMSE results indicated that GSR models are better performers in ATL and DEN datasets, as opposed to the DFW dataset, where the best-of-50-runs GP-ARMA model appeared superior.

Given that in time-series modelling it is often practical to assume a *deterministic* and a *stochastic* part in a series' dynamics, this result can well corroborate on the ability of standard symbolic regression models to effectively capture the deterministic aspect of a time-series, and successfully forecast future values in the case of time-series with a weak stochastic or volatile part. Another interesting observation is that there is a difference in the generalisation performance between GP-AR models of different order, suggesting that the higher the order of the AR process the better its performance on seasonal forecasting. Statistical significant differences (unpaired t-test, $p < 0.0001$, $df = 98$) were found in mean test RMSE between GP-AR models of order 12 and those of order 36, in all three datasets.

During the learning process, we monitored the generalisation performance of the best-of-generation individual, and we adopted a model selection strategy whereby the best-generalising individual is designated as the outcome of the run. In the context of *early stopping* for counteracting the marked tendency of model overtraining, Figures 1(g), (h), (i) illustrate the distributions of the generation number where model selection was performed, for the three datasets. It can be seen that GSR models are less prone to overtraining, then follows

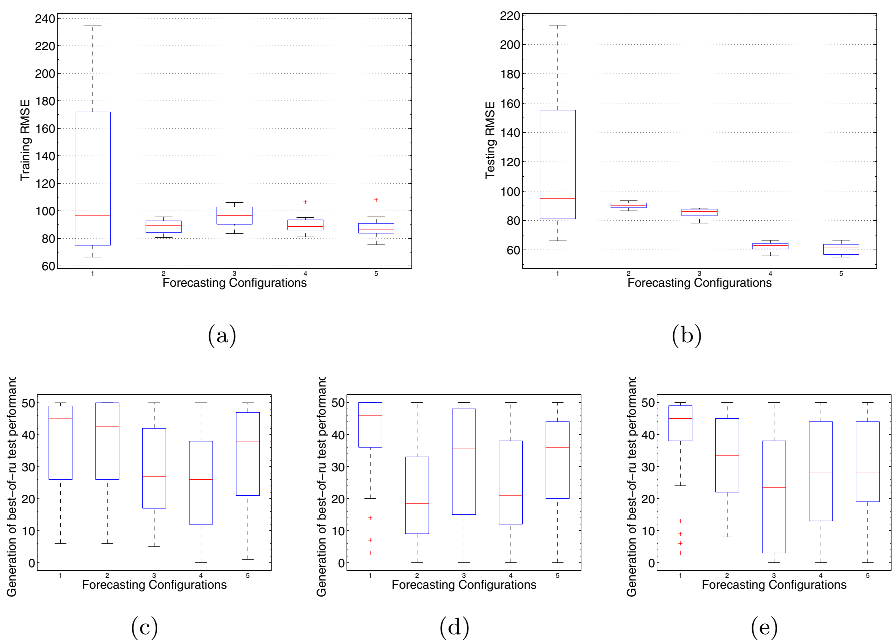
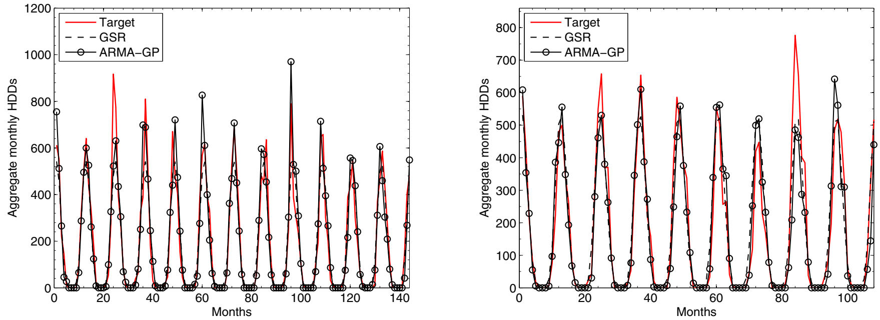


Fig. 1. Distribution of best-of-run training and test RMSE accrued from 50 independent runs. Figures (a), (b) for DFW. Figures (c), (d), (e) show the distribution of generation number where each best-of-run individual on test data was discovered for the cases of ATL, DEN, and DFW respectively.



(a) (b)

Fig. 2. Target vs. Prediction for best-performing models of GSR and GP-ARMA for the DFW dataset. (a) training data; (b) test data.

GP-ARMA, and finally it can be noted that GP-AR models of high order are the most sensitive to overfitting the training data. Interestingly is the fact that this observation is consistent in all three datasets.

Finally, Figures 2(a), (b) show the target and predicted values from best-performing GSR and GP-ARMA models for the DWF dataset, for training and testing data respectively. Both models achieved a good fit for most of the out-of-sample range. Equation 5 illustrates the parabolic GP-ARMA model that generated the predictions in Figure 2.

$$f(t) = \sqrt{HDD_{t-12} * \left(HDD_{t-36} + \sqrt{HDD_{t-12} * \left(\frac{HDD_{t-26}}{-0.92 + (HDD_{t-7} * log(HDD_{t-21}))}\right)}\right)} \tag{5}$$

5 Conclusion

This paper adopted a time-series modelling approach to the production of a seasonal weather metric forecast, as part of a general method for pricing weather derivatives. Two GP-based methods for time series modelling were used; the first one is based on standard symbolic regression; the second one is based on autoregressive time-series modelling that is realised via an iterated single-step prediction process and a specially crafted terminal set of past time-series information.

Results are very encouraging, suggesting that GP is able to successfully evolve accurate seasonal forecasting models. More specifically, for two of the three time-series considered in this study, standard symbolic regression was able to capture the deterministic aspect of the modelled data and attained the best test performance, however its overall performance was marked as unstable, producing some very poor-generalising models. On the other hand, the performance of search-based autoregressive moving average models was deemed on average the most stable in out-of-sample data. On a more general note, experiments also revealed a marked tendency of the GP-AR models to overfit the most, with GSR being the most resilient program representation in this problem domain. Whether this is due to a slower learning curve in the case of GSR, or to a very sensitive to overfitting representation in the case on GP-AR models is left to be seen in future work. B

Acknowledgement. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant Number 08/SRC/FM1389.

References

1. Agapitos, A., Dyson, M., Kovalchuk, J., Lucas, S.M.: On the genetic programming of time-series predictors for supply chain management. In: GECCO 2008: Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation (2008)
2. Bollerslev, T.: Generalised autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327 (1986)
3. Garcia, A., Sturzenegger, F.: Hedging Corporate Revenues with Weather Derivatives: A case study. Master's thesis, Universite de Lausanne (2001)
4. Koza, J.: Genetic Programming: on the programming of computers by means of natural selection. MIT Press, Cambridge (1992)
5. Cao, M., Wei, J.: Equilibrium valuation of weather derivatives, working paper. School of Business, York University, Toronto (2002)
6. Campbell, S., Diebold, F.: Weather forecasting for weather derivatives. Journal of the American Statistical Association 100(469), 6–16 (2005)
7. Jewson, S., Brix, A., Ziehman, C.: Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations. Cambridge University Press (2005)
8. Makridakis, S., Wheelwright, S., Hyndman, R.: Forcasting: Methods and Applications. Willey, New York (1998)
9. Sprundel, V.: Using weather derivatives for the financial risk management of plant diseases: A study on Phytophthora infestans and Fusarium head blight. Ph.D. thesis, Wageningen University (2011)
10. Weigel, A., Baggenstos, D., Liniger, M.: Probabilistic verification of monthly temperature forecasts. Monthly Weather Review 136, 5162–5182 (2008)