Meteorol. Appl. 00, 1–10 (2003) DOI:10.1017/S1350482703001105

Seasonality in the statistics of surface air temperature and the pricing of weather derivatives

Stephen Jewson[1] & Rodrigo Caballero[2]
[1] *Risk Management Solutions, 10 Eastcheap, London, EC3M 1AJ, UK*
Email: x@stephenjewson.com
[2] *Department of the Geophysical Sciences, University of Chicago, 5734 S. Ellis Ave., Chicago IL 60637, USA*
Email: rca@geosci.uchicago.edu

The pricing of weather derivatives motivates the need to build accurate statistical models of daily temperature variability. Current published models are shown to be inaccurate for locations that show strong seasonality in the probability distribution and autocorrelation structure of temperature anomalies. With respect to the first of these problems, we present a new transform that allows seasonally varying non-normal temperature anomaly distributions to be cast into normal distributions. With respect to the second, we present a new parametric time-series model that captures both the seasonality and the slow decay of the autocorrelation structure of observed temperature anomalies. This model is valid when the seasonality is slowly varying. We also present a simple non-parametric method that is accurate in all cases, including extreme non-normality and rapidly varying seasonality. Application of these new methods in some realistic weather derivative valuation examples shows that they can have a very large impact on the final price when compared to existing methods.

1. Introduction

Weather derivatives are contracts that allow entities to insure themselves against the financial losses that can occur due to unfavourable weather conditions. For instance, there are many retailing companies which have sales that are adversely affected by either warmer or colder than normal temperatures. These companies could all potentially benefit from hedging their exposure with temperature-based weather derivatives.

The payoff from a weather derivative is determined by the outcome of a weather index such as mean summer or winter temperature. Pricing of weather derivatives is mainly based on estimates of the distribution of possible values of the index. Before any forecasts are available historical weather data must be obtained to make this estimate. Commonly a simple method is used in which the values of the index for past years are determined. Using 30 years of historical data would give 30 historical index values (for example, 30 mean winter temperatures) and these would be taken as 30 independent samples from the distribution to be estimated. In most cases, a trend would be removed prior to estimating the index distribution, either from the daily weather data or from the 30 annual historical index values. This method for deriving the index distribution is known as (detrended) *burn analysis*.

The most obvious disadvantage of burn analysis is that it does not sample the possible extreme outcomes of the index very well. This can be overcome by fitting an appropriate distribution to the historical index values, which smooths the distribution and extrapolates it to higher and lower levels of probability. This is known as *index modelling*.

However, neither of these so-called *index-based* approaches for determining the distribution of the index achieve the highest possible accuracy because much of the historical data is discarded in the calculation of the historical index values. For example, when considering a one-week contract, 51/52 of the data is not used by these methods. If some of the data from the other 51 weeks of the year contain statistical information about the behaviour of weather during the one week of the contract, as is likely, then it could be more accurate to use some of these data too.

These considerations lead one naturally to consider the possibility of using statistical modelling of temperatures on a daily basis, and considerable work has gone in to trying to build such models. In the most common approach, a deterministic seasonal cycle is removed from the mean and/or standard deviation of temperature, and the residuals (which we will call *temperature anomalies*) are then modelled using continuous or discrete stochastic processes. The difficulty lies in finding stochastic processes that accurately capture the observed behaviour of the temperature anomalies. Dischel (1998), Cao & Wei (2000), Torro et al. (2001) and Alaton et al. (2002) have suggested using AR(1) (first-order autoregressive) models or continuous equivalents. Others (Dornier & Querel 2000; Diebold & Campbell 2001) have suggested more general models that all lie within the larger class of autoregressive moving-average (ARMA) models (Box & Jenkins 1970). Caballero et al. (2002) (henceforth CJB) have shown that all these models fail to capture the slow decay of the temperature autocorrelation function, and hence lead to significant underpricing of weather options. CJB and Brody et al. (2002) (henceforth BSZ) have suggested Gaussian discrete and continuous stochastic processes, respectively, that overcome this problem by using models with long memory (i.e. power law decay of the autocorrelation function). CJB model the variability of temperature anomalies using a stationary autoregressive fractionally integrated moving average (ARFIMA) process (Granger & Joyeux 1980), while BSZ use a fractional Ornstein-Uhlenbeck (fOU) process. ARFIMA and fOU models are mutual analogies in the discrete and continuous domains. These models work well for locations where the assumptions of normality and stationarity are accurate. However, as we shall see below, temperature anomalies at many locations show marked departures from normality and stationarity. In these cases, the CJB and BSZ models should not be used for pricing weather derivatives as they are likely to give misleading results.

This paper tackles the question of how to model temperatures at these locations. In Section 2 we describe the data sets to be used. In Section 3 we examine the evidence for non-stationarity in the autocorrelation structure of surface air temperatures. In Section 4 we model it using a new class of parametric statistical models that we have developed specifically for this purpose. In Section 4.2 we also describe a simple non-parametric model that provides an alternative method for cases where parametric models entirely fail. In Section 5 we investigate seasonally varying non-normality of temperature variability and we present a new method that can be used to model such non-normality. In Section 6, we compare the performance of the various approaches in pricing a specific weather derivative. In Section 7 we draw some conclusions.

2. Data

The studies of temperature variability described in this paper are based on two data sets. The first data set originates from the US National Weather Service and consists of daily minimum and maximum temperatures measured between midnight and midnight (local time) for 40 years (1961–2000) at 200 US weather stations (in this study we only present results for eight of them, listed in Table 1). These data are not directly suitable for analysis because of (a) gaps in the data due to failures in measurement equipment, recording methods or data transmission and storage, and (b) jumps in the data due to changes in measurement equipment or station location. To avoid these problems we use data that have been processed by Earth Satellite Corporation to fill in such gaps and remove such jumps (Boissonnade et al. 2002). These processed data were provided to the authors by Risk Management Solutions Ltd. Even after such processing the data are still not representative of the current climate because of trends due to urbanisation and global warming. Removing such trends is extremely difficult, as the trends vary by location, by year and by season. Attempting to model the variations by year and season can only by done rather subjectively, and so we have restricted ourselves to the simple and transparent approach of removing a non-seasonal linear trend, estimated over the entire data period, for daily mean temperature at each station. The linear trends were fitted by the method of least squares, and removed in such a way that the detrended data values are consistent with the last day of the data (in other words, where there is a warming trend we increase the earlier values up to present-day levels). Most weather contracts depend on daily mid-point temperature, calculated as the mid-point between the daily maximum and minimum, and so all our modelling will focus on these values.

Table 1. *The eight US weather stations used in the modelling described in the text, with the optimum lengths of the four moving averages as selected automatically as part of the fitting procedure for the AROMA model.*

Station	m_1	m_2	m_3	m_4
Chicago Midway	1	2	3	17
Miami	1	2	4	28
Los Angeles	1	2	9	33
Boston	1	2	5	32
New York Central Park	1	2	4	18
Charleston	1	2	4	22
Detroit	1	2	3	24
Atlanta	1	2	7	27

The second data set is the US National Centre for Environmental Prediction (NCEP)'s 40 year reanalysis (Kalnay et al. 1996). This data set is obtained by assimilating observations into a dynamical model and produces a gridded representation of the climate over the whole globe. We only use surface temperature values from these data. The reason we use these data in addition to the station data described above is simply that the reanalysis data are conveniently presented on a spatial grid, which makes it much easier to produce maps of spatial fields.

Both temperature data sets are converted to anomalies by removing deterministic seasonal cycles in the mean

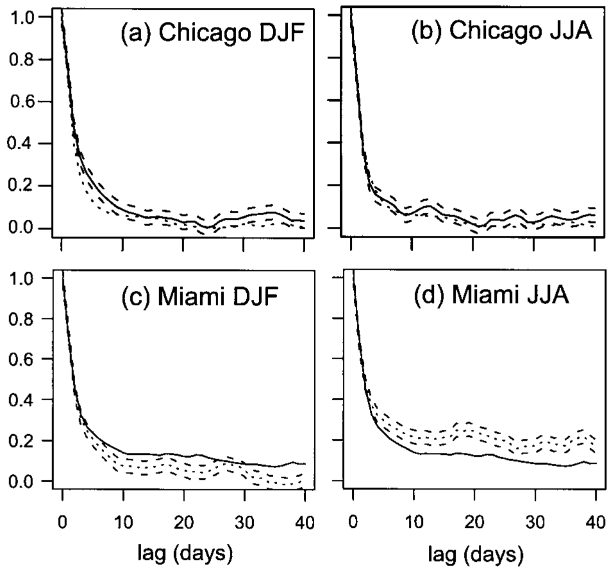

Figure 1. *Observed ACFs for Chicago (top row) and Miami (bottom row) in winter (defined as December–January, DJF) and summer (June–August, JJA). In each panel, the solid black line is the annual ACF (i.e. the ACF computed using the entire data set) and the dotted line is the ACF for the season specified (computed as the average of the seasonal ACFs in the 40 individual years). The dashed lines show the 95% confidence intervals around the observed estimate, calculated using the method of Moran (1947).*

and the standard deviation: this was achieved by regression onto three sines and cosines in each case.

3. Seasonality in the autocorrelation structure of surface air temperatures

In Figure 1 we compare the seasonal autocorrelation functions (ACFs) at two locations, Chicago and Miami. Chicago shows essentially no seasonality, with no statistically significant difference between summer, winter and annual ACFs. The situation is strikingly different in Miami. Persistence of temperature anomalies is clearly much higher in summer than in winter. It is clear from this that a stochastic model (such as those used in CJB or BSZ) that assumes stationarity of the ACF will severely underestimate the memory in summer, and will overestimate it in winter. If such a model is used to derive the distribution of a weather index such as cumulative temperature, it will severely underestimate the standard deviation of the index in summer (see CJB, Equation 11)

We turn to the NCEP data set to examine how widespread seasonality of the ACF is. We define an index s as

$$s = \sum_{k=1}^{40} \rho_k \qquad (1)$$

where ρ_k is the ACF for a particular season at lag k. The higher the value of s, the greater the persistence of the temperature anomalies. Figure 2 shows maps of s for summer and winter over North America and Europe. In both seasons, persistence is greater over the oceans. In winter, persistence is relatively uniform over North America, but shows a banded pattern over Europe, with high-persistence bands over Scandinavia and the Mediterranean and lower persistence over the southern European mainland. The situation changes markedly in summer. The biggest change is over the oceans, where persistence increases almost everywhere. There are also changes over the continents. In North America, persistence increases over the Gulf states and southwestern USA, but decreases over coastal California. In Europe, the banded pattern described above is still in place but the relative amplitudes change, with persistence decreasing over Scandinavia and central Europe and increasing over the Mediterranean. In summary, we find strong seasonality of the ACF over large (and economically significant) parts of the USA and Europe. This motivates the search for a time series model capable of capturing such seasonality.

4. Modelling of seasonality in the autocorrelation structure

We now proceed to the main topic of this article, which is the modelling of seasonality in the autocorrelation structure of temperature. A first approach might be to try to extend the ARFIMA model of CJB to include seasonality. This could be attempted by allowing the parameters of the model to vary with time of year, or perhaps by fitting the model to data from only one part of the year. Both such approaches are theoretically possible. However, they are also rather complex. The ARFIMA model represents the slow decay of the ACF using a long memory parameter d. A model that allows the long memory parameter d to vary with the time of year is hard to fit, and fitting d on data for only one season is also difficult.

CJB suggested that the long memory of surface air temperatures may simply result from the aggregation of several processes with different time-scales, such as internal atmospheric variability on short time-scales, land-surface processes on medium time-scales, and the interaction of the atmosphere and ocean on long time-scales. Indeed, a simple statistical model consisting of a sum of 3 AR(1) processes was shown to give results that were indistinguishable from long memory over the length of data and number of lags used. This $3 \times$ AR(1) model is not, however, practical for simulating artificial temperatures because the parameters cannot be easily estimated. It does, nevertheless, motivate the search for other simple models that might perform as well as ARFIMA and yet avoid the difficulties introduced by the long memory parameter d.

It was shown in CJB, and subsequently in more detail in Brix et al. (2002) that the well-known ARMA

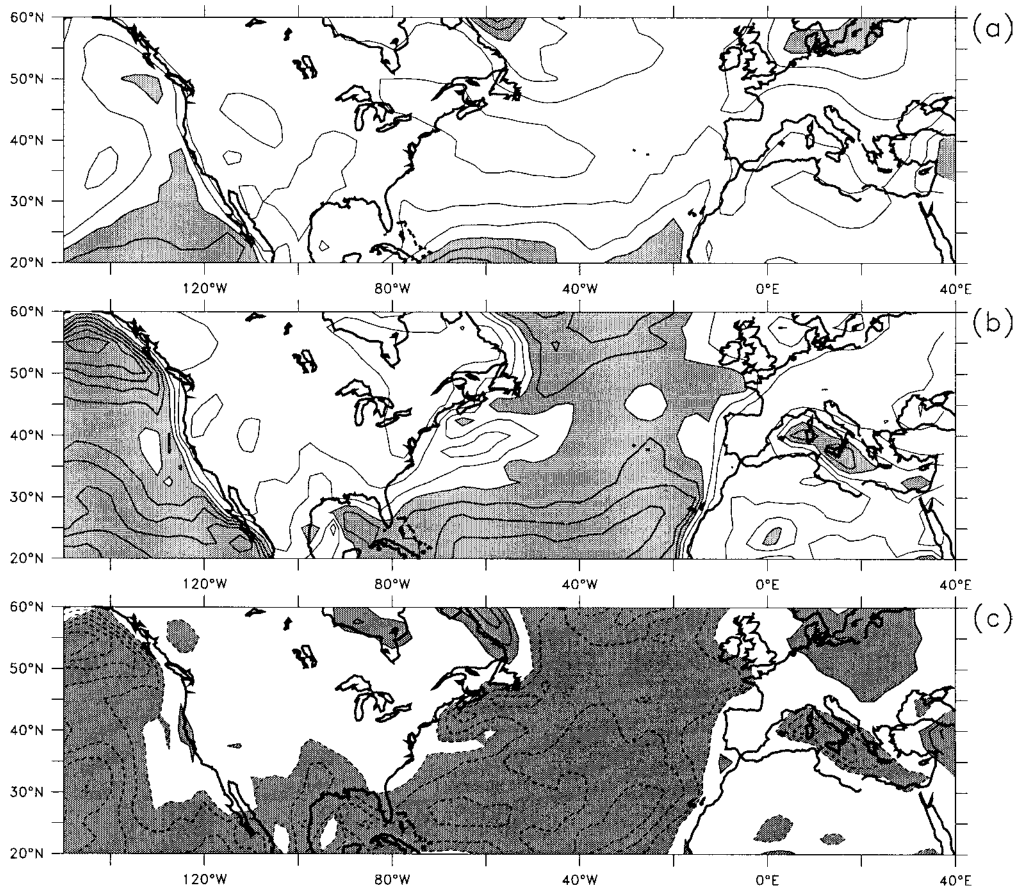

Figure 2. *Persistence of surface air temperature anomalies as quantified by the index s (defined as an average of the ACF at each point over the first 40 lags; see text) in (a) winter (DJF) and (b) summer (JJA). Contour interval is 0.05. Light shading shows values* > 0.1, *dark shading values* > 0.2. *(c) Difference (DJF–JJA). Contour interval is 0.05. Dashed lines show negative values. Shading shows absolute values* > 0.05. *Data from NCEP Reanalysis, 1950–1999.*

processes (Box & Jenkins 1970) do not work well for surface temperature anomalies because they do not capture the observed slow decay of the ACF. For an AR process to capture this shape of ACF out to 40 lags, 40 parameters are needed. Such a model is extremely non-parsimonious, and the parameters cannot be estimated with any accuracy. Indeed, it would be impossible to distinguish most of the parameters from zero. This is unfortunate, because it would be relatively straightforward to generalise the AR process to have seasonally varying parameters, and hence solve the problem of modelling seasonality.

We now present a new statistical model that maintains the simplicity of the AR processes, but is as accurate and parsimonious as the more complex ARFIMA model. Initially, we present a non-seasonal version of the model, but extend it to include seasonality in Section 4.1. The model is patterned after AR(p):

$$x_n = \alpha_1 x_{n-1} + \alpha_2 x_{n-2} + \cdots + \alpha_p x_{n-p} + \xi_n, \qquad (2)$$

where x_n is the value of the process at step n, ξ_n is a Gaussian white-noise process, and α_i are constants, but rather than using individual temperatures for days $n-1, n-2, \ldots$ as predictors, we use *moving averages*

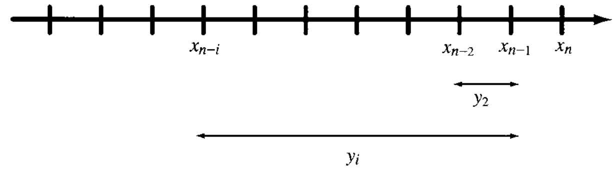

Figure 3. *Averaging intervals used in the AROMA model.*

of past temperatures:

$$x_n = \alpha_1 y_{m_1} + \alpha_2 y_{m_2} + \cdots + \alpha_r y_{m_r} + \xi_n, \qquad (3)$$

where

$$y_m = \frac{1}{m} \sum_{i=1}^{m} x_{n-i} \qquad (4)$$

Note that all the averages start from day $n-1$. This is sketched graphically in Figure 3. Since this model consists of autoregressive terms *on* moving averages, we give it the name AROMA. An AROMA($m_1, m_2, \ldots, m_r$) process can be rewritten as an AR(m_r) process, but it can accurately capture the observed temperature autocorrelation structure with much fewer than m_r independent parameters (see below).

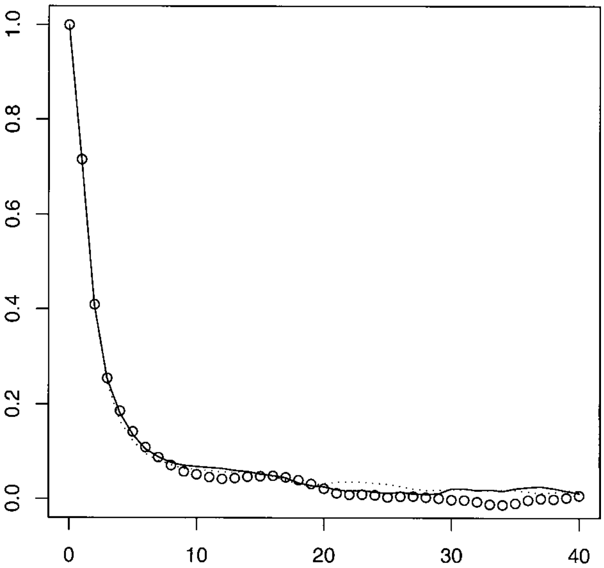

Figure 4. *The observed (solid line) and modelled ACFs for Chicago. The modelled ACFs were produced using an ARFIMA model (dotted lines) and the AROMA model (circles).*

How are we to choose the number and length of moving averages to use in the model, and calculate the coefficients α_i? As for the number of moving averages, this should be chosen to be as small as possible, so that the parameters can be well estimated. Experiments on temperature anomalies for our 8 stations suggest that four moving averages (i.e. $r = 4$) are typically enough to capture the observed ACF well out to 40 lags: this is a great improvement on the number of parameters required by the AR model for the same accuracy. Given values for m_1, m_2, m_3 and m_4, it is straightforward to calculate the weights $\alpha_1, \ldots, \alpha_4$ using linear regression. All that remains is to decide which moving averages to use. Experiments were performed on our eight stations in which all combinations of values of m_1, m_2, m_3 and m_4 up to 35 were tested. The results were ranked according to the root mean square error between the model ACF and the observed ACF (an alternative method would be to rank the results using the likelihood of each model). Results are shown in Table 1. Interestingly, all locations were modelled optimally using $m_1 = 1$ and $m_2 = 2$. Values of m_3 and m_4, were, however, different for different stations. This suggests that a simple way to choose the lengths of the moving averages to be used is to fix $m_1 = 1$ and $m_2 = 2$, and optimise the ACF by varying the other two lags. This is a simple optimisation problem and can be solved in a matter of seconds on a personal computer.

Figure 4 shows the observed and modelled annual ACF for Chicago using the AROMA and ARFIMA models. We see that both the AROMA and the ARFIMA models give a very good fit to the observed ACF. The advantage of the AROMA model is that it can be fitted much more simply (as we have seen above) and (as we shall see below) can be extended very readily to include seasonality. From the standpoint of meteorological interpretation, AROMA is more attractive than the ARFIMA model, appealing as it does to the idea of different timescales in weather variability, corresponding to each of the moving average terms. If temperature today is related to an average of temperature over the last two days, then this is probably just a reflection of the impact of small scale weather systems. If it is related to an average of temperatures over the last 20 days then this may be because of the memory in soil moisture, for instance.

4.1. Extending AROMA to include seasonality

We now extend the AROMA model to include seasonality. First, we fit the model to annual data, as described above. This fixes the lags m_1, m_2, m_3, m_4 that are to be used. We do not use different lags at different times of year. For each day of the year, we then fit a different model with different regression parameters α_i. This is done be selecting a window of data either side of that day, and performing the regression using data within that window. The regression parameter for a moving average of length m can only be fitted if the length of the fitting window is significantly larger than m, in order that the window contain a sufficient number of regression pairs that a sensible regression can be performed. In our case, we limit the lengths of the moving averages to 35 days, and set the window length to 91 days (an ad hoc choice designed to give enough data to allow accurate fitting of the model for each day while still allowing for seasonal variability). Thus even for the longest moving average of 35 days we have 66 pairs of values from each year of data with which to calculate the appropriate α_i. The data used for adjacent days of the year overlap almost entirely, bar one day at each end of the 91-day window, and so we would expect the regression coefficients to vary only slowly during the year.

Figure 5 shows observed and modelled ACFs for different seasons for Miami. In each panel of this figure we show the annual ACF for reference, which is the same in all four figures. The solid line in each case shows the observed ACF for that season. Because these seasonal ACFs are calculated using fewer data than for the annual ACF, they show more sampling variability and are hence less smooth. The circles show the seasonal ACFs from the seasonal AROMA (SAROMA) model fitted to the Miami daily temperatures as described above. We see that the model captures the seasonal variation in the ACF well. In particular it shows a slow decay of memory in summer, and more rapid decay in the other seasons.

Figure 6 shows the seasonal variation of the four regression parameters in the SAROMA model. We see that they vary slowly with the time of year. The slow decay of the ACF in summer corresponds to a summer

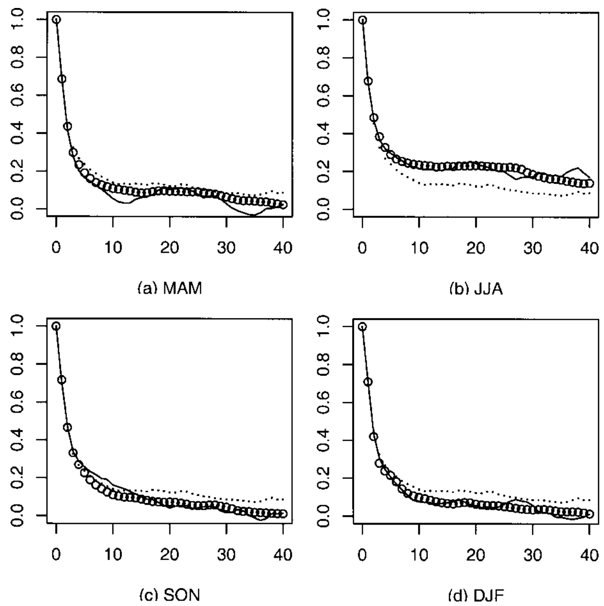

Figure 5. *The four panels show observed and modelled ACFs for Miami for the four seasons. The observed data are the same as in Figure 1. In each panel the dotted line is the annual ACF which is included for reference. The solid line is the observed ACF for that season, and the circles are the modelled (SAROMA) ACF for that season. Confidence limits are omitted for clarity.*

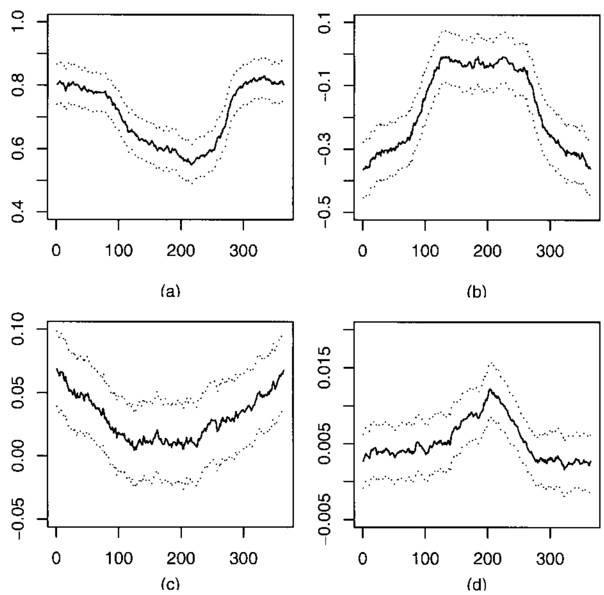

Figure 6. *Seasonal variation of the four regression parameters for the SAROMA model for Miami. The solid lines show the estimated parameter values, while the dotted lines show the 95% error bounds. We see that each of the parameters shows a strong seasonal cycle, corresponding to the strong seasonal cycle seen in the observed and modelled ACFs.*

peak in the fourth parameter which applies to a moving average of length 28 days. The SAROMA model could be extended by smoothing these regression parameters in time (either parametrically or non-parametrically) to preserve only the long time-scale variations and remove the short time-scale fluctuations which are presumably only due to sampling error.

4.2. A non-parametric method: extended burn analysis

We have discussed non-seasonality in the ACF, and shown that it is strong in some cases. We have also introduced a simple parametric model that captures this seasonality well. There are, however, potential limits to how well the model can work. In particular, if the timescales of change of the observed ACF are rapid, then one would want to use a short window to fit the model. A short window, however, prevents accurate estimation of the parameters because the regression has insufficient data. The model is thus limited to representing ACFs that change over timescales somewhat longer than the longest lag in the model. Another limitation is that the model uses the same lags at different times of year, while it could be that the timescales of memory in the physical system actually change with time of year. Finally, the model assumes that the (distribution adjusted) temperature anomalies are governed by linear dynamics, which may not be entirely correct. It is possible that one could make a different parametric model that would solve some of these problems in a satisfactory way. However, given the richness and complexity of climate variability, we believe it will always be the case that there will be some location somewhere for which the observed temperature variability will not fit a given parametric form. Because of this, it would seem useful to also explore non-parametric methods that make the fewest possible assumptions about the data, and hence are likely to be generally applicable.

We present here one such simple method that is essentially an extension of the burn analysis described in the Introduction. It relies on the simple idea that there may be some information from outside the contract period which is relevant to the contract period itself. Consider for instance a one-week contract: we may expect data from the weeks preceding and following the contract period to have statistics similar to that of the contract period itself. There is an implicit assumption here that the ACF and distribution of variability vary only fairly slowly (i.e. do not change much from week to week), and that the inaccuracy introduced by the small week-to-week changes in the ACF and distribution can be outweighed by the benefits of having more data to work with. Note, however, that this model can work with ACFs that change more rapidly than the SAROMA model can accommodate.

The method works as follows. We define a period that extends either side of the contract period and captures the data to be used. In the above example, we might employ a window of two weeks on either side of the contract period. This gives us five times as much historical data to work with than using only the contract period (as in standard burn analysis), which would be expected to increase the accuracy of our estimates by more than a factor of 2. We now slide a window of the same length as the contract period along the relevant data. For each window position, we add back the seasonal cycle in variance and mean appropriate for the contract period, and calculate a historical index value. The end result is many more historical index values than are obtained in the index-based methods. In our example we have a seven-day contract and a 35-day relevant data period which means that the seven-day window can take 29 different positions. Each year of data thus gives us 29 historical index values. This is 29 times as many as if we only used the contract period itself (although these 29 values are not independent). The advantage of sliding the window rather than jumping it (even though the underlying data used are the same) is that (a) it creates a smoother estimate of the final distribution, and (b) it uses more possible combinations of days of daily weather, which can result in more extreme values for the index.

The only remaining issue is to decide the length of the relevant data period. Too long a period will start including data that are not relevant because the statistics of weather variability have changed. Too short a period will not reap the potential benefits of using more data. The optimum window length is clearly dependent on the station: with Chicago we might be tempted to use data for the whole year, while with Miami that would clearly be wrong since the summer data are markedly different from those of the other seasons. One way to choose the window length would be to analyse seasonal variations in higher moments of temperature variability and seasonal variations in autocorrelations at, or averaged over, certain lags.

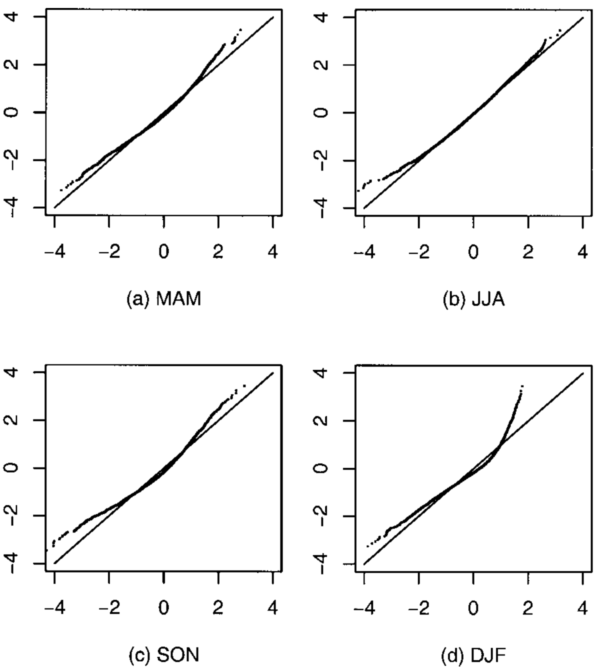

Figure 7. *The four panels show QQ plots for temperature anomalies in Miami for the four seasons. The horizontal axis shows the observed quantiles, while the vertical axis shows the modelled quantiles. We see that in all seasons the cold tail of the distribution is heavy tailed (cold events are more likely than predicted by the normal distribution) while the warm tail of the distribution is light tailed (warm events are less likely than predicted by the normal distribution). The most significant departure from normal is the warm tail in winter.*

5. Seasonal non-normality of surface air temperatures

By construction, our temperature anomalies are stationary in the mean and the variance. It is still possible, however, that they are non-stationary in higher moments of variability. Figure 7 shows the cumulative distribution functions (CDFs) of observed temperature anomalies for the four seasons for Miami. We have plotted these distributions against a fitted normal distribution in the form of a QQ plot. If the observed data are normally distributed, they will lie along the diagonal. If they are skewed they will tend to lie at an angle to, and cross, the diagonal. If there are heavy tails at the warm end of the distribution, the data will lie below the diagonal and if there are heavy tails at the cold end of the distribution, the data will lie above the diagonal. We see that all seasons show deviations from a normal distribution with light warm tails and heavy cold tails. The largest deviations are the light warm tails in summer, showing that extreme warm events are much less likely than would be supposed from a normal distribution.

The levels of non-normality seen in many locations are sufficiently large that making the mathematically convenient assumption of normality will degrade the accuracy of the temperature simulations significantly, and lead to mis-pricing of weather derivatives. The errors will be particularly large for contracts which depend heavily on extreme values of temperature, but will also be important for standard contracts. It is thus sensible to attempt to model this non-normality directly.

There are a number of possible methods that can be used to model non-normality in temperature variability. An initial observation we make is that although temperatures are clearly not normally distributed at many locations, the distribution is, at the same time, still reasonably close to normal. This suggests that if we approach the problem using transformations that convert the data to a normal distribution, then the dynamics will not be affected too strongly, and the same time series models as are used in the normal case may still work.

The first choice is whether to model the non-normality before or after attempting to model the autocorrelation. The first of these methods would involve applying a transformation to the temperature anomalies that renders them more or less normal. Time series modelling methods that assume normality, such as ARFIMA or SAROMA, can then be applied to the transformed anomalies. Simulated values are passed back through the transformation to get back to the original distribution.

A second method would be to apply a time series model to the temperature anomalies directly, and model non-normality in the residuals. We choose to apply the first of these methods because it allows us to re-use algorithms such as the ARFIMA or SAROMA models simply by applying a transformation to the inputs and outputs of those models.

The second choice is whether to use parametric or non-parametric models for the transformation of the anomalies. Box–Cox transformations (Box & Cox 1964) are a commonly used parametric distribution transform, and can be extended to vary seasonally. They are not, however, completely general. Non-parametric methods, on the other hand, can cope with any temperature anomaly distribution. For this reason we choose a non-parametric method, and the method we present, is, as far as we are aware, new.

The method works as follows. We derive a separate estimate of the cumulative distribution of the temperature anomalies for each day of the year (this step is discussed in detail below). These cumulative distributions are used to convert the temperature anomaly for each day (over all years) into a probability, and that probability is then converted, using the inverse of the standard normal CDF, into a value sampled from a normal distribution.

It remains to specify how to estimate the distributions used for each day. They could be estimated using only data from that day of the year: however, this would give a poor estimate because using 50 years of data would only give 50 points on the distribution. Instead we assume that the anomaly distribution only varies slowly with time and that we can estimate this distribution more accurately by using additional data from surrounding days. We do this by taking temperature values from a window surrounding the actual day, with a window length of 91 days (arbitrarily chosen to be long enough to give a smooth distribution, but short enough not to smooth out seasonal variations). Thus for each day of the year, we form the estimate of the anomaly distribution for that day using 91 days per year $\times$ 50 years = 4550 days of data. This gives a smooth estimate. Since the distributions for adjacent days are based on almost the same data because of substantial overlapping of the windows, the transform only changes gradually with the time of year, which seems realistic.

The results of the application of this transform to Miami temperatures are shown in Figure 8. We see that most of the non-normality has been removed. The transformed temperatures can now be modelled using a Gaussian process, and simulations of the transformed temperatures can be converted back to the appropriate distribution using the inverse of the distribution transform. One caveat for this method is that as long as we use an empirical distribution as described above, it would not be possible for final simulated temperature anomalies to exceed historical values. This is unlikely to be a problem in most cases. However, if we have a particular interest in extreme events then it would be advisable to extend the distribution used in the transformation using extreme value theory (Reiss & Thomas 1997).

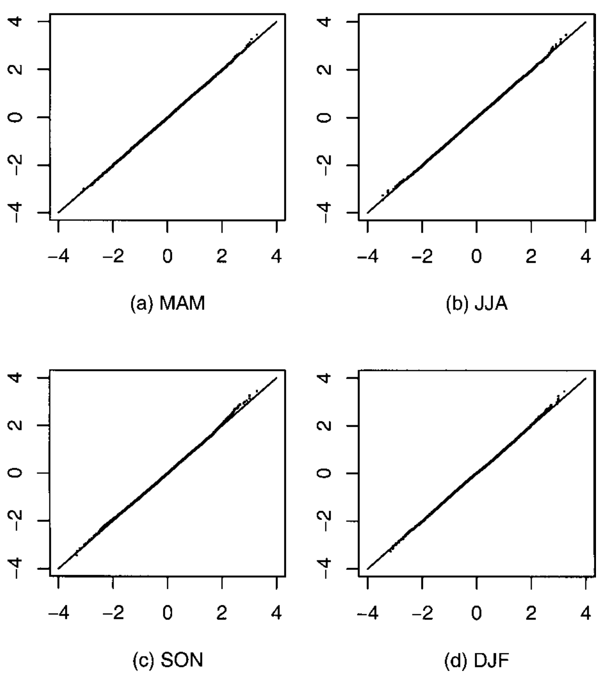

Figure 8. *The four panels show QQ plots for temperature anomalies in Miami for the four seasons, after having been transformed using the non-parametric seasonally varying transform described in the text. We see that most of the non-normality has been removed.*

6. Examples

We now present some examples to illustrate the effects of improved modelling of the seasonality in the distribution (Section 5 above) and the ACF (Section 4 above) on the calculation of weather derivative prices. Our examples are all based on Miami, since that location has strong levels of seasonality in both the distribution and the ACF and so is likely to show the greatest benefits of using the more advanced modelling methods.

The type of weather derivative we consider for our examples are ‘unlimited call options’. These require one party (the seller of the option) to pay another (the buyer) a certain amount of money if the final weather index is above a specified level known as the *strike*. The amount paid (the payoff) is proportional to the difference between the index and the strike, with a constant of proportionality known as the *tick*. This effectively insures the buyer against high values of the index. The question is: given the strike and tick, what should the premium (i.e. the price paid by the buyer) be? The simplest answer is that the premium should equal the expected payoff of the contract. Then, in the

Table 2. *The contract structures for the examples.*

	Summer contract	Winter contract
Strike	1750 CDDs	90 HDDs
Tick	\$5000/CDD	\$5000/HDD

long run, neither the buyer nor the seller will make profit. In practice, the seller may add a *risk loading* on top of the premium to compensate for the risk taken in underwriting the derivative. For the present discussion we will ignore this issue and focus on the estimation of the expected payoff. Given the payoff structure and the index distribution, we can easily estimate the expected payoff using numerical integration. We will do this for both ARFIMA and SAROMA models, both with and without the distribution transform (giving four cases in all). We will consider two examples: a winter contract (December to February) and a summer contract (June to August).

The details of these contracts are shown in Table 2. The winter contract is based on heating degree days (HDDs), which are defined as the sum of the excursions *below* 18 °C (65°F) during the contract period, while the summer contract is based on cooling degree days (CDDs) which are defined as the sum of the excursions *above* 18 °C during the contract period (these definitions are standard in the energy industry). HDDs are a measure of how cold the season is, and relate to use of heating; CDDs are a measure of how warm the season is and relate to the use of cooling.

Table 3 shows the results for the summer contract. The index means for the different models are virtually identical in all cases. This is because summer in Miami is very warm and the temperature rarely drops below 18 °C. As a result, the mean number of CDDs is fixed by the seasonal cycle (see Eq. 10 in CJB) which is modelled in the same way for all the models presented. The small differences are due to sampling error, caused by the use of a finite number of years of simulation. The differences between the models appear in the index standard deviations. The ARFIMA model gives lower standard deviations that the SAROMA model, for both normal and transformed distributions. This is explained by the higher autocorrelation in summer (see Figure 5), which is captured by SAROMA but not by ARFIMA. Higher autocorrelations directly lead to a higher index standard deviation (see Eq. 11 in CJB). As a result, SAROMA prices are over 25% higher than ARFIMA prices. A seller using ARFIMA pricing would, on average, lose over \$20,000 on this contract.

In the summer example above, it makes little difference whether one uses a normal or transformed distribution in the models: clearly, the distribution is always close to normal in this case. Things are quite different in the winter (Table 4). The mean index is now much higher for the models with the distribution transform than not. This can be explained as follows. Miami in winter is quite warm, and temperatures below 18 °C are uncommon. It is only the cold tail of the temperature distribution that creates HDDs, and hence modelling of this cold tail is crucial. Without transforming the temperature distribution, the cold tail is modelled as being too thin, and fewer HDDs occur in the model than in reality. This causes the normally distributed models to underestimate the mean number of HDDs. The transformed-distribution models are more accurate since they take into account the fatter cold tail in winter (Figure 7). This has a dramatic impact on the expected payoff: prices given by the transformed-distribution models are almost three times higher than those of the normally-distributed models. In this case, it hardly matters whether one used ARFIMA or SAROMA, since most of the price difference is due to the change in the index mean rather than its standard deviation.

Table 3. *Results for the summer contract example. Expected payoff values have been rounded to three significant figures.*

Model	Distribution	Index mean (CDDs)	Index std. dev. (CDDs)	Expected payoff (\$)
ARFIMA	Normal	1727.1	62.1	75,900
ARFIMA	Transformed	1727.5	62.5	76,400
SAROMA	Normal	1727.0	73.5	96,200
SAROMA	Transformed	1727.4	73.6	96,900

Table 4. *Results for the winter contract example. Expected payoff values have been rounded to three significant figures.*

Model	Distribution	Index mean (HDDs)	Index std. dev. (HDDs)	Expected payoff (\$)
ARFIMA	Normal	56.1	45.1	15,000
ARFIMA	Transformed	85.2	61.0	43,200
SAROMA	Normal	57.1	45.8	16,100
SAROMA	Transformed	87.1	60.7	44,300

Taken together, these examples emphasise the importance of modelling both the distribution and the ACF of temperature correctly in the pricing of weather options.

7. Summary

The advent of weather derivatives has created significant interest in the understanding and statistical modelling of surface air temperature variability. Weather derivative pricing methods based on modelling of daily temperatures have certain potential advantages over simpler methods. Such modelling is not, however, easy because of the richness and complexity of climate variability, and in particular, because of long memory and seasonality. The CJB and BSZ models were the first to capture the observed slow decay of memory of surface air temperature variability. They are, however, limited by assumptions of normality and stationarity and, as we have shown, many locations do not conform to these restrictions.

We have presented a relatively simple framework in which the non-normality and seasonality of temperature variability can be accommodated, as long as it is reasonably slow in varying: changes in probability distribution and ACF from season to season can be captured, but much more rapid changes cannot. The model for seasonal variation in the ACF that we present can also be interpreted more simply than the long memory models. The latter capture the slow decay of the ACF using a slightly mysterious long memory parameter d. Our model, however, presents temperature today as the sum of components of temperature variability on different timescales, some short (presumably due to small scale weather variability), and some longer (presumably due to either atmospheric long waves, or soil or ocean processes). Finally, we present a non-parametric model that can be applied to the pricing of weather derivatives in those cases where the parametric methods presented still do not appear to give a good fit to the observed variability.

The models we have presented should lead to the more accurate pricing of weather derivatives, especially for contracts based on locations that show strong seasonality. Furthermore, since they deal with the difficult problem of modelling non-normal time series with slowly decaying autocorrelations and seasonally varying dynamics in novel ways, they may find applications in other areas of geophysical and environmental modelling.

Acknowledgments

The authors would like to thank Jeremy Penzer, Anders Brix and Christine Ziehmann for useful discussions on the subject of the statistical modelling of daily temperatures. Rodrigo Caballero was supported by Danmarks Grundforskningsfond.

References

Alaton, P., Djehiche, B. & Stillberger, D. (2002) On modelling and pricing weather derivatives. *Appl. Math. Fin.* **9**(1): 1–20.

Boissonnade, A., Heitkemper, L. & Whitehead, D. (2002) Weather data: cleaning and enhancement. *Climate Risk and the Weather Market*, Risk Books, pp. 73–98.

Box, G. & Cox, D. (1964) An analysis of transformations. *J. Roy. Stat. Soc. B* **26**: 211–243.

Box, G. E. P. & Jenkins, G. M. (1970) *Time Series Analysis, Forecasting and Control* Holden-Day.

Brix, A., Jewson, S. & Ziehmann, C. (2002) Weather derivative modelling and valuation: a statistical perspective. *Climate Risk and the Weather Market*, Risk Books, pp. 127–150.

Brody, D., Syroka, J. & Zervos, M. (2002) Dynamical pricing of weather derivatives. *Quant. Finance* **2**: 189–198.

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

Cao, M. & Wei, J. (2000) Pricing the weather. *Risk* **13**(5): 14–22.

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

Diebold, F. & Campbell, S. (2001) Weather forecasting for weather derivatives. University of Pennsylvania Institute for Economic Research, Tech. Rep. 8.

Dischel, R. (1998) Black-scholes won't do. *Energy Power and Risk Management Weather Risk Special Report*.

Dornier, F. & Querel, M. (2000) Caution to the wind. *Energy Power and Risk Management Weather Risk Special Report*, pp. 30–32.

Granger, C. W. J. & Joyeux, R. (1980) An introduction to long memory time series models and fractional differencing. *J. Time Ser. Anal.* **1**: 15–29.

Kalnay, E. et al. (1996) The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**: 437–471.

Moran, P. (1947) Some theorems on time series. *Biometrika* **34**: 281–291.

Reiss, R. & Thomas, M. (1997) *Statistical Analysis of Extreme Values*. Birkhauser.

Torro, H., Meneu, V. & Valor, E. (2001) Single factor stochastic models with seasonality applied to underlying weather derivatives variables. European Financial Management Association, Tech. Rep. 60.