ENVIRONMETRICS
Environmetrics 2008; **19**: 818–835
Published online 25 January 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.905

Generation of synthetic sequences of half-hourly temperature

L. Magnano[1*,†], J. W. Boland[1] and R. J. Hyndman[2]

[1] *Centre for Industrial and Applied Mathematics, University of South Australia, SA 5095, Australia*
[2] *Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia*

SUMMARY

We present tools to generate synthetic sequences of half-hourly temperatures with similar statistical characteristics to observed historical data. Temperatures are generated using a combination of daily and half-hourly temperature models which account for intra-day and intra-year seasonality, as well as short- and long-term serial correlations. Details of the model estimation are given as well as a description of the synthetic generation.

KEY WORDS: temperature data; time series; Fourier series; ARMA models; seasonal block-bootstrap; synthetic generation

1. INTRODUCTION

Artificial temperature data are useful whenever it is necessary to simulate the effect of temperature on a system. For example, such data are useful in studying electricity demand to assess the effect of different possible temperature patterns on demand or to construct a probability distribution for demand (McSharry *et al*., 2005; Magnano and Boland, 2007). In this paper, we present powerful models that can be used as tools for generating synthetic sequences of half-hourly temperatures.

The initial aim of this work was to build a model to generate synthetic sequences of half-hourly temperatures for summer months—December, January and February. However, due to the structure of the temperature dynamics, we required an additional model that generates daily mean temperatures. This allows the modelling of long-term serial correlations. For example, in our observed summer temperature data for Adelaide, Australia, the temperature typically increases over consecutive days leading up to a cool change. Therefore, there are successive days in which the temperature is above the general average until it drops below the average and then increases steadily again. We call this behaviour *thermal inertia*. This thermal inertia is stochastic since the number of days that the temperature remains above or below the general average is not fixed.

*Correspondence to: L. Magnano, Centre for Industrial and Applied Mathematics, University of South Australia, SA 5095, Australia.
†E-mail: luciana.magnano@unisa.edu.au

Received 7 December 2006
Accepted 27 November 2007

To capture this temperature behaviour, we propose a modification to the conventional methods used to generate temperature data. Boland (1995) and Hansen and Driscoll (1977) used Fourier series to model annual and daily seasonality, and modelled the different variability and skewness of the data, before modelling the residuals. In contrast, we propose eliminating the daily average for each day at the beginning of the estimation procedure before modelling the daily seasonality and analysing the residuals. For this reason, a model that generates daily mean temperature was also developed and is presented in this paper.

This new approach considerably enhances the distribution of the final outcome. The autocorrelation function (ACF) of the generated data resembles more closely the ACF of the actual data. The successful merging of the effects of persistence at the two time scales, half-hourly and daily, is a significant advance. The principles described here can be transferred to other situations where there is the mixing of effects on various time scales. Differences in the distributions of half-hourly temperature among the months were identified in the estimation process. Thus, they were modelled separately using similar procedures. Descriptive statistics, graphs and statistical tests were performed to compare the generated and the real data. We found that the new model performs very well in generating both daily and half-hourly temperature.

This paper is structured as follows: Section 2 describes the model and the data used in the estimation. Section 3 describes the daily mean temperature analysis while Section 4 presents the half-hourly temperature analysis. These sections also include model estimation, synthetic generation and validation of the model. Section 5 discusses an application of the model to the analysis of electricity demand. Finally, Section 6 presents a summary of our main findings.

2. OVERVIEW OF THE MODELS AND DATA STRUCTURE

We construct daily and half-hourly temperature generation models representing different components of temperature dynamics and describing the statistical characteristics of temperature data. The daily and half-hourly models were developed using temperature data from Adelaide, Australia, from July 1996 to June 2005 (Magnano, 2007). The data from leap days were omitted from the analysis to avoid difficulties with variable seasonal periods.

Exploratory analysis showed discrepancies in the distribution of the half-hourly temperatures between the three months. These discrepancies are more evident between December and the other two months as is shown in Figure 1. Although the difference between January and February is relatively small, the slight difference identified led us to perform the analysis of each month separately.

Let x_t denotes the temperature observed in period t where t is measured in half-hour intervals. We also construct the time series of daily mean temperatures which we denote by $\{\bar{x}_1, \bar{x}_2, \ldots\}$ where each $\bar{x}_j$ is calculated by taking the mean of the 48 half-hourly observations in day j.

Our daily mean model is given by

$$\bar{x}_j = \mu(j) + \sigma(j)\bar{x}_j^*$$

where $\mu(j)$ describes the annual cycle for the mean temperature and $\sigma(j)$ describes the annual cycle for the standard deviation of temperature. These cycles are given by first-order Fourier series

DOI: 10.1002/env

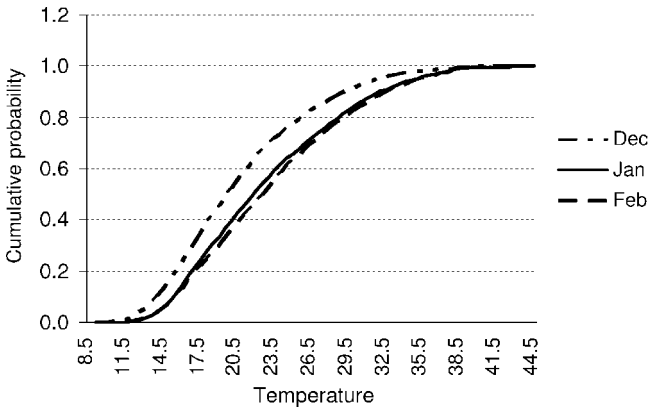

Figure 1. Cumulative distribution functions for each month

approximations:

$$\mu(j) = a + \delta_1 \cos\left(\frac{2\pi j}{365}\right) + \delta_2 \sin\left(\frac{2\pi j}{365}\right) \tag{1}$$

$$\sigma(j) = b + \gamma_1 \sin\left(\frac{2\pi j}{365}\right) + \gamma_2 \cos\left(\frac{2\pi j}{365}\right) \tag{2}$$

The standardised daily mean temperatures $\{\bar{x}_j^*\}$ are assumed to follow a zero-mean ARMA process after suitable transformation. (We use a separate ARMA process for each month of observation.)

The daily half-hourly model is given by

$$x_t = \bar{x}_j + \zeta(t) + \varepsilon_t \tag{3}$$

where j is the day in which period t falls and $\zeta(t)$ describes the intra-day cycle. We also use a Fourier series approximation for $\zeta(t)$, but delay giving it until later. The residual, ε_t, is assumed to include some autocorrelation, and we simulate it using a seasonal block-bootstrap.

We now explain the modelling procedure in more detail, using the Adelaide temperature data to illustrate our approach.

3. DAILY MEAN TEMPERATURE MODEL

First we compute the daily mean temperatures by simply taking the mean of the 48 half-hourly observations for each day giving 3285 daily means for the 9 years of data. We also compute the standard deviation of the daily mean temperatures for each day of the year giving 365 standard deviations.

We estimate Equations (1) and (2) using ordinary least squares applied using the daily means and daily standard deviations, respectively. One year of daily temperatures and the estimated yearly cycle $\hat{\mu}(t)$ are shown in Figure 2. The lower panel of Figure 2 shows the fitted curve $\hat{\sigma}(t)$ along with the daily standard deviations.

DOI: 10.1002/env

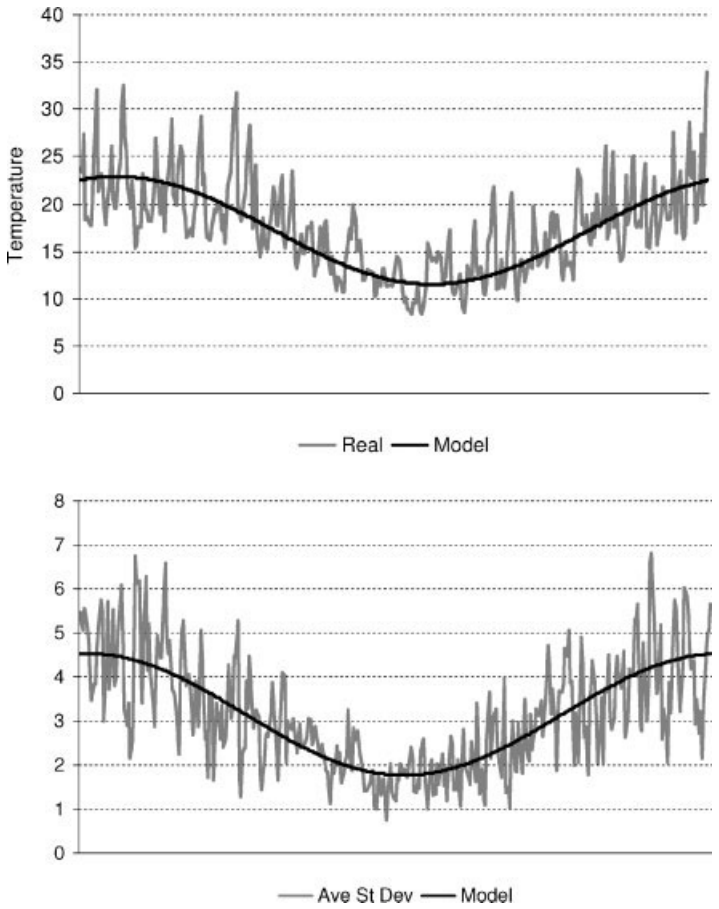

Figure 2. Top: Daily mean temperatures for January–December 2005, and the estimated annual cycle $\hat{\mu}(t)$. Bottom: Daily standard deviations, and the estimated annual cycle $\hat{\sigma}(t)$

Then we construct the standardised residual series, $\hat{x}_t^* = (x_t - \hat{\mu}(t))/\hat{\sigma}(t)$. Exploratory analysis showed that the distributions of $\hat{x}_t^*$ were different for each month. Thus the analysis of the standardised residuals was performed for each month.

For each month separately, we transform the standardised residuals to normality using the transformation $z_t = \Phi^{-1}(\hat{F}(x_t^*))$ where $\hat{F}$ is the empirical distribution function and Φ^{-1} is the inverse of the standard normal cumulative distribution function. Then z_t will have a standard normal distribution.

These residuals (again taking each month separately) were then modelled using ARMA(p, q) models:

$$\phi_p(B)z_t = \theta_q(B)a_t \tag{4}$$

where $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$ and $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)$, B is the *backward shift operator* which is defined by $Bx_t = x_{t-1}$ and a_t is Gaussian white noise with mean zero and variance σ_a^2.

DOI: 10.1002/env

Table 1. ARMA coefficients for daily temperature models

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
ϕ_1	1.37	0.50	0.41	0.46	0.55	0.49	0.56	0.37	0.73	0.75	0.51	0.72
ϕ_2	−0.82								−0.22	−0.25		−0.26
ϕ_3	−0.28											
θ_1	−0.50	0.48	0.43	0.45	0.23	0.30		0.46			0.43	0.24

After analysing the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the transformed data, different ARMA models were fitted to z_t. For each month, the selection of the best model was done based on Akaike's information criterion defined as

$$\mathrm{AIC}(m) = n \ln\left(\hat{\sigma}_a^2\right) + 2M \tag{5}$$

where n is the effective number of observations (which is equivalent to the number of residuals that can be calculated from the series), $\hat{\sigma}_a^2$ is the maximum likelihood estimate of σ_a^2, and $m = p + q$ is the total number of parameters in the model. The optimal model is that for which AIC(m) is minimum. The estimated coefficients for each ARMA model are shown in Table 1.

The synthetic generation of daily mean temperatures is formed by reversing the above modelling procedure.

$$\bar{x}_j = \hat{\mu}(j) + \hat{\sigma}(j)\hat{F}^{-1}(\Phi(z_t)) \tag{6}$$

where z_t are simulated from the ARMA process given by Equation (4). The errors for the ARMA model were randomly generated from a normal distribution.

3.1. Validation of the model

To validate the model, descriptive statistics and graphs were used to compare real and generated values. The descriptive statistics in Table 2 show that both the real and generated data have similar characteristics. This similarity is also shown in the empirical distribution functions of real and generated daily mean

Table 2. Descriptive statistics for real and generated daily mean temperatures by month

	Dec		Jan		Feb	
	Real	Gnr	Real	Gnr	Real	Gnr
Mean	20.9	20.5	23.2	23.4	23.2	23.5
Median	19.7	19.1	22.1	22.3	22.4	22.8
St.dev	4.5	4.3	4.8	5.0	4.7	4.8
Skewness	0.9	0.9	0.6	0.4	0.5	0.4
Kurtosis	0.2	0.1	−0.7	−0.9	−0.5	−0.7

DOI: 10.1002/env

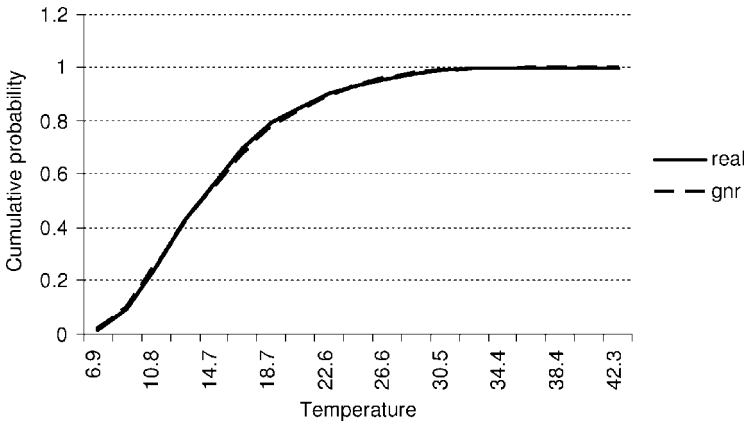

Figure 3. Cumulative distribution functions of real and generated daily mean temperatures

temperature presented in Figure 3. Finally, we compared the ACF of the real and generated daily mean temperature in Figure 4. Both autocorrelations follow a similar pattern.

Based on the descriptive statistics and on the graphs presented, we conclude that the daily model presented generates traces of daily mean temperature with statistical properties similar to the ones of the real data. These traces daily mean temperature generated with the procedure explained above were used to generate synthetic half-hourly sequences of temperature.

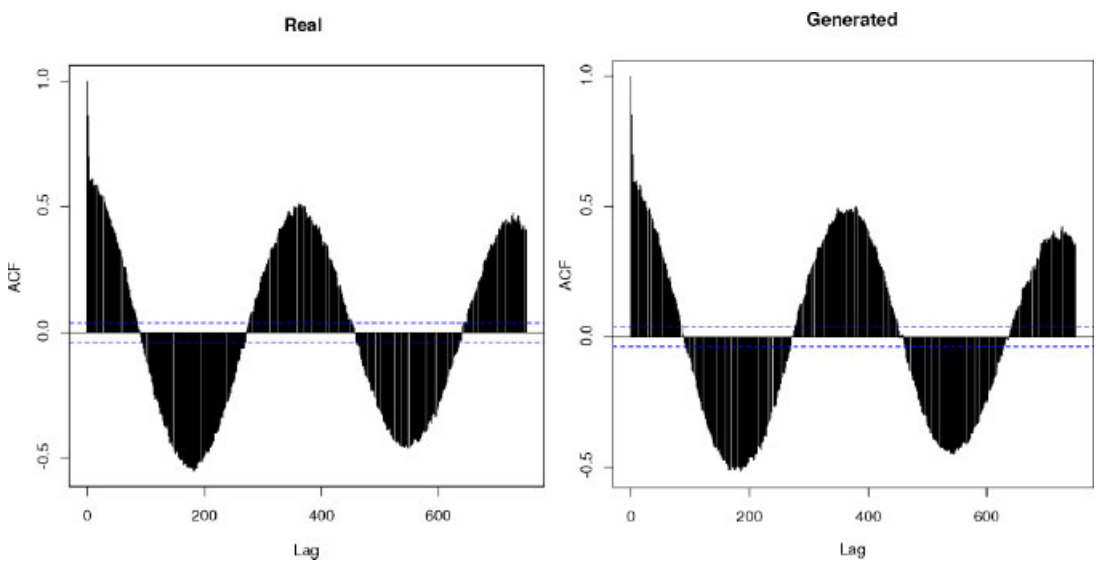

Figure 4. ACF of real and generated daily mean temperatures

DOI: 10.1002/env

4. HALF HOURLY TEMPERATURE MODELS

The estimation of the half-hourly temperature model was based on the same 9 years of half-hourly temperatures in Adelaide from 1996 to 2005. An exploratory analysis of the data showed that there is a relationship (Figure 5) between daily mean temperature and standard deviation within the day—in general, as daily mean temperature increases so does the variability.

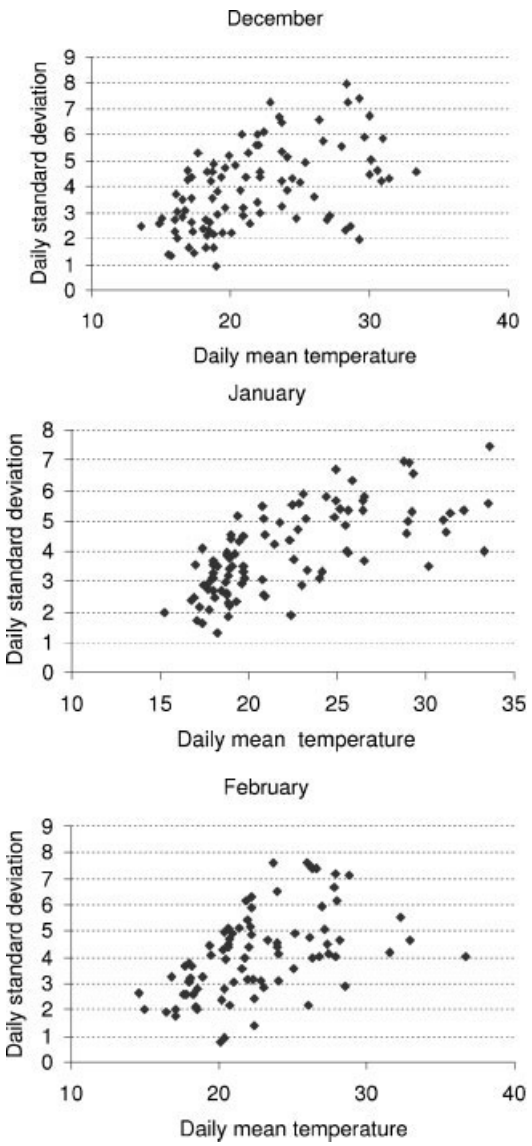

Figure 5. Daily standard deviations vs daily mean temperatures

DOI: 10.1002/env

This variation in variability leads to different intra-day profiles. In most cases, the differences between the maximum and minimum are larger on days with higher daily mean temperature. To describe this behaviour, the analysis was performed dividing the days into three groups: moderately cold days (MC) when daily mean temperature is below or equal to 20° C; mild days (MD) when the daily mean temperature is greater than 20° C and less than or equal to 25° C; and hot days (HD) when the daily mean temperature is greater than 25° C.

4.1. Half hourly results per month

The model estimation is performed in the following steps. First, the daily average is eliminated from the original observations.

$$\tilde{x}_t = x_t - \bar{x}_j \tag{7}$$

where j is the day in which period t falls and $\bar{x}_j$ is the average temperature on the jth day. These 'de-meaned' temperatures will still retain the seasonality within each day, but the annual variation in mean will be eliminated. Through this operation, the thermal inertia described earlier is removed and the remaining data fluctuates around zero.

We let $\tilde{x}_t = \zeta(t) + \varepsilon_t$, where the intra-day cycle is given by

$$\zeta(t) = \sum_{s=1}^{2} \left[\alpha_{s\ell i} \cos\left(\frac{2\pi ts}{48}\right) + \beta_{s\ell i} \sin\left(\frac{2\pi ts}{48}\right) \right] \tag{8}$$

Here, i denotes the month in which period t falls and ℓ denotes the group of days (ℓ = MC, MD, HD) in which period t falls. Two cycles per day were identified in each group of days in each month. The first cycle was highly significant while the contribution of the second was considerably smaller. The amplitude coefficients, $\alpha_{s\ell i}$ and $\beta_{s\ell i}$, are estimated from the data using ordinary least squares applied separately for each combination of the ℓth group and the ith month.

Figure 6 shows the estimates of the intra-day cycles for the three groups of days in December. The estimated parameters are shown in Table 3.

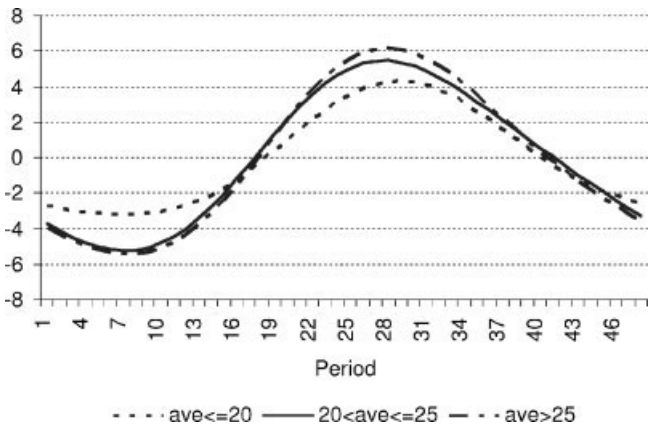

Figure 6. Intra-day cycles: December days

DOI: 10.1002/env

Table 3. Fourier series coefficients for intra-day cycles

Group		Dec		Jan		Feb	
		$s=1$	$s=2$	$s=1$	$s=2$	$s=1$	$s=2$
$\bar{x}_{hi} \leq 20$	α_s	−2.85	0.29	−2.91	0.36	−2.73	0.50
	β_s	−2.40	0.54	−2.50	0.65	−2.61	0.77
$20 < \bar{x}_{hi} \leq 25$	α_s	−3.99	0.68	−3.50	0.47	−3.25	0.64
	β_s	−3.33	−0.01	−3.39	0.82	−3.40	0.76
$\bar{x}_{hi} > 25$	α_s	−4.34	0.77	−4.25	0.83	−4.25	1.05
	β_s	−3.61	0.27	−3.96	0.36	−3.72	0.61

The residuals from these models are given by $\hat{\varepsilon}_t = \tilde{x}_t - \hat{\zeta}(t)$. A graphical analysis of $\hat{\varepsilon}_t$ showed that the data fluctuates around zero but still there was daily seasonality left that was not captured by $\hat{\zeta}(t)$. The analysis of $\hat{\varepsilon}_t$ was also performed by groups of days (MC, MD, HD). Figure 7 shows $\hat{\varepsilon}_t$ for December 2004 as an example.

The ACF and PACF for each group showed remaining seasonality and a slightly different seasonal pattern among the groups (Figure 8). To generate residuals that maintain this seasonal pattern, we used the seasonal block-bootstrap method (Politis, 2003). The block-bootstrap is based on the idea of drawing blocks of consecutive observations with replacement from a set of blocks. This is performed to capture the dependence structure of the data. The seasonal block-bootstrap uses blocks with length equal to a multiple of the seasonal period.

In most of the PACFs, the partial autocorrelation of order 3×48 is still significant (Figure 9). For this reason we used sets of nonoverlapping blocks of length 3×48.

4.2. Synthetic generation of half-hourly temperature

The synthetic generation of half-hourly temperatures is summarised below.

Step 1 Daily mean temperatures are generated by using Equation (6).

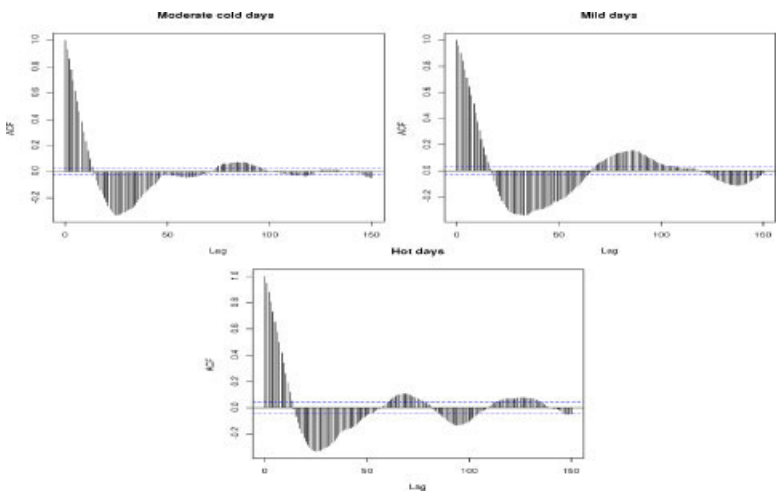

Figure 7. $\hat{\varepsilon}_t$ for December 2004

DOI: 10.1002/env

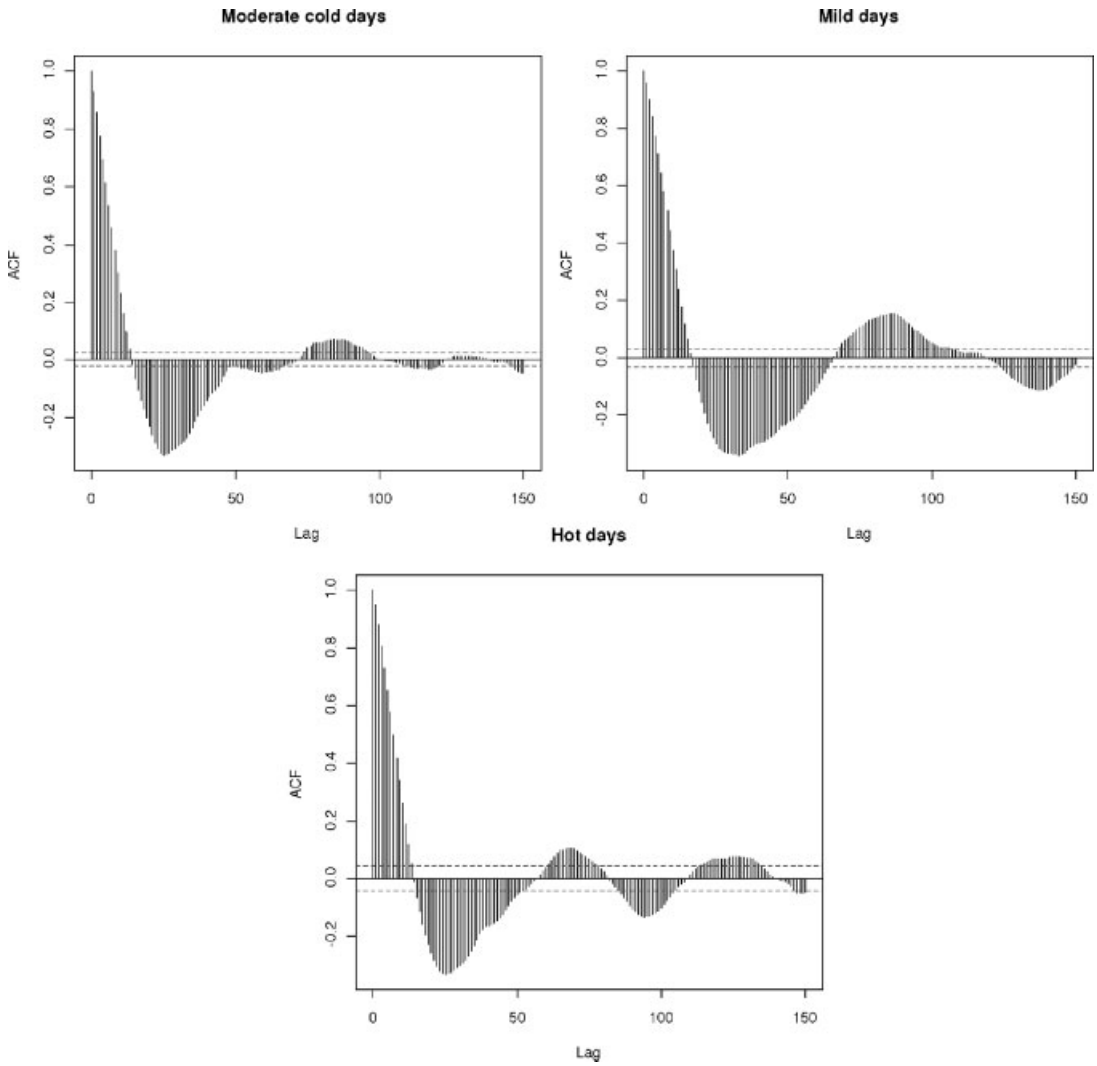

Figure 8. ACF of $\hat{\varepsilon}_t$ for December by groups

Step 2 The parameters estimated in Section 4.1 are used in the generation of the intra-day cycles, $\zeta(t)$.

Step 3 As was done in the estimation of the intra-day cycles, the generation of sequences of ε_t was performed by group of days and by months. Within each of the nine groups, the data was divided in blocks of 3×48 periods to capture the seasonal patterns that remained in the data. Then, the blocks within each group were shuffled. Figure 10 shows an example of actual and simulated ε_t values for hot days in January.

Step 4 The half-hourly observations were then obtained by

$$x_t = \bar{x}_j + \hat{\zeta}(t) + \varepsilon_t \tag{9}$$

DOI: 10.1002/env

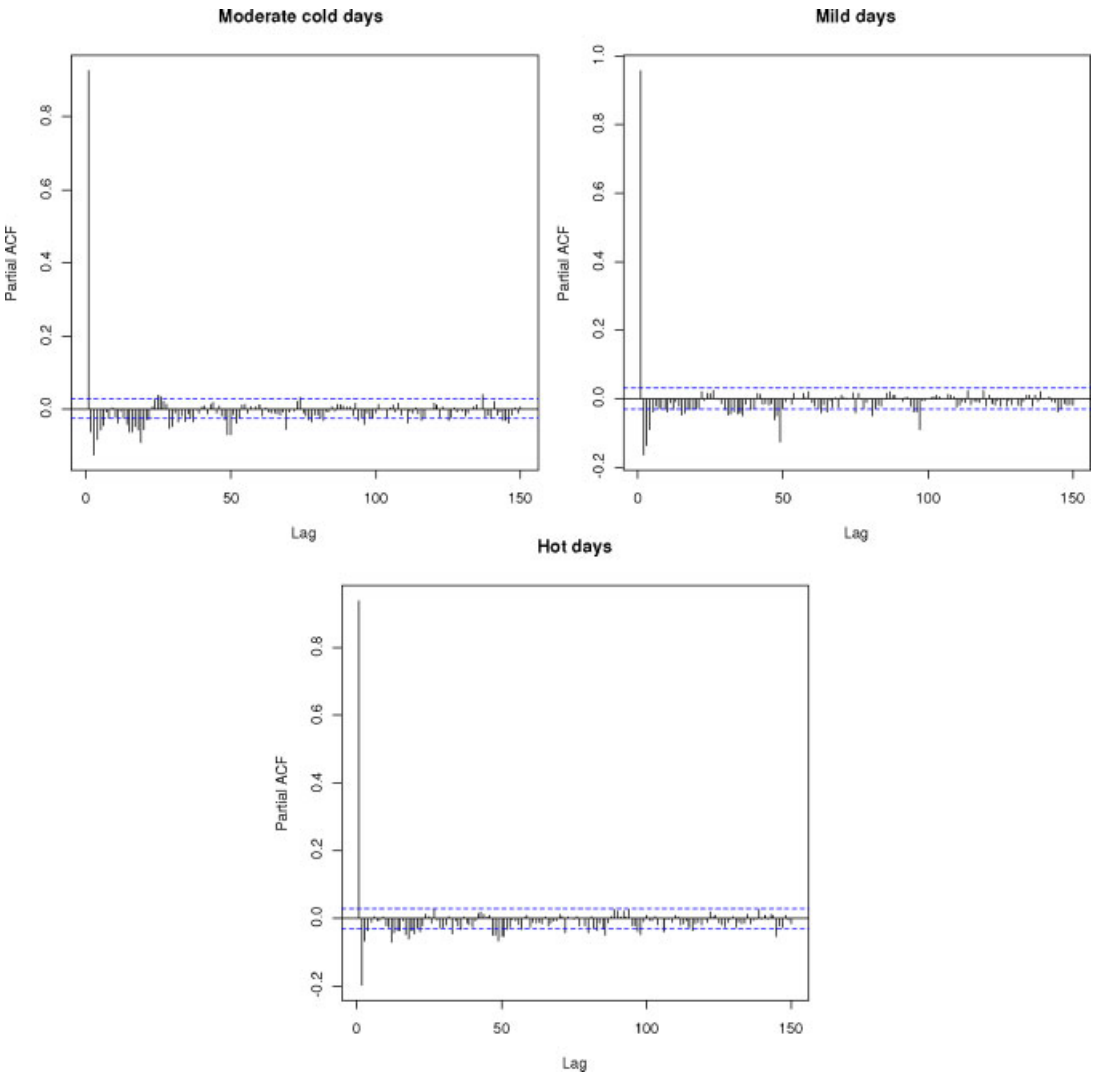

Figure 9. PACF of $\hat{\varepsilon}_t$ for January by groups

where $\{\varepsilon_t\}$ are generated by the seasonal block-bootstrap described in Step 3.

It is worth mentioning that although the ε_t were generated in blocks of three consecutive days, this periodicity is only preserved when the generated daily mean temperatures belong to the same group of days for three or more consecutive days. For instance, suppose the generated daily mean temperatures for six consecutive days are 18.5, 18.0, 19.5, 20.2, 22.0 and 25.0. Then the $\hat{\varepsilon}_t$ values for the first 3×48 periods belong to the MC group, the $\hat{\varepsilon}_t$ values for the next 2×48 periods belong to the MD group and the final 48 periods belong to the MH days. To preserve the seasonal pattern, it is necessary to divide the $\hat{\varepsilon}_t$ within each group into blocks of 3×48 periods. Then partial blocks may be sampled if there are fewer than three consecutive means belonging to the same group.

DOI: 10.1002/env

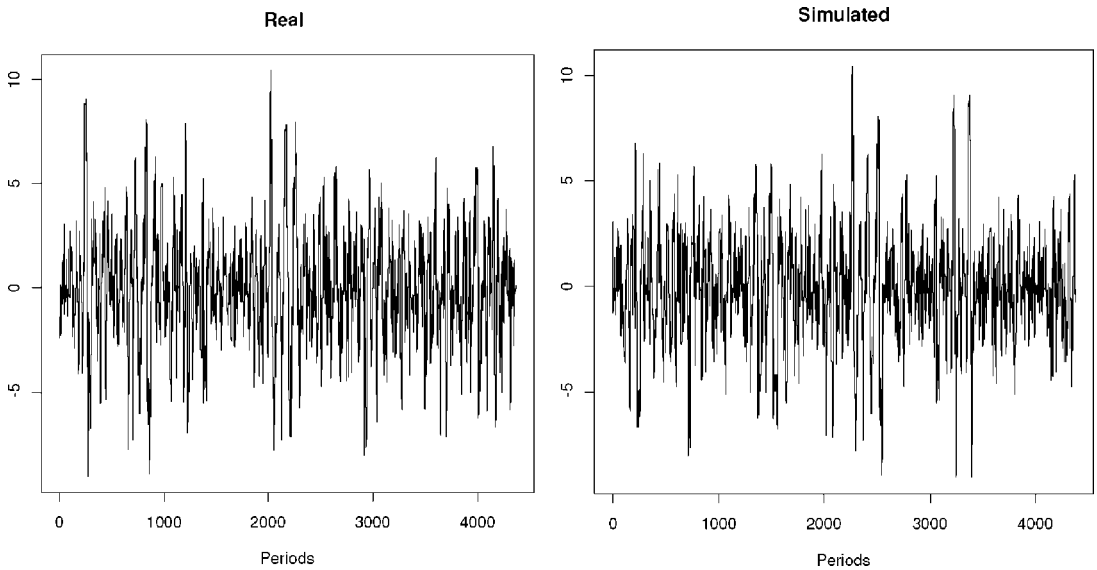

Figure 10. Real and simulated ε_{ijt}: hot days in January

Some censoring may be necessary in the generation process. For instance, when the generated temperature is below the historical minimum it should not be considered as a possible realisation.

Finally, given that there could be discontinuities introduced in the generation caused by the series of daily mean temperatures generated in Step 1, cubic splines can be used to smooth the final temperature in the periods that link one day with the next one.

Table 4. Descriptive statistics for half-hourly temperatures by month

	Dec		Jan		Feb	
	Real	Gnr	Real	Gnr	Real	Gnr
Mean	20.7	20.9	22.9	23.1	23.3	23.2
Median	19.7	20.0	21.9	22.2	22.6	22.6
St.dev	5.9	6.0	6.4	6.3	6.4	6.2
Skewness	0.7	0.7	0.6	0.6	0.5	0.5
Kurtosis	0.1	0.2	−0.4	−0.3	−0.5	−0.2
5%	13.1	12.8	14.3	14.4	14.4	14.4
10%	14.0	14.0	15.4	15.7	15.6	15.7
25%	16.1	16.4	17.6	18.2	18.2	18.5
75%	24.4	24.3	27.3	27.2	27.7	27.3
90%	29.2	29.4	32.2	32.2	32.6	32.0
95%	31.8	32.4	35.0	34.9	35.2	34.6

DOI: 10.1002/env

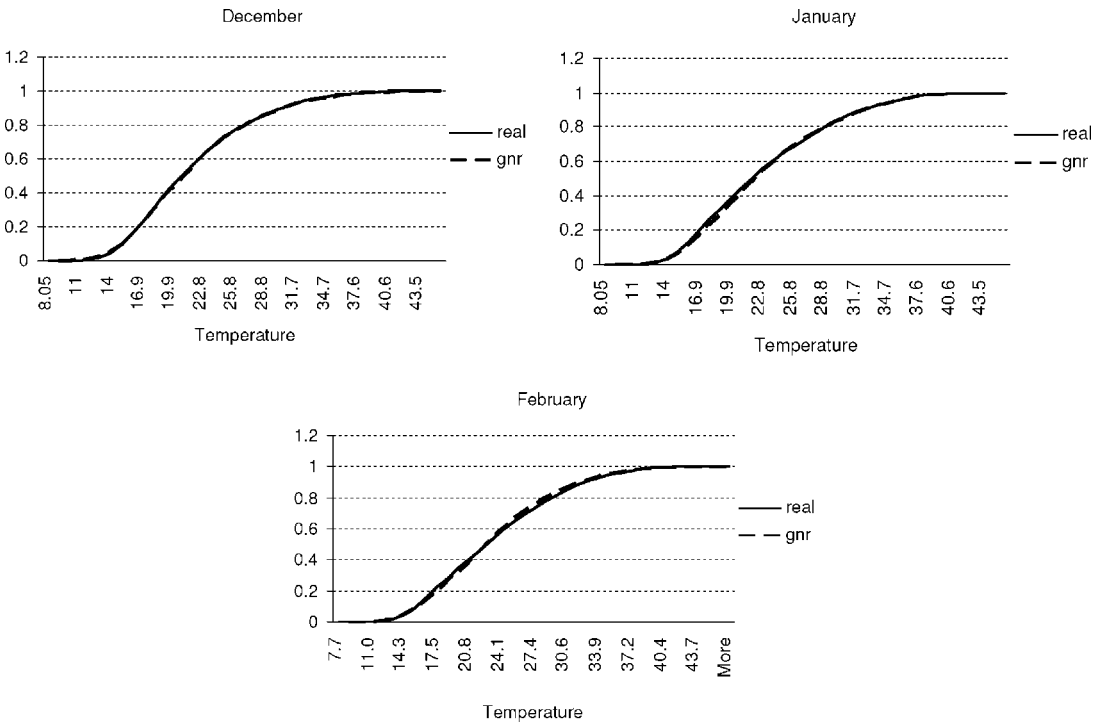

Figure 11. Cumulative distribution functions of real and generated half-hourly temperatures by month

4.3. Validation of the models

Descriptive statistics, graphs and statistical tests were used to compare real half-hourly temperature against 250 synthetic generated sequences. Table 4 shows that the means, medians, standard deviations, coefficients of skewness and kurtosis and percentiles for the real and simulated data are similar. This similarity is also shown in the cumulative distribution functions presented in Figure 11.

Figures 12 and 13 show the ACF and PACF of the real and generated half-hourly temperatures. The generated traces replicate the autocorrelations and partial autocorrelations of the real series quite closely. The most obvious difference is noted in the PACF at lag 96. Even though this partial autocorrelation is significant in both PACFs, they show opposite sign. This value measures the degree of association between x_t and x_{t-96} when the effects of the other time lags are removed. This association is negative for the real data and positive for the generated one. However, all the partial autocorrelations around lag 96 are negative for both sets of data.

We also compared the ACFs of the real and the generated data obtained using the approach presented in (Hansen and Driscoll, 1977). It models the yearly cycles, intra-day cycles and the different variability within the period with Fourier series, and makes correction for skewness. However, in contrast to the model presented in this paper, the approach in (Hansen and Driscoll, 1977) does not eliminate the daily mean temperature before analysing the daily cycles. When the daily mean temperature is eliminated, the thermal inertia is eliminated making the analysis of the intra-day cycles more accurate. From Figures 12 and 14 it can be seen that the model presented in this paper shows more similarity with the ACF of the real data than the model in (Hansen and Driscoll, 1977).

DOI: 10.1002/env

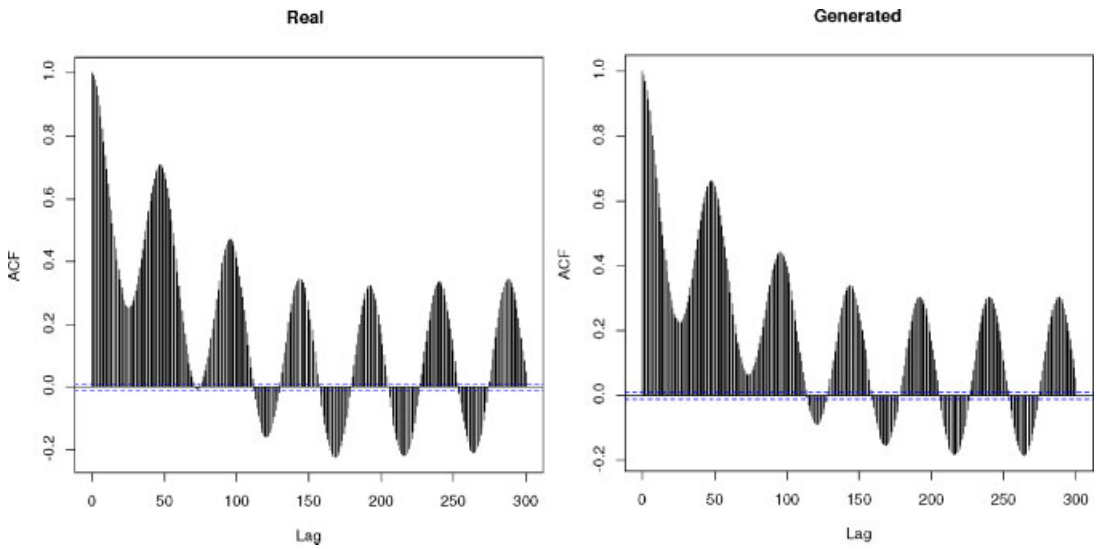

Figure 12. ACF of real and generated temperatures

Following the same criteria used by Hansen and Driscoll (1977), we compared the number of times that M consecutive hours occur below or above a specified threshold. For example, we calculated the number of times in a month that the temperature was above 34° C for $M = 3$, 6, 9 and 12 consecutive hours. These M periods were overlapping periods so there are $N - M + 1$ separate but overlapping periods of M consecutive hours, where N is the total number of hours of observed data. We compared

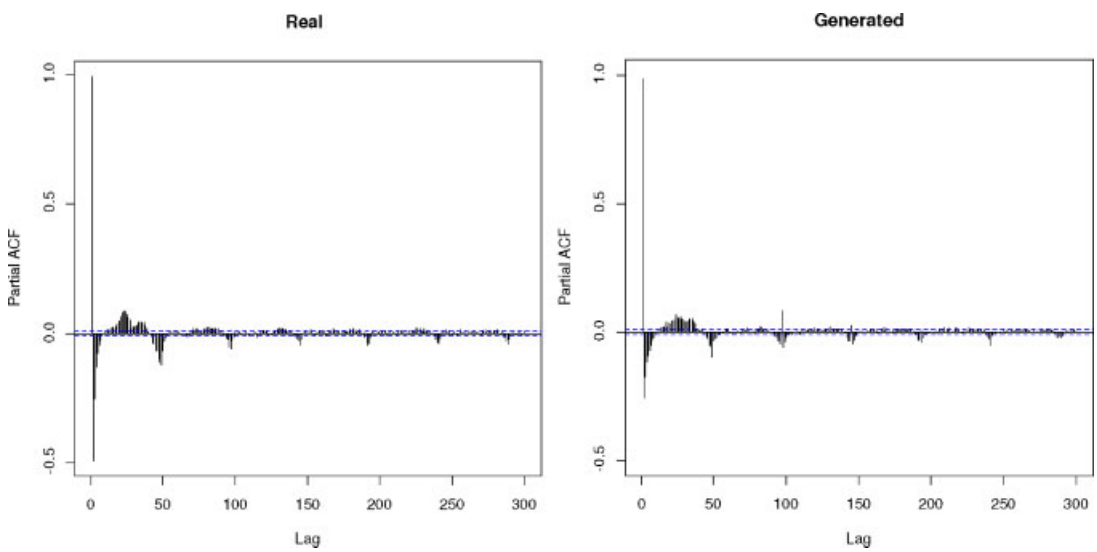

Figure 13. PACF of real and generated temperatures

DOI: 10.1002/env

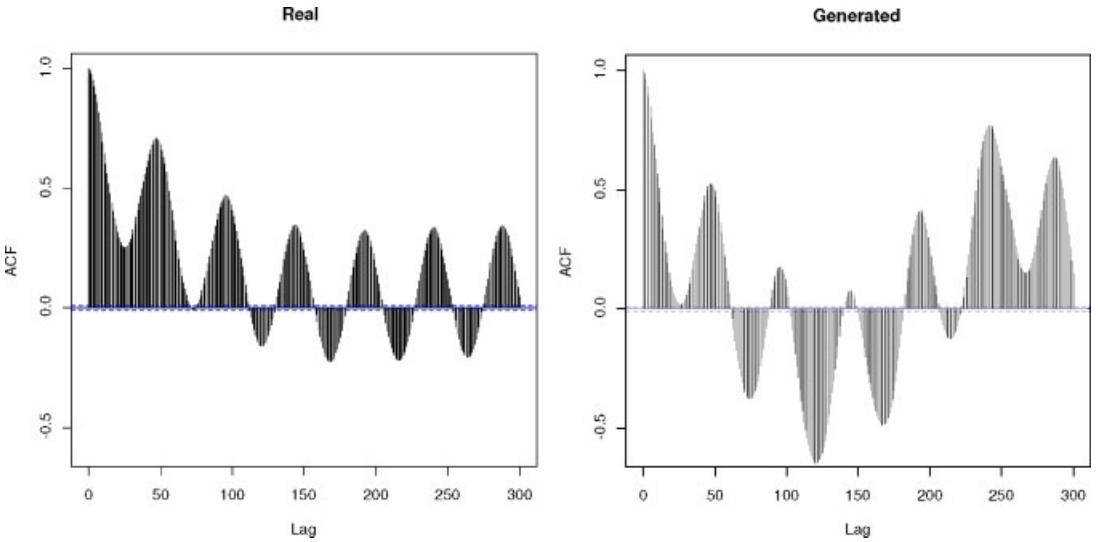

Figure 14. ACF of real and generated temperatures without eliminating the daily mean temperatures

the average frequency of occurrence for each duration M at different temperature thresholds: 18° C and below, 25° C and above and 34° C or above for both real and generated data. Then t-tests were performed to test the hypotheses that the mean of occurrence of real and generated runs of each duration in a specific month are equal. The null hypothesis was only rejected (at the 1% level) in 1 of the 36 tests performed, namely in the December group when the threshold is 34° C and $M = 12$.

Finally, from the half-hourly sequences generated, the daily minima, maxima and averages were calculated and compared with the 9 years of historical daily temperature data used in the estimation of the model. Descriptive statistics were calculated for real and generated data for each month and the results for December–February are shown in Tables 5–7. In general the statistics show good agreement. The largest discrepancies are observed in the average in January when the median is overestimated by about 1.4° C and in the minimum in February when the median is overestimated by about 1.3° C. The other statistics show no large discrepancies between real and generated data. We can conclude that the daily statistics obtained from the generated half-hourly temperature are close to the daily statistics obtained from the historical data.

5. APPLICATION

The half-hourly traces obtained from the model presented in this paper were used to perform analyses in the electricity market in South Australia (Magnano and Boland, 2007). In this region, the electricity demand in summer is very volatile and highly dependent on temperature. Summers can be very hot and a high percentage of South Australian electricity demand is driven by air conditioners. Analyses were performed to test whether, under different temperature scenarios, there will be enough capacity to satisfy the demand. This analysis needs to be performed using half-hourly data since it is of particular interest to analyse how the demand could be supplied in critical times of the day such as in the afternoon when the peak usually occurs.

DOI: 10.1002/env

There are different temperature behaviours that can be identified. For example, the summer can be mild on average without extremely hot days; mild on average with only a few extremely hot days; or it can be a hot summer when there are several days when the temperature is high. However, it may happen that the highest temperature of a milder summer is above the highest temperature of a generally hot summer. These different behaviours, and others that are not mentioned here, affect the behaviour of both peak and energy electricity demand.

In practice it might occur that there is only one historical occurrence of one specific scenario. For example, from the 9 summers of half-hourly temperature in the period from 1996 to 2005, only the summer 2000–2001 showed an average temperature above 24° C. Although this scenario is unlikely when it occurs, it can have a high impact in the level of demand for this summer.

When a specific temperature scenario occurred only a few times in the observed historical data, there is little randomness associated with the historical temperature data for future analysis. For instance, if the hottest day was on a Saturday, it cannot be known how the demand behaviour would have been

Table 5. Descriptive statistics for minimum, maximum and average daily temperatures: December

	Min		Max		Ave	
	Real	Gnr	Real	Gnr	Real	Gnr
Mean	15.7	16.2	26.5	26.7	20.7	21.4
Median	14.7	15.2	25.2	25.3	19.4	20.2
St.dev	4.0	4.4	5.5	5.5	4.3	4.5
Skewness	1.1	1.0	0.6	0.7	0.8	0.9
Kurtosis	0.8	1.0	−0.5	−0.3	0.0	0.2
5%	10.8	10.5	19.4	18.7	15.6	15.9
10%	11.7	11.4	20.0	20.6	16.0	16.5
25%	12.8	13.0	22.1	22.3	17.3	18.0
75%	17.6	18.5	30.7	30.6	23.5	24.1
90%	21.9	22.4	34.2	34.9	27.6	28.2
95%	24.2	24.9	36.5	37.6	29.3	30.4

Table 6. Descriptive statistics for minimum, maximum and average daily temperatures: January

	Min		Max		Ave	
	Real	Gnr	Real	Gnr	Real	Gnr
Mean	17.6	18.1	29.2	29.3	22.9	23.7
Median	16.2	17.2	28.6	28.7	21.6	23.0
St.dev	4.4	4.4	6.0	5.8	4.8	4.7
Skewness	0.8	0.7	0.3	0.3	0.6	0.5
Kurtosis	−0.2	−0.1	−0.1	−0.9	−0.8	−0.7
5%	12.4	12.2	21.1	21.4	16.8	17.4
10%	13.1	13.0	21.9	22.3	17.6	18.1
25%	14.3	14.7	23.8	24.1	18.9	19.6
75%	20.3	21.1	34.0	33.8	26.3	27.0
90%	24.4	24.3	37.7	37.5	30.0	30.5
95%	26.5	26.4	39.0	39.3	31.6	32.3

DOI: 10.1002/env

if the hot day had occurred on a Tuesday. Furthermore, the effect of high temperature on demand is different after many consecutive days of high temperature than for a single hot day. For this reason, it is important to be able to generate temperature traces that represent all the possible behaviours of the temperature. This means it is necessary to use temperature traces that present the same statistical characteristics of the real traces but that include some randomness. Our proposed model allows such analysis to be carried out.

To analyse the possible patterns that electricity demand can follow under different scenarios of temperature, a half-hourly electricity demand model was developed (Magnano and Boland, 2007). The temperature traces were used as input in the electricity demand model and generation of traces of half-hourly electricity demand under different scenarios of temperature were obtained.

The generated traces of demand can also be used as input data in market simulation software to analyse different aspects such as supply availability or electricity price under different temperature scenarios. The temperature data used in these analyses are important to determine the different demand profiles.

In (Magnano and Boland, 2007), we present an example of how much the energy and peak change because of different average temperatures. For this particular example, we chose two different temperature profiles. In the first profile, the average summer temperature is around 23.8° C and the maximum is 42.1°. A second profile was chosen with the summer average temperature being 21.6° C and the maximum 38.8° C. Half-hourly electricity demand was generated for each of the two profiles. The simulations gave the following results: in the hot summer scenario the expected energy use is 3472.8 GWh with a peak of 2.96 GW while in the mild summer the expected energy use is 3339.6 GWh with a peak around 2.75 GW.

As mentioned above, this model can be used for simulation purposes when there are insufficient data. Unfortunately, sometimes temperature data is scarce, especially on this finer time scale. In this case, the generated traces play a crucial role in the analysis.

Another possible use of generated traces of ambient temperature is in simulation models which assess the performance of photovoltaic cells given that the power obtained from solar cells is a function of temperature and global solar radiation. For crystalline solar cells in particular, the efficiency of conversion decreases with increase in temperature.

Table 7. Descriptive statistics for minimum, maximum and average daily temperatures: February

	Min		Max		Ave	
	Real	Gnr	Real	Gnr	Real	Gnr
Mean	17.6	18.3	29.6	29.4	23.3	23.8
Median	16.2	17.5	29.3	28.8	22.4	23.2
St.dev	4.4	4.7	5.8	5.7	4.9	4.8
Skewness	0.8	0.4	0.2	0.4	0.5	0.5
Kurtosis	−0.2	−0.3	−1.0	−0.7	−0.6	−0.5
5%	12.4	11.7	20.9	21.5	16.8	17.4
10%	13.1	12.8	22.3	22.5	17.7	18.2
25%	14.3	14.6	24.7	24.5	19.5	19.9
75%	20.3	21.6	34.3	33.6	27.0	27.2
90%	24.4	24.7	38.0	37.2	30.5	30.7
95%	26.5	26.7	39.0	39.3	32.0	32.5

DOI: 10.1002/env

6. CONCLUSIONS

This paper presents models to generate synthetic sequences of daily and half-hourly temperatures. Our approach involves an innovative characteristic: it accounts for the thermal inertia that occurs when the temperature is above the monthly average for successive days until a cool change occurs and the temperature is below the monthly average for a short period. We modelled this feature by adding to the half-hourly model the generated daily mean temperatures that contain these longer autocorrelations.

The real and generated data were compared using descriptive statistics, graphs and statistical tests. These comparisons showed that, in general, the characteristics of real and generated traces are similar. The ACFs and PACFs between both sets of data are close with the exception of the partial correlation of order 96 that shows an opposite sign. We also compared the number of times that consecutive values are above or below specific thresholds for 3, 6, 9 and 12 consecutive hours for each month. There was only 1 significant difference in the 36 tests performed.

Also daily measures calculated from the half-hourly generated sequences were compared against historical daily measures. The largest discrepancies were observed in the average in January and in the minimum in February. In both cases the generated data overestimated the real one. The other daily measures showed a good agreement between the real and generated data.

An application was presented where the temperature model was used to generate different patterns of electricity demand. The temperature model allowed detailed analysis of how electricity demand varies with different temperature profiles.

The half-hourly temperature model presented in this paper deals with the two time scales. The temperature depends not only on recent half-hourly values but also on the thermal inertia of the previous days. This model deals with a combination of both variables. It successfully generate traces of temperature with similar characteristics to real data.

We plan to generalise our model to the whole year. Moreover, there is a hypothesis that climate change will be reflected in an increase in the minimum temperature and a lower increase in the maximum temperature. Using Fourier series and adding some constraints, the general temperature profile under climate change can be obtained. We will analyse this feature for synthetic generation in a future paper.

ACKNOWLEDGEMENT

We are grateful for the financial support and supply of data of the Electricity Supply Industry Planning Council.

REFERENCES

Hansen JE, Driscoll DM. 1977. A mathematical model for the generation of hourly temperatures. *Journal of Applied Meteorology* **16**(9): 935–948.

Boland JW. 1995. Time series analysis of climatic variables. *Solar Energy* **55**(5): 377–388.

Magnano L, Boland JW. 2007. Generation of synthetic sequences of electricity demand. *Energy* **32**: 2230–2243.

Magnano L. 2007. Mathematical models for temperature and electricity demand. *PhD Thesis*, University of South Australia.

McSharry PE, Bouwman S, Bloemhof G. 2005. Probabilistic forecasts of the magnitude and timing of peak electricity demand. *IEEE Transactions On Power Systems* **20**(2): 1166–1172.

Politis DN. 2003. The impact of bootstrap methods on time series. *Statistical Science* **18**(2): 219–230.

DOI: 10.1002/env