The seasonal forecast of electricity demand: a simple Bayesian model with climatological weather generator

Sergio Pezzulli, Patrizio Frederic, Shanti Majithia,
Sal Sabbagh, Emily Black, Rowan Sutton
and David B. Stephenson

12th October 2004

Abstract

In this paper we focus on the seasonal prediction of the electricity peak-demand daily trajectory during the winter season in Central England and Wales. We define a Bayesian hierarchical model for predicting the winter trajectories and present results based on the past observed weather (climatology). Thanks to the flexibility of the Bayesian approach, we are able to produce the marginal posterior distributions of all the predictands of interest. This is a fundamental progress with respect to the classical methods. The results are encouraging in both skill and representation of uncertainty. Further extensions are straightforward, at least in principle. The main two of those consists in conditioning the weather generator model with respect to additional information like the knowledge of the first part of the winter and/or the seasonal weather forecast.

1 Introduction

In Central England and Wales, winter peak-demand usually occurs between 5.00 pm and 5.30 pm. These data, among others, are routinely collected and investigated by the National Grid Trasco group (NGT), operating the high-voltage electricity transmission system.

NGT operates the high voltage network and is responsible for residual short term system balancing 24 hours a day. To this aim, demand's prediction is essential. In particular the *seasonal* prediction, that is the forecast from one month up to one year ahead, is crucial for medium-term planning of energy production and trade. In collaboration with NGT, in this paper we focus on the seasonal prediction of the peak-demand daily trajectory during the winter season.

Different approach have been taken for electricity demand forecasting, including time-varying splines (Harvey and Koopman, 1993), artificial neural networks (Connor, 1996) and multiple regression models (Ramanhathan et al. 1997, Taylor J.W. and Buizza R.(2003)). The standard NGT seasonal model regresses the demand against a calendar cycle, three weather transforms and the Service sector index, a main component of the National Gross Product which is issued quarterly by the National Statistical Office. Obviously, services and weather series are unknown regressors in extrapolating the demand prediction. One year ahead, the standard NGT seasonal forecast assumes an average (*climatological*) weather and the last year services. Later on, until six months ahead, a constant value is periodically added to the estimated demand's trajectory in order to take account of the expected growth of the services and of other economical factors.

A main drawback of this method is that the resulting forecasts cannot express the uncertainty on weather and services. Some classical strategies may be adopted for this, like using climatological quantiles or bootstrapping. We adopt instead a Bayesian approach, which we found more straightforward and clear. We define a Bayesian hierarchy as a generalization of the NGT regression model. For the weather, we assume a probability model based on past data: the climatological weather generator described in section 3.1. For the Service series, finally, because it grows approximately linearly in time, we avoid the explicit use of the regressor by including a random effect following a linear pattern (as shown in section 3).

While for day-to-day forecasting the engineering constraints produce a quadratic loss function, NGT seasonal forecasting, instead, is much more complex and subjected to change from year to year. The general aim is multipurpose. Several predictands can be relevant to the industrial planning. The most important ones are the general level of the winter trajectory, the highest demand's value and its location in time. Thanks to the flexibility of the Bayesian approach, we are able to produce the marginal posterior distributions of those predictands. This is a fundamental progress achieved by the Bayesian procedure with respect to the classical methods. Also, we will show that a more realistic representation of uncertainty is produced.

After introducing the standard NGT regression model in the next section, we define, in section 3, the Bayesian model. Results and conclusions are in sections 4 and 5, respectively.

2 The standard NGT prediction method

2.1 The data

The available data include the daily peak demand fluxes (in MW) collected by NGT for $k = 17$ consecutive winters, from 1986-1987 to 2002-2003. Those trajectories start from about the last Sunday of October to the last Saturday of March. The last observed trajectory is represented in figure 1(a). A weekly cycle can be clearly spotted, with less demand on the weekend due to less industrial activity. For the same reason there is a low during Christmas holidays.

Because high demand is more important, weekends, bank holidays, Christmas holidays and other special days are omitted from the fitting process. Some anomalous day, like in case of severe storms causing network interruptions, and consumption fall, is also excluded in order to avoid erroneous contaminations of the results. The data actually used correspond to the bold part of the curve in Fig. 1(a). These are also represented in Fig. 1(b) for all available winters. Here, the general growth of the winter peak demand can be clearly observed.

On the other hand, even the form of the seasonal trajectory has changed in the last decades. Electricity demand is subjected to fast structural changes in time, because the consumer can choose different options within the energy sources. The weekly cycle is also changing because of the increasing activity of the service sector during the weekends. Therefore, the fitting process is usually restricted to four consecutive winters only.

2.2 The NGT model

The NGT model is a linear model whose explicative variables are divided into 3 groups:

i) the calendar component C: including dummies for Thursday and Friday and a winter cycle described by a cubic polynomial.

ii) the economic component S: that is the series of the Service Sector Index.

iii) the weather component W: including three opportune transforms of ground temperature, wind speed and ground solar radiation.

The original weather variables, provided by the UK Met Office, are weighted averages at few selected stations. The transforms are called effective temperature: TE; cooling power of the wind: CP; and (solar) effective illumination: EI. These transforms aim to depict how the external weather can affect the energy needs inside the buildings (see also Taylor J.W. and Buizza R.(2003)). The effective temperature, TE, is an exponentially smoothed form of another variable, TO, which is the mean spot temperature during the 4 hours before the peak time. The wind's cooling power, CP, is a nonlinear transform of TO and wind speed. Finally, the effective illumination, EI, was originally a complex function of visibility, number and type of cloud and type of precipitation, but recently it has been replaced by a cubic function of the ground solar radiation. Calling:

$d_j(t)$	Electricity demand
$Th_j(t)$	Dummy for Thursday
$Fr_j(t)$	Dummy for Friday
$TE_j(t)$	Effective Temperature
$CP_j(t)$	Cooling Power of the wind
$EI_j(t)$	Effective Illumination

for day $t = 1, 2, ..., T$ $(T = 183)$, of winter $j = 1, 2, ..., k$ $(k = 17)$, the NGT model has the form:

$$\begin{aligned} d_j(t) &= \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \alpha_4 Th_j(t) + \alpha_5 Fr_j(t) + \qquad (1)\\ &\quad +\beta S_j(t) \\ &\quad +\gamma_1 TE_j(t) + \gamma_2 CP_j(t) + \gamma_3 EI_j(t) + \\ &\quad +u_j(t) \\ &= \alpha_0 + \boldsymbol{\alpha}'\mathbf{C}_j(t) + \beta S_j(t) + \boldsymbol{\gamma}'\mathbf{W}_j(t) + u_j(t) \qquad (2) \end{aligned}$$

where $\mathbf{C}_j(t) = (t, t^2, t^3, Th_j(t), Fr_j(t))'$ is the *cycle* vector, $S_j(t)$ is the Service Index, and $\mathbf{W}_j(t) = (TE_j(t), CP_j(t), EI_j(t))'$ is the weather vector; α_0, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_5)'$, β, and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)'$ are the regressor coefficients; and $u_j(t)$ is an ARMA modelled 'error' variable.

2.3 NGT basic and shifted forecasts

The *basic* NGT forecast is produced in late March, when the previous winter is just finishing. At that moment, not only weather and services are unknown for the next year, but the latter are unknown even for the current winter. In fact, the winter services series are issued in May by the Statistical Office. In order to solve this additional problem, the basic (one year ahead) forecast uses a simple trick that manages to use approximately the weather average and the last winter services as follows.

Let $\overline{TE}(t)$, $\overline{CP}(t)$ and $\overline{EI}(t)$ be the average-trajectories of the weather variables for the winter day t. Any statistical measure of location can in principle be chosen, like the daily means or medians. Here we prefer the cubic splines shown in the next section.

Let $\overline{\boldsymbol{W}}(t) = (\overline{TE}(t), \overline{CP}(t), \overline{TE}(t))'$. If k is the current winter and $\hat{\boldsymbol{\gamma}}$ is the current estimate of the weather coefficient $\boldsymbol{\gamma}$ (fitted on the winters from $k-1$ to $k-4$), let us consider the estimator of $d_{k+1}(t)$:

$$D_{k+1}(t) = d_k(t) + \hat{\boldsymbol{\gamma}}'(\overline{\boldsymbol{W}}(t) - \boldsymbol{W}_k(t)) \tag{3}$$

Assuming model 1 and $\hat{\boldsymbol{\gamma}} \approx \boldsymbol{\gamma}$, it is easy to show that:

$$D_{k+1}(t) \quad \approx \quad \alpha_0 + \boldsymbol{\alpha}'\mathbf{C}_k(t) + \beta S_k(t) + \boldsymbol{\gamma}'\overline{\boldsymbol{W}}(t) + u_k(t)$$

Thus D_{k+1} is based on $S_k(t)$ although that series is unknown. Moreover, the weather component is fixed to the climatology and the last year error term replaces the next year error. Also, both the cyclic and economic coefficients enter (3) with their *true* coefficient values, that is a quite appealing property for an estimator.

Note, finally, that the winter cycle is aligned with the last winter, instead of the $(k+1)$-th. In other words, there is $C_k(t)$ instead of $C_{k+1}(t)$. However, this is only a minor drawback because $D_{k+1}(t)$ can be shifted for few days in order to align the weekdays in C_k to the next year calendar. This final re-alignment gives the basic forecast: NGT0, say. It approximately assumes no economical growth, that is persistence (or stagnation) of economy, and average weather conditions. Then the inter-annual growth of the demand is periodically estimated by separate econometric modelling. More precisely, an expected (constant) inter-annual shift is predicted and summed up to the basic forecast.

Here we will consider the basic forecast NGT0 and the last shifted forecast, NGT1 say, which is computed just before the start of the winter, that is 6 month ahead in our terminology. Thus we retain the NGT standard methods under the worst and the best possible information states. Pezzulli et al. (2004) review the standard NGT procedures, where the ARMA modelling of the error term was not found much relevant. This motivates the following model.

3 The Bayesian Hierarchical Model

As shown in figure 2 , the Service index growth approximately linearly in time. Therefore the linear part of the calendar cycle can adjust for the S contribution to the demand and made S itself become redundant. This will be achieved by means of a random winter effect α_{0j} in our model. Therefore we assume:

$$(d_j(t)|\theta_j(t), \sigma) \sim N(\theta_j(t); \sigma^2) \tag{4}$$

with

$$\theta_j(t) \;=\; \alpha_{0j} + \boldsymbol{\alpha}'\mathbf{C}_j(t) + \boldsymbol{\gamma}'\mathbf{W}_j(t)$$

and where α_{0j} represents the random effect for winter j:

$$(\alpha_{0j}|\lambda_0, \lambda_1, \sigma_0) \sim N(\lambda_0 + \lambda_1 \cdot j, \sigma_0^2)$$

This assumption introduces three additional unknowns λ_0, λ_1 and σ_0, called hyperparameters.

As usual in Bayesian Statistics we assume that the vectors of regression coefficients $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ have noninformative normal prior distributions while the scale parameters σ and σ_0 have noninformative inverse gamma priors.

By the previous settings we can infer on all the unknown parameters and hyperparameters of the model. However, if we want to predict the electricity demand for the next year (e.g. for year $k+1$), we need to define a probability distribution for the three weather vectors $\boldsymbol{TE}_{k+1}$, $\boldsymbol{CP}_{k+1}$, $\boldsymbol{EI}_{k+1}$ which are currently unknown. This further distribution will depend on all the past available information and is defined in the next subsection. The predictive distribution for the next winter trajectory is then evaluated by averaging the demand formula 4 over the unknown quantities, that is the parameters, the hyperparameters and the weather vectors.

3.1 The Climatological Weather Generator

First, define the $3T$-dimensional composite-weather vector $\boldsymbol{W}_j$ such that

$$\boldsymbol{W}_j' = (\boldsymbol{TE}_j', \boldsymbol{CP}_j', \boldsymbol{EI}_j').$$

We assume the $\boldsymbol{W}_j$ exchangeable for all $j = 1, 2, ..., k+1$. This means, for example, that the effect of climate change is ignorable in such a short period of time, which is quite reasonable.

Secondly, we assume the $\boldsymbol{W}_j$ to be multivariate normally distributed:

$$\boldsymbol{W}_j \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

The normality assumption is forced by the computational burdens, but it is also in agreement with our previous investigations (Pezzulli et al. 2003). Actually, this is not a strong constrain, because it allows for a very complex structure in terms of means, variances and correlations. In fact, the number of parameters is so huge $p = 3T + 3T(3T+1)/2 = 151,524$ that the probabilistic structure is even too much complex. Since we have observed just 17 winters, the maximum likelihood estimate of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is strongly ill-conditioned. Moreover, this great flexibility of the multivariate normal family is against our prior knowledge. For example, it allows very different mean and variances for the same weather transform (e.g. TE) on two consecutive days. It also supports very different correlations when moving from the pair (t, t') of days to the pair $(t+1, t'+1)$. We know, instead, that some sort of quasi stationarity in the mean, in the variances and correlations should hold. Thus, in order to take account of the temporal structure of the weather trajectories, we used the following smoothed approach.

For the mean vector $\boldsymbol{\mu}$, we compose the 3 smoothing splines of the daily mean trajectories (with subjective choice of the smoothing parameters). The result is shown in figure 3 , where the raw daily means (e.g. the maximum likelihood solutions) can be compared to the smoothed estimates. Of course we have prior reasons to believe much more in the latter. The same method has been used for the variances. For the correlations, working separately on each of the 6 variance-blocks $\boldsymbol{TE}$ vs $\boldsymbol{TE}$, $\boldsymbol{TE}$ vs $\boldsymbol{CP}$, ..., $\boldsymbol{EI}$ vs $\boldsymbol{EI}$, the same smoothing spline estimation is used to 'correct' the empirical correlations of lag 1, then the lag 2 correlations and so on. At the end of this rather extensive process, positive definiteness is achieved by the Higham algorithm on the estimated correlation matrix (Higham 2002).

In conclusion, we found that smoothing improves greatly the definiteness of the variance matrix. While the raw estimate has $k = 17$ positive eigenvalues out of $3T = 459$ dimensions of the matrix, the smoothed estimate identify more than 300 dimensions. The eigenvectors interpretation (not shown here) is also more clear and credible in the latter case. A comparison between raw and smoothed approaches is shown in figure 4(a) and 4(b), respectively, for the $\boldsymbol{TE}$ vs $\boldsymbol{TE}$ correlation block. As shown in Parzen (1961), the assumption of continuity of the covariance kernel is essential for the predictability of the stochastic process. This is equivalent to impose regularity conditions on the weather trajectories (see e.g. Wahba (1990)).

For simplicity, the weather generator has been fitted by means of all the available winters and then used to evaluate either the standard NGT and the Bayesian models in forecasting some of the last observed trajectories. This is not completely correct because we should use the weather observed in the previous winters only, excluding of course the targeted one. However, there are two main reasons for accepting this procedure. First, the (smoothed) weather generator does not sensibly change if we exclude the last few winters. Secondly, we are comparing the models under the same conditions, so that the relative values are likely unchanged. The model parameters, on the other hand, are estimated over the last four winters that precede the targeted trajectory, either for the standard NGT methods and the Bayesian forecast.

Note, finally, that in order to forecast the next winter trajectory, the climatological weather generator enter the Bayesian model as a known weather distribution. On the other hand, we could also model our uncertainty about the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of the weather generator itself, thus depicting a more realistic position. This is computationally very expensive, however, and probably will have only a minor effect on the results. For this reasons it is not implemented in this paper.

4 Results

We compared standard and Bayesian models on the last 6 winters. For each one, we used Splus for computing the standard NGT predictions and WinBUGS for the Bayesian one. For the latter, we checked the convergence of the MCMC algorithm by the ANOVA based-method running in WinBUGS (Brooks and Gelman, 1998). After this, a posterior sample of 1,000 predicted trajectories has been collected and compared to the actual observations. The expected trajectory (composed by the daily averages of the posterior sample) has been used as the Bayesian forecast.

In Table 1, we compare the Bayesian forecast with the basic NGT forecast (NGT0) and the shifted one (NGT1) over the last six observed winters. The bias error corresponds to the error in guessing the average winter demand. To this aim, we conclude that both the Bayesian and the shifted-NGT methods are slightly improving the basic NGT forecast. The performances of those two are rather close in terms of mean square error. Although relative efficiencies find some differences, those are very small in term of MW. This is probably due to the similarity between the underlying models.

In order to assess the uncertainty around the expected trajectory, we computed some reference quantile trajectories and then checked the observed coverages. For example, for any day we compute a 50% credibility interval and trace the upper and lower line for all the days of the winter, thus obtaining a 50% credibility band. Then the model is validated if the observations included in the band are about 50%. In figure 5 , for the year 2002-03 we show the expected trajectory and the 50, 90 and 98 percent bands (lines). The actual percent of observations (points) inside those bands are 55.5, 90.0 and 98.2 respectively. Thus the predicted fractions are quite similar to the observed ones, and we can conclude that the spread of the posterior sample is reasonable. This is not so if we assume a fixed climatological weather, that

Winter	NGT0		NGT1		Bayes			
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Eff0	Eff1
1997-98	0.3	1.1	0.0	1.1	-0.4	1.1	96	93
1998-99	-0.2	1.0	-0.6	1.2	-0.7	1.2	82	93
1999-00	0.5	0.9	0.6	1.0	-0.2	0.9	97	103
2000-01	1.3	1.8	0.9	1.5	0.8	1.5	124	106
2001-02	0.7	1.2	0.3	1.0	0.3	1.0	115	97
2002-03	0.9	1.4	0.4	1.1	0.3	1.2	117	96

Table 1: Forecasting performances of NGT0, NGT1 and Bayesian models. Bias and root mean square error (RMSE) units in GW. Efficiencies Eff0 and Eff1 of the Bayesian model versus NGT0 and NGT1 models are ratios of the form NGT-RMSE/Bayesian-RMSE in percent values

essentially corresponds to use the confidence intervals on the standard NGT models. The 98% credibility band (not shown here) is about halved in size and contains less than 60% of the data.

Two important predictands for NGT regard the maximum of the trajectory. These are the intensity and the time of the winter maximum, which are useful for planning the peak production load. The marginal posterior densities for winter 2002-03 are shown in figure 6(a) and 6(b) for intensity and time, respectively. In order to improve the exposition we used the Rosemblatt's kernel density method. Those pictures are rather straightforward to present to the decision maker, who can probably make sensible use of the conveyed information. As shown by the circles, the predictions are rather accurate for that winter.

5 Conclusions

We propose a Bayesian hierarchy with normal likelihood, noninformative prior on the parameters and hyperparameters and a weather generator model based on climatology.

The results are encouraging in both skill and representation of posterior uncertainty. The credible intervals for the demand trajectory show to cover a percentage of observations that is enough close to the expected percentage. Compared with the standard operative methods we do not need either the Service Index series and the separate econometric modelling that is used to update the basic NGT-operational forecast.

Also, the Bayesian forecast is available as soon as the previous winter is finished, that is one-year ahead, and can easily provide sensible predictions regarding any functional of the targeted trajectories, like the maximum value and its location in time.

The Bayesian approach is also extremely flexible for using new information. For example, after the starting of the winter, the weather generator can easily be modified in order to condition on the observed part of the weather trajectories.

A further development of our project is to find the opportune modification of the weather generator for conditioning on the seasonal forecast of the weather variables.

6 Acknowledgement

Funded by NERC COAPEC project, grant number NER/T/S/2002/00154: H1792500. We thanks Krystyna Zietak (Wroclaw University of Technology) for providing suggestions and references on how to achieve the positive definiteness of the correlation matrix.

References

Brooks S. P. and Gelman A. (1998) Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics. 7, 434-455.

Connor J. T. (1996) A robust neural network filter for electricity demand prediction. J. of Forecasting, 15, 436-458.

Harvey A. C., Koopman S. J. (1993) Forecasting hourly electricity demand using time-varying splines, JASA 88, 1228-1236.

Higham, N. J. (2002) Computing the nearest correlation matrix: a problem from finance, Journal of Numerical Analysis, 22, 329-343.

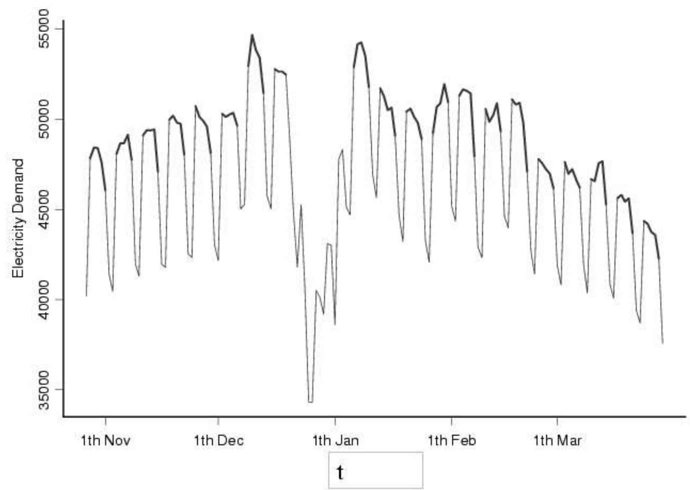
Parzen, E. (1961) An approach to time series analysis, AMS, 32, 951-989.

Pezzulli S. Sabbagh S. and Majitia S. (2003). Electricity Demand Seasonal Forecast at the National Grid Company: basic procedure and model validation (NGT Internal Report).

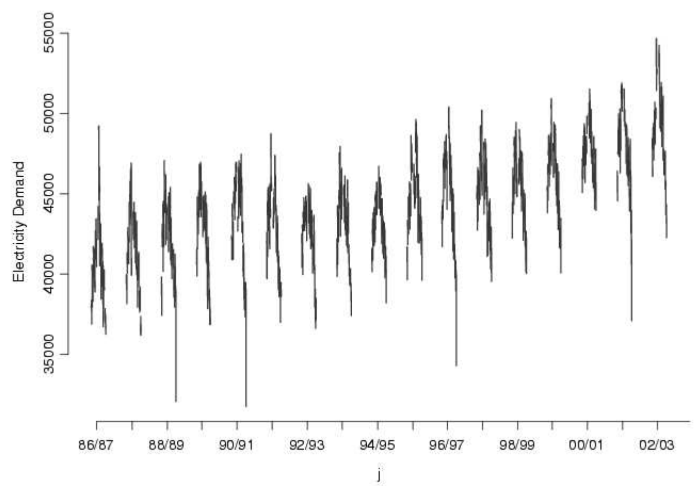
Ramanathan R., Engle R., Granger C. W. J. Vahid-Araghi F. and Brace C. (1997) Short-run forecast of electricity loads and peaks. Int. J. of Forecasting, 13, 161-174.

Taylor J.W. and Buizza R. (2003) Using wether ensamble predictions in electricity demand forecasting, Int. J. of Forecasting, 19, 57-70.

Wahba, G. (1990) Spline Models for Observational Data. SIAM, Philadelphia.



(a)



(b)

Figure 1: a. The 2002-2003 winter trajectory of the daily peak electricity demand (MW). b. All the winter trajectories (used data only).

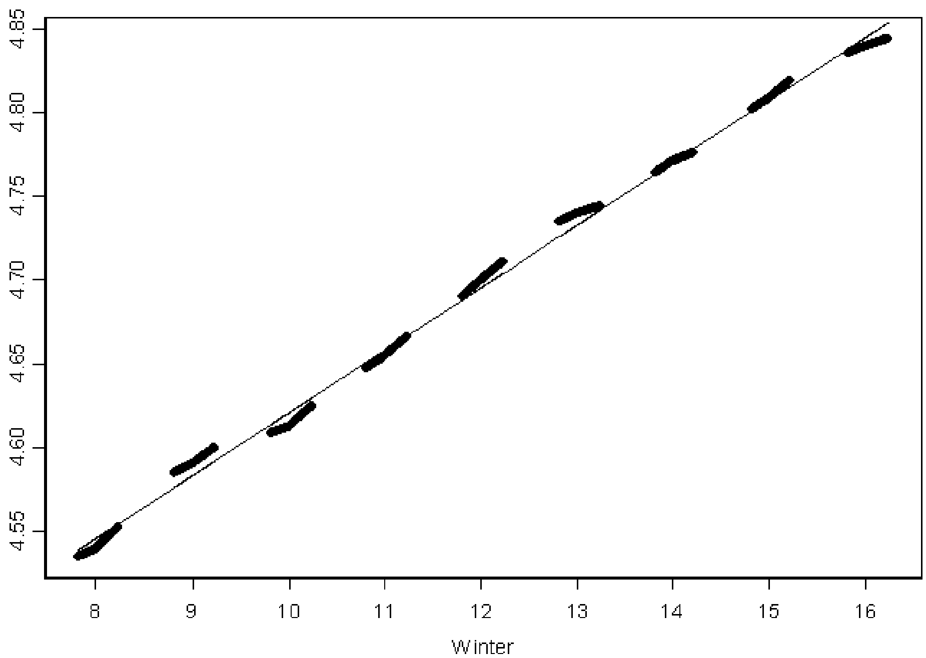


Figure 2: The Service Index trajectories for the last 9 observed winters (broken lines) versus an interpolating line (continuous line)

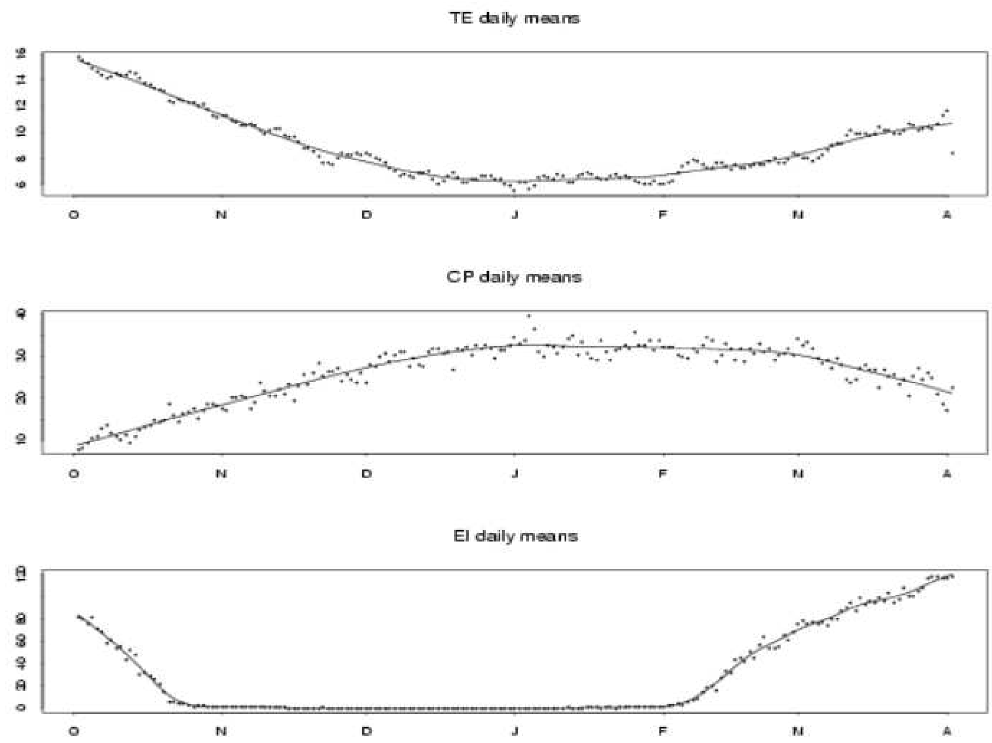
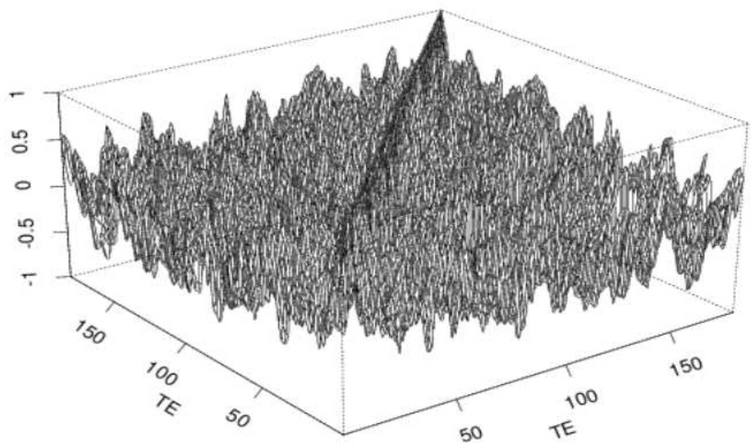
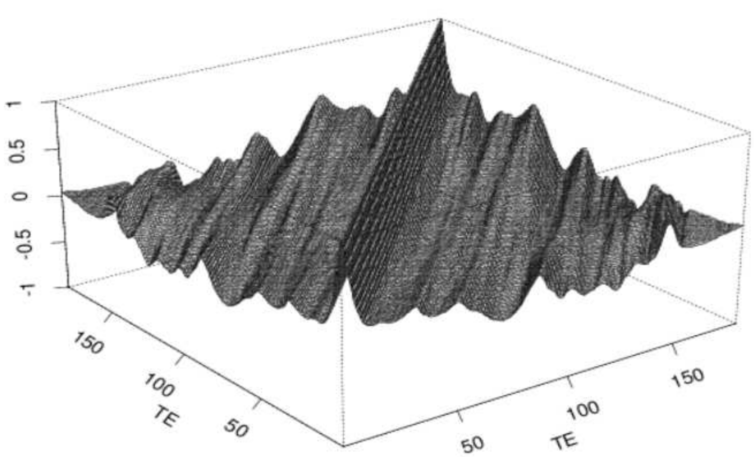


Figure 3: Observed daily means of effective temperature (TE), cooling power of the wind (CP) and effective illumination (EI), denoted by points, in comparison with the corresponding smoothing splines (lines)



(a)



(b)

Figure 4: Raw correlation surface (a) and smooth correlation surface (b) for winter effective temperature (TE). Winter days on X and Y axees.

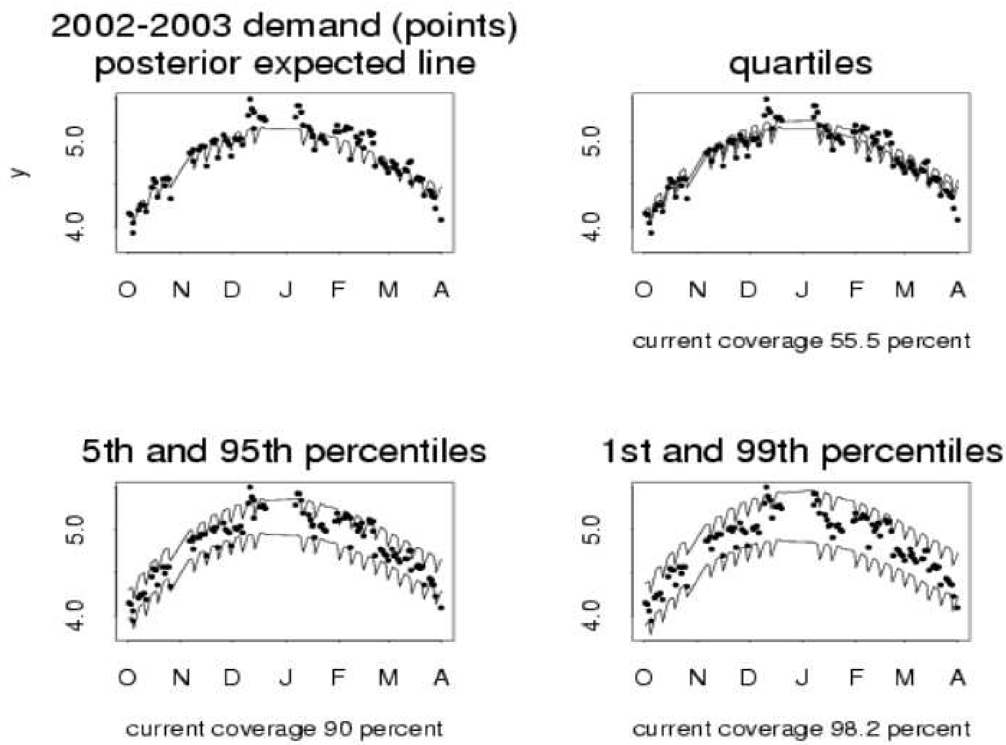
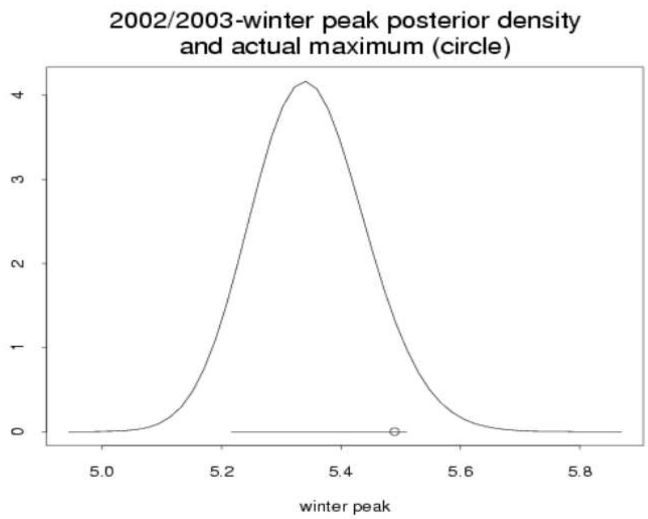
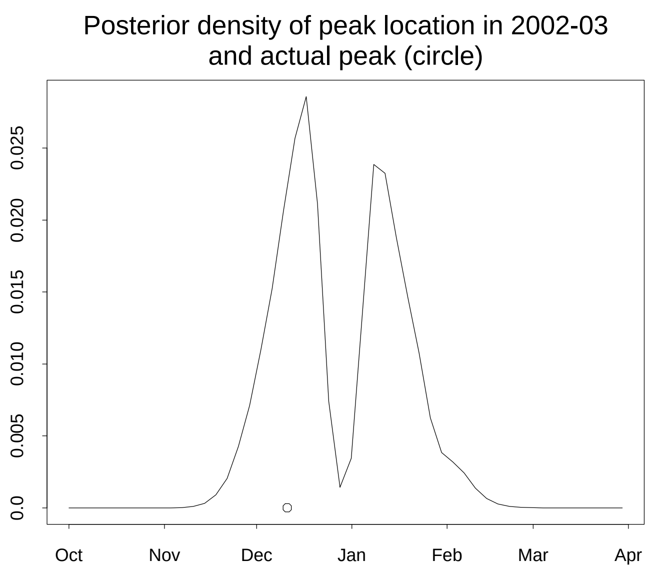


Figure 5: Bayes forecast and 50%, 90% and 98% credibility bands (lines) against actual trajectory (points). Units in 10GW.



(a)



(b)

Figure 6: Kernel density representations of the marginal posterior distributions of the maximum winter demand intensity (a) and time (b), compared to the occurred values. In (a), the horizontal line indicate the 95% credibility interval (10GW units)