Local Load Analysis with Periodic Time Series and Temperature Adjustment

Marcelo Espinoza, Bart De Moor
K.U.Leuven-ESAT-SCD
Kasteelpark Arenberg 10
3000 Leuven, Belgium
marcelo.espinoza@esat.kuleuven.ac.be

Caroline Joye
ELIA
Bd. de l'Empereur 20
1000 Brussels, Belgium
caroline.joye@elia.be

Ronnie Belmans
K.U.Leuven-ESAT-ELECTA
Kasteelpark Arenberg 10
3000 Leuven, Belgium
ronnie.belmans@esat.kuleuven.ac.be

Abstract

Accurate modelling tools for TSO planners for the problems of peak load temperature adjustment, short-term forecasting, and customer identification, are presented in this paper. The results are derived from the analysis of intra-day (hourly) load records from local substations of the Belgian high-voltage grid, as provided by Elia (Belgian National Transmission System Operator - TSO). Using time series of hourly load values over a 5 years period, the short-term forecasting problem is addressed by a Periodic Autoregressive (PAR) model that leads to customer identification; the task of temperature adjustment is tackled by a multi-equation system with autocorrelated residuals. Satisfactory results are obtained for a large sample of substations in the Belgian high-voltage grid.

Keywords - Short-Term Load Forecasting, Periodic Time Series, Load Profiles, Temperature Sensitivity, Weather Adjustment

1 Introduction

The quantitative analysis of the electric load is currently a key research area [1, 2] with important implications for grid managers. Not only accurate forecasts are needed for the short-term operations and mid-term scheduling, but also network managers need to have insight in the type of customers they supply as a support for long-term planning. In order to deal with the everyday process of planning, scheduling and unit-commitment [2], the need for accurate short-term forecasts has led to the development of a wide range of models based on different techniques, with different degrees of success. In recent literature, some interesting examples are related to traditional time series analysis [1, 3, 4], and neural networks applications [5, 6, 7, 8, 9]. At the same time, the unbundling between generation, transmission, distribution and sales induced by market liberalization strenghtened the blindness of network managers beyond a certain substation level with respect to the power input and the final customers. In these cases, it is required to use indirect techniques to assess the type of demand they face [10, 11] in order to support their long-term investment planning. In this context, categories of residential, business and industrial customers have been documented for some locations [12, 13] and are recognized usually by their "typical" load pattern in a day. On the other hand, for long-term analysis and planning purposes, it is required to determine how much of the peak load observed in any year is due to weather effects, and how it is possible to adjust the observed peaks taking this information into account. Weather adjustment of historical load data and normalization of future load prospects to standard temperature conditions have indeed become highly critical in planning procedures: one of the consequences of liberalization is a higher financial pressure on the transmission system operator, notably leading to higher use than before [14].

Within this context, this paper presents research results and novel techniques derived from the analysis of intra-day (hourly) load records from local substations in the Belgian high-voltage grid, as provided by Elia (Belgian National Transmission System Operator - TSO). The goal is to provide accurate statistical and modelling tools to TSO planners for the tasks of peak load temperature adjustment, short-term load forecasting and customer identification. The data available consists of a set of time series of hourly load value over a 5 years period, derived from metering at local grid nodes. We have developed methodologies based on traditional econometrics that can tackle these problems. On the one hand, the task of peak load temperature adjustment is addressed [15] by estimating a multi-equation model for residential loads, where each hour of the week is modelled separately, as a function of temperature, weekly and monthly periodicities plus a deterministic trend factor. On the other hand, the problem of individual load modelling is tackled using a vector autoregression structure, based on a Periodic Autoregressions (PAR) system. By a simple extension, this model can be used to identify typical daily shapes. This paper is structured as follows. The problem of peak load temperature adjustment it discussed in Section 2, and the methodology applied for the problems of short-term forecasting and customer identification is described in Section 3.

2 Peak Load Temperature Adjustment

This section addresses the problem of temperature adjustment of the (peak) load. The TSO grid development team is currently testing new tools dedicated to temperature effect correction. This new method attempts to tackle some shortfalls present in the existing technique, namely, the volatility of temperature sensitivity coefficients from one planning year to another and its partial and/or asymmetrical implementation.

2.1 Procedure

The procedure of temperature adjustment can be structured in the following 3 steps:

1. Weather-load relationship identification: the optimal[1] model structure must be determined, capable to isolate with high precision the specific effect of temperature on the load;
2. Model check: once the adequate model structure is determined and estimated, it has to be assessed whether the obtained results are suitable for further data treatment, notably temperature adjustment;
3. (Peak) Load data adjustment: finally, the estimates of the relevant temperature effect are used to adjust observed peak load values for their extreme temperature component.

It should be stressed here that, while the global model structure is the same for all local loads, the estimation is performed for each load individually, e.g. "tailored" estimates of the temperature sensitivity for each analyzed local load are found. The following drivers traditionally determine the local load behaviour[16]:

1. A temperature sensitive part: in Belgium, for residential consumers, it is observed a clear negative relationship between load and temperature, due to the heating use; some loads do also experience a positive relationship with temperature, notably in locations were there is a high concentration of commercial and office buildings (usually equipped with air-conditioning);
2. A non-temperature sensitive part:
   - The periodicity inside the week and inside the year: socio-economical parameters (timetables, holidays, business cycle etc.) clearly produce significant electric load variations;
   - A monthly trend factor.

2.2 *The Model*

The model developed consists of a system of 24 independent equations (one equation for each hour of the day):

$$\begin{aligned} y_{h,d} \quad &= \quad \alpha_h + \beta_{1,h}\mathrm{HDD}_d + \beta_{2,h}\mathrm{CDD}_d + \sum_{j=1}^{6}(\delta_{j,h}\mathrm{WD}_j) \\ &\quad + \sum_{m=1}^{11}(\gamma_{m,h}M_m) + \tau_h\mathrm{MT}_d + u_{h,d}, \qquad (1) \end{aligned}$$

with $\quad u_{h,d} = \rho_h u_{h,d-1} + \varepsilon_{h,d}$

where $E(\varepsilon_{h,d}) = 0, Var(\varepsilon_{h,d}) = \sigma^2, Cov(\varepsilon_{h,d}, \varepsilon_{h,d-1}) = 0$, $h = 1, \ldots, 24$, and where:

- $y_{h,d}$: Load at hour h of day d;
- HDD_d : Heating degree-days = $\max(0, 16.5 - \mathrm{ET}_d)$;
- CDD_d : Cooling degree-days = $\max(0, \mathrm{ET}_d - 16.5)$;
- ET_d: Equivalent temperature = $0.6T_d + 0.3T_{d-1} + 0.1T_{d-2}$;
- T_d: Average temperature at day d;
- WD_j: Dummy variable for the day j in a week;
- M_m: Dummy variable for the month m;
- MT_d: Monthly trend variable;
- ρ_h: AR(1) coefficient of the residuals for equation of hour h.

The temperature variables finally elected[2] are actually temperature-derived indicators. They correspond to the usual degree-day index that measures the degree of difference between ambient temperature and outside temperature producing the best comfort inside buildings. The particularity of the degree-days used here is that the ambient temperature measure usually considered is replaced by the "equivalent temperature", namely a weighted average of the temperature over the 3 last days. This allows taking into account the inertia that characterizes the reaction of power consumption to temperature fluctuations (resulting from buildings isolation). Since the ordinary least squares (OLS) residuals are found to be autocorrelated, the model is estimated with the Yule-Walker [17] estimation method for autoregressive error models (first order).

2.3 *Estimation Results*

Model (1) has been estimated on 448 time series of residential load records measured at substation level, on an hourly basis, from January 1999 to December 2003. For evident space constraint, illustrations reported here will involve only two specific cases: one substation for which the yearly peak takes place in winter (LOAD1), and another one for which the yearly peak takes place in summer (LOAD2). The general quality of the fit is good: only 7% of the regressions estimated (24 $\times$ 448) show a "weak" quality of fit (R-square measures $< 75\%$). The heating temperature effect is most of the time significant: only few of the estimated HDD parameters are declared statistically insignificant. On the other hand, the cooling temperature effect is much less widespread, which is not surprising since for the moment the penetration of air-conditioning in Belgium remains fairly low and limited to specific locations. As far as the non-temperature part of the load is concerned, most of the parameters for periodicity and for trend are declared significant and hence useful for explaining the load. Figure 1 illustrates the temperature sensitivity parameters for each hour of the day, respectively for LOAD1 (winter peak) and LOAD2 (summer peak). As expected, LOAD1 is mainly characterized by heating temperature sensitivity, while LOAD2 is mainly characterized by a cooling temperature effect.

[1] Optimal must be understood as the best structure from both a statistical and applied viewpoint: one must be capable to precisely separate the temperature effect from the non-temperature sensitivity and a trade-off has also to be made between the degree of technical sophistication of the model and the resultant practical constraints.

[2] Several alternative temperature variables were tested; the final choice was based on both practical and statistical considerations.

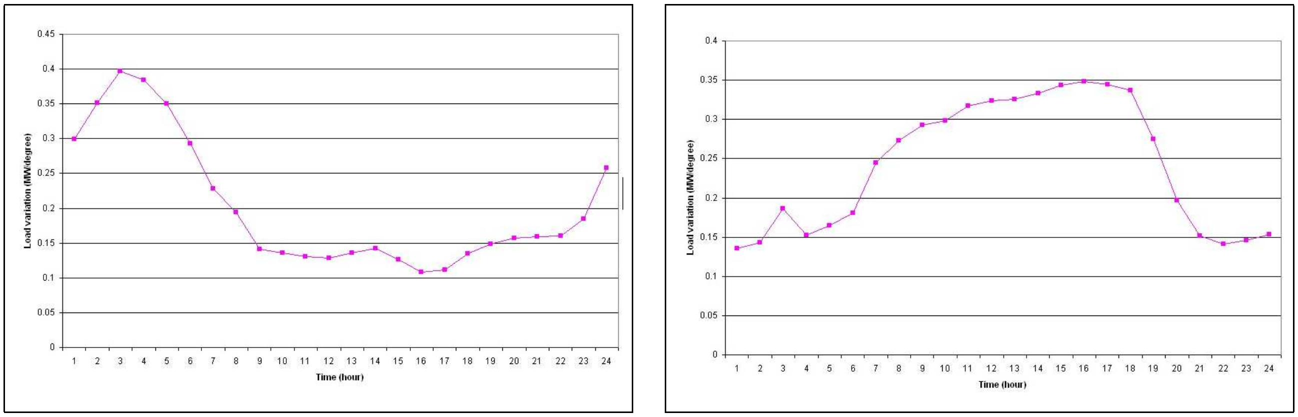

Figure 1: Heating temperature sensitivity during the day for LOAD1 (Left). Cooling temperature sensitivity during the day for LOAD2 (Right).

Temperature sensitivity coefficients differ during the day, which is the particularity of the model. Temperature indeed influences the load differently according to the hour of the day. The effect of temperature on LOAD1 is higher at night than during the day, which corresponds to the usual behavior of a load supplying places with a high penetration of storage electric heating. For LOAD2, the temperature sensitivity is the highest during the day, being typical of power consumption in locations where there is an important concentration of air-conditioning equipment.

2.4 Application: Temperature correction of Peak Load

Once the system (1) is estimated, the parameter estimates of heating degree-days or cooling degree-days can be used to adjust peak load (PL) records for their extreme temperature part. The actual procedure consists of performing the following operation (equation (2) for a winter peak correction and equation (3) for a summer peak correction):

$$\mathrm{PL}_{\mathrm{adj}} = \mathrm{PL}_{\mathrm{obs}} + \hat{\beta}_{1,h_p}(\mathrm{HDD}_{\mathrm{norm}} - \mathrm{HDD}_{\mathrm{obs}}), \quad (2)$$

$$\mathrm{PL}_{\mathrm{adj}} = \mathrm{PL}_{\mathrm{obs}} + \hat{\beta}_{2,h_p}(\mathrm{CDD}_{\mathrm{norm}} - \mathrm{CDD}_{\mathrm{obs}}), \quad (3)$$

where $\mathrm{PL}_{\mathrm{adj}}$ is the peak load corrected by temperature; $\mathrm{PL}_{\mathrm{obs}}$ is the observed peak load value (at the peak hour h_p); $\mathrm{HDD}_{\mathrm{obs}}$ (resp. $\mathrm{CDD}_{\mathrm{obs}}$) is the observed HDD (resp. CDD) on the day of the peak; $\mathrm{HDD}_{\mathrm{norm}}$ (resp. $\mathrm{CDD}_{\mathrm{norm}}$) is the median value of historical HDD (resp. CDD) for the corresponding month; $\hat{\beta}_{1,h_p}$ (resp. $\hat{\beta}_{2,h_p}$) is the estimated parameter of heating (resp. cooling) temperature sensitivity at the hour of the peak. Figure 2 gives an overview of the load behavior for LOAD1 and LOAD2, during the week of their respective annual peak (Monday to Sunday). Both observed load profile and the one adjusted to normal temperature conditions are presented.

3 Short Term Load Forecasting using Periodic Time Series

This section briefly describes the implementation of the short-term forecasting problem using Periodic Autoregression (PAR) models [18].

3.1 PAR Models, Implementation and Estimation

In simple terms, an autoregression is said to be periodic when the parameters are allowed to vary across seasons. Consider the case of a univariate time series y_t, $t = 1, \cdots, N$, (in this case, the hourly load measurements) available for a sample of $N_d = N/24$ days, corresponding to the N hours. The general form of a periodic autoregressive model of order p (PAR(p)) is:

$$y_t = C_s + \phi_{s,1} y_{t-1} + \phi_{s,2} y_{t-2} + \cdots + \phi_{s,p} y_{t-p} + \varepsilon_{s,t} \quad (4)$$

where C_s is a seasonally varying intercept term, the $\phi_{i,s}$ are the autoregressive parameters up to the order p, varying across the N_s seasons ($s = 1, 2, \cdots, N_s$). The choice of N_s depends on the frequency of the data and the seasonal pattern under scrutiny. The error term $\varepsilon_{t,s}$ can be a standard white noise with zero mean and variance σ, or it can be allowed to have a variance σ_s corresponding to seasonal heteroskedasticity. It is worth to note that (4) gives rise to a system of N_s equations that can be estimated using Ordinary Least Squares (OLS). For further details, the interested reader is referred to [19, 20, 21].

For the implementation discussed here, an approach similar to [22] is followed. Monthly and weekly seasonal patterns are modelled by dummy variables, and the intraday seasonal pattern is assumed to be captured by PAR parameters, i.e. $N_s = 24$ is the number of different "seasons" (here, hours) to be identified using the PAR model.

Denote by $y_{h,d}$ the value of the load measured in hour h of day d, with $h = 1, 2, \cdots, 24$ and $d = 1, 2, \cdots, N_d$. A formulation is built where the hourly load $y_{h,d}$ is a function of the last 48 hourly values, thus defining a PAR(48) model.

The PAR(48) model applied to the hourly load forecasting problem defines the following set of equations:

$$\begin{aligned} y_{1,d} &= C_1 + \phi_{1,1} y_{24,d-1} + \phi_{1,2} y_{23,d-1} + \cdots + \varepsilon_{1,d} \\ y_{2,d} &= C_2 + \phi_{2,1} y_{1,d} + \phi_{2,2} y_{24,d-1} + \cdots + \varepsilon_{2,d} \\ y_{3,d} &= C_3 + \phi_{3,1} y_{2,d} + \phi_{3,2} y_{1,d} + \cdots + \varepsilon_{3,d} \\ &\vdots \\ y_{24,d} &= C_{24} + \phi_{24,1} y_{23,d} + \phi_{24,2} y_{22,d} + \cdots + \varepsilon_{24,d} \end{aligned} \quad (5)$$

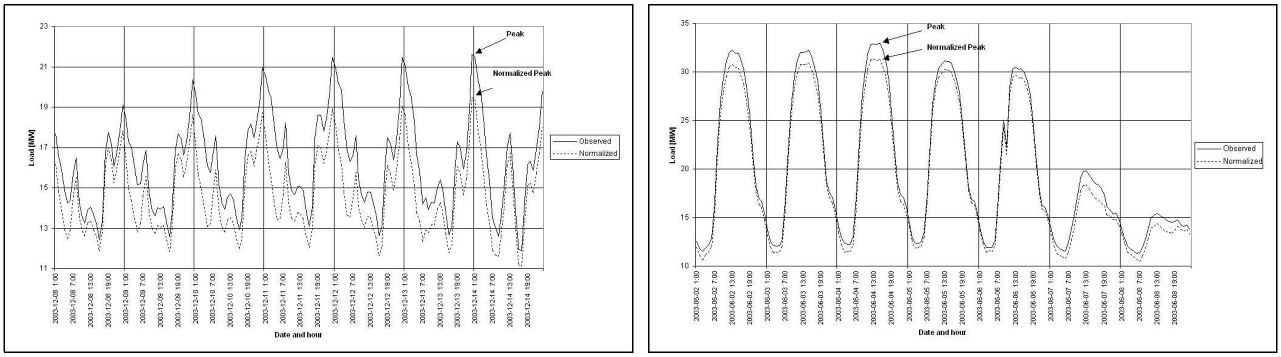

Figure 2: Load patterns (Observed and Adjusted) during the week of the yearly peak for LOAD1 (Left) and LOAD2 (Right).

This basic PAR template consists of $24 \times 49 = 1176$ parameters. It is further extended to include exogenous variables to account for temperature effects as well as monthly and weekly dummy variables. The temperature variables included in the model are heating and cooling degree-hours, computed analogously as those defined on Section 2 for degree-days. It is important to stress on that the PAR system is different from that used on Section 2, even when both are made of a set of 24 hourly load equations. In the case of the system (1), each equation uses a load for a particular hour of the day. In the case of the PAR system (5), all equations use information from *previous* hours, no matter what hour of the day it is.

3.2 Results

This methodology is applied to 245 load series, from a sample of substations containing residential, industrial and commercial customers. It is found that the model can describe quite well most substations behaviour, with only 20 substations having an adj-R^2 lower than 0.90[3]. Adding more lags and/or including more external variables may be required to improve the accuracy for these substations. The identified coefficients for all dummy and temperature variables allow for interesting comparisons [18] between substations, although not shown here because of space constraints.

In order to quantify the performance of the PAR(48) model template over the different 245 time series, the Mean Absolute Percentage Error (MAPE) and the Root-Mean Squared Error (RMSE) are computed for the following out-of-sample forecasting exercises:

- Case I: One-step-ahead prediction, for a 7 days-period.
- Case II: Iterative prediction with update every 24 hours, for a 7 days period.

Table I shows how many substations are included within each category of different error levels. Clearly, the PAR(48) model template can produce excellent forecasting results (less than 3% error in this setting) for 238 load series when working with iterative predictions with updates every hour (one-step-ahead prediction, Case I). If the update is made every 24 hours (Case II), then 166 time series have their errors below 3% . As stated above, the model performance can be improved by adding more terms into the PAR formulation. In this setting, $p = 48$ gives a satisfactory performance while keeping model parsimony.

An example of the forecasting performance is presented in Figure 3 for 3 selected substations with very different behavior. Each row represents a substation, where the left panel shows the observed load series for a period of 96 hours starting on a Sunday. The center panel shows the forecasts and confidence intervals using the "one-step-ahead" prediction mode. The observed load series (dashed line) is compared with the forecasted values (thick line). The confidence intervals are also indicated (thin lines). The right panel shows the situation for an "iterative-prediction" mode with updates every 24 hours.

Clearly the best performance is obtained when using the one-step-ahead mode, which implies an update every hour with the actual observations. The iterative-forecasting with updates every 24 hours is less optimal, but depending on the substation, it can still provide good predictions. As mentioned above, the performance for specific substations can be improved by increasing the lags in the PAR(p) model or by adding external information.

3.3 From PAR models to Typical Daily Profiles

The stationarity properties of the PAR system (5) are exploited in order to produce a Typical Daily Profile vector from which all calendar and temperature effects have been removed. Writing the model (5) in Vector-AutoRegression (VAR) form, with $\boldsymbol{Y}_d = [y_{1,d} \quad y_{2,d} \quad y_{3,d} \quad \cdots \quad y_{23,d} \quad y_{24,d}]^T$, yields

$$\Phi_0 \boldsymbol{Y}_d = \boldsymbol{C} + \Phi_1 \boldsymbol{Y}_{d-1} + \Phi_2 \boldsymbol{Y}_{d-2} + \Phi_3 \boldsymbol{X}_d + \boldsymbol{\varepsilon}_d \quad (6)$$

with the definitions of Φ_0, Φ_1, Φ_2 and Φ_3 using the coefficients ϕ of the system (5) [19]. The matrix $\boldsymbol{X}_d$ contains all exogenous variables for temperature and calendar information. The next-day forecasts can simply be written

[3]Note that the threshold of 90% here, compared to the 75% threshold considered for model (1) validation, may suggest that the latter is worse; nevertheless, this higher reference level arises from the fact that regression models with AR terms use to produce better R^2 (adjusted) than those without AR terms.

	Number of Substations for which prediction error is less than					
	< 1%	<3%	<5%	<8%	<10%	<20%
Case I						
MAPE	206	238	241	241	242	245
RMSE	201	238	241	242	242	245
Case II						
MAPE	13	166	189	205	214	245
RMSE	8	166	200	229	234	245

Table 1: Cumulative number of substations with MAPE and RMSE below a certain level for different forecasting modes.

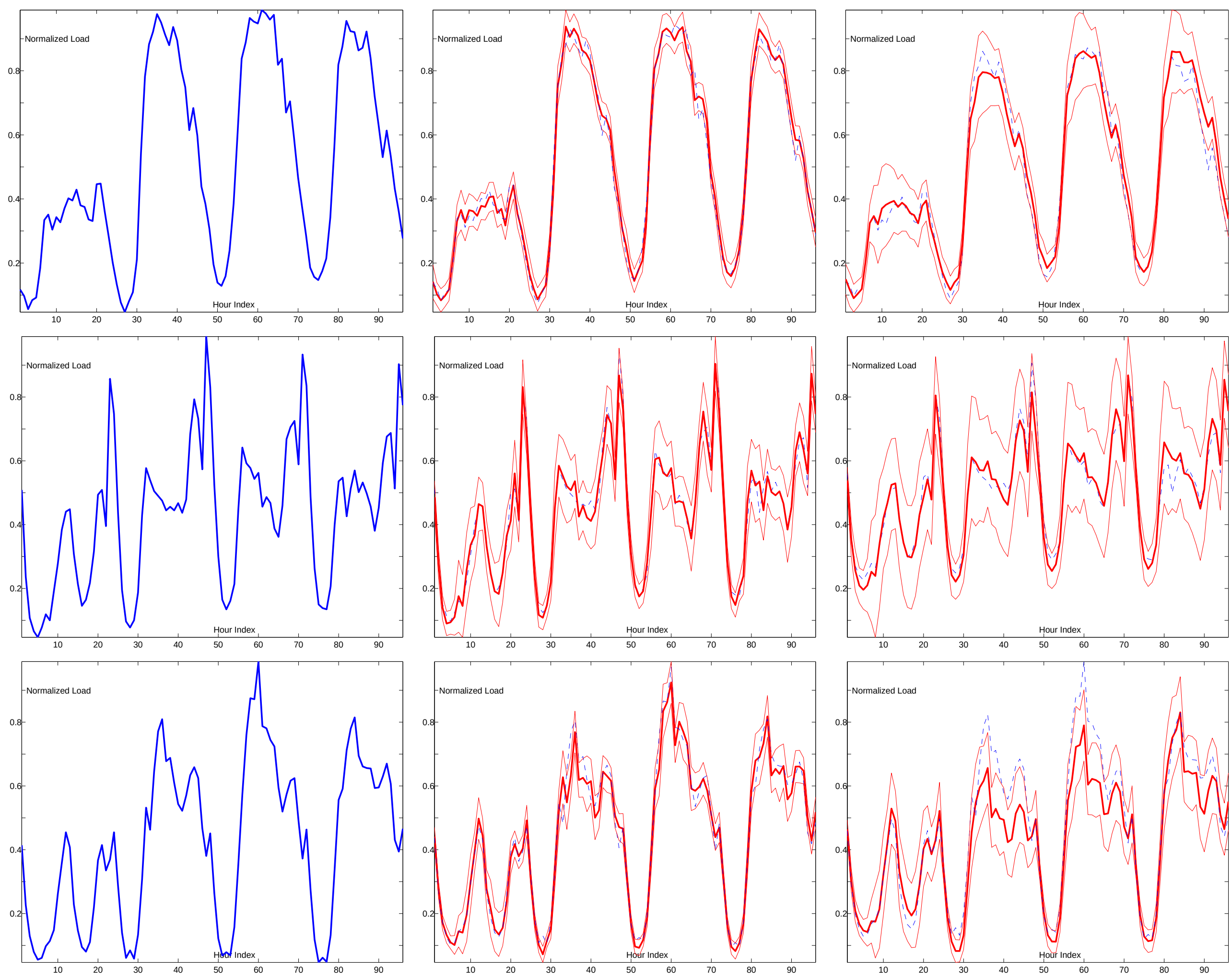

Figure 3: Out-of-sample predictions for 3 selected substations. Within each row, representing a substation, the observed load series (left) and its forecasts under different updating modes: One-step-ahead (center) and Iterative-forecasts with update after 24 hours (left). On each panel, the forecasts (thick line), the observed value (dashed) and the 95% confidence interval (thin lines) are depicted.

as

$$E_d[\boldsymbol{Y}_{d+1}] = \Phi_0^{-1}\{\boldsymbol{C} + \Phi_1 \boldsymbol{Y}_d + \Phi_2 \boldsymbol{Y}_{d-1} + \Phi_3 \boldsymbol{X}_{d+1}\}$$

where E_d is the expectation taken at time d. Now, removing all calendar and temperature effects is equivalent to setting $\boldsymbol{X}_d = 0$, where the system becomes

$$\Phi_0 \boldsymbol{Y}_d = \boldsymbol{C} + \Phi_1 \boldsymbol{Y}_{d-1} + \Phi_2 \boldsymbol{Y}_{d-2} + \boldsymbol{\varepsilon}_d, \qquad (7)$$

which is a Vector-Autoregression of order 2 (VAR(2)) with convergence vector

$$\boldsymbol{Y}^* = \{\Phi_0 - \Phi_1 - \Phi_2\}^{-1}\boldsymbol{C}. \qquad (8)$$

which exists if and only if the VAR system is stationary, i.e., all the roots of $|\Phi_0 - \Phi_1 z - \Phi_2 z^2| = 0$ are outside the unit circle.

In the dataset considered here, every load series has its own convergence vector $\boldsymbol{Y}^*$, computed from the estimated model coefficients contained in Φ_0, Φ_1, and Φ_2. As each vector $\boldsymbol{Y}_d$ includes daily information on the load, the $\boldsymbol{Y}^*$ convergence vector, computed after all seasonal effects have been removed, can be interpreted in terms of daily load information as a Typical Daily Profile (TDP). These profiles can be used for further analysis into the clustering of types of customers behind the substation level [18]. This definition of Typical Daily Profiles requires the obtained VAR system to be stationary, a condition attainable in the process of defining the order of the PAR(p) process. It is also an empirical definition, as it is based on a statistically sound procedure which is the estimation of a set of autoregressions. Finally, it is important to note that each one of the TDPs are “virtual”, in the sense that they can not be observed empirically as all observations are affected by the corresponding calendar and temperature effects for a particular date and hour.

Figure 4 shows 6 examples of Typical Daily Profiles $\boldsymbol{Y}^*$ computed from selected substations. Each one of the TDPs contains features relevant to the load at those locations, when all the seasonal and temperature effects have been removed. It is easily visualized that the daily behavior of these substations are not the same, with peaks located at different hours of the day. Using TDPs is a simple and powerful procedure for comparing the profiles of substations.

4 Conclusion

This paper presents research results and novel techniques derived from the analysis of intra-day (hourly) load records from local substations of the Belgian high-voltage grid, as provided by Elia, the Belgian National Transmission System Operator - TSO. The techniques shown in this paper are aimed at producing accurate statistical and modelling tools to TSO planners for the tasks of short-term load forecasting, customer identification and temperature adjustment for long-term forecasts.

The first technique described permits the analyst to normalize (peak) load data to average temperature conditions. A multi-equation model is developed. Each hour of the day is modelled separately in two main parts, the first capturing the temperature-related part and the second the non-temperature one. The main interest of this specification is to allow for an accurate estimate of the temperature-influenced part of the load (varying during the day), while at the same time controlling for its fundamental periodical and trend drivers. Once estimated, the temperature part coefficients could be used in order to adjust historical load records for severe temperature conditions.

The second technique described allows the analyst to produce accurate short-term forecasts, and to compare substations by identifying coefficients for temperature and calendar effects. Furthermore, the technique based on Periodic Autorregressions (PAR) allows to compute a Typical Daily Profile from the same model, thus obtaining a clean and efficient way to represent and compare substations. These Typical Daily Profiles can be used for further analysis, e.g. clustering of customer types.

Acknowledgments This work was supported by grants and projects for the Research Council K.U.Leuven (GOA-Mefisto 666, GOA-Ambiorics, several PhD/Postdocs & fellow grants), the Flemish Government (FWO: PhD/Postdocs grants, projects G.0240.99, G.0407.02, G.0197.02 (power islands), G.0141.03, G.0491.03, G.0120.03, G.0452.04, G.0499.04, ICCoS, ANMMM; AWI;IWT:PhD grants, GBOU(McKnow)Soft4s), the Belgian Federal Government (Belgian Federal Science Policy Office: IUAP V-22; PODO-II (CP/01/40), the EU(FP5-Quprodis;ERNSI, Eureka 2063-Impact;Eureka 2419-FLiTE) and Contracts Research/Agreements (ISMC/IPCOS,Data4s,TML,Elia,LMS,IPCOS,Mastercard). B. De Moor and R. Belmans are full professors at the K.U.Leuven, Belgium. The scientific responsibility is assumed by its authors.

REFERENCES

[1] R. Ramanathan, R. Engle, C. Granger, and C. Vahid-Aragui, F.and Brace, “Short-run forecasts of electricity load and peaks,” *International Journal of Forecasting*, no. 13, pp. 161–174, 1997.

[2] E. Mariani and S. Murthy, *Advanced Load Dispatch for Power Systems*, ser. Advances in Industrial Control. Springer-Verlag, 1997.

[3] N. Amjady, “Short-term hourly load forecasting using time-series modeling with peak load estimation capability,” *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 798–805, 2001.

[4] S.-J. Huang and K.-R. Shih, “Short term load forecasting via ARMA model identification including non-gaussian process considerations,” *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 673–679, 2003.

[5] H. Steinherz, C. Pedreira, and R. Castro, “Neural networks for short-term load forecasting: A review and evaluation,” *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, 2001.

[6] A. Khotanzad, R. Afkhami-Rohani, and D. Maratukulam, “ANNSTLF-artificial neural network short-term load forecaster-generation three,” *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1413–1422, 1998.

[7] K.-H. Kim, H.-S. Youn, and Y.-C. Kang, “Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method,” *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 559–565, 2000.

[8] L. Mohan Saini and M. Kumar Soni, “Artificial neural network-based peak load forecasting using conjugate gradient methods,” *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 907–912, 2002.

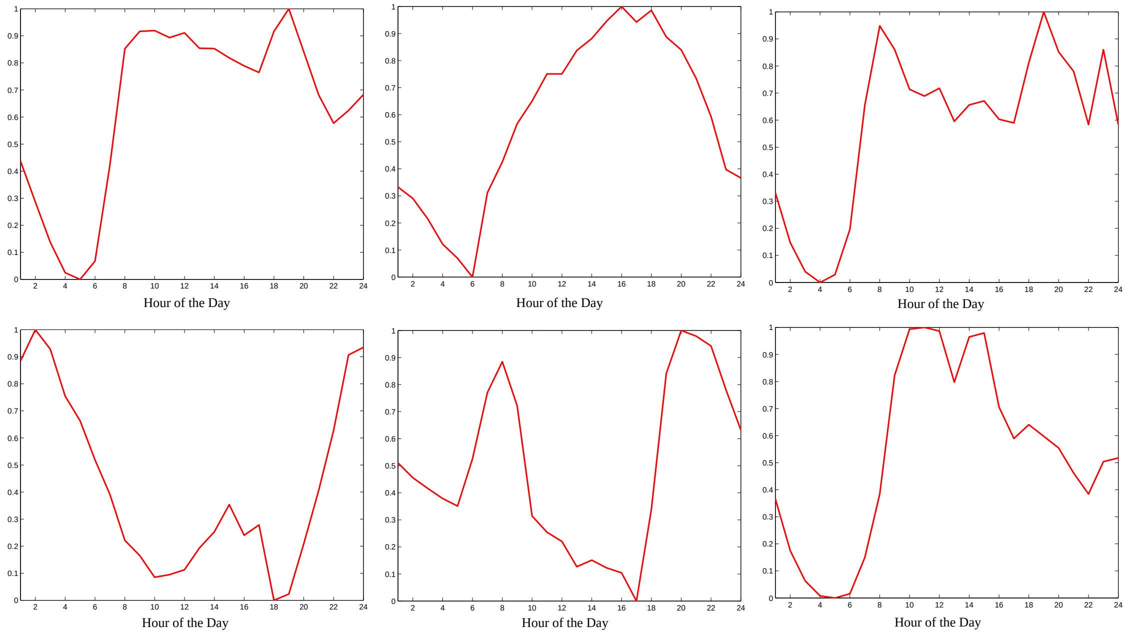

Figure 4: A set of 6 Typical Daily Profiles computed from selected substations. The scale has been normalized, in order to compare shapes rather than magnitudes.

[9] D. Fay, J. Ringwood, M. Condon, and M. Kelly, "24-h electrical load data-a sequential or partitioned time series?" *Neurocomputing*, no. 55, pp. 469–498, 2003.

[10] H. Liao and D. Niebur, "Load profile estimation in electric transmission networks using independent component analysis," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 707–715, 2003.

[11] S. Heunis and R. Herman, "A probabilistic model for residential consumer loads," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 621–625, 2002.

[12] J. Jardini, C. Tahan, M. Gouvea, and S. Ahn, "Daily load profiles for residential, commercial and industrial low voltage consumers," *IEEE Trans. Power Delivery*, vol. 15, no. 1, pp. 375–380, 2000.

[13] E. Carpaneto, G. Chicco, R. Napoli, and M. Scutariu, "Customer classification by means of harmonic representation of distinguishing features." IEEE Bologna Power Tech Conference, 2003.

[14] D. Bunn, "Forecasting load and prices in competitive power markets," *Proceedings of the IEEE*, vol. 2, no. 88, pp. 163–169, 2000, invited Paper.

[15] H. Willis, *Spatial Electric Load Forecasting - Second Edition*. Marcel Dekker, Inc., 2002.

[16] R. Engle, C. Granger, J. Rice, and A. Weiss, "Semiparametric estimates of the relation between weather and electricity sales," *Journal of the American Statistical Association*, vol. 81, no. 394, pp. 310–320, 1986.

[17] J. Hamilton, *Time Series Analysis*. Princeton University Press, 1994.

[18] M. Espinoza, C. Joye, R. Belmans, and B. De Moor, "Short term load forecasting, profile identification and customer segmentation: A methodology based on periodic time series," *IEEE Trans. Power Syst.*, to appear.

[19] P. Franses and R. Paap, *Periodic Time Series Models*. Oxford University Press, 2003.

[20] B. Troutman, "Some results in periodic autoregressions," *Biometrika*, vol. 66, pp. 219–228, 1979.

[21] A. McLeod, "Diagnostic checking of periodic autoregression models with applications," *The Journal of Time Series Analysis*, vol. 15, no. 2, pp. 221–223, 1994.

[22] G. Guthrie and S. Videbeck, "High frequency electricity spot price dynamics: An intra-day markets approach," New Zealand Institute for the Study of Competition and Regulation, Tech. Rep., 2002.