Sujet de TER 2018-2019

Electricity load forecasting

jairo.cugliari@univ-lyon2.fr

Context

Electricity load forecasting is a necessary task permitting utilities to anticipate needs in terms of electrical load. With this, system operators take decision on how the electricity should be produced and transmitted over the electrical grid.

Objective

In this project, the students will study the problem of electrical load forecasting using data either from the Ukrainian system (if available) either from the Uruguayan system.

A number of prediction methods have been implemented, for instance in the R packages enercast, prophet or forecast. The aim of the project is to set a fully functional web environment based on shiny in order to make the prediction task as automatic as possible.

https://shiny.rstudio.com/

Wiley StatsRef:
Statistics Reference Online

DOI: 10.1002/ISBN.stat00999.pub9

Electricity Demand Forecasting

Jairo Cugliari[a]*and Jean-Michel Poggi[b]

Abstract: In this article, the focus is short-term demand forecasting at some aggregate level (e.g. zone or nation wide demands) from data with at least hourly sampled data. The main salient features of the load curve are first highlighted. Some of the common covariates used in the prediction task are also discussed. Then some basic or now classical methodological approaches for electricity demand forecasting are detailed. Finally, a presentation of new and open problems is postponed to the last section.

1. Introduction

Unlike the most common demand forecasting cases, electricity cannot be stored, or it is poorly done. Thus, injections to and subtractions from the electrical grid must be balanced. Traditionally the equilibrium was assessed by monopolistic and government-controlled power sectors that were responsible of the whole electrical chain. But with the opening to competitive electricity markets a plethora of participants on the market (e.g. utilities, sellers) now share this equilibrium responsibility in order to avoid blackouts and financial penalties. The statistical forecast of load demand is crucial and several forecast horizons are useful: from the short-term (up to a few days) to the long-term (several years) through the medium-term (a week to a year). The last two are needed for sizing, managing and maintaining the network as well as the production units. The first one aims to optimize the load to be produced the next day, make the ultimate purchases and call for customers' erasure among those benefiting from discounts against the possibility of cutting their consumption a few days a year.

To this purpose, the main inputs are the historical load levels of the electrical system as well as some important exogenous information. Past records are usually available at different time resolutions and may come from different and homogeneous sources (e.g. from production units, substations, smart meters). The electrical load plot as a time function is usually called the load curve.

Several technical and scientific communities participate in the forecasting task, including engineers, economists, statisticians and computer scientists, which yields on a wide range of methodological approaches to predict the load. Figure 1 summarizes some of the common terms. The joint objective is to produce an accurate anticipation of the

[a] **Université Lumière Lyon2, Lab ERIC**
[b] **LMO, Université Paris Sud**
***Email: Jairo.Cugliari@univ-lyon2.fr,Jean-Michel.Poggi@math.u-psud.fr**

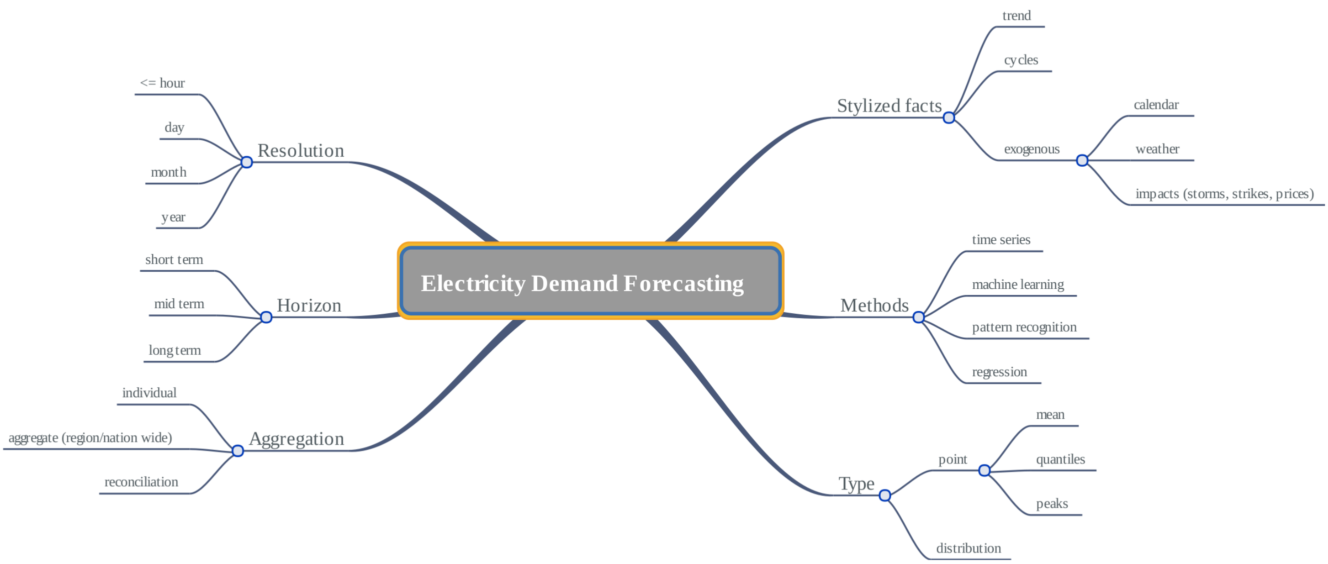

Figure 1. Common terms on Electricity Demand Forecasting (abbreviated EDF).

load needed at some aggregation level and at some resolution level. Predictions can concern both point loads (e.g. the maximum, median or mean load) and the whole prediction law (or at least some prediction interval).

In what follows, the focus is short-term demand forecasting (STDF) at some aggregate level (e.g. zone or nation wide demands) from data with at least hourly sampled data. The main salient features of the load curve are highlighted in Section 2. Some of the common covariates used in the prediction task are also discussed. Section 3 presents basic or now classical methodological approaches for electricity demand forecasting (EDF). A presentation of new and open problems is postponed to the last section.

2. Salient features

While the specific data features ultimately depend on the particularities of each electrical system and market (i.e. number of clients, degree of industrialization, repartition between household and industrial consumers) some common conclusions can be drawn. We have chosen to illustrate these features with data from the Uruguayan market[1] which is a relatively small system with a low industrialized economy.

Figure 2 represents different views of the data. Let us first concentrate on the leftmost plot. On the left, three time series show the evolution of mean and standard deviation of load as well as the mean temperature, all three on a weekly resolution level. Vertical lines separate years. The first element to analyse is the global trend of the load demand (upper panel). Since residential use explains a major part of the electrical consumption in Uruguay, most of the causes that explicate the features encountered come from household usage of electricity. The time series present a clear upwards long-term trend on the 8-year period. Two known causes of this trend are population increase and the electrical devices intensification usages. Different cycles are also present. Annual cycles reflect the seasonality induced by both economical activity and meteorological phenomena. However, this annual pattern changes through the span of the analysis window. While in the first years only one high demand period is clearly present, on the last years

[1]Data kindly provided by the TSO at *Administración Nacional de Usinas y Trasmisiones Eléctricas* (UTE), `portal.ute.com.uy`

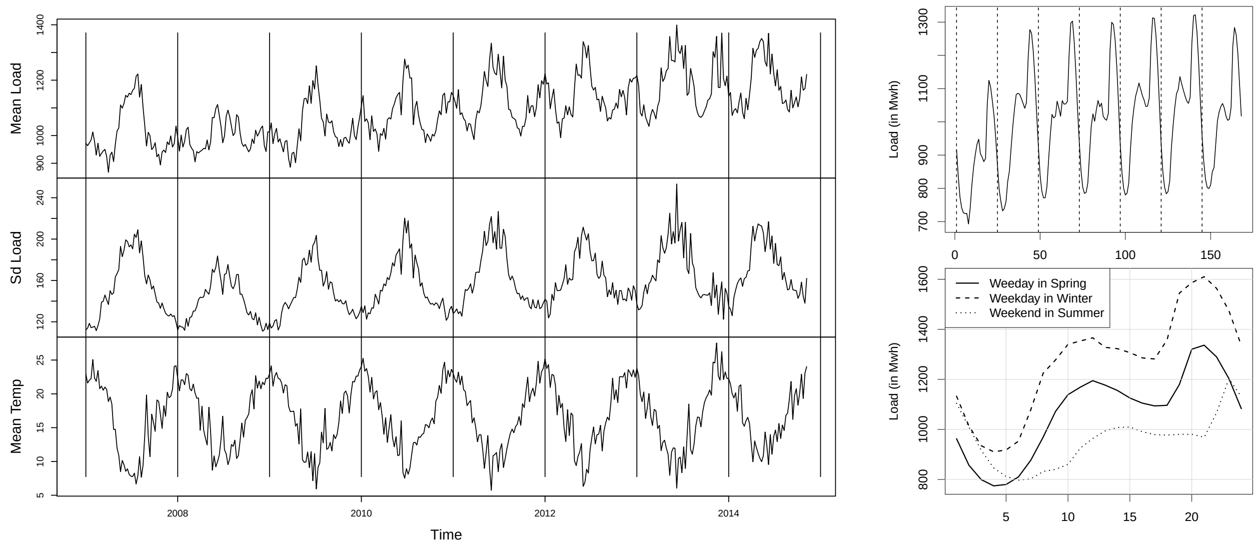

Figure 2. On the left: weekly mean load (upper panel), load standard deviation (middle pannel) and mean temperature (bottom panel) for the Uruguayan electrical system. On the top right: a weekly load curve beginning on Sunday. On the bottom right: three daily load curves. Load is measured in Mwh and temperature in Celsius degrees.

two important modes are present. This shift from unimodal to bimodal pattern is explained by the popularization of electrical cooling systems which has a major impact, for instance on maintenance schedules. This shows the importance of having data analysis tools that can adapt to a changing structure of the electricity demand or consider only shorter and more recent datasets.

The link between temperature and demand can be quite complex. The bottom level of Figure 2 shows the annual cycle of the temperature time series as well as the absence of a long-term trend. A relation is easily marked between temperature and demand. High and low temperatures, at least beyond some threshold, produce an increase of the demand due to the use of cooling and heating electrical systems. However, the link is clearly stronger if we look at the variability of the daily electrical demand. For this, the middle left panel of Figure 2 represents the daily standard deviation of the demand. We can see that an annual cycle is present and matches with the daily mean temperature with an inverse association. This is to say that the lower the temperatures, the higher the variability of the electrical demand.

On the top right a weekly pattern is shown, beginning on a Sunday. The demand is clearly higher from Mondays to Fridays and lower during the weekends. The economic structure induces its cycle of working and non-working days on the demand. Other special days like bank holidays, School vacation or extra days off may also induce special consumption patterns.

On the bottom right, we represent examples of the variability which can be observed on daily load patterns. In general, load levels are lower at night and higher during day. The morning ramp shows the pick-up load on the system, which varies from a sharpen rise in cold workable days to very slow gradual rise on summer weekends. The highest demand is observed during evenings. The peak load shifts latter or earlier in the day with the sunlight cycle of the year. These daily features can be explained by the social and economical behaviour of customers. The amplitude and position of these features are important landmarks of the load curve.

Among the external factors affecting the load demand, the meteorological ones are the most commonly used for EDF. It is usually said that the electrical demand is thermosensitive, that is that it responds to variations on temperature. Electrical heating and cooling of buildings are a large part of the demand. A way to look at the thermosensitivity of the load demand is to represent the demand as a function of the temperature as done in Figure 3. The temperature gradients, i.e. the slope of these plots, show how the dependence can be complex to model. The dependence structure is affected by the the position on the year, the type of day, and the hour of the day.

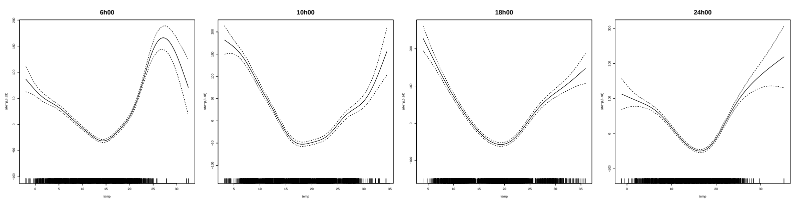

Figure 3. Temperature gradients estimated at different hours.

Other meteorological variables such as cloud coverage which may have an impact on building or street lightening can be of interest. Combined with wind speed and direction, humidity and dew point, they may alter the real feel of the temperature. Developments on meteorological services that provides forecast of these variables permit their incorporation on forecasting methods. It is worth to mention that one usually uses weather predictions only to obtain EDF but not to train or estimate the models used to obtain those forecasts.

3. Methods

Though the methods used to build models to forecast the peak load of the day, the total load of the day, the hourly load or the sequence of hourly (or half- hourly) load during a day could be different, it appears that they are essentially the same, so we prefer not to specify the description of a particular forecasting objective (see Alfares & Nazeeruddin (2002)). Similarly, the forecasting horizon impacts the choice of predictors and the metrics used to design or, more often, to assess the models, but not really the methods involved. We will then focus on this section on STDF or middle-term demand forecast (MTDF) for the a highly to moderate aggregate load curve.

Structure of a typical model. In general, one of the key points is to use some time series decomposition approaches to define several components, each component representing different kinds of pattern (see Figure 2). According to the nature of the considered load series, possible components are long term trend, yearly, weekly, daily cycles, temperature effect, calendar events, and so on. So, a generic model for STDF could be written as follows:

$$Y_{t+1} = F(Y_t, Y_{t-k}, \dots, Y_{t-K}; X_t, X_{t-k'}, \dots, X_{t-K'}; C_t) + \epsilon(t) \qquad (1)$$

where F is general and distinguishes three components: endogenous variables (instantaneous and lagged values of Y), exogenous ones related to meteorology (X) and calendar effects (C). As consumption habits depend hardly on the hour of the day, very often one model per instant is fitted. For instance, if the data are measured every 30 minutes and if we are interested in forecasting the next day curve, we then fit 48 models corresponding to each of the sampled instants of the day.

Except endogenous variables, useful for STDF, meteorological covariates are crucial due to their large influence on electricity demand (see Figure 3) and they are generally selected from a dictionary of tens of covariates constructed from the initial ones and considering simple transformations as the daily minimum, maximum and mean. Smoothed climate variables could also play an important role in electricity load model. To fix ideas, let us consider a model with a function F supposed to be additive:

$$\begin{aligned} Y_t = &\sum_{k=1}^{K} f_{0,k}(Y_{t-k\cdot 48}) + f_1(\mathrm{DayType}_t, \mathrm{Offset}_t) + \sum_{i=1}^{12} f_{2,i}(T_{t-12})\mathbb{1}_{M_t=i} + f_3(\mathrm{ToY}_t) + f_4(t_{15}) + f_5(t) \\ &+ f_6(\mathrm{Cloud}_t) + f_7(T_t) + f_8(W_t) + f_9(\theta_t) + f_{10}(\theta_t^{Min}) + f_{11}(\theta_t^{Max}) + \epsilon_t, \end{aligned} \tag{2}$$

where at time t, Y_t is the electric demand, ToY_t is the time of the year of observation t, $\mathrm{DayType}_t$ and Offset_t are categorical variables indicating the type of day and the daylight saving time, M_t is the Month, several lagged and smoothed variables related to temperature T_t and θ_t an exponential smoothing of T_t, Cloud_t and W_t are the cloud cover and the wind; and $t_{15} = t\mathbb{1}_{T_t \leq 15}$ estimating a heating trend.

Methods to design the forecasting model. A lot of methods are available to build prediction models (see Weron (2007)). The most classical models, including those of the SARIMAX family (an extension of the ARMA processes), constitute an important baseline and are a favourite time series model. Corresponding to F linear in Equation (1), they lead to simple and explicit models allowing to easily interpret the role and importance of each of the predictors involved. They can achieve excellent results even if the price to pay is sometimes the complexity (a lot of parameters to estimate) and a certain difficulty to be adaptive. An interesting and conceptually simple extension of these models is to consider additive non linear models, an example is given by Equation (2). Of course, today black box-type models forgetting the interpretation of the role of variables in favor of the sole objective of forecasting are particularly in vogue, thanks to the machine learning era. Several types of methods coexist, but one that is most frequently used is undoubtedly neural networks (see Park et al. (1991)), mainly used by engineers and computer scientists, with its last (complex) avatar: deep learning methods. Finally, let us mention other forecasting approaches, out of the scope of this paper focused on statistical methods, like for instance expert system based ones.

To be a little bit more specific, let us focus on methods for forecasting the national consumption developed recently by the main French electricity provider (*Electricité de France*). Models based on regression analysis make possible to use prior knowledge to model the internal structure of consumption and dependence on exogenous factors. In Bruhns et al. (2005) this dependence is modeled by a non-linear regression on the temperature that depends on the month, the day of the week and the time of day. This model is the baseline model and reaches a mean absolute percentage error (MAPE) of 1.5%. A nonparametric version of this approach, similar to the one given in Equation (2), is proposed in Pierrot & Goude (2011), improving performance and adaptability. Another idea is to consider the daily multivariate time series model by 24 or 48 regression equations, which leads to too many coefficients possibly varying across equations and over time. In this case, a state-space factor model is developed imposing common dynamic factors for these parameters that drive the dynamics across different equations (see Dordonnat et al. (2012)). An innovative application of machine learning methods is proposed by Devaine et al. (2013) and Gaillard & Goude (2015) using the ideas of the sequential prediction of arbitrary sequences based on specialized experts. The principle is to combine different models, designed using various hypotheses and historical data, leading to elegant and robust algorithms improving performances. Finally, a Bayesian approach is proposed in Launay et al. (2015) to transfer knowledge at different scales or perimeters. The idea is to fit a parametric model on a short historical dataset in which the lack of data is compensated by the construction of an informative hierarchical Bayesian prior based on another longer dataset assumed to be quite similar to the original one.

Extensions. The type of models to consider is also influenced by the nature of the objective: point or distribution forecasting. So, let us examine two possible extensions of point forecast of consumption. First, the traditional problem of the distribution of this forecast by allowing uncertainty to be taken into account, is of interest and can be tackled explicitly (for ARMA-type models for example) or implicitly through bootstrap schemes in nonparametric cases. A second problem, less classical, is density forecasting, especially useful for medium and long term peak electricity demand forecasting, providing estimates of the full probability distributions of the future values of the electricity demand (see Hyndman & Fan (2010)).

Another interesting extension is forecasting the shape of the load curve which can be taken into account as a whole. This problem is essential because one can easily see the limits of a naive approach, still often used, to consider independently each moment of discretization and thus to ignore the functional character of the data. While this strategy may remain relatively efficient for aggregate consumption, it is unable to adapt it to new, less aggregated, less stationary, more uncertain data. Other approaches are possible, for example, to reduce the functional problem to a set of uncorrelated problems by means of a Principal Component Analysis or to perform a projection on a basis independent of the data such as the wavelet basis that allows us to represent the shapes of the consumption curves.

4. New topics

At a aggregated level, demand can be roughly approximated by a long-term trend plus some seasonal components due to the interaction of economic and social activities and climate. However, over the recent years, three major developments have affected the structure of consumption and thus complicated the forecasting task. First, competition in the market increases gains and losses of customers for the supplier portfolio. Then, the usages change rapidly, and this is accelerated by the energy transition: low energy bulbs, air conditioners, flat screens, smartphones, better insulated buildings, electric cars. Finally, new measurement technologies structure intelligent networks, with more local and high resolution information. The installation of smart meters makes available, in real time, individual load curves, and thus of all kinds of aggregates to be defined according to a geographical, electrical or socio-demographic point of view, opening new perspectives for the control of electricity demand on various scales. The first two developments lead to lack of stationarity, greater volatility in aggregated consumption and the necessity to consider more flexible strategies to adapt to changes. The latest development is a more radical challenge, which requires not only other methods and their adaptation to the Big Data context, but also their complete automation.

These new data present all the characteristics of the Big Data: massive to be stored, transmitted and processed in real time, and they are of quality not yet quantified and confidentiality is mandatory as for any individual data. Users could control their consumption and suppliers could define dynamic pricing and incentives to better manage daily consumption, better manage consumption erasures, integrate local and distributed production of renewable energies and improve global forecasts at regional and national levels. The operational models of consumption forecasting, if they achieve excellent performances for aggregated consumption, regional or national (in France for example) are complex and depend on many parameters, which are not very adaptive or difficult to adapt. Methods, both more general and more flexible, are needed. For example, at the individual level, one can develop tools for predicting the erasure deposit in residential customers. At a very local level, but already aggregated, the automation of the design of several thousand models linked to different scales of aggregation of individual consumption requires the development of new methods of variable selection. Another example is to build forecast by looking in the past for contexts similar to the present situation and forecasting the future by a weighted average of their futures that are most similar to the present. The notion of similarity is based on the division of the time series into characteristic segments, such as the consumption of a day, assimilated to functions, making it possible to adapt to both classical non-stationarities and more subtle ones resulting from losses or customer gains under various scenarios. A last example is the disaggregation of national or regional consumption by grouping, again by means of a notion of similarity, the thousands or millions of individual consumptions that are most homogeneous to define a few thousand super-consumers. We then construct a hierarchy of partitions among which we select the best one for the disaggregated prediction (see Cugliari et al. (2016)).

Some other new topics could have some significant impact, like the increase of renewables, even if this new challenge is much more related to electricity supply forecasting, it has also some very hard problems from the demand side like future trends in smart grids, microgrids and smart buildings (see Hernandez et al. (2014)). It is remarkable that new and classical tools, namely classification or clustering methods together with sophisticated variable selection or nonparametric methods may be combined to tackle the new problems and to design automatic systems of electricity demand forecasting. Automatic approaches are necessary to deal with new data granularity, new frequency rate, new spatial resolutions and their associated consequences: instability, irregularity, sensitivity and data quality issues.

References

Alfares, HK & Nazeeruddin, M (2002), 'Electric load forecasting: Literature survey and classification of methods,' *International Journal of Systems Science*, **33**(1), pp. 23–34, doi:10.1080/00207720110067421.

Bruhns, A, Deurveilher, G & Roy, JS (2005), 'A non-linear regression model for mid-term load forecasting and improvements in seasonality,' in *Proceedings of the 15th Power Systems Computation Conference*, pp. 22–26.

Cugliari, J, Goude, Y & Poggi, JM (2016), 'Disaggregated electricity forecasting using wavelet-based clustering of individual consumers,' in *2016 IEEE International Energy Conference (ENERGYCON)*, pp. 1–6, doi:10.1109/ENERGYCON.2016.7514087.

Devaine, M, Gaillard, P, Goude, Y & Stoltz, G (2013), 'Forecasting electricity consumption by aggregating specialized experts,' *Machine Learning*, **90**(2), pp. 231–260.

Dordonnat, V, Koopman, S & Ooms, M (2012), 'Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling,' *Computational Statistics & Data Analysis*, **56**(11), pp. 3134–3152.

Gaillard, P & Goude, Y (2015), 'Forecasting electricity consumption by aggregating experts; how to design a good set of experts,' in *Modeling and Stochastic Learning for Forecasting in High Dimensions*, Springer, pp. 95–115.

Hernandez, L, Baladron, C, Aguiar, JM, Carro, B, Sanchez-Esguevillas, AJ, Lloret, J & Massana, J (2014), 'A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings,' *IEEE Communications Surveys Tutorials*, **16**(3), pp. 1460–1495, doi:10.1109/SURV.2014.032014.00094.

Hyndman, RJ & Fan, S (2010), 'Density forecasting for long-term peak electricity demand,' *IEEE Transactions on Power Systems*, **25**(2), pp. 1142–1153, doi:10.1109/TPWRS.2009.2036017.

Launay, T, Philippe, A & Lamarche, S (2015), 'Construction of an informative hierarchical prior for a small sample with the help of historical data and application to electricity load forecasting,' *Test*, **24**(2), pp. 361–385.

Park, DC, El-Sharkawi, MA, Marks, RJ, Atlas, LE & Damborg, MJ (1991), 'Electric load forecasting using an artificial neural network,' *IEEE Transactions on Power Systems*, **6**(2), pp. 442–449, doi:10.1109/59.76685.

Pierrot, A & Goude, Y (2011), 'Short-term electricity load forecasting with generalized additive models,' in *Proceedings of ISAP power*, pp. 593–600.

Weron, R (2007), *Modeling and forecasting electricity loads and prices: A statistical approach*, vol. 403, John Wiley & Sons.

Related Articles

Forecasting, Forecasting, Business Forecasting Methods, Bayesian Forecasting, Time Series Analysis, Prediction and Forecasting, Computer codes for electrical consumption, Extreme events modelling in energy companies

Further Reading