Electricity Demand Forecasting: the Uruguayan Case

Andrés Castrillejo   Jairo Cugliari   Fernando Massa   Ignacio Ramirez

May 14, 2018

1 Introduction

When the balance between electrical supply and demand is broken, the resulting grid damage or supply outages may affect the quality of the service. Electricity load anticipation is therefore an important task for the transmission system operator of an electrical grid as it helps to reduce the risk of this happening; larger prediction horizons (greater than one year) help to anticipate the needs on production means and distribution, while shorter ones (hours, weeks) are employed to decide the production and distribution plans. In general, more accurate predictions result in lower production costs.

Recent years have seen a diverse array of new technologies which allow generation of electrical energy from new renewable sources (*renewables* for short). This diversity increases the complexity of the scenario to be managed by the agent or agents responsible of the equilibrium balance on the electrical grid. Such is the case of Uruguay, whose plan for minimizing its dependence on fossil fuel has led to a very rapid increase in power generation from a wide array of renewables DNETN [2008]. Such complex scenario calls for new, sophisticated methods for power management.

A comprehensive survey on traditional electricity load forecasting methods can be found in Weron [2006]; the reader is further referred to [Cancelo et al., 2008] and [Taylor et al., 2006] for an extensions and updates of the methods described therein. Broadly speaking, forecasting methods employ statistical methods for capturing the salient features of the load demand (we discuss them in Section 2), the main difference lying in the underlying hypothesis assumed in each case for justifying and providing validity to the proposed models. (Clearly, any systematic departure of the data from the assumed hypothesis results in a degradation in prediction performance.) We classify these methods into four large groups: time series analysis, pattern methods, regression analysis, and machine learning. We elaborate on such methods below.

Time series analysis methods form their prediction by combining past load information using linear models. Typical approaches are for example ARIMA (Auto Regressive Integrated Moving Average, see for example Cancelo et al. [2008], Lanzilotta and Rodriguez [2014] or Taylor et al. [2006]. An important drawback of such methods is their rigidity (that is, non-adaptability), leading to large prediction errors during highly unstable load demand periods. Such issues are alleviated for example by the use of heavy tailed error models (instead of the traditional Gaussian noise assumption) Weron et al. [2001], or allowing for adaptive linear parameters and *state space models* as in Dordonnat et al. [2008].

Pattern methods are based on self-similarity within the data, that is, they assume that future load curves are well modeled as a combination of previously observed patterns in the data. A typical embodiment of this idea is given in Poggi [1994], where predictions result from the weighted mean of past loads, the weights given by a similarity measure between the last observed time frame (e.g., the last observed daily curve) and all past daily curves in the recorded past. In Antoniadis et al. [2012, 2014], daily curves are modeled as continuous functions, and tools from functional theory are used to determine their similarity to past daily curves. A variant of the previous idea is to look for similarities not in the whole past but within a small *dictionary* of typical patterns which combine past curves and exogenous covariates Mougeot et al. [2015].

Regression analysis methods predict the future load one sample at a time by decomposing the load curves into various explicative effects, usually driven by exogenous factors. For instance, the predictions obtained in Bruhns et al. [2005], Cancelo et al. [2008], Soares and Medeiros [2008] are the result of a non-linear regression involving two components: one due to meteorological conditions and the other due to calendar structure. Predictions obtained with this approach can be very accurate but suffer from the same rigidity issues as those observed in time series methods. Various strategies have been proposed to alleviate such limitations. Examples include adaptive regression parameters Lloyd [2014], generalized additive models [Nedellec et al., 2014], and Bayesian models [Launay et al., 2012].

Machine learning techniques usually employ models which are generally more flexible, albeit less interpretable, than the previously described ones. Examples include [Lloyd, 2014] which uses *kernel methods* combined with Gaussian processes, and [Ben Taieb and Hyndman, 2014] which employs *gradient boosting*. See the aforementioned references for the description of such techniques.

A very successful strategy derived from this field is the idea of combining the output of several predictors, often of different type (i.e., different algorithms may be used to obtain the different predictions), to produce the *final* prediction of the system [Devaine et al., 2013]. This is known as *experts aggregation*; this is one of the methods we adopt in our framework, as will be seen later.

The goal of this work is to obtain a forecasting framework that can be used to anticipate the load needs of the Uruguayan grid. In particular, we need to account for the fact that meteorological data is scarce in the Uruguayan scenario. Our work is thus focused on three models, two of which do not rely on meteorological data.

The reminder of this document is structured as follows. Section 2 describes both the electrical and the meteorological data from the Uruguayan system. Section 3 discusses the models we adopt. A series of experiments illustrate the resulting prediction framework in Section 4. A short discussion in Section 5 concludes the work.

2   Data

This section explores the typical features encountered in Uruguayan demand data. The electricity demand data was kindly provided by *Administración Nacional de Usinas y Trasmisiones Eléctricas* (UTE)[1]; the dataset in question contains hourly electrical load
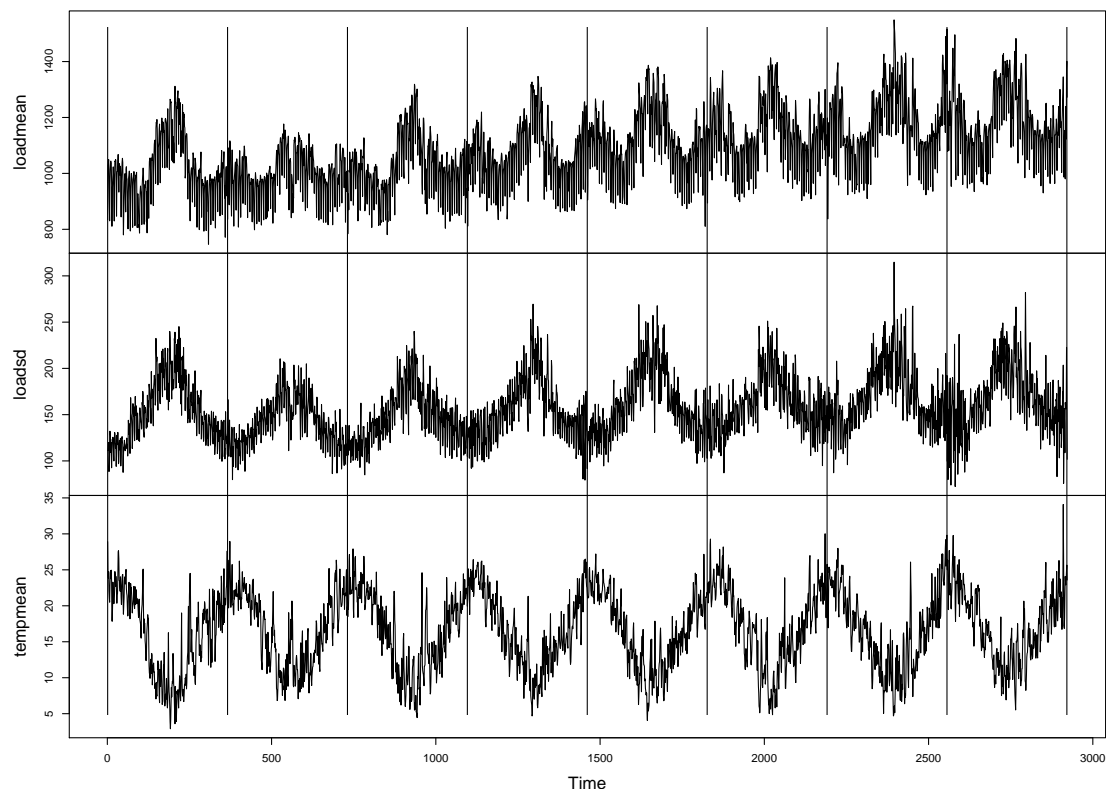
---

[1] `portal.ute.com.uy/`

Figure 1: Mean (above) and standard deviation (middle) of daily load, and mean temperature (bottom) for the Uruguayan electrical system; vertical lines separate years.

measurements between January 1st 2007 and 31 December 31st 2014. Meteorological data covering the same time period, with the same sampling frequency, was kindly provided by the *Instituto Uruguayo de Meteorologia* (INUMET)[2].

Figure 1 aggregates the three available curves (daily load mean, daily load standard deviation and daily mean temperature). We first focus on load demand (upper panel). As residential electricity consumption represents a major portion of the total Uruguayan demand, most salient features in the curves are derived from domestic human activity patterns. First of all, a clear upwards long-term trend can be observed which is linked to population increase and increasing use of electrical household devices, in particular high powered ones. Them a number of cyclical components can also be observed. The annual cycle reflects the seasonality induced by both economical activity and meteorological phenomena. It is important to note that this annual pattern has gradually evolved across the years. While the first years are characterized by one strong mode in winter, later years show another important mode in summer, which is due largely to the recent widespread adoption of electrical cooling systems. Note that this change has a major impact in electricity supply planning, for instance on maintenance schedules. The latter feature shows the importance of having models capable of adapting themselves to major dynamic changes in the structure of the demand.

Being the domestic demand a major portion of the total, and given that cooling and
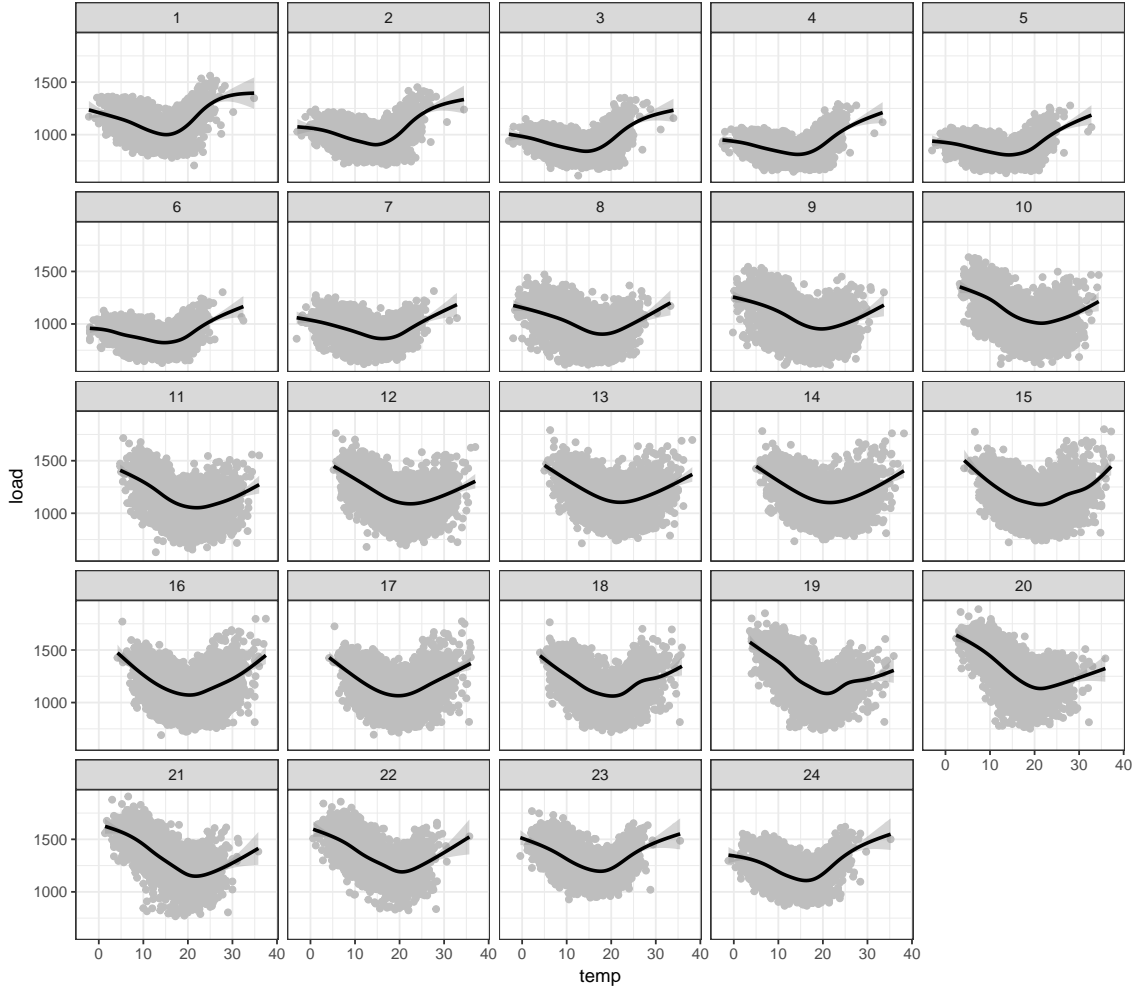
---

[2]`inumet.com.uy/`

3

Figure 2: Temperature gradients estimated at different hours. Each panel represents the scatter of the load (in Mwh) as a function of the temperature (in Celsius degrees) at one specific hour of the day. A smooth non parametric estimation of the link is added as a black line.

heating devices usually require high power to operate, it is to be expected that the temperature (bottom curve of Figure. 1) has a strong effect in the load curves. This dependency is however more complex than what one might observe at a first glance. For instance, air conditioning devices are triggered on by extreme temperatures, either low or high; this is more evident at the daily demand scale. However, this is not a simple matter of thresholds, and this is evidenced by the marked increase in standard deviation (middle curve of Fig. 1) during cold seasons.

Further insight into the complex dependency between temperature and demand can be observed when one plots the load curve as a function of the temperature. This representation allows one to estimate how much the electricity demand changes when the temperature changes and thus it is usually called the *temperature gradient*. Graphics shown in Figure 2 represent a way to estimate the temperature gradient (obtained actually as a nonparametric fit of the curves not explained here). Each curve fit corresponds to a specific hour of the day, which allows one to see how this dependency evolves inside the day presenting in some situation very important slopes on both high and low temperatures.
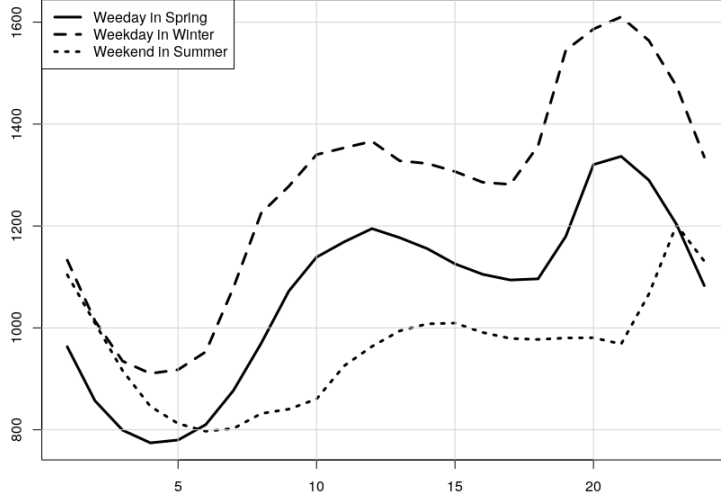
4

Figure 3: Three daily load curves.

Figure 3 exemplifies the variability found in daily load patterns. Despite such variabilities, common features can still be identified, such as lower demands during night time, higher ones during the day, a steady increase during the morning with a first peak in mid afternoon. The amplitude and position of these features are important landmarks of the load curve.

3   Methods

As mentioned in 1, the proposed framework combines the output of various models through an experts aggregation strategy.

We now describe describe the individual models which make up the experts to be aggregated. For clarity of presentation we separate them into *exogenous models* (using meteorological data) and *endogenous models* (relying only on past demand patterns). We then present and discuss the expert aggregation strategies applied to obtain the final prediction.

3.1   Exogenous models

The following models use external information that must be available at the moment of the prediction.

Hong's Vanilla Benchmark (HVB)   Hong et al. [2011]. This is a multiple linear regression model that incorporates the effect of temperature on the load as a third order polynomial; the parameters of the polynomial are indexed by calendar features such as day, month, and hour, resulting in the following formulation

$$
\begin{aligned}
\mathbb{E}[\mathrm{load}_t] = {} & \beta_0 + \beta_1 t + \beta_2 \mathtt{Day}_t \times \mathtt{Hour}_t + \beta_3 \mathtt{Month}_t + \beta_4 \mathtt{Month}_t \times \mathtt{Temp}_t \\
& + \beta_5 \mathtt{Month}_t \times \mathtt{Temp}_t^2 + \beta_6 \mathtt{Month}_t \times \mathtt{Temp}_t^3 + \beta_7 \mathtt{Hour}_t \times \mathtt{Temp}_t \\
& + \beta_8 \mathtt{Hour}_t \times \mathtt{Temp}_t^2 + \beta_9 \mathtt{Hour}_t \times \mathtt{Temp}_t^3.
\end{aligned} \tag{1}
$$

The form of (1) is simple and easy to interpret, its main feature being the thermo-sensitivity being allowed to vary according to a given calendar hour, day and month. This model was used as a benchmark model in the GEFCOM 2012 competition Hong et al. [2014], hence its name and popularity. In that competition, gains of up to 40% were obtained using more flexible models. Notice that Hong's Benchmark is a purely mid-term model, as it does not incorporate past load records; this may be useful such data is not available. It does need, however, enough temperature and load data as to adjust its parameters, something which is done using plain least squares regression. Point and interval predictions are then obtained by simple evaluation of the model using the current temperature and temperature forecasts respectively.

State space models (SSM). This family of models includes an inertial term in the form of an intrinsic state α of the whole system being modeled [Durbin and Koopman, 2001]. The observed output (the demand) is a function of the state and external variables represented in ϵ, and the state evolution over time is modeled as a linear equation involving the previous state and other observed variables summarized in a vector η. The general formulation is given by,

$$\begin{cases} y_t &= x_t\alpha_t + \epsilon_t \\ \alpha_{t+1} &= T_t\alpha_t + R_t\eta_t \end{cases},$$

where y_t is the target variable observed at time t, $x_t \in \mathbb{R}^{m+1}$ is a vector of *predictors*, the state at time t is represented as $\alpha_t \in \mathbb{R}^{m+1}$, T_t and R_t are known matrices, and ϵ_t and η_t are the noise and disturbance processes usually assumed to be independent Gaussian with zero-mean and with *unknown* covariance matrices.

One particular case is described in Dordonnat et al. [2008] for describing evolving parameters on the load forecast task. The first equation, which provides the current prediction in terms of the state and exogenous variables, is given by

$$\text{load}_t = X_t^f\beta + X_t^e\gamma_t + w_t, \tag{2}$$

where the time dependent parameters are included in the state γ_t and the fixed ones are in β. The regressors X_t^e associated to the time evolving parameters are meteorological variables such as as temperature which is the only meteorological information we use. To represent the inertia of the weather, temperature is represented as 3 variables. First, the observed temperature at each moment of the day. Second, the lowest temperature record in the last 24 hours. Third, the highest temperature record in the last 24 hours. The last two series are smoothed using splines. Behind this construction we follow the strategy of capturing effects due to extreme weather which are well know by the electrical engineers in Uruguay. Extreme observations usually arrive in clusters called respectively hot and cold waves, that is sequences of days where weather is extreme. Keeping the information of the extreme in important to detect these days. During winter, low highest temperatures are an evidence of the presence of a cold wave. Reciprocally, high lowest records evidences the presence of a hot wave.

The regressors X_t^f are associated to the fixed parameters, in this case calendar variables such as day, month, etc..

In our implementation of this method we choose two components: a *trend*, which is a linear function of the discrete time index $t = 1, 2, \ldots, n$, and a seasonal component which is modeled as a superposition of three pairs of sines and cosines,

$$\sum_{j=1}^{j=3} a_j \sin\left(\frac{2\pi t}{s_j}\right) + b_j \cos\left(\frac{2\pi t}{s_j}\right),$$

6

where the coefficients $\{a_j, b_j\}_{k=1,2,3}$ are to be estimated. With respect to the model in Dordonnat et al. [2008] this is a difference since in that work calendar effects are time-varying. The second equation of the model, which describes the state transition function, is given by

$$\alpha_t = \alpha_{t-1} + \eta_t, \tag{3}$$

that is, a multivariate random walk; the noise magnitude (the variance of η_t) is to be estimated.

The parameters of the model are estimated using Kalman filter theory [Kalman, 1960]. Given an initial state, the following states are obtained using closed-form equations that involve the inversion of possibly large matrices. The dimension of such matrices depends on the respective dimensions of the state space and the observation vector. Since this is an effective computational bottleneck, we follow Cancelo et al. [2008] and break the state space into 24 separate SSMs which can be treated in parallel, providing a huge decrease on the computation time.

Random Forest (RF). Originally proposed in [Breiman, 2001], a random forest is comprised of a set of *maximal decision trees* which involve randomness in their construction two ways. First, each maximal tree is estimated using a bootstrap sample of the data. Second, at each training step of a tree, which involves splitting one of its nodes, the splitting decision is made on a random subset of the feature variables. These random subsamplings help in reducing statistical dependency between the output of each tree in the ensemble. Finally, whereas each maximal tree may exhibit a large variance, the aggregation of many such hopefully independent trees yields a predictor with significantly smaller variance. This method is applied to load demand prediction in, for example, Dudek [2015], using past loads and past temperatures as predictors for the current load. Another example is Papadopoulos and Karakatsanis [2015] where the difference between the current temperature and the one observed some hours ago (that is, with a given lag) is used for prediction. In our work, we adopt the latter method using a lag of 72 hours to take into account important changes in the meteorological conditions.

3.2 Endogenous models

The following models predict future loads based solely on past load data and calendar information; no exogenous information is used. For these methods to work, however, very recent load data must be readily available.

Time series benchmark (TSB). Well known in econometrics, Autoregressive Moving Average (ARMA) models are a family of stationary linear stochastic processes given by the following general form [Brockwell and Davis, 1991],

$$y_t = \sum_{k \leq t} \theta_k \epsilon_k, \tag{4}$$

where ϵ_k is a sequence of i.i.d Gaussian variables of zero mean 0 and variance σ^2. The linear coefficients $\{\theta_k\}_k$ are fixed parameters which are a priori unknown.

ARMA processes can be viewed as the output of finite response filters (FIR) where the input is white Gaussian noise. The parameters are usually estimating using the Box–Jenkins method [Box et al., 2015]. First the empirical autocorrelation function is defined

as $\hat{\rho}_\ell = \hat{\gamma}_\ell / \hat{\gamma}_0$, where ℓ identifies the lag and

$$\hat{\gamma}_\ell = \sum_t (y_t - \hat{y})(y_{t-\ell} - \hat{y}). \tag{5}$$

As many real-life time series cannot be assumed to be stationary, the Box–Jenkins method proposes a number of possible transformations for obtaining time series which fall closer to that hypothesis. In our case, we first apply a discrete differentiation, $y_t' = y_t - y_{t-\ell}$, with ℓ typically 1 or 2. The differentiated series y_t' is then modeled as an ARMA process; we call this process an integrated ARMA (ARIMA for short) of order ℓ. A second differenciation is then performed. The second transformation is a *seasonal* differentiation, $y_t'' = y_t - y_{t-\ell'}'$, which is identical in form as the first one but with a lag ℓ' corresponding to the period of a given season in the data (e.g., hours in a day, days in a week, months in a year. This two step transformation is known as a SARIMA (Season Auto-Regressive Integrated Moving Average). The process of identification, estimation and validation of such SARIMA model is usually done by an expert and can be quite time consuming, especially in the case of load demand as it typically exhibits two seasonal components Instead, we propose a new estimation scheme based on mimickingthe empirical autocorrelation function $\hat{\rho}_\ell$ to the theoretical one, ρ_ℓ, as it is done by econometrics.

For a particular SARIMA model one can write

$$S(\theta) = \sum_{\ell=0}^{L} (\rho_\ell(\theta) - \hat{\rho}_\ell)^2, \tag{6}$$

where θ is a vector of unknown parameters and L is a sufficiently large lag. The estimator $\hat{\theta}$ is obtained as $\arg\min_\theta S(\theta)$. (In Guang-Te and Singh [1994], a similar idea is used to estimate autoregressive models. The above estimator can be viewed as a generalized method of moments for time dependent observations. While the theoretical properties of the proposed estimator are not as good as those of the maximum likelihood estimator (MLE), its computation is significantly less demanding than the latter, which makes it our choice of estimator within our framework.

Point predictions are obtained using the last observed values of load_t and the last prediction errors. Prediction intervals can be obtained using either a normal approximation or bootstrap procedures.

Kernel Wavelet Functional (KWF) As discussed in Section 2, the shape of the load curves carries valuable information about the context of the observed load (e.g. position of the year, type of the day, meteorological conditions). Assuming that similar past conditions induce similar future conditions, one can construct an easy-to-interpret predictor [Poggi, 1994]. A modern version of this method, which exploits the functional nature of time series curves, is used in Antoniadis et al. [2006]. There, the predictor is written as

$$\widehat{\text{Load}}_{n+1}(\tau) = \sum_{m=1}^{n-1} w_{m,n} \text{Load}_{m+1}(\tau), \qquad \tau \in [0,1]$$

where $\text{Load}_n(\tau)$ is the load curve for day n at instant $\tau \in [0,1]$ and the weight $w_{m,n}$ is proportional to the similarity between the load curves Load_m and Load_n (more on the similarity criterion is mentioned below). Note that the curves Load_n are not observed directly, but rather estimated from the corresponding (discrete) hourly values, For instance, Load_n is estimated from $\text{load}_t, t = (n-1) \times 24 + j, j = 1, \ldots, 24$.The preceding estimation

is performed using *wavelets*, which also serve for computing the similarity weights $w_{m,n}$ in terms of the wavelet coefficients of both curves. Non stationary patterns are treated by means of corrections applied on the wavelet coefficients (see Antoniadis et al. [2012] for details).

Prediction intervals can be obtained using a bootstrap strategy where the bootstrap sampling is determined by $\{w_{m,n}\}$ the weight vector (see Antoniadis et al. [2016]).

3.3 Online mixture of experts

The general idea of mixing experts is to combine the outputs of different individual predictors in order to produce a better overall prediction. If this mixture is properly implemented, the mixed output is theoretically guaranteed to be at least as good as the best of all the individual predictors [Devaine et al., 2013]. This carries on to practice, where significant improvements are consistently obtained with this strategy in a wide variety of settings beyond time series analysis. The idea is simple, and can be summarized as follows. Consider the observations up to time T, $y_{1:T} = \{y_t, t = 1, \ldots, T\}$, and a set of K individual predictors (experts). For each $k = 1, \ldots, K$, the *performance* of the k-th predictor up to time T is computed in terms of some loss function, in our case squared loss, between its output $e_{k,1:T} = \{e_{k,t}\}$ and the observed output,

$$l_{k,T}(y_{1:T}, e_{k,1:T}) = \sum_{t=1}^{T}(y_t - e_{k,t})^2.$$

Given $l_{k,T}, k = 1, \ldots, K$, the aggregated prediction at time $T + 1$, $\hat{y}_{T+1}^{\mathrm{mix}}$, is computed as a weighted sum of the output of all K predictors at time $T + 1$,

$$\hat{y}_{T+1}^{\mathrm{mix}} = \sum_{k=1}^{K} w_{k,T+1} e_{k,T+1}$$

where the weights $w_{k,T+1} = \omega(l_{k,T})$ with $\omega(\cdot)$ some decreasing function. In our case we use polynomial potential aggregation rules (ML-Poly) which computes the mixture as a weighted average of experts using polynomial weights and allowing different learning rates on each expert.

4 Experiments

In this section we report on a number of experiments performed in conditions which are close to the operational ones. The goal of these experiments is to evaluate the performance of each of the implemented predictors, as well as that of the aggregation strategy. We use the open source R environment [R Core Team, 2017] and the package `enercast`[3] which was developed by the authors. This package implements the proposed individual models, whereas the on-line mixing is done with the `opera` R package [Gaillard and Goude, 2016].

The performance of each model is measured through the one-day-ahead daily prediction of load demand curves during the whole last year. That is at the end of day J we use all the records of that day to predict the following 24 records. The only exception is the data used as temperature predictions which is the effectively observed values of day $J + 1$ as it is usual to estimate the model. Forecasting performance is measure using error measures.

---

[3]`github.com/cugliari/enercast`

More specifically, we use two metrics to compare predicted against actual values. One is the daily Mean Absolute Proportional Error (MAPE), which is given as a percentage,

$$\text{MAPE}(y, \widehat{y}) = \frac{100}{24} \sum_{h=1}^{24} \left| \frac{y_h - \widehat{y}_h}{y_h} \right| \tag{7}$$

where $y = (y_1, ..., y_{24})$ is the effective load, and $\widehat{y} = (\widehat{y}_1, ..., \widehat{y}_{23})$ is the daily prediction. The other measure is the Root Mean Square Error (RMSE),

$$\text{RMSE}(y, \widehat{y}) = \sqrt{\frac{1}{24} \sum_{h=1}^{24} (y_h - \widehat{y}_h)^2}, \tag{8}$$

which is expressed in the same units as the target variable, Megawatts per hour (MWh). Note that this measure is more consistent with the criterion used by the sequential aggregation strategy (square loss) for weighting the contribution of each expert.

Model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1. HVB	2.12	4.26	5.49	5.93	6.87	30.85
2. KWF	1.83	3.40	4.30	4.61	5.31	20.17
3. TSB	0.65	2.23	3.14	4.13	5.11	22.03
4. SSM	0.58	2.29	3.42	4.12	4.98	25.84
5. RF	0.76	2.09	2.97	3.66	4.34	20.30
Mixture	0.76	1.76	2.40	3.06	3.58	17.60

Table 1: Summary statistics of daily MAPE (%) over year 2014. Best results are shown in blue, while worst are shown in red.

Model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1. HVB	31.83	61.05	73.38	81.27	92.32	305.30
2. KWF	25.66	48.69	60.11	65.96	74.71	278.50
3. TSB	7.75	32.94	46.37	58.92	72.94	287.60
4. SSM	11.88	32.96	46.97	57.34	68.42	288.70
5. RF	9.87	29.41	42.32	51.36	62.82	206.30
Mixture	11.19	25.19	36.08	44.35	53.08	196.04

Table 2: Summary statistics of daily RMSE (in MWh) over year 2014. Best results are shown in blue, while worst are shown in red.

Tables 1 and 2 show summary yearly statistics for the MAPE and RMSE prediction performances for each of the individual methods (experts) as well as that of the mixture model which combines all of them. To begin with HVB, an exogenous model, comes out clearly as the worst method in all the statistics, in both tables. Interestingly, the other purely exogenous model, the Random Forest, comes out as the best individual expert in almost all cases, the only exceptions being the minimum and maximum MAPE, and the minimum RMSE. We also note that the performance of TSB, which is purely endogenous, is very close to that of SSM, which also employs exogenous (temperature) information.

Finally, as can be clearly seen in both Tables 1 and 2, the mixture strategy significantly outperforms the rest of the models in all but one of the statistics, the minimum, which

is also arguably the less relevant one. In particular, we obtain a 15% reduction in RMSE w.r.t the best individual model. The following discussion aims at providing more insight into these results.
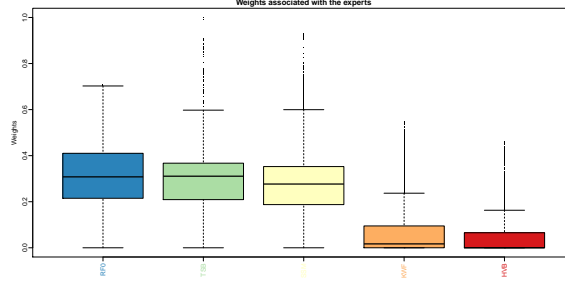

Figure 4: Box plot of the empirical distribution of the weights assigned during the year 2014 to each expert.

The box plotshown in Figure 4 depicts the empirical distribution of the weights assigned to each individual expert along the 365 days of 2014. The figure shows two clear groups among the experts. On one side, RF, TSB and SSM exhibit a similar median weight of about 33%. On the other side, the median weight of both KWF and HVB are sensibly smaller, barely above 10%, and yet they both peak at values over 40%.
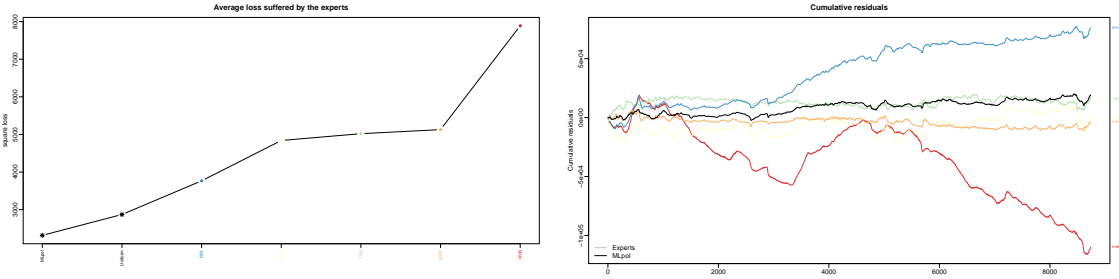

Figure 5: Average square loss (left) and cumulative residuals (right) for the individual experts and the online mixture.

Figure 5 provides more insight into the benefits of the weighting mechanism. On the left, we show the average squared loss by predictor in increasing order, whereas the right plot shows the cumulative prediction error across the year for each method. Here we can see that both endogenous models, HVB and RF, exhibit strong biases during long periods, even though RF appears to be the best single expert among the ones tested. The mixture model however does not exhibit such bias, which shows its ability to temporarily shift its weight to other predictors, even if they are overall worse, when such systematic deviations occur.

4.1 A closer look at the results

TBeing a yearly summary, the performance statistics shown in tables 1 and 2 are insufficient to characterize the appropriateness of any given individual model, as they do not reflect the relative importance of different seasons, days (such as holidays), or specific hours of the day, in what respects to energy generation planning. We now focus on two of these aspects, namely, the behavior across the year (monthly performance) and across the day (hourly performance). The first is shown in Figure 6 for the individual predictors as well
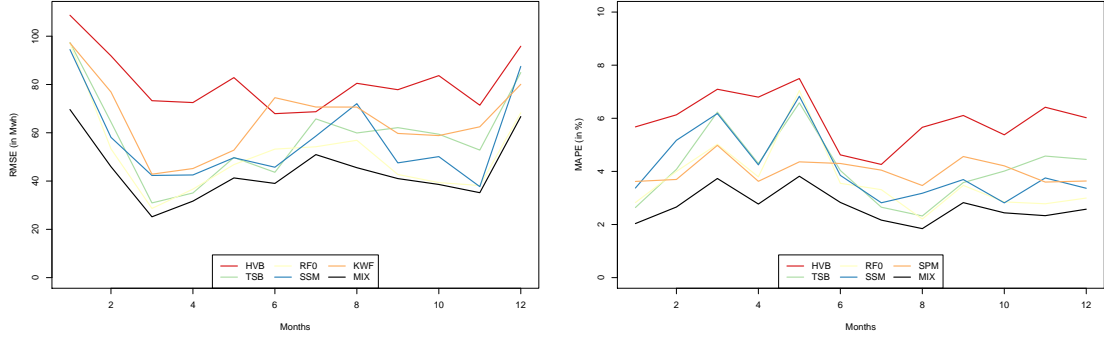
11

Figure 6: Mean RMSE (left) and MAPE (right) for each expert and mixture (right).

as for the mixture. Although the mean monthly RMSE and MAPEs evolve differently for each expert, some common elements can be observed. First we look at RMSE. Here the worst performances occur during (Austral) summer (December, January and February); winter (June, July and August) is generally well predicted; autumn is better predicted than spring by all the methods except HVB. These conclusions however do not carry on to the MAPE case, that is, when the relative error is evaluated. Here the picture is quite different. The only similarity is winter, where again the best performance is obtained in general. The worst season in this case appears to be autumn, peaking in May and then March (this can be explained by the fact that May, for example, is a month with relatively low load, thus amplifying relative errors).
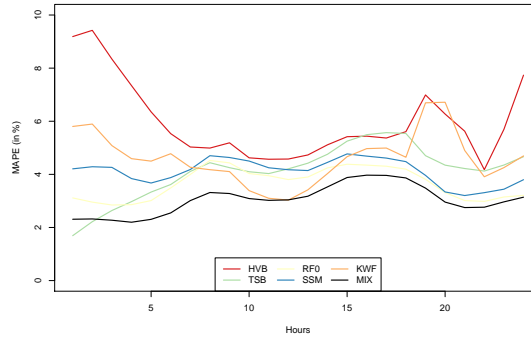

Figure 7: Mean MAPE for each expert and mixture by hours.

The hourly MAPE for all predictors is shown in Figure 7. Note that the performance of endogenous hour-by-hour forecast methods, which can only rely on the previous day loads, becomes worse for larger prediction horizons; this effect is superimposed to the particular difficulties of the load pattern itself. We note that the RF and SSM curves are similar in shape to the mixture one. The shape of these predictors reflect the higher difficulty of predicting the *morning ramp* and the *afternoon plateau* (see Figure 3). As expected, the performance of TSB worsens TSB for larger horizons and yet, it is the best predictor for the first hours of the day, surpassing the performance of the mixture on the first two hours of the day. The afternoon peak (between 18h00 and 20h00) is difficult for KWF and HVB, but KWF has a good performance during midday, matching the mixture during those hours.

12

Figure 8 shows the contribution of the individual experts to the global prediction as well as the weights distribution. The time axis are the same as the one on the left panel of Figure 5. Clearly, the best three experts (RF, TSB and SSM) share the contributions most of the time, that is, when one of the main three experts reduces its contribution, the other two take on larger weights (cf. February, around time step 1000 where the RF contribution is relatively small and SSM and TSB take over). However, exceptions to this behavior are of interest. Such is the case of three periods of the year: Winter, the month of May and Spring. During these periods, HVB and KWF have point-wise larger participation, meaning that the other three experts are not able to follow the consumption structure during these periods. The month of May is of particular interest because the Uruguayan weather in this period is highly erratic and difficult to anticipate. In other terms, the additional information provided by the temperature in the exogenous models is of less quality. The online mixing algorithm quickly detects this phenomenon, thereby adjusting the weigths in favour of the endogenous experts during these periods.
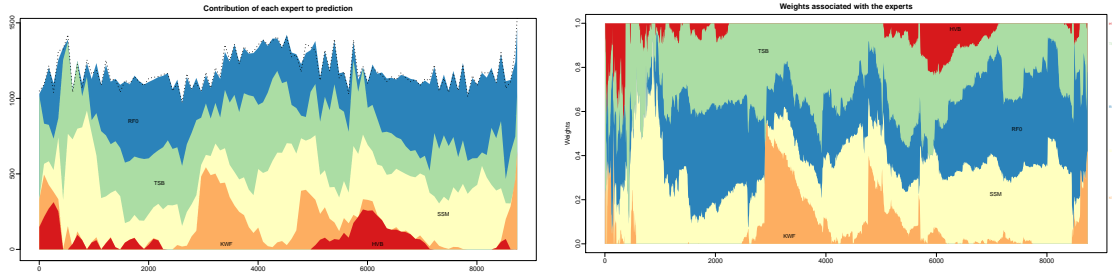

Figure 8: Contribution of the individual experts to the mixture prediction (left) and weight distribution per expert by time (right).

5 Discussion

aggregar una discusion, ideas:
- resumen tonto
- [done] impacto del uso de meteo efectiva en lugar de predicha
- uso de KWF y TSB como mas genericos

Anticipation of load demand is a crucial element to help decision taking on dispatch. The introduction of new challenging contexts such as renewables, changes and intensification on electronic devices usages or the increasing availability of new data sources, pushes the domain towards the use of new automatic approaches to produce forecast. In this work we show the construction a few very models very different in their nature. Instead of choosing the best of them to produce the final forecast, we use a sequential prediction, that is an ensemble method adapted to time series in order to enhance the prediction quality.

While we tried to be as close as possible to operational conditions some important issues are to be raised. Unfortunately, forecast measurements were not available during this work, as such information would have provided a more realistic operational scenario in which to evaluate the different predictors. For instance, the performance of the exogenous models (HSV, SSM, RF) has to be considered an optimistic estimation, since predictions from these experts were obtained from the *true* measurements of the day being predicted rather than from daily forecast, which introduces its own (significant) errors and biases. Note however that he conclusions on what respects to the purely endogenous experts as well

as the benefits of using mixtures are expected to hold in the presence of such information. Also, the prediction obtained in our experiments are computed at midnight with all the data of the that day available. However, in real condition one would have to predict many hours in advance in order to be able to use this information in the decision taking procedure for the dispatch.

A natural perspective would be to add more experts on the prediction mix. One inexpensive way to obtain this is to create ? Two additional perspective are

N1

N1: No se puede comentar sobre la predicción de curvas de temperatura usando por ej KWF y luego usar eso como pronóstico para HVB, SSM, RF??

References

A. Antoniadis, E. Paparoditis, and T. Sapatinas. A functional wavelet-kernel approach for time series prediction. *Journal-Royal Statistical Society Series B Statistical Methodoloty*, 68(5):837, 2006.

A. Antoniadis, X. Brosat, Cugliari, J, and J.-M. Poggi. Prevision d'un processus a valeurs fonctionnelles en presence de non stationnarites. Application a la consommation d'electricite. *Journal de la Societe Française de Statistique*, 153(2):52–78, 2012.

A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi. Une approche fonctionnelle pour la prévision non-paramétrique de la consommation d'électricité. *Journal de la Société Française de Statistique*, 155(2):202 – 219, 2014.

A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi. A prediction interval for a function-valued forecast model. *International Journal of Forecasting*, 32(2):939–947, 2016.

S. Ben Taieb and R. J. Hyndman. A gradient boosting approach to the kaggle load forecasting competition. *International Journal of Forecasting*, 30(2):382–394, April 2014. ISSN 01692070. doi: 10.1016/j.ijforecast.2013.07.005.

G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley, 2015. ISBN 9781118674925. URL https://books.google.fr/books?id=rNt5CgAAQBAJ.

L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

P.J. Brockwell and R.A. Davis. *Time Series: theory and methods*. Statistics. Springer, 2nd edition, 1991.

A. Bruhns, G. Deurveilher, and J.S. Roy. A non linear regression model for mid-term load forecasting and improvements in seasonality. In *Proceedings of the 15th Power Systems Computation Conference*, pages 22–26, 2005.

J. Cancelo, A. Espasa, and R. Grafe. Forecasting the electricity load from one day to one week ahead for the spanish system operator. *International Journal of Forecasting*, 24: 588–602, 2008.

M. Devaine, P. Gaillard, Y. Goude, and G. Stoltz. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning*, 90(2):231–260, February 2013. ISSN 0885-6125, 1573-0565. doi: 10.1007/s10994-012-5314-7.

DNETN. Politica energética 2005-2030. Technical report, Ministerio de Industria, Energia y Mineria, 2008. URL `http://www.miem.gub.uy/documents/49872/0/Pol%C3%ADtica%20Energ%C3%A9tica%202030?version=1.0&t=1352835007562`.

V. Dordonnat, S.J. Koopman, M. Ooms, A. Dessertaine, and J. Collet. An hourly periodic state space model for modelling French national electricity load. *International Journal of Forecasting*, 24(4):566–587, 2008. ISSN 01692070. doi: 10.1016/j.ijforecast.2008.08.010. URL `http://linkinghub.elsevier.com/retrieve/pii/S0169207008000988`.

G. Dudek. Short-term load forecasting using random forests. In D. Filev, J. Jablkowski, J. Kacprzyk, M. Krawczak, I. Popchev, L. Rutkowski, V. Sgurev, E. Sotirova, P. Szynkarczyk, and S. Zadrozny, editors, *Intelligent Systems'2014*, volume 323 of *Advances in Intelligent Systems and Computing*, pages 821–828. Springer International Publishing, 2015. ISBN 978-3-319-11309-8. doi: 10.1007/978-3-319-11310-4_71. URL `http://link.springer.com/chapter/10.1007/978-3-319-11310-4_71`.

J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series. Clarendon Press, 2001. ISBN 9780198523543. URL `https://books.google.fr/books?id=XRCu5iSz\_HwC`.

P. Gaillard and Y. Goude. *opera: Online Prediction by Expert Aggregation*, 2016. URL `http://pierre.gaillard.me/opera.html`. R package version 1.0.

W. Guang-Te and V. P. Singh. An autocorrelation function method for estimation of parameters of autoregressive models. *Water Resources Management*, 8(1):33–55, Mar 1994. ISSN 1573-1650. doi: 10.1007/BF00872278. URL `https://doi.org/10.1007/BF00872278`.

T. Hong, P. Wang, and H. L. Willis. A naïve multiple linear regression benchmark for short term load forecasting. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–6. IEEE, 2011. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6038881`.

T. Hong, P. Pinson, and S. Fan. Global energy forecasting competition 2012. *International Journal of Forecasting*, 30(2):357–363, April 2014. ISSN 01692070. doi: 10.1016/j.ijforecast.2013.07.001. URL `http://linkinghub.elsevier.com/retrieve/pii/S0169207013000745`.

R. Kalman. A new aproach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, Marzo 1960. doi: 10.1115/13662552.

B. Lanzilotta and S. Rodriguez. Modelos de predicción de energía eléctrica con datos horarios para Uruguay. Informe de investigación, Fondo Sectorial de Energía. ANII, marzo 2014.

T. Launay, A. Philippe, and S. Lamarche. Construction of an informative hierarchical prior distribution. application to electricity load forecasting. *arXiv*, 1109.4533, 2012.

J. R. Lloyd. GEFCom2012 hierarchical load forecasting: Gradient boosting machines and gaussian processes. *International Journal of Forecasting*, 30(2):369–374, April 2014. ISSN 01692070. doi: 10.1016/j.ijforecast.2013.07.002.

M. Mougeot, D. Picard, V. Lefieux, and L. Maillard-Teyssier. *Forecasting Intra Day Load Curves Using Sparse Functional Regression*, pages 161–181. Springer International Publishing, Cham, 2015. ISBN 978-3-319-18732-7. doi: 10.1007/978-3-319-18732-7_9. URL `https://doi.org/10.1007/978-3-319-18732-7_9`.

R. Nedellec, J. Cugliari, and Y. Goude. GEFCom2012: electric load forecasting and backcasting with semi-parametric models. *International Journal of Forecasting*, 30(2): 375–381, April 2014. ISSN 01692070. doi: 10.1016/j.ijforecast.2013.07.004.

S. Papadopoulos and I. Karakatsanis. Short-term electricity load forecasting using time series and ensemble learning methods. In *2015 IEEE Power and Energy Conference at Illinois (PECI)*, pages 1–6, Feb 2015. doi: 10.1109/PECI.2015.7064913.

J.-M. Poggi. Prévision nonprametrique de la consommation électrique. *Rev. Statistiqué Appliquée*, XLII(4):93–98, 1994.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

L. J. Soares and M. Medeiros. Modeling and forecasting short-term electricity load: A comparison of methods with an application to brazilian data. *International Journal of Forecasting*, 2008.

J. W. Taylor, L. M. de Menezes, and P. E. McSharry. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22(1):1–16, 2006. ISSN 01692070. doi: 10.1016/j.ijforecast.2005.06.006.

R. Weron. *Modeling and forecasting electricity loads and prices: a statistical approach*, volume 396 of *Wiley finance series*. John Wiley and Sons, 2006.

R. Weron, B. Kozlowska, and J. Nowicka-Zagrajek. Modeling electricity loads in California: a continuous-time approach. *Physica A: Statistical Mechanics and its Applications*, 299 (1):344–350, 2001.