Univariate *versus* Multivariate Models for Short-term Electricity Load Forecasting

Guilherme Guilhermino Neto[1], Samuel Belini Defilippo[2], Henrique S. Hippert[3]

[1] IFES – Campus Linhares. guilherme.neto@ifes.edu.br
[2] PPG Modelagem Computacional, UFJF. samueldefilippo@outlook.com
[3] Depto. Estatística, UFJF. henrique.hippert@ufjf.edu.br

Abstract. Online short-term load forecasts are needed for efficient demand management on power systems. To model the load, univariate and multivariate forecast approaches were developed: while the first consider the load as a linear function of its time series, the other also takes in account the nonlinear effects of weather-related variables (mainly the air temperature). Despite the wide recent literature on multivariate models, some authors state that univariate ones are sufficient for short-term purposes, claiming that including temperature variables unnecessarily elevates the model complexity, putting parsimony and robustness at risk. In this study, we compare the forecasts produced, for real data, by several univariate and multivariate time series and neural network-based load curve models. We then use a nonparametric hypothesis test to compare the daily mean errors of the best forecaster of each kind and, so, verify if considering the air temperature leads to any statistically significant improvement in the forecasting.

Keywords: short-term load forecasting; load curve models; exponential smoothing; neural networks.

1 Introduction

Hourly short-term load forecasts, with lead times that range from one hour to seven days, have been an important input for the operating of power systems operating for many years. Due to the facts that the setup of generators may take hours and it is not possible to keep large energy safety stocks in order to meet demand peaks, forecasting is a vital task for these systems.

Until the nineties, electricity load forecasting had reached a comfortable state of art. Adaptive models worked sufficiently and companies used to manage the uncertainty via high capacity. However, the deregulation of the power industry, followed by the arrival of free electricity markets, brought the expectation of a greater consumer and shareholder participation. This scenario would translate the physical risk of inadequate capacity to a financial risk of high prices [1], for what the need of accurate forecasting models became much more of interest, since any marginal improvement would now lead to a great rise on profits.

Literature on short-term load forecast show numerous experiments using statistical and computational intelligence methods. These methods may be applied to univariate models, in which the load is written as a linear function of the past loads and usually forecast by time series techniques [2], or multivariate, which also consider the effects of so called exogenous variables, such as social, economic and, mainly, weather-related (air temperature, wind speed, cloud covers, for example).

Even though there are reports of low-error forecasters based on both approaches, there seem not to be an agreement on which may be more appropriate for short-term purposes. [3] and [4], for example, state that the relationship of load and weather is the most important factor to be considered for a reliable short-term forecast. This idea, along with the development of high-processing capacity computers, made computational intelligence multivariate weather-based methods, specially artificial neural networks-based (ANN-based), one of the most common way of carrying out the task nowadays [5]. However, some authors, as [6], assert that estimate the load by analyzing its dependence on meteorological parameters and then to predict the weather renders doubles the task unnecessarily. Also, [7] states that univariate methods are easier to ensure robustness and considered to be sufficient for short lead times because weather variables tend to change in a smooth fashion over short time frames, what will be captured in the demand series itself.

In this paper, we compare the performance of univariate *versus* multivariate models by the means of computer simulations over a real American dataset. We first try time series methods over univariate models and choose the best of them to add a weather component (air temperature). The weather component is forecast using artificial ANN. We then use a nonparametric hypothesis test to compare the mean daily errors produced by the best univariate and multivariate models. This allows us to verify if is there any statistically significant improvement (and, thus, any possibility of cause positive impact on the profits of the organization) when considering the weather component.

2 Material and methods

2.1 The American load data

In this study, the computational experiments are made over time series of loads (in MWh) and temperatures (ºC), measured by Puget Sound Power and Light Company (USA, state of Washington), and originally made available by the Electric Power Institute [8]. The data consist of seven full years (1985-1991), plus 282 days (1992), of hourly observations for each variable (thus, 406 weeks / 68,208 observations).

Special days, such as bank holidays, are usually forecast offline. As the forecasting for these days is not the focus of this study, we decided to smooth the loads on special days, replacing them by the average of the observations for the same hours on the weeks before and after.

Figure 1 shows the daily mean load for all years covered by the data. This plot suggests the existence of an intra-year seasonal cycle, with loads higher during the winter and lower during the summer.

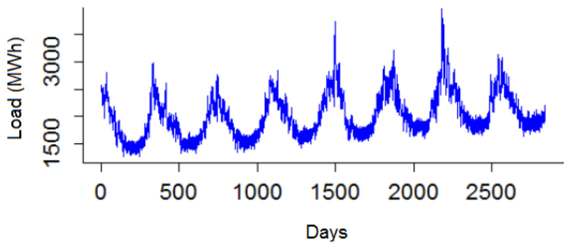

Fig. 1. Daily mean load for 1985 to 1992 (Puget Sound Power & Light Co. data)

Figure 2 shows the loads for a typical summer fortnight. We can notice another two superimposed seasonal cycles across the same data: loads are always lower at night (a daily cycle), and they are lower in the weekends (a weekly cycle).

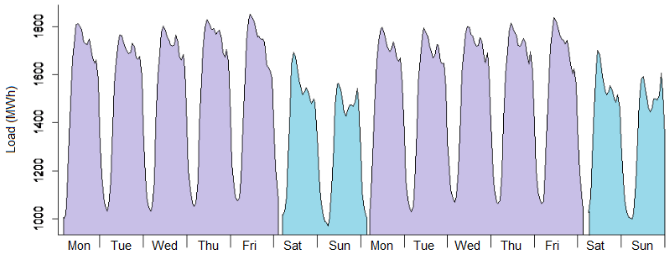

Fig. 2. Loads for a typical summer fortnight (Puget Sound Power & Light Co. data)

The scatterplot on Figure 3 suggests that, for the analyzed data, the loads hold a nonlinear correlation with the temperatures. The graph suggests that the higher the temperature, the lower the load. We believe that this happens due to the higher use of air heating during colder days.

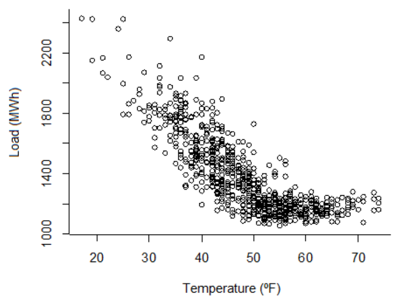

Fig. 3. Load *versus* temperature at 1 a.m. (Puget Sound Power & Light Co.)

2.2 Standard load curve models and forecasting strategies

In this paper, we model the load via the so called "standard load curve model" [2]. One of the most traditional short-term load forecasting devices, this model describes the functional form of the average daily load profile (the standard load curve), and the deviations from its mean, by a mathematical equation.

The model has two main formulations. The first one is the univariate in eq. (1), where L(t), the load in an hour t in a certain day, is said to be a function of a base component B(t), which is related to the history of the loads, its trends and seasonality; and of a random error component R(t).

$$L_t = B_t + R_t \tag{1}$$

As the model is the same for all the hours of the day, we need to adopt a strategy to consider all three seasonal cycles mentioned in 2.1. There are two major possibilities: to create sub-models, one for each of the 24 hours of the day, as in [9], or one for each of the 168 hours of the week, as in [10]; or one single model that can accommodate the three seasonal cycles at once, as, recently, in [6].

The second formulation for the load curve model is the multivariate one in eq. (2). In this equation, a weather-related component W(t) is added to the model. Since the relationship of load and weather usually appears to be nonlinear (as shown for our data), function approximation devices, such as ANN, are frequently used for these models nowadays.

$$L_t = B_t + W_t + R_t \tag{2}$$

2.3 Forecasting procedure for the univariate model

In our experiments, the first step is to model the load via the univariate model in eq. (1), forecasting B(t) by means of classical linear time series methods [11].

For each of the strategies mentioned above, we try the following methods:

24 sub-models (one for each hour of the day): Holt-Winters seasonal method and SARIMA - because of the weekly seasonal cycle for each hour of the day.

168 sub-models (one for each hour of the week): Holt linear trend method and ARIMA - because, in short-term, there is no seasonal cycle for each hour of the week, but trends may occur;

One model for all the hours: Holt-Winters-Taylor triple-seasonal method - because it can accommodate three seasonal cycles in a same model for all hours.

For each strategy, we also use a Naïve Forecaster [11] as a benchmark.

2.4 Forecasting procedure for the multivariate model

In order to forecast with the multivariate model, it is first necessary to forecast B(t) and, then, W(t), for which, in this study, we use an ANN (for more on ANN, see [12].)

The chosen ANN architecture is a Multi-Layer Perceptron, with a single hidden layer. The activation functions are sigmoidal, and the weights are estimated using a conjugate gradient learning rule. We chose this architecture because, in our test, it lead us to results with better accuracy.

The ANN's input variables are:

24 temperatures for the forecast hours
24 differences between the hourly temperatures measured today, and those measured yesterday
24 differences between the hourly temperatures measured today, and those measured last tomorrow

The outputs are the deviations of the forecasted base load from the actual base load, which, in eq. (2), must be explained by W(t).

2.5 Error function

We compare the post-sample forecasting accuracy of the different methods analyzing the mean absolute percent error (MAPE), defined as:

$$MAPE = 100 \times \frac{1}{n} \sum_{t=1}^{n} \left| \frac{L_t - \hat{L}_t}{L_t} \right| \tag{3}$$

where L(t) is the actual load and $\hat{L}_t$ is the forecast load. We chose this measure because of its easy interpretation, and also because of its widespread use in the short-term load forecasting literature.

2.6 Model comparison procedure

To verify if there is a significant improvement when adding a weather-related component to our model, we proceed with the following steps:

(i) Choose the best forecaster for the univariate model: From the time series methods mentioned in 2.3, we choose the one that forecasts with the best accuracy. For this, we use weeks 1 to 198 of the dataset (approximately 4 years) to estimate the parameters of the methods, and weeks 199 to 302 (2 years) to test post-sample accuracy.

(ii) Obtain forecasts for the multivariate model: we use the forecaster with the best performance for the base load in (i) and add the component W(t), which is calculated using an ANN, as described in 2.4.

(iii) Compare (i) and (ii): In order to compare the univariate and multivariate models, we use a nonparametric test to see if the differences between their post-sample accuracies are statistically significant (that is, if we can say, with some probability, that the forecasting errors came from different distributions). Although this is not commonly done, we believe it is a more interesting procedure than purely comparing the MAPEs, because it allows us to conclude if the differences between the forecasting errors may hold to other samples. The test we chose was the nonparametric Wilcoxon's signed-rank test [13], because of its application to paired samples. Using weeks 1 to 302 of the dataset to estimate the time series method parameters and train the neural networks and the rest to calculate the post-sample accuracy, we use the test to compare the daily MAPEs of the models. A similar procedure was tried in [14], but the author does not give any details on, for example, which samples were used to calculate the error measure. This is an important choice, since there is no evidence that Wilcoxon's test works properly for autocorrelated time series and, so, comparing, for example, the hourly errors (which are heavily autocorrelated) could lead to non-reliable results. That is, in fact, why we chose to work with the daily MAPEs – we expect the MAPE in one day not to be correlated with the one from the day before.

3 Results and discussion

All tests in this paper were made with the free software R.

Table 1 shows the MAPEs obtained by using the time series methods in 2.3 to forecast B(t) for the univariate standard load curve model.

Strategy	Method	MAPE (%)
24 sub-models	Naive Forecaster	5,44
	Holt Linear Trend Method	6,54
	SARIMA	3,76
168 sub-models	Naive Forecaster	6,56
	Holt-Winters Seasonal Method (HW)	3,91
	ARIMA	6,47
Triple-seasonal method	Holt-Winters-Taylor Triple-seasonal Method (HWT)	**3,04**

Table 1. MAPES for the time series methods (post-sample test; weeks 199-302)

Forecasts made by HWT were the best, according to the table.

Figure 3 shows the boxplots for the absolute errors (APEs), given by $\left|\hat{L}_t - L_t\right|$. We notice, in this figure, that HWT was also the method that produced lower variance for the APEs, which suggests that it is not just a more accurate method, but also of more robustness than the others tested.

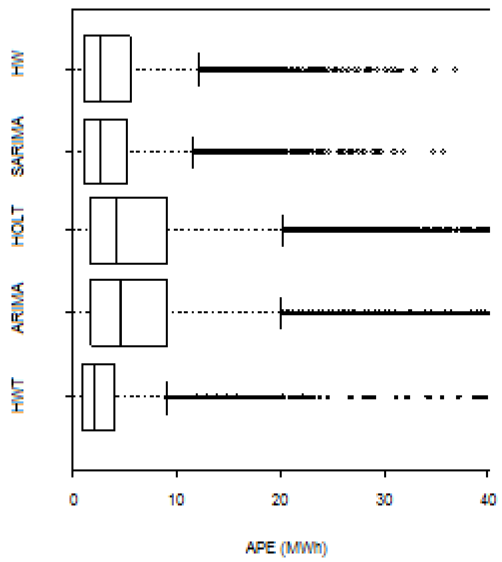

Fig. 4. APEs for the linear methods (weeks 199-302)

With B(t) forecast by HWT, we added a weather-related component, and forecast it by using the ANN specified in 2.4. We simulated 100 runs of the ANN.

Considering the relationship between load and temperature, in this case, lead to a slight reduction in the MAPE for all the runs of the ANN, as we can see in Table 2.

Model	Forecaster(s)	MAPE (%)
Univariate	HWT	3,04
Multivariate	Worst HWT+ANN	3,03
	Mean HWT+ANN	2,98
	Best HWT+ANN	**2,96**

Table 2. MAPES for the multivariate models (post-sample test; week 302 ahead)

In order to see if the differences observed between the two models are statistically significant (that is, if the errors come from different distributions), we did a Wilcoxon signed-rank test for paired samples for the samples of daily MAPEs of the univariate model and the best multivariate model. We chose this test because a previously done Shapiro-Wilk test pointed no evidence that the samples are normally distributed.

For a two-tailed test where the null hypothesis H_0 is that the performance of the models does not differ (in terms of the Wilcoxon test, the sum of the positive ranks equals the sum of negative ranks) and H_1, the alternative hypothesis is that the performance of the two models differ, our decision was to reject H_0 in favor to H_1.

The forecasts for the best multivariate model for the first days of the test sample can been seen and compared to the actual measured loads in Figure 5.

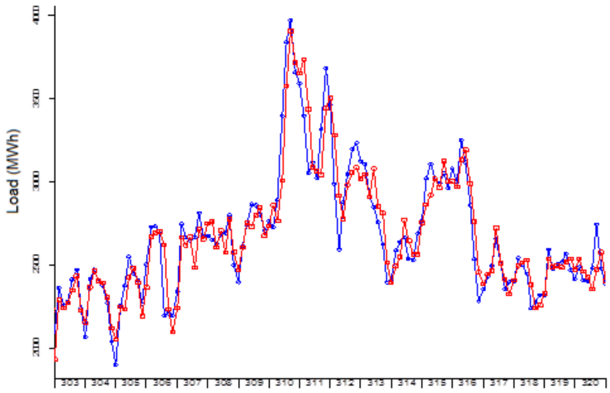

Fig. 5. Actual load (blue) and forecasts by HWT + best ANN (weeks 303-320)

4 Conclusion

The use of computational tools, such as ANN, for short-term electric load forecasting with a multivariate weather-based approach has been recommended by many authors in recent publications. However, others suggest that, for short-term purposes, the effect of weather-related variables are minimum, which justifies that univariate linear methods (such as time series') may be a more interesting alternative to complex and non-parsimonious neural networks.

In this paper, we tested a standard load curve univariate model, based in the auto-regressive behavior of the data, and compared its results with those produced by a multivariate formulation, which also considers information about changes in air temperature.

The simulations over real data showed that adding the weather-related lead, for the data, to slight improvement on the forecasting accuracy. The results of nonparametric test showed us that this difference is statistically significant. Since [5] says that even little improvement in forecasting may imply a huge impact on the organization's profits on deregulated markets, we believe that it is worth to choose the multivariate model in this case.

For future works, we suggest trying the same methodology on different data, as well as experiment other computational intelligent methods (such as genetic algorithms) to forecast the weather-related company. We also suggest trying combinations of linear time series methods to forecast the base load and apply statistical tests that can compare the results from multiple models at the same time, since Wilcoxon test can only deal with paired samples.

5 References

1. Bunn, D. W. Forecasting loads and prices in competitive power markets. Proceedings of the IEEE, 2, 163-169. (2000)
2. Bunn, D.W.; Farmer, E.D. Comparative Models for Electrical Load Forecasting. 1. ed. New York: Wiley (1985).
3. Gupta, P.C.; Yamada, K. Adaptive short-term forecasting of hourly loads using weather information. IEEE Transactions on Power Apparatus and Systems, 5, 2085-2094. (1972)
4. Rahman, S; Hazim, O. A generalized knowledge-based short-term load forecasting technique. IEEE Transactions on Power Systems, 2, 508-514. (1993)
5. Hippert, H.S.; Pedreira, C.E.; Souza, R.C. Neural networks for short-term load forecasting: A review and evaluation. IEEE Transactions on Power Systems, 1, 44-55. (2001)
6. Farmer, E.D.; Potton, M. Development of online load prediction techniques with results from trials in the south-west region of the CEGB. Proceedings of the Institution of Electrical Engineers, 10, 1549-1558. (1968)
7. Taylor, J.W. Triple seasonal methods for short-term electricity demand forecasting. European Journal of Operational Research, 1, 139-152. (2010)
8. Electric Power Institute. Probabilistic Methods in Forecasting Hourly Loads. San Diego: QUERI (1993).
9. Ramanathan, R.; Engle, R.; Granger, C.W.J.; Vahid-Araghi, F.; Brace, C. Short-run forecasts of electricity loads and peaks. International Journal of Forecasting, 14, 161-174 (1997).
10. Papalexopoulos, A.D.; Hesterberg, T.C. A regression-based approach for short-term load forecasting. IEEE Transactions on Power Systems, 4, 1535-1550 (1990).
11. Makridakis, S.G; Wheelwright, S.C; Hyndman, R.J. Forecasting: Methods and Applications. 3. ed. New York: Wiley (1997).
12. Principe, J. C.; Euliano, N. R.; Lefebvre, W. C. Neural and Adaptive Systems: Fundamentals Through Simulations. New York: John Wiley & Sons, 1999
13. Siegel, S. Nonparametric Statistics for The Behavioral Sciences. 1. ed. New York: McGraw Hill (1956).
14. Hong, W. Application of chaotic ant swarm optimization in electric load forecasting. Energy Policy, 38, 5830-5839 (2010)