On the similar-day approach to energy load forecasting

Francesco Cordoni[a]

June 2, 2017

Abstract

In the present paper we review some important fact of energy markets, focusing in particular on the problem of energy load forecasting. This topic is a central one in recent energy market related literature for various reasons. Different methods have been proposed to tackle the above problem, we propose here a method simply based on similar days. This methods can be seen to have comparable accuracy to standard methods used but also it is incredibly easy to implement and fast to calibrate.

1 Introduction

Since the deregulamentation of energy markets, *energy load forecasting*, and strictly related to it *spot prices forecasting*, has become a key ingredient of any energy market related operation. A consistent part of the literaute concerning energy markets has been devoted to the modellization and to the consequent forecasting of short term quantities, either energy loads or spot prices, see, e.g. [7, 19]. The present is a technical paper which aims at focusing on the problem of energy load forecasting, with particular references to *similar-day techniques*.

In recent literature several methods have been proposed in order to tackle the non-trivial task of produce an accurate forecast of future energy loads. The most typical methods consists in ARIMA(X) models whereas more recently advanced technique of neural networks have been succesfully applied to the current task, see, e.g. [2, 17] or also [6] for neural network applied in diffrent areas of modern financil mathematics. In general it has been reported that both the aforementioned methods have performance of around $2-3\%$ *Mean Absolute Percentage Error* (MAPE). We stress that, as regard the importance of being able to produce accurate load forecasting, altough it is difficult to give a precise estimate of reveniews an accurate load forecast can give, as reported in [11], savings from a 1% reduction in the (MAPE) of short-term load forecasting can yied up to 300.000 $ per year.l

[a] HPA s.r.l., Via Guglielmo Marconi, 28, 38052, Caldonazzo (TN); francesco.cordoni@hpa8.com

[a] Department of Computer Science, University of Verona, Strada le Grazie, 15, 37134, Verona; francescogiuseppe.cordoni@univr.it

Clearly these methods have good performances, but they usually have the drawback to be costly to be properly calibrated. Nevertheless it always appears in literature how a naive method, that takes only into account day of the week, can produce accurate forecasts whose results can be compared to results of previous methods. One can also show how even the more sofisticated statistical methods, if not well calibrated, can be beaten by the previously mentioned naive method. This naive method simply consists in predicting monday, resp. saturday, resp. sunday, with the quantity of previous monday, resp. saturday, resp. sunday, whereas one predict tuesday, wednesday, thursday and friday with the quantity of previous monday. This method implies the idea that day of the week considerably affect the energy load, so that days that are considered *similar* will most likely have *similar* loads.

We aim here at extending and formalizing the previously introduced notion of *similar day* taking into account not just the day of the week but also holidays and most important our method will include some relevant weather factors. Of course including weather factors implies that one has to properly understand how different weather factors affect the effective loads and most important one has to find which factors most influence the next days load. Notice that this factors might not be universal in the sense that different geograpical locations might have different driving factors, we refer to [18] for a deep study on the selection of suitable weather factors.

2 On the main weather factors

In electricity load forecasting, climatic factors have always been recognized among the major variables that affect the load curve. Among these factor let us mention for instance temperature, being the most simple to measure and perhaps also to forecast, or also rainfall, humidity, wind speed and solar irradiation. We have to stress also that dependending on different geographical areas, electricity loads can be influenced by different climatic conditions, so that it is impossible to give a universal list of the most important weather factors.

Besides weather effects electricity load exibhits a clear *calendar effect*, in the sense that electricity load changes according to the day of the week or it also changes between working days and holidays. This fact makes electricity loads difficult to forecast, since some days have a peciliar curve which is different from any other day, with a consequent significant presence of outliers.

Before introducing in details our method, let us see in more details the relation between electricity loads and some relevant weathar factors, in particular we will consider temperature. In fact, despite we have just said that a list of main weather factors universally valid is impossible to draw up, we can also safetly say that temperature is perhaps the most important weather factor, accounting for more than 70% of the load varianc, we refer to [5, 9].

Next analysis focus on the norther zone in Italy during the year 2012. Also we have choosen as the main factors that may affect the final loads temperature, rainfall and working days.

Figure 1 shows the 2012 hourly times series of energy load compared with temperatue and rainfall during the year 2012. A correlation between the three time series can be immediately noticed and one can also see how previous discussion is in fact reflected in electricity loads curve.

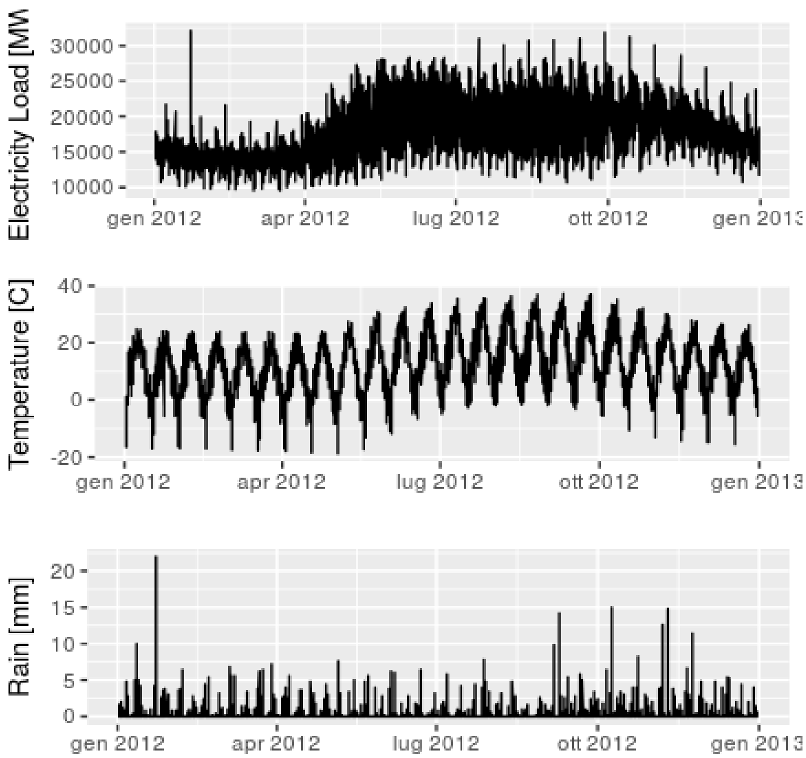


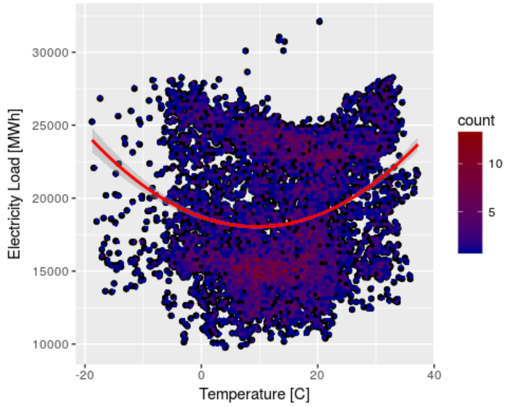
Figure 1: Comparison on the same period of electricity load (top panel), temperature (central panel) and rainfal (bottom panel)

We remark that as standard in energy load forecasting we will not consider the hourly time series but instead we will consider 24 dinstinct time series, each for a given hour of the day, so that in what follows we will perform a 1 step ahead forecast for any hour, and then we will reconstruct the whole load curve for the next day. This procedure can be seen to outperform techniques that cosndier the whole hourly load time series. We also stress that one usually have to consider some seasonal pattern in load time series, as can be clearly seen in the above figure 1, nevertheless we will not enter in this details here, leaving the delicate argument to a future work.

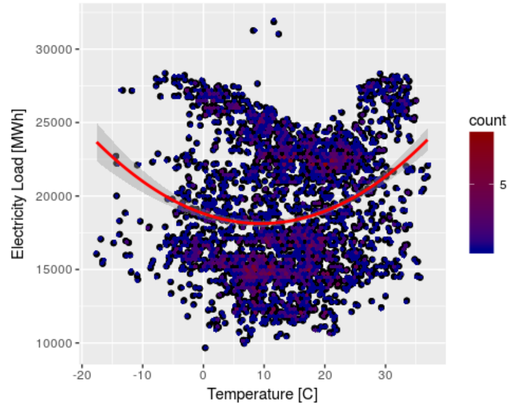
Next figure 2 shows the scatter plot of electricity loads against temperatures for working days and holidays. It can be clearly seen a non-linear relation, where instead the typical $V-$ form of the scatter plot suggest a quadratic response seems. This fact implies that when one is to forecast future electricity loads, one has to carefully calibrate the model with particular regard to this non-linear response. Also the same figure highlights the presence of some outliers, other key fact that one has always to carefully consider in electricity forecasting.

3 The similar-day method

As mentioned above we have choosen here temperature and rainfall as factors that may affect hourly electricity load. There are many other factor that clearly has an impact on



(a)



(b)

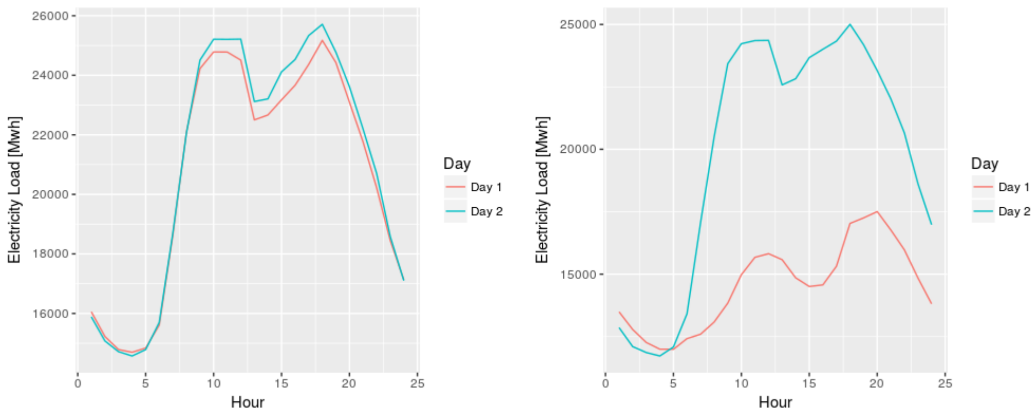
Figure 2: Scatter plot of Temperature against loads for working days (left panel (a)) and holidays (right panel (b))

loads, one of the most used is perhaps solar radiation. The best choice of external variables is a key topic in order to have the best possible forecast. This point is skipped in the present work, but we want to stress once more that, in order to have a robust forecasting procedure, the choice of regressors variable is a delicate point that one cannot neglet, we refer to [18] for more details on the topic.

We are now to briefly introduce our forecasting procedure based on *similar day*. We first select a suitable time window of N days which will constitute the training data set. For each day d of the training period we create a vector c_d representing the characteristic of the day d. In our analysis c_d will be a three-dimensional vector where the first component is a dummy variable which consider if d is a working day or not, the second variable represents the temperature whereas the third component takes into account rainfall. We stress that, due to the non-linear response of loads and temperatures one has to carefully choice how to translate the real temperature for day d into a suitable variable that consider also the quadratic response.

Thus the same characteristic vector is computed for day h once whishes to predict the load. Then, considering the characteristics of each day, we calculate how much every day d in the training window is different from day h. Therefore one can immediately see which days are more similar to the day to be forecasted. We have to stress that in order to see how similar two days can be, we have to properly decide how a given variable may affect the final outcome, in the sense that if one believes that temperature, rather than rainfall, is more important in the final determination of load, one has to properly assign a weight to the temperature variable so that the forecasting value will depends more on similar temperature than similar rainfall.

Once the *similarity* for each day is computed one choose a suitable selection of the most similar days in order to compute the electricity load for day h as a weighted average of the most similar days among the N–long training set; in particular the weights are decided in accord to the similarity value previously calculated. Figure 3 shows the load curve for days which are considered similar and the load curve for days that on the contrary differs. In



(a) (b)

Figure 3: Load curve for two *similar* days (left panel (a)) and load curve for two *different* days (right panel (b))

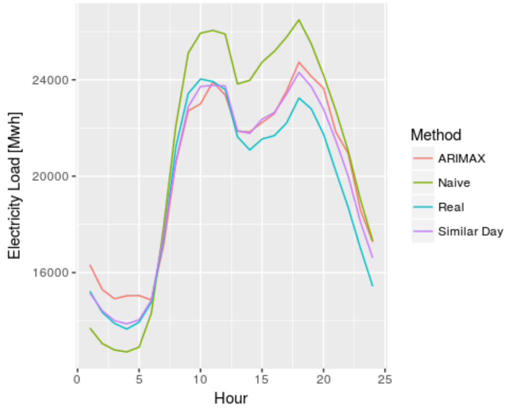
particular in the right panel 3a we can see the load curve for two different days d_1 and d_2 that are both working days and with no rainfall, with a difference in the registered temperature of $1.1\,C$. On the right panel 3b on the contrary it can be seen the curve load for two days that differs in all the three component of the characteristic vector.

Clearly 3 shows a strong dependance in the resulting load curve with respect to some weather variables. This suggests in particular that a forecasting method based on weather and daily variables might have a good performance.

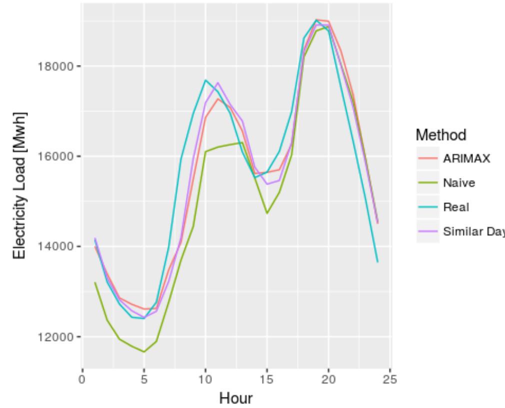
Peformances evaluation

We are now to evaluate performances of the *similar day* method just introduced in Section 3. In particular we compare our *similar day method* with a classical ARIMAX model, where external regressors are the same used to compute similar day forecast, i.e. temperatures, rainfall and working days. Althoug its simplicity, it can be easily seen how ARIMAX model has a good performance, see, e.g. [19]. In what follows we will measure the accuracy of the forecast using the *Mean Absolute Percentage Error* (MAPE), which is defined as the daily average percentage error. Also we will compare our method to the naive method mentioned above, which we think it represents the real benchmark for our method. We recall that the naive method works as follows, we predict monday, resp. saturday, resp. sunday, with the quantity of previous monday, resp. saturday, resp. sunday, whereas one predicts tuesday, wednesday, thursday and friday with the quantity of previous monday.

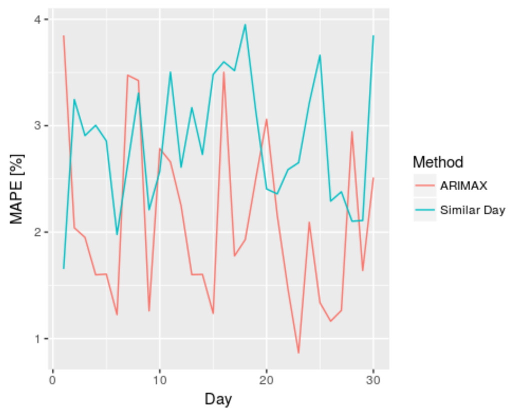
Figure 4 shows the performances of the methods under analysis. In particular in figures 4a–4b it can be seen the load curve forecasted compared to the real load curve. It emerges how the naive method is the worst performing one, whereas ARIMAX model and similar day model have similar predicted curve. Similar argument can be also evinced from figure 4c, where the MAPE for the ARIMAX model and the similar day method is plotted over a 30 days rolling period.



(a) Load curve forecast



(b) Load curve forecast



(c) MAPE for a rolling period of 30 days of ARIMX method and similar day method

Figure 4

Conclusions

The present paper has to be intended as a quick introduction to electricity load forecasting focusing mainly on the similar day methods. In particular it can be seen how different weather factors may affect the energy load, fact that suggests that using such weather variables in the forecasting procedure may lead to accurate results. In particular it has been shown how a forecasting procedure based on a weighted average of suitable past days that are considered to be *similar* to the day one wishes to forecast, has performances that can be compared to more sofisticated statistical models such as the ARIMAX model.

Nevertheless the present paper aims at showing how a forecasting method based on similar can be highly useful. First, it does not require any specific package or software, but it can be easily implemented in any program one wishes. This has the immediate result of being incredibily fast to compute, so that it works easily also on high dimensional data set. Also, and most important, one can calibrate the similar day methods taking into account specific holidays so that the method is accurate also when forecasting the load curve in some particular holidays, where other methods usually show weakness.

To conclude we think that the above mentioned advantages of the similar day method can be exploited mainly in two field of energy forecasting. Firt it provides a really good benchmark, proving to be incredibly more accurate than standard naive methods, so that one can use the similar day method to compare the goodness of different implemented methods. Second, and major utility of the present techniques, it can be used in forecast combination. In fact it appears that usually the best performing method is not a single one but rather a suitable combination of different accurate methods. The easiness of implementation and the fast calibation make the *similar day* method a perfect candidate to be added in a combination of different forecasts. We stress that this latter point is a central point of recent investingation, see, e.g. [20]. Empirical evidences show that often combining different forecast has the advanges of stabilizing the curve prediction with the immediate consequence of a drop in the overall percentage error. Given the importance of the argument we aim at treating the argument of *forecast combination* in more details in a future work.

References

[1] R. Aïd; Electricity derivatives. Springer International Publishing, 2015.

[2] A.G. Bakirtzis, V. Petridis, S.J. Kiartzis, M.C. Alexiadis and A.H. Maissis; *A neural network short term load forecasting model for the Greek power system.* IEEE Transactions on power systems 11.2 (1996): 858-863.

[3] M. Bierbrauer, S. Trück and R. Weron; *Modeling electricity prices with regime switching models*. Computational Science-ICCS 2004 (2004): 859-867.

[4] A. Cecchin, F. Cordoni and L. Di Persio; *Novel approaches to the energy load unbalance forecasting in the Italian electricity market.* Journal of Mathematics in Industry 7.1 (2017): 5.

[5] T.H. Dang-Ha, The-Hien, F.M. Bianchi and R. Olsson; *Local Short Term Electricity Load Forecasting: Automatic Approaches.* arXiv preprint arXiv:1702.08025 (2017).

[6] L. Di Persio and O. Honchar; *Artificial Neural Networks architectures for stock price prediction: comparisons and applications.* International Journal of Circuits, Systems and Signal Processing 10 (2016): 403-413.

[7] T. Hong; *Energy forecasting: Past, present, and future.* Foresight: The International Journal of Applied Forecasting 32 (2014): 43-48.

[8] T. Hong and S. Fan; *Probabilistic electric load forecasting: A tutorial review.* International Journal of Forecasting 32.3 (2016): 914-938.

[9] T. Hong, P. Pinson and S. Fan; *Global energy forecasting competition 2012.* (2014): 357-363.

[10] T. Hong et al. *Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond.* (2016): 896-913.

[11] T. Hong; *Crystal ball lessons in predictive analytics.* EnergyBiz Mag (2015): 35-37.

[12] R.J. Hyndman and S. Fan. *Density forecasting for long-term peak electricity demand.* IEEE Transactions on Power Systems 25.2 (2010): 1142-1153.

[13] E. Lorenz, et al. *Regional PV power prediction for improved grid integration.* Progress in Photovoltaics: Research and Applications 19.7 (2011): 757-771.

[14] J. Janczura and R. Weron. *Efficient estimation of Markov regime-switching models: An application to electricity spot prices.* AStA Advances in Statistical Analysis 96.3 (2012): 385-407.

[15] J. Janczura and R Weron. *An empirical comparison of alternate regime-switching models for electricity spot prices.* Energy economics 32.5 (2010): 1059-1073.

[16] A. Misiorek, S. Trueck and R. Weron. *Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models*. Studies in Nonlinear Dynamics & Econometrics 10.3 (2006): 2.

[17] D.C. Park, et al. *Electric load forecasting using an artificial neural network*. IEEE transactions on Power Systems 6.2 (1991): 442-449.

[18] B. Uniejewski, J. Nowotarski and R. Weron. *Automated variable selection and shrinkage for day-ahead electricity price forecasting*. Energies 9.8 (2016): 621.

[19] R. Weron. Modeling and forecasting electricity loads and prices: A statistical approach. Vol. 403. John Wiley & Sons, 2007.

[20] R. Weron. *Electricity price forecasting: A review of the state-of-the-art with a look into the future*. International journal of forecasting 30.4 (2014): 1030-1081.

[21] R. Weron, M. Bierbrauer and S. Trück. *Modeling electricity prices: jump diffusion and regime switching*. Physica A: Statistical Mechanics and its Applications 336.1 (2004): 39-48.