Studies on Several Issues of Electricity Price Forecast

Wei LIU, Huanyue LIAO
China-EU Institute for Clean And Renewable Energy
Huazhong University of Science and Technology, Wuhan 430074, China
liuw1204@hust.edu.cn

Abstract. The main models used in electricity price forecast have been reviewed in this paper. Several statistical models based on machine learning algorithms have been tested and compared to forecast the next-day hourly electricity price in Nordpool electricity market. Stationarity of electricity price has also been studied and it is concluded that electricity price is daily seasonal. Many features of both local region and neighboring regions from the past week have been respectively analyzed in terms of the importance to electricity price forecast. Results show that local zone, compared with neighbor zones, exerts the major influence on electricity price. Meanwhile, price is the most important category other than categories like demand, wind generation or transmission congestion rate. Hydro reserve is also an important parameter to electricity price forecast.

1. Introduction
As a very special commodity, electricity differs from other commodities in many aspects. Firstly, there must be a constant balance between production and consumption [1]. Moreover, both electricity consumption and renewable production are highly dependent on weather conditions and human activities. These unique and specific characteristics lead to price dynamics not observed in any other market, exhibiting seasonality at the daily, weekly and annual levels, and abrupt, short-lived and generally unanticipated price spikes. At the same time, over-contracting and under-contracting will both lead to a giant economic loss, which makes electricity price forecast so important.

Seeing the importance of electricity price forecast, numbers of models have been developed to forecast the hourly electricity price in a day-ahead market [2–11]. Some of them achieved a decent forecast accuracy with relatively small errors. However, authors think it’s also necessary to study the influence of different features on electricity price forecast so that researchers can be aware which features should be taken into consideration when forecasting electricity price.

This paper is organized as follows: section 2 introduce the methodology of this paper including algorithms comparison in section 2.1 and features importance in 2.3. Section 3 express the case and results analysis for each part in section 2. Finally, several conclusions are made in section 4.

2. Methodology

2.1. Model Selection
In order to analyze which machine learning algorithms are better suitable for hourly electricity price forecast in a day-ahead market and decide which algorithms to use to study the stationarity and importance of features, the common machine learning algorithms are respectively applied to make the forecast and then the performances are compared and analyzed. In order to evaluate the algorithms’
performance, several metrics are selected which are root of mean square error (RMSE), mean absolute error (MAE) and executing time.

So as to reduce coincidences and make the results more convincible, each algorithms were tested with three different inputs. For each input, each algorithm was tested three times and the average was calculated to evaluate the performance. In order to compare the performances quantitatively, a new score which is the linear combination of RMSE, MAE and executing time is created, shown in Equation 1.

\[
\text{Score} = MAE + RMSE + \frac{\text{Time}}{100}
\]

Equation 1

2.2. Stationarity analysis
Stationarity implies that the time series remains at a fairly constant level over time. If a trend exists, like constantly increasing or decreasing, then the series is not stationary. The data should also show a constant variance in its fluctuations over time. This is easily seen with a series that is heavily seasonal and growing at a faster rate. In such a case, the ups and downs in the seasonality will become more dramatic over time. Without these stationarity conditions being met, many of the calculations associated with the process cannot be computed.

In order to prove that electricity price is daily seasonal, two kinds of time series inputs were set and the performances were compared. One input consists of all the hours in the past week while the other one is only involved with the corresponding hour in the past week (24h, 48h, 72h, 96h, 120h, 144h and 168 h ago). For example, the electricity price at 14:00 tomorrow is being predicted, the second time series input is the data only at 14:00 every day in the past week, rather than the data at every hour in the past week.

2.3. Features importance
In order to test the relative importance of each of these relevant and available features, the algorithms are tested repeatedly with one of the features removing each time. For each feature, the algorithm is tested three times as before, then average of MAE and RMSE are computed and the sum of them was taken as the metric to evaluate the performance. The features in sequence are hourly features (price, load, congestion ratio of each transmission line, wind generation) of each zone and weekly features (hydro reserve in Norway and Sweden), from seven days ago to yesterday in order.

After studying the relative importance of each feature among these features, the importance of the categories, zones and lagging days are further analyzed by removing all the features from one category or one zone.

3. Case study and result analysis
In Nordpool electricity market, each country may be divided into several zones and there is a hourly market clearing price (MCP) for each zone. Datasets from 2014-05-12 to 2016-12-31 are chosen as the train sets to train the models while datasets from 2017-01-01 to 2017-07-23 are chosen as test sets to evaluate the performance of algorithms. All the datasets are available on the official website of Nordpool [12]. Hourly demand, price, transmission capacity, power flow, wind generation for each zone are available as well as weekly hydro reserve for each country. The study zone is Oslo, Norway and the information from Oslo and interconnected zones are all taken into consideration in this paper.

When choosing forecast models, only features in the studied zone (Olso) are taken into consideration. In terms of categories, only price and demand are taken as studied features. And the data in the whole past week are chosen and the inputs are historical prices, results are shown in Table 1. LR represents linear regression while RR represents ridge regression. Neural network is abbreviated as NN and decision trees is abbreviated as DT. SVM is the acronym of support vector machine. Bagging at the suffix means the ensemble method bagging and suffix Ada means adaptive boosting. Moreover, DT_Forest stands for model random forests.

Overall, it can be concluded that RR_Bagging, RR, SVM and DT_Forest are the best models with relatively small scores.
Table 1 Comparison of algorithms (€/MWh)

| Algorithms    | Score | Algorithms    | Score |
|---------------|-------|---------------|-------|
| SVM           | 4.12  | DT Ada        | 4.61  |
| RR            | 4.03  | LR Bagging    | 5.24  |
| RR Bagging    | 4.00  | LR            | 5.56  |
| DT Forest     | 4.30  | NN            | 6.17  |
| DT Bagging    | 4.40  | RR Ada        | 6.10  |
| LR Ada        | 9.74  | DT            | 5.48  |

Ridge regression bagging is again tested with only relevant hours on week ago. The score for forecast with relevant hours is only 3.91(€/MWh), which is 2.3% better than 4.00 (€/MWh), the score for forecast with all the hours. Therefore, it can be concluded that only relevant hours are necessary features for making forecasts.

Features can be classified by several aspects including lagging days, zones and categories (price, demand, ratio, hydro and wind). Note that, zone 1 refers to Oslo while zone 2, zone 3, zone 5, zone 8 refers to respectively Kristiansand, Molde, Bergen and Stockholm. Hydro reserve in Norway or Sweden are classified into Hydro in categories, NO or SE in zones and 0 in lagging days.

Firstly, in terms of the categories (price, demand, ratio, wind and hydro), all the scores are computed respectively when removing one feature from each category. After that, the average value of all the scores is computed for each category. The average score for each category is drawn in Figure 1 where category All means the error with all the features, which means no feature is removed.

![Figure 1 Removing one feature in each category](image1.png)

![Figure 2 Removing all features in one category](image2.png)

Figure 1 indicates that removing the hydro or price is enlarger the errors while removing ratio, wind generation or load is improving the algorithm in various degree. To verify this conclusion, all the features relevant to each category are removed and retested and the results are exhibited in Figure 2. Clearly, removing price or hydro reserve is worsening the performance while removing other categories is improving the performance in various degree, verifying that the previous conclusion is correct. Specifically, compared to other categories, removing the price will vitally worsen the performance of electricity price forecast model. It can therefore be concluded that price is the most important category compared to other categories.

The same as categories, the average score which removes one feature that belongs to each zone is also computed and the results are shown in Figure 3. Similarly, zone All means no zone is removed. Zone NO and SE stand for removing hydro reserve in Norway or Sweden.

![Figure 3 Removing one zone](image3.png)
Figure 3 indicates that the average value of all the errors those are computed when feature that belongs to zone 1 is removed one by one is much higher than any other zones. It can be therefore concluded that zone 1, which represents local zone where the electricity price is being predicted, is much more important than any other zones. Moreover, compared to the case with all the zones considered, removing zone 1 or hydro reserve in Norway is deteriorating the algorithm while removing other zones is actually strengthening the algorithms. To verify this conclusion, all the features relevant to each zone are removed and retested and the results are exhibited in Figure 4. Clearly, removing features in zone 1 or hydro reserve in Norway is worsening the performance while removing other zones is improving the performance, verifying that the previous conclusion is correct.

4. Conclusions
By comparing twelve machine learning algorithms, it can be concluded that ridge regression, SVM, random forests and ridge regression bagging are the best suitable algorithms for electricity price forecast. In particular, ridge regression bagging is the best one.

With respect to the stationarity of electricity price, electricity price in Oslo indeed exhibits a seasonality in daily level, proving its stationarity in time series.

After carefully studying the relative importance of different categories and zones, several conclusions can be made. In Oslo, the electricity price is mainly relevant with the information from local zone. The information from neighboring regions are exerting very limited influence on the electricity price in Oslo. Moreover, in terms of categories, electricity price in Oslo is principally effected by previous electricity price and very slightly influenced by other features like wind generation or load. Apart from historical price, hydro reserve is also an important feature for electricity price forecast.

References
[1] M. Shahidehpour, H. Yamin, and Z. Li, “Frontmatter and Index,” in Market Operations in Electric Power Systems, John Wiley & Sons, Inc., 2002, pp. i–xiv.
[2] S. K. Aggarwal, L. M. Saini, and A. Kumar, “Electricity price forecasting in deregulated markets: A review and evaluation,” Int. J. Electr. Power Energy Syst., vol. 31, no. 1, pp. 13–22, 2009.
[3] M. Stevenson, “Filtering and Forecasting Spot Electricity Prices in the Increasingly Deregulated Australian Electricity Market,” Quantitative Finance Research Centre, University of Technology, Sydney, Research Paper Series 63, Sep. 2001.
[4] E. Ni and P. B. Luh, “Forecasting power market clearing price and its discrete PDF using a Bayesian-based classification method,” in 2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194), 2001, vol. 3, pp. 1518–1523 vol.3.
[5] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, “ARIMA models to predict next-day electricity prices,” IEEE Trans. Power Syst., vol. 18, no. 3, pp. 1014–1020, Aug. 2003.
[6] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, “Day-ahead electricity price forecasting using the wavelet transform and ARIMA models,” IEEE Trans. Power Syst., vol. 20, no. 2, pp. 1035–1042, May 2005.
[7] A. M. Gonzalez, A. M. S. Roque, and J. Garcia-Gonzalez, “Modeling and forecasting electricity prices with input/output hidden Markov models,” IEEE Trans. Power Syst., vol. 20, no. 1, pp. 13–24, Feb. 2005.

[8] O. Abedinia, N. Amjady, M. Shafie-khah, and J. P. S. Catalão, “Electricity price forecast using Combinatorial Neural Network trained by a new stochastic search method,” Energy Convers. Manag., vol. 105, no. Supplement C, pp. 642–654, Nov. 2015.

[9] J. Che and J. Wang, “Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling,” Energy Convers. Manag., vol. 51, no. 10, pp. 1911–1917, Oct. 2010.

[10] X. Qiu, P. N. Suganthan, and G. A. J. Amaratunga, “Short-term Electricity Price Forecasting with Empirical Mode Decomposition based Ensemble Kernel Machines,” Procedia Comput. Sci., vol. 108, no. Supplement C, pp. 1308–1317, Jan. 2017.

[11] R. Deb, R. Albert, L.-L. Hsue, and N. Brown, “How to Incorporate Volatility and Risk in Electricity Price Forecasting,” Electr. J., vol. 13, no. 4, pp. 65–75, May 2000.

[12] “Historical Market Data.” [Online]. Available: http://www.nordpoolspot.com/historical-market-data/. [Accessed: 11-Oct-2017].