

Improve forecasting of daily load

Boris Bizjak

Faculty of Electrical Engineering and Computer Science
University of Maribor
Koroška cesta 46, 2000 Maribor, Slovenija
E-pošta: boris.bizjak@um.si

Improve forecasting of the daily load

Abstract. In this paper, we discuss the daily hourly load forecasting of load for an industrial complex using ARIMA methodology with one predictor. For predicting the daily load, we need to make a forecast of 24 steps in the case of the 1-hour prediction interval. The story of predicting in the article is due to trading in electricity. The basis for the forecasting is the time series of energy consumption data, which is a time series from 15 min values.

The presented load forecasting model needs at the learning phase two input historic time series: Energy consumption and real production data. We get production data (predictor) from the Production Planning Department. We wanted to use a simple and robust predictor, which not only has a theoretical value for forecasting, but also a practical utility to be used in practice. So, the predictor has a value of 1 when the furnace is running and 0 when the furnace is not in operation. For predictor type 1, the time resolution at the predictor is 1 hour, and for predictor type 2 the time resolution at the predictor is 15 min. The predictor also follows production with different dynamics. Type 1 has a constant value of 1 for all the duration interval of the arc furnace working, and predictor type 2 follows the arc furnace cycle more dynamically with fluctuation between 1 and 0.

To find the optimal predictor we used a similar spectral density method of two time series. We called it predictor type 2. In the case of the 1-hour prediction step with predictor type 2, by linear regression, the independent variable of production explains 84% of the variance in the load, MAPE is 23% and RMSE 3389. Additionally, energy consumption at the industrial plant had a seasonal day, hours feature. Therefore, we improved forecasting performance - statistical indicators, using ARIMA methodology, by approximately 15%.

Finally, the ARIMA model, with one predictor type 2, explains 95% (R^2 0.95) of the variance in the load, with MAPE 9% and RMSE 1982.

1 Introduction

Our example of predicting the daily power flow for the industrial complex has a specific large electric arc furnace (UHP). In addition, other companies within the

industrial complex operate according to the seasonal model. Therefore, three prediction methods should be used in order to make good predictions: ARIMA methodology, seasonal model and predictor of the operation of the arc furnace. In the article, we first show how to determine a predictor that illustrates the operation of an arc furnace. The final result is a forecast of the daily power flow using the selected predictor, ARIMA and the seasonal model.

2 Industrial complex and steel production

The beginnings of the ironworks in Koroška go back to the year 1620. Today, there are several companies in the area of the former Ravne Ironworks, employing approximately 3,000 people. The largest company on the site is Metal Ravne with approximately 900 employees. Metal Ravne is the largest, and the largest consumer of electricity. The company consists of a steelwork, a rolling mill and electro-smelting under slag. In the steel plant, the basic unit is a 45-tonne electric UHP oven and a vacuum refill kiln for castings of classical ingots. In the Electro-under-slip section under the slag, 36-tonne and 3-tonne ESR devices are in use. At present, it operates on the scale of an industrial complex of 20 companies.

The melting process [1] is always carried out in an arc furnace with reduced voltage, since the conditions for burning the arc are poor in a cold cartridge; the ignition of the arc is carried out in such a way that the graphite electrode is lowered to the cartridge until it touches it, and until contact is reached with the other electrodes with the cartridge. At the discharge of the electrode, then an electric arc is triggered - like the firing of the arc during manual arc welding. Because of this, the current size changes from the short-circuit current through the rated power to the zero current at the end of the arc. We say that the arc furnace is operating restlessly at the beginning of melting. Due to the formation of the first melt at the bottom of the furnace, the conditions for burning the arc are improved due to good ionization conditions, so we increase the voltage of the arc gradually and the power of melting to full power: This is always the largest when melting the cartridge when there is already a melt on the bottom of the furnace. We say that we are melting with a hidden arc, which radiates at full power in the crater, which the boulder has drilled into the plunged insert of old iron. In the further heating of the melt, or in maintaining its temperature, the power of the furnace is

significantly lower. The characteristics of the electric arc must be different in this situation, since the arc can now freeze to the walls and the furnace vane.

3 Forecasting daily power consumption per hour interval

Due to the way of trading, the forecast of the daily flow of electricity flows requires a forecast for at least 24 hours, even better in 48 steps (48 hours). We want to predict power flow on the common energy supply of the industrial complex Ravne. The structure of total energy consumption for the Ravne industrial complex is the sum of 20 different companies. The biggest, 75% of the energy consumers, is the Metal Ravne Steel Mill. The dominant energy consumer in the steel industry is the UHP electric arc furnace, the second largest consumer is the oven for overheating LF + VD, and then the other technological line of the steel mill, which follows the melting process of old iron. The remaining 25% of consumption, on a common energy connection, is represented by the other companies, with a typical spending profile of 7 days a week, of which 5 are working days. Change of 25% of consumption over a day is relatively low, and therefore it can be predicted very well.

The steel plant does not operate statistically on a random basis, but over the day according to the orders and the price of electricity. Since its consumption is 75% of the entire energy of the industrial complex, it has the greatest influence on the consumption profile on the total energy supply of the industrial complex. The issue is the production time; real production will be moved to the late afternoon, next to the evening (cheap electricity), over midnight, until 9:00 AM the next day. Occasionally, when there is a large volume of orders, the steel plant operates for 24 hours continuously, regardless of the daily price of the electric power unit over a working day. Due to cheap electricity, Saturdays and Sundays also operate for 24 hours. Once a year, the steel works carry out a repair that lasts about a month. Of course, the electric arc furnace does not work then. Power flow in a UHP arc furnace during steel production is more or less constant. The difference between individual loads at the arc furnace is only in what happens after we have finished melting in the arc furnace. Sometimes overheating occurs and sometimes it does not, which means a short leap in the consumption of electricity to the maximum.

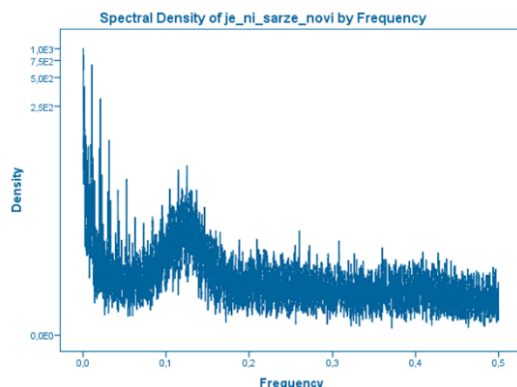


Figure 1: Spectral density for 15 minutes' power flow.

The spectral density of the for 15 minutes' power flow for one year indicates additional periodicity of the signal at frequency 0.12 (Figure 1). Given the spectral density fluctuation for 1-hour average power flow at one year, we have had monotonic falls from low frequencies to the highest frequencies. In other words: 1-hour time series have less frequency content.

The prediction was started with predictor type 1. The predictor has a value of "1" if the hourly average power consumption is > 4500 and "0" if the hourly average power consumption is < 4500 . The predictor's time is coincident with the progress of the blue step line in Figure 2. We reached: The ARIMA model, with one predictor, explains 85% (R^2 0.85) of the variance in the load, with MAPE 13%. The results of predicting with the type 1 predictor were not satisfied.

Therefore, we have introduced a new type 2 predictor with only two amplitude values: 0 and 1. Predictor type 2 is more time-divergent than type 1, so it is better to follow the operation of the electric arc furnace and other technological lines of the steel plant. The predictor has a value of "1" if the 15 minutes' power consumption is > 6500 and "0" if the 15 minutes' power is < 6500 .

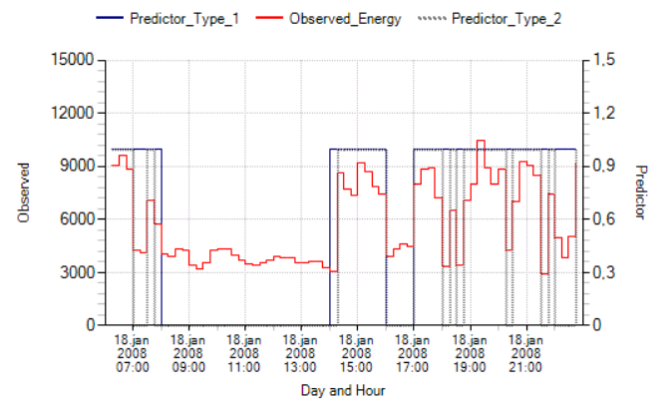


Figure 2: Predictor type 1 and predictor type 2.

To find the optimal predictor we used a similar spectral density method of two time series. Predictor type 2 is an optimal predictor. This is confirmed by the similarity of the density of the frequency spectrum of the time series of electricity consumption and the frequency of the type predictor type 2 for the same time period (Figure 1,3). Also, there were excellent results of predictions using the type 2 predictor.

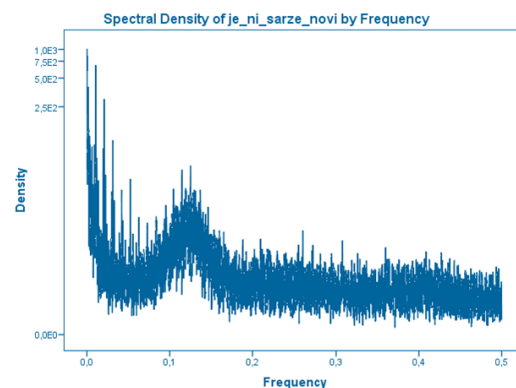


Figure 3: Spectral density for predictor type 2.

In the time series of electricity consumption, the outliers are also noticed; these are values when the consumption of electricity escapes from the standard Gaussian distribution and is close to zero. Such values are expressed on May 1 and December 31. These two cases were not addressed specifically in the forecast itself, but, in any case, they have a negative impact on forecasting performance statistics.

Various statistics are used to evaluate the performance of forecast models. We decided to use in the article only MAE, MAPE, RMSE and R^2 to comparison between individual solutions.

The Mean Absolute Percentage Error (MAPE), is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses accuracy as a percentage, and is defined by the formula:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{Ai - Fi}{Ai} \right|$$

where Ai is the actual value and Fi is the forecast value. Problems can occur when calculating the MAPE value with a series of small denominators. A singularity problem of the form 'one divided by zero' and/or the creation of very large changes in the Absolute Percentage Error can occur, caused by a small deviation in error.

The Root-Mean-Square Error (RMSE) is a frequently used measure of the differences between values predicted by a model and the values observed. The RMSE represents the sample Standard Deviation of the differences between predicted values and observed values. These individual differences are called residuals when the calculations are performed over the data sample that was used for estimation and are called prediction errors when computed out-of-sample. RMSE is a measure of accuracy, to compare forecasting errors of different models for data and not between datasets, as it is scale-dependent. RMSE is sensitive to outliers.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Ai - Fi)^2}{n}}$$

R squared in statistics, the coefficient of determination, denoted R^2 , is the proportion of the variance in the predictable variable Fi that is from the actual value Ai :

$$SS_{tot} = \sum_{i=1}^n (Ai - \bar{Ai})^2$$

$$SS_{res} = \sum_{i=1}^n (Ai - Fi)^2 = \sum_{i=1}^n e_i^2$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

For the described type of industry complex we started forecasting continuously with linear regression [2]. This is a robust method that has a weakness in the limited level

of confidence in the forecast. We calculate the 1-hour step predictor type 2 from the 15-minute time series with aggregate functions. The predictor's time is coincident with the progress of the grey step line in Figure 2. In the case of the 1-hour prediction step with linear regression, the independent variable of production data at the arc furnace now explains 84% of the variance in the load, which is highly significant, and the F-test says we can trust β_0 and $\beta_1 > 99.9\%$, MAPE 23% and R square 0.84.

Table 1: Daily load forecasting at linear regression.

R square	MAE	MAPE	RMSE	Number of predictions
0,84	2700,62	22.84 %	3388,97	8543

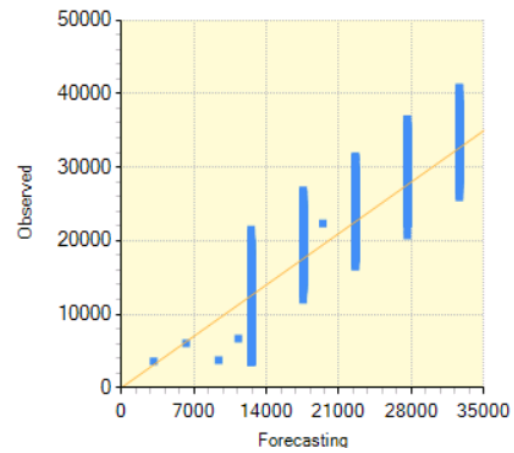


Figure 4: Daily load forecasting at linear regression.

Figure 4 show a typical chart of a prediction with linear regression. A forecast with linear regression could be improved by changing the predictor's timing or amplitude, which means that the on / off principle would be adapted more often to actual production or that the amplitude would follow a production value with not only 0 or 1, but, for example, "0.9", "1" and "1.1". We achieved this by proceeding with the forecast in 15-minute steps, and then we aggregate the 15 min prediction to a 1 hour prediction step. We do this because we believe that such a predictor can be expected from the Production Planning Department.

Table 2: Forecasting daily power consumption per hour at ARIMA methodology.

Hour +	R square	MAE	MAPE	RMSE	Number of predictions
1	0,95	1386,93	6.81 %	1965,98	504
2	0,94	1455,65	7.33 %	2060,92	504
3	0,94	1534,3	7.8 %	2106,36	504
4	0,94	1616,43	8.26 %	2149,73	504
5	0,93	1675,69	8.78 %	2224,52	504
6	0,93	1711,12	9.08 %	2238,19	504
7	0,93	1748,47	9.28 %	2268,13	504

8	0,93	1756,52	9.51 %	2280,33	504
9	0,93	1759,68	9.47 %	2280,7	504
10	0,93	1789,77	9.76 %	2317,68	504
11	0,93	1828,01	9.83 %	2360,4	504
12	0,93	1837,03	9.91 %	2364,44	504
13	0,93	1824,9	9.89 %	2330,4	504
14	0,92	1855,75	10.02 %	2387,83	504
15	0,93	1816,1	9.87 %	2330,17	504
16	0,93	1845,7	10.03 %	2352,68	504
17	0,93	1795,34	9.73 %	2276,53	504
18	0,93	1794,97	9.77 %	2286,87	504
19	0,93	1813,72	9.91 %	2295,4	504
20	0,93	1814,17	9.77 %	2322,71	504
21	0,93	1758,78	9.7 %	2262,13	504
22	0,93	1781,22	9.81 %	2309,08	504
23	0,93	1780,48	9.86 %	2296,16	504
24	0,93	1786,35	9.99 %	2298,05	504

From linear regression results will be improved with ARIMA metrology [3] and one predictor. Energy consumption at the industrial plant had a seasonal day, hour (7/24) feature and one dominant predictor. The forecasting results in Table 2 show much better: For "24 hour - 24 steps" forecasting the MAPE statistic was 9% and R^2 0.95.

Linear regression has an RMSE constant value of 3388 for all predictions. With ARIMA methodology, for the same historical time series, RMSE ranges between 1386 and 1859, which is a significant improvement. So, the model ARIMA plus predictor is a better prediction model, because the lower RMSE value means a better forecast model. Looking in Table 2 at the 24-hour forecasts, we notice that better forecasts are for the values near 1+ and 24+ hours, which shows the 24-hour periodicity of our time series.

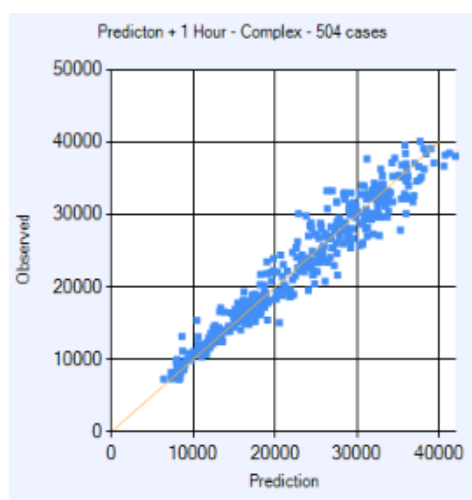


Figure 5: Forecasting daily power consumption per hour with ARIMA methodology, + 1 hour.

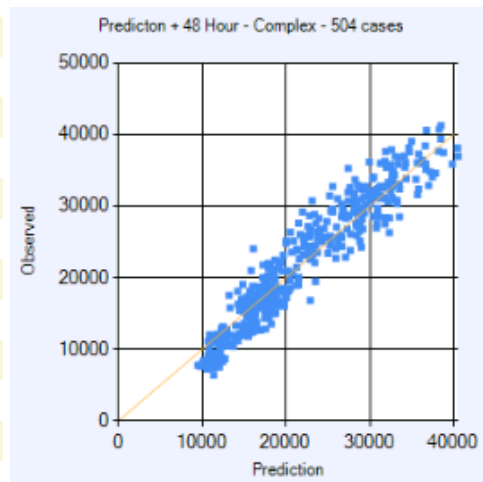


Figure 6: Forecasting daily power consumption per hour with ARIMA methodology, + 48 hour.

Figure 5 and Figure 6 shows that we have improved the forecast for low power consumption dramatically. There is also an improvement in high consumption, but it's not as good as we would like.

3 Conclusion

The test forecasting was carried out over 3 weeks from December 1 to December 21. We predict 48 times for each hour, which means first 48 hours ahead ... and the last announcement is 1 hour ahead. The results are shown in Figure 5,6 (+1 hour, +48 hours) and in Table 2 for each hour plus separately. The advantage of the presented forecast system is its small forecast error. Finally, forecasts for the Ravne industrial complex are implemented with ARIMA methodology, seasonal models and extended with one predictor. The solution is based on modified IBM SPSS and Microsoft software products, and the theory from [2] and special [3]. Models of forecasts are learning dynamically. The prediction models are, thus, optimized constantly according to the current time series. To learn the model, we use the time series of the common energy meter of the Ravne industrial complex and the optional energy meter before the steelmaker Metal Ravne. The prediction database is a SQL server. All images and Tables are part of an existing WEB site application.

References

- [1] Janez Bratina, Elektroobložna peč, Ravne na Koroškem, 1994.
- [2] Douglas C. Montgomery, George C. Runger, Applied statistic and probability for engineers, Wiley, 2003.
- [3] George E. P. Box Gwilym M. Jenkins Gregory C. Reinsel. Time series analysis forecasting and control, Wiley, 2014.