

Wind Power Forecasting: Using JMP® Time Series Analysis for Hourly Prediction of Power Generation

Srinivas Reddy Busi Reddy¹ and Srikar Rayabaram²

¹Management Information Systems, Oklahoma State University, Stillwater, Oklahoma 74078.

²Associate Manager, HR Analytics, Walmart.



Introduction

Forecasting wind power generation is crucial for the power system management or energy trading. Wind power forecasts also serve as key inputs for deciding on the use of conventional power plants and for the optimization of scheduling of these plants. Bids for energy to be supplied on a given day are usually required during the morning of the previous day.

This poster focuses on the use of JMP Time Series Analysis model to predict hourly power generation, based on past power observations and available meteorological wind forecasts for that period. It first explains the various steps involved in preparing and identifying patterns in time series data, the various models build using JMP time series analysis, evaluating the models and finally discuss the results.

Methods

Data Preparation and Analysis:

There were 2 data sets, one containing the historical hourly power generation and the other containing the wind forecasts for the same period. They had to be joined based on the hour of wind forecast with the corresponding power generated at the same hour. Anomalous data were removed from the training set (a) low power with strong wind, (b) high power with low wind. The target variable is log transformed to stabilize the variance and it also helps in making the series more stationary.

Trend and Seasonality Analysis: Trend represents a general systematic linear or nonlinear component that changes over time and does not repeat or at least does not repeat within the time range captured by the data, there is considerable amount of noise in our data so we performed simple moving average smoothing to remove the noise and identify the trend. Seasonality has a formally similar nature as trend, however, it repeats itself in systematic intervals over time usually known as lag. We examined the auto correlation and partial autocorrelation graphs to observe seasonal patterns.

Removing Serial Dependency: Used differencing to remove some of the autocorrelations, this will change other auto correlations, that is, it will eliminate them or it may make some other seasonalities more apparent. It also helps in making the series stationary which is necessary for Arima and other techniques.

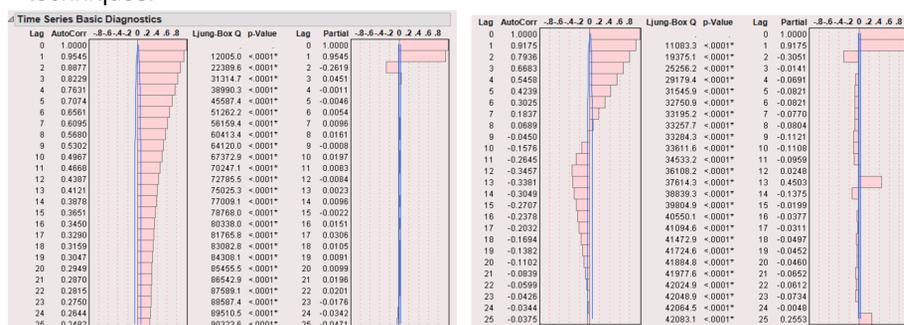
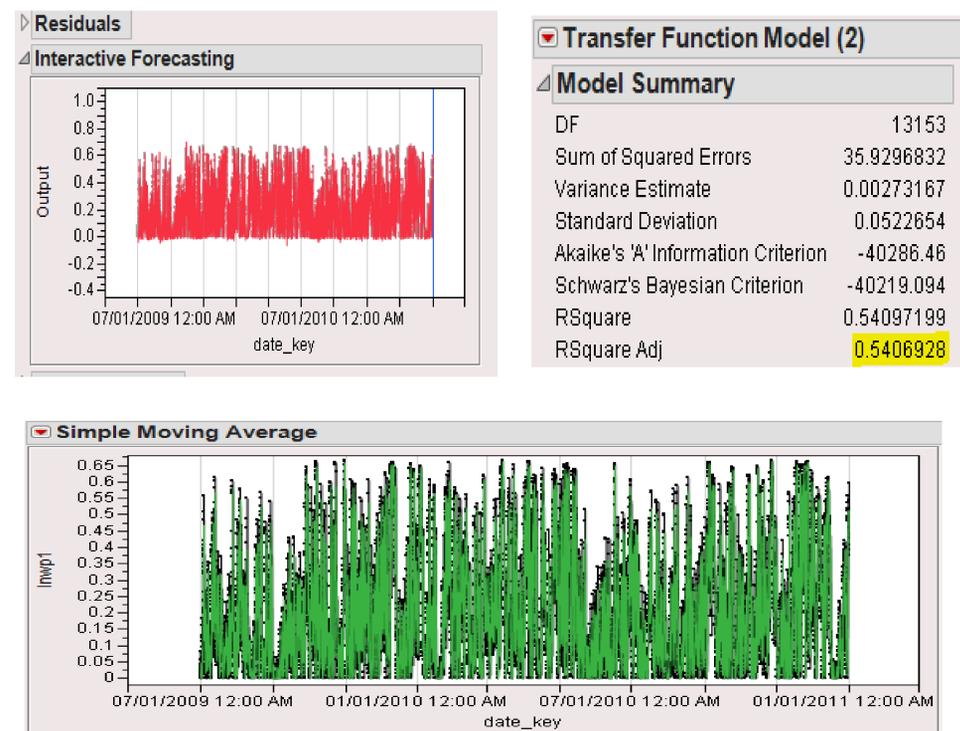


Figure 1. Auto correlation and partial auto-correlation graphs before and after differencing

Models:

After completing data analysis and the next step is the identification process for parameter estimates(auto regressive parameter, moving average parameter and differencing parameter) of the model, major tools used in the identification phase are plots of the series, correlograms of auto correlation (ACF), and partial autocorrelation (PACF). ARIMA transfer function model has been built by including the effects of other input variables, that is, the wind forecasts available for that period.

Results



Discussion

•The period between 2009/7/1 and 2010/12/31 which contains 13175 observations is used as the training period, while the remainder of the dataset, that is, from 2011/1/1 to 2012/6/28, which contain 12745 observations is used for evaluation. Make SAS data step option in JMP creates a SAS code that can be used to score a new data set.

• The RSquare Adjusted value for the model is 0.54, which suggests that the regression line explained variance in the data points reasonably well.

•Another way to evaluate the model performance is to test if the residuals are statistically independent, that is, they contain only noise and no systematic components. We tested it by plotting the residuals and inspect them for any systematic trends and by examining the auto-correlation of residuals, there were no systematic trends and no serial dependencies between residuals.

• When identifying a transfer function model with multiple input variables, if there is feedback from the response variable to an input variable then response variables should be pre-whitened. PROC ARIMA cannot handle feedback models, STATESPACE and VARMAX procedures are more appropriate for models with feedback.

Reference

[HTTP://JMP.COM/SUPPORT/HELP/INDEX.SHTML/](http://jmp.com/support/help/index.shtml)

[HTTPS://STATSOFT.COM/TEXTBOOK/TIME-SERIES-ANALYSIS/](https://statsoft.com/textbook/time-series-analysis/)

[HTTPS://KAGGLE.COM/](https://kaggle.com/)

[HTTP://SUPPORT.SAS.COM/DOCUMENTATION/CDL/EN/ETSUG/60372/HTML/DEFAULT/VIEWER.HTM#ETSUG_ARIMA_SECT034.HTM](http://support.sas.com/documentation/cdl/en/etsug/60372/html/default/t/viewer.htm#etsug_arima Sect034.htm)

Acknowledgement

We would like to thank Prof. Dr. Goutam Chakraborty, founder of SAS and OSU Data Mining and Business Analytics Program at Oklahoma State University, for his continuous guidance and support throughout the project.

Organizers of Data Mining Competitions website Kaggle, for providing the data needed for analyzing the problem.