

PROJECT EXERCISE #3

Data Bosses
Anne Christine Domercant
Tina Truc Hoang
Tasnim Quayum

Contribution:

Christine	Ses model, code merge, report
Tina	Arimax, report
Tasnim	Arima, Arfima, report

Course: Industrial Informatics (14:540:485:01)
Instructor: Dr. Ahmed Aziz Ezzat
Rutgers, The State University of New Jersey

Task 1: Preliminary analysis:

	wind_speed	wind_power	temperature	humidity	solar_irradiance
wind_speed	1.0000000	0.8885186	0.274028779	-0.169861218	0.2765546
wind_power	0.8885186	1.0000000	0.257044590	-0.067523396	0.1958399

Table 1: Correlation between variables

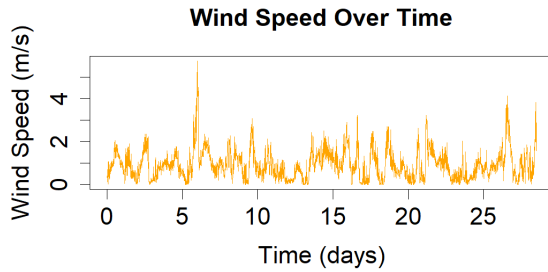


Figure 1: Wind Speed Time Series Plot

Table 1 shows the correlation between wind speed/wind power and other environmental variables. This can help us observe which variables can be used in time series models like ARIMAX. In this case, the value of temperature and humidity could be beneficial in capturing the seasonal pattern of wind speed in the forecast model. The wind speed over time plot identifies seasonal data trends, as seen in Figure 1.

Sliding window forecast is applied to estimate the most accurate RMSE based on the average of RMSEs from each window segment. The data of each window is 130, and the window size was changed between 20 and 30 to compare the average RMSE. From those error metrics, we can decide which model is the best.

Task 2: Model building:

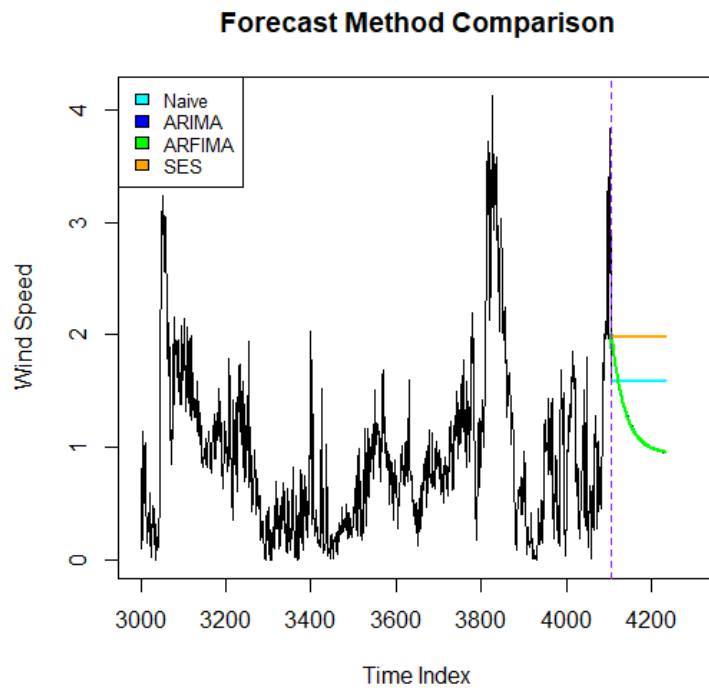
Forecast Model	RMSE _ Wind speed	RMSE_PowerScale (Linear Regression)	RMSE_PowerScale (Polynomial regression)
Naive	0.6380118	0	0
ARIMA	0.5599901	0.0596	0.0544
ARFIMA	0.5599882	0.0596	0.0544
Exponential Smoothing	0.6157661	0.0605	0.0603
ARIMAX	0.6009415	0.06	0.058

Table 2: RMSE of wind speed from different time series forecast models and its RMSE of Wind Scaled power by Polynomial regression model.

The method to predict wind power based on the weather data consisted of two parts. The first part was forecasting wind speed based on environmental data points. The second part was creating a regression model to predict wind power based on the forecasted wind speed values. Four forecasting models were trained and tested to forecast wind speed: ARIMA, ARFIMA, Simple Exponential Smoothing, and ARIMAX. These models were tested using a test train split of 130 test points and 3975 train points. A rolling horizon was used to cross-validate the models

since time series data cannot be randomized. To determine which model performed the best, the root mean square error (RMSE) was calculated, and the model with the lowest error was deemed the best fit.

The first model tested was the ARIMA model. This was chosen because ARIMA uses historical wind speed data to forecast observations and considers errors. The second model tested was the ARFIMA model. This model was selected as it is shown to be able to handle non-stationary data well. Since the historical data we are working with is non-stationary, the team thought it might better predict wind speed. The next model tested was the simple exponential smoothing model.



and ARFIMA models perform

Figure 2: Wind Speed Forecasts

This model was chosen to see how accurate predictions would be if they were just based on historical data without including errors. The final model tested was the ARIMAX model. The ARIMA model had a low RMSE value, so we wanted to explore if other weather variables outside historical wind speed values could help accurately predict the wind speed. Temperature and solar irradiance were factored into the model.

Based on *Table 2*, due to its low RMSE value, the ARFIMA model was chosen as the best model. Figure 2 shows the historical wind speed data and forecasts based on the abovementioned models. The ARIMA

similarly, as seen in *Table 2* and the nondifferentiable lines in Figure 2.

The second part of the method consisted of the regression model. When creating our models, we used a k-fold cross-validated with $k = 5$ since it is common practice to use $k = 5$. The regression models tested were linear, polynomial, random forest, and KNN. The linear model was tested due to the high correlation values calculated between wind speed and power. The exponential model was tested as the plotted data appeared to be exponential. The polynomial model was tested after as it was determined that there might be a non-linear relationship between environmental variables and wind power. The model was created utilizing temperature and wind speed data and their second and third-order interactions. The next model tested was the random forest. Temperature and wind speed were used as predictors for the model. The random forest

model was chosen for its ability to potentially capture complex relationships between two variables and their joint effect on wind power. The final model tested was the k-nearest neighbor model. We wanted to try a non-parametric approach to see if it was more successful than the parametric approaches previously tested. Temperature and wind speed were used in this model.

Table 3: Regression Model RSME

Regression Model	Train_RMSE	Test_RMSE
Linear	0.042	0.042
Exponential	0.064	0.073
Polynomial	0.029	0.029
Random Forest	0.015	0.031
Knn	0.025	0.032

Since the polynomial model had the lowest RMSE values of all regression models tested, as seen in Table 3, it was chosen as the final method for predicting wind power. A linear model relating wind speed to wind power was used as a benchmark to strengthen our conclusion of the polynomial model being the best to predict wind power now that wind speed forecast

errors have been added to the factors that need to be considered when choosing the best model. The polynomial model performed better than the benchmark for all the forecasting models used to forecast wind speed, thus validating our initial decision to use the best model from part 2 of the project.

Task 3: Final Prediction

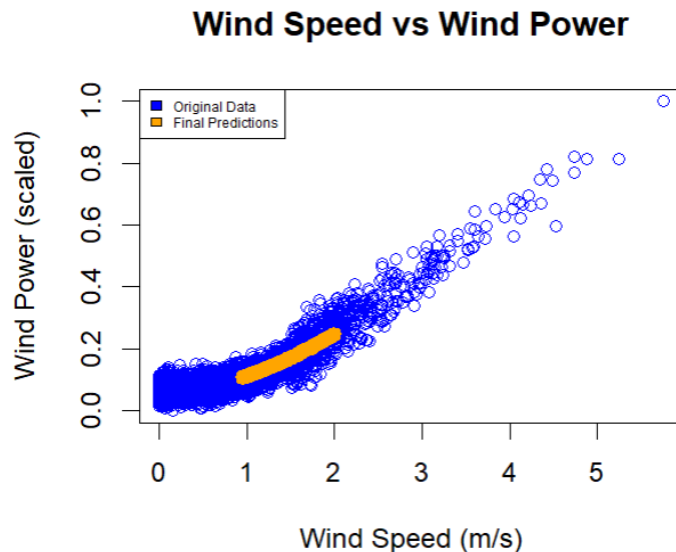


Figure 3: Final Wind Speed and Wind Power Prediction

After comparing the RMSE for the forecast methods and regression models, the polynomial method was trained on the historical data containing wind speed and other weather variables. The predictions were obtained using the function `predict()` and the test data set where the forecasted wind speed values for ARFIMA were inserted into the empty wind speed column. Figure 3 visualizes the final predictions.