31761 - Renewables in electricity markets
Assignment 3

Krishna Mohinani (s192182),
Martin Sturdy (s192462),
Aditya Surendran (s192530)

20th of May 2020

Contents

1 Introduction **3**

2 Data inputs **3**

3 Analysis **4**

3.1 Stage 1 4
3.1.1 Approach 4
3.1.2 Improvements and creativity 4
3.2 Calculating RMSE and bias 5
3.3 Stage 2 5
3.3.1 Approach 6
3.3.2 Improvements and creativity 6
3.3.3 Calculating RMSE and bias 6
3.4 Stage 3 7
3.4.1 Approach 7
3.4.2 Improvements and creativity 7
3.4.3 Calculating RMSE and bias 8
3.5 Stage 4 8
3.5.1 Approach 8
3.5.2 Improvements and creativity 9

4 Discussion **10**

5 Appendix **11**

1. Introduction

After having analyzed the market participation strategies of an energy producer in the electricity market, it has become clear that there is a need for forecasting. The forecast can be for generation quantities, load, and power flows, for example, it is, therefore, a decision-making tool that helps to be more confident about what will occur in the future.
In this report, the forecasting of wind power generated by the "Horns Rev" wind farm in Western Denmark is developed for different periods in 2018: the nominal power of this wind farm is of 160 MW, but the power output will be normalized therefore as a value between 0 and 1. The forecasts generated and described in the following sections were utilized in a forecasting competition amongst other groups of the 31761 - Renewables in electricity markets course. The team name for this stage was Co-Wind and the logo for the team is represented in figure 1.

Figure 1: Team's logo

Weather forecasts inputs for three years previous to 2018 were given to understand the relationship between the forecasted quantities and the power generated by the wind farm: after having learnt from the past, the implemented forecasting techniques can then be applied to the 2018 data to find the future wind power production. The input quantities for 2018 are wind speed predictions given at two heights (10m and 100m with respect to the ground) in their zonal and meridional components: some data processing was therefore required to obtain the optimal values to output the forecast.
Four different forecasts were made for four different periods of 2018: initially, a forecast for the first day of 2018 was made. Secondly, the whole of January (except for 1/1/2018) was asked to be forecasted; thirdly, in a similar way, February was analyzed; finally, March was the last case taken into consideration.

2. Data inputs

This section provides a basic overview of the inputs that were provided at each stage of the study and the different techniques that were applied to get to the final result.

The initial data provided represents the wind speeds, zonal (U10 & U100) and meridional components (V10 & V100), for the respective period depending on the stage. These values were used to estimate the basic inputs, the relative speeds at 10m and 100m height. A further estimation was done to plan a connection between wind speeds and force. These relationships varied for each stage depending upon the regression techniques that were used. The initial data provided were:

Power : The measured power of Horns Rev wind farm normalized
U10 : Zonal component of the wind forecast (West-East projection) at 10m.
V10 : Meridional component of the wind forecast (South-North projection) at 10m.
U100 : Zonal component of the wind forecast (West-East projection) at 100m.
V100 : Meridional component of the wind forecast (South-North projection) at 100m.
The following data was estimated using the aforementioned inputs.
WS10 : The relative speed of the wind at 10m, calculated as the length of the two vectors U10 and V10.
WS100 : The relative speed of the wind at 100m, calculated as the length of the two vectors U100 and V100.
Also, the models were developed using Python and excel for all the 4 stages.

3. Analysis

3.1. Stage 1

For this first forecasting analysis, the first day of January 2018 has been taken into consideration. The wind values at the heights of 10m and 100m were calculated using the zonal and meridional components: for this first stage, the wind at a height of 100m was used to determine the optimal power production of the wind turbine. Similar procedures were done for the years of 2015, 2016, and 2017 in order to make an analysis and apply the results to the forecasted year of 2018. After having normalized the power, the wind speeds were displayed on a scatter plot and a regression line was derived.

3.1.1. Approach

After having normalized the power produced in the three years prior to 2018, the wind speeds corresponding to each specific power value were displayed as a scatter plot: from these, a regression line was then derived. For this first stage, considering the small number of data and the inexperience with the forecasting process, a simple linear regression technique was used: a linear trendline was therefore used to represent the set of power/wind data and their relationship. From this equation, by substituting the wind values, one obtains the forecasted power quantities for the "Horns Rev" wind farm in the given period.

3.1.2. Improvements and creativity

When empirically analyzing electricity market data, it is quite unlikely that simple linear regressions models are sufficient for describing the relationship between input and response variables. Therefore it's inaccuracy increases when the number of values and data input start becoming large. It is therefore advisable to use a better and more suitable regression technique: this is what was done after the submission. The improvement of the forecast was the objective: a logarithmic regression was the chosen method in order to achieve such an upgrade. The logarithmic curve fits the plotted wind speeds in a more precise manner, resulting in power values that come closer to the actual production as shown in figure 2.

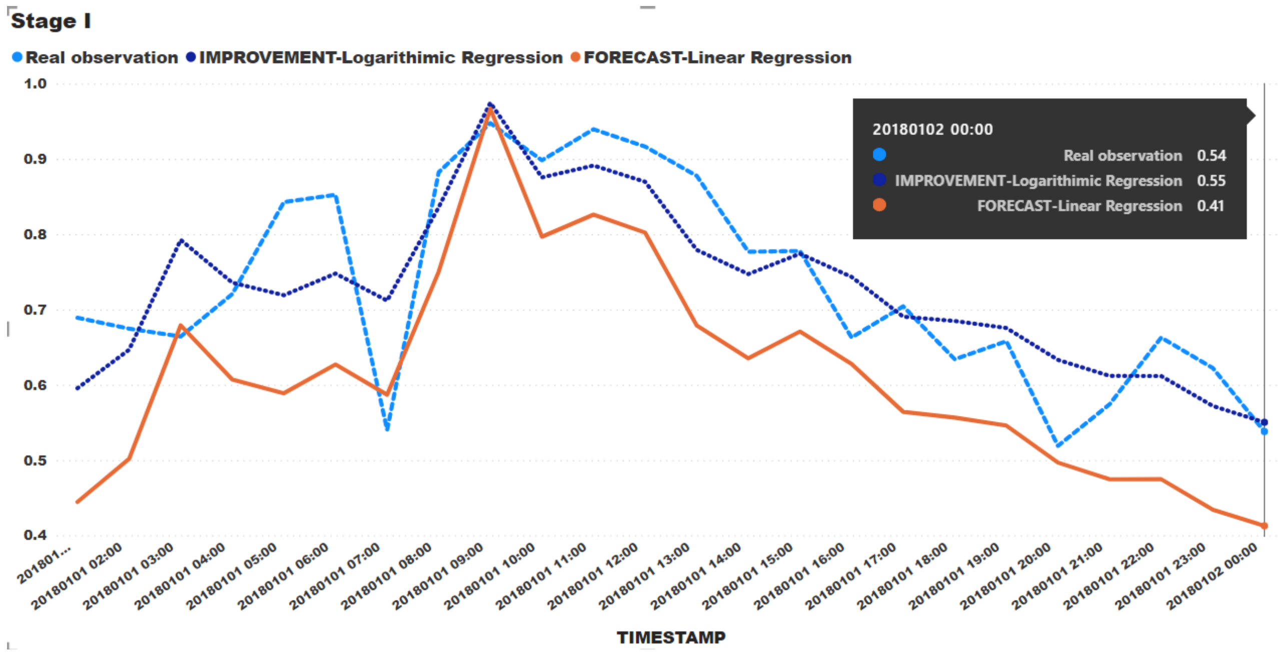

Figure 2

3.2. Calculating RMSE and bias

In order to evaluate the performance of the model, different techniques were studied and implemented, alongside some evaluation tools like the Root Mean Square Error (RMSE) and BIAS. The RMSE yields a more straightforward understanding of the effectiveness of a forecast since the lower its value, the more accurate the prediction; the other two tools presented are still valid to assess if the forecast is effective or not, but unlike the RMSE, one cannot fully grasp the essence of the forecast. The formulas for each of these different tools are shown in the Appendix. The RMSE score obtained using the simple linear regression, yields an acceptable 14.69: this is still far from good value and tells that the approximations are given with the linear trendline affect the forecast. The simple application of a logarithmic regression technique lowers the RMSE score by a large amount, confirming that a better approximation of the wind speeds, outputs a power closer to the one that is actually produced.

Tool	Forecast	Improved Forecast
RMSE	14.69	7.35
BIAS	-0.117	-0.004

Looking at the BIAS scores one can see that the linear regression technique outputs a higher value compared to the improved logarithmic case. This can be expected since in the first case the inaccurate fitting given by the linear trendline results in increased discrepancies between the forecasted and real values; the logarithmic regression instead fits the data much better and the deviation from the actual production is little and so is the BIAS score as a consequence.

3.3. Stage 2

For the second stage of the forecasting competition, each team was required to predict the energy production of the wind farm for the remaining days of January 2018. The data processing operation was similar to the one done for the first of January, and also, in this case, the wind speed at a height of 100m was utilized. Backtesting on the historical data was performed to gain more insight into the forecasting process: this time, data from 2015 was omitted since in some cases it proved to be inconsistent and not easy to work with. In this stage, like for each subsequent case, the team's objective was to improve the forecast.

3.3.1. Approach

A polynomial regression was implemented: wind values associated with the normalized power ones were once again scatter plotted and a trendline was obtained: the order of the trendline was the to be determined. When increasing the order of the polynomial fitting, the trendline shows some curves due to overfitting: this results in an accurate visual effect but the power values obtained from higher-order equations result inaccurate. For this reason, a second-order polynomial equation was chosen since the best compromise is obtained with such a curve.

3.3.2. Improvements and creativity

The results obtained from the forecast were very representative of the actual power production, in accordance with the low RMSE value of 14.68%. This value could still be reduced to get an improved forecast: one way this could be done is by taking a more thorough analysis of the input data, by interpolating the previously wind speeds further to achieve a wind speed at 70m (WS70). The reason for this was that the wind turbine's hub is placed at this height, possibly making for slightly more precise wind values. The new approach as showed in figure 3 performs slightly better.

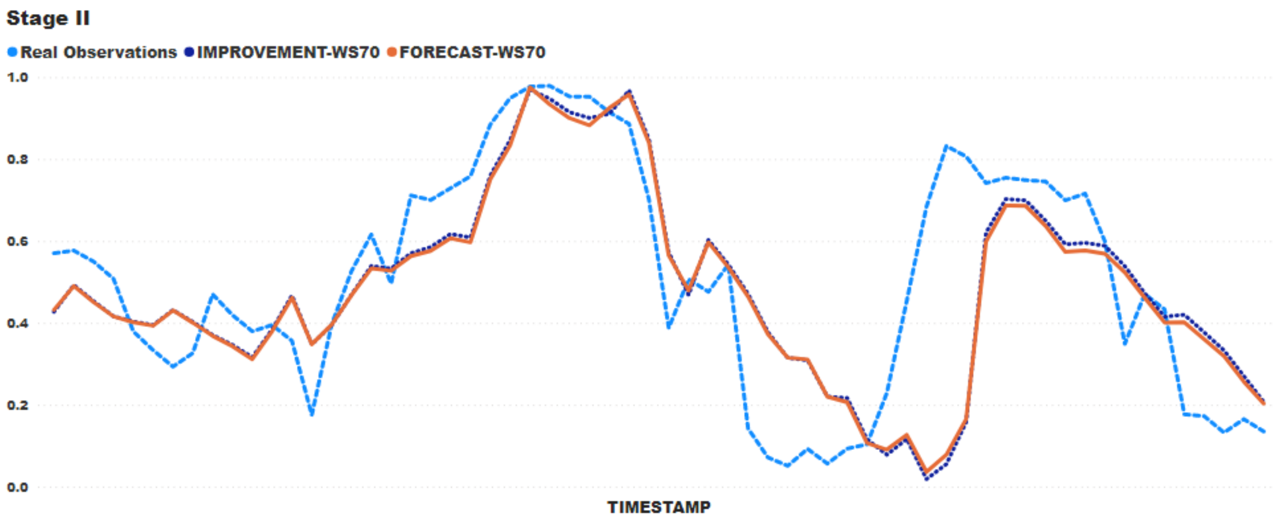

Figure 3: Curve for initial days of January

Figure 3 only shows the initial days of the forecasting in order to display how the real observation, the forecast and its improvement develop. The full set of data is represented in figure 5 in the appendix. From this image one can see that the forecast and the real observation are quite similar overall: the trend is very similar but some errors in terms of amplitude and more significantly of lag in prediction can be seen. Furthermore, the improvement is very slight as can be seen from the figure, but still getting closer to the real observation.

3.3.3. Calculating RMSE and bias

Table 1 provides the performance of the techniques applied in stage 2. Since the data to be processed increased from one day to the rest of the month, the RMSE increased by a significant margin compared to the RMSE provided for the previous improvement in stage 1. The improvement strategy for stage 2 still reduces the RMSE slightly, proving that the utilization of wind speed at the hub height of 70m is minutely effective; and there is certainly more room for additional processing of data to attain a better RMSE score.

Tool	Forecast	Improved Forecast
RMSE	14.68	14.54
BIAS	0.0049	0.011

Table 1: Forecast verification for stage II

Comparing the two BIAS score estimated for stage 2, a higher BIAS score is observed for the improved version. This proves the previously mentioned statement of BIAS not being up to the standard of checking the performance of the forecast techniques (3.2). A low BIAS score does not necessarily mean a better forecast: since the summation of the individual differences between forecasted and real observation values is taken, the positive and negative may cancel themselves out leading to a lower score, without considering the data discrepancy. The observed BIAS score was also a bit low compared to the BIAS score from stage 1, the reason for this can again be attributed to the fact that this time around the data processed was significantly larger than before.

3.4. Stage 3

The objective for stage 3 was to forecast the power output during the month of February. After having processed the data in order to be used for the analysis, the usual wind speed at 100m was utilized and trendlines were determined from the scatter plotted wind data.

3.4.1. Approach

For this stage's forecast, a polynomial regression of sixth order was implemented: although the fitting equation's order is quite high, this yielded the best approximation to the observed power produced, when compared to the lower degree of polynomial regressions, it was seen that the polynomial curve inclined further away from the power-wind speed curve at the edges(GRAPHS). For this stage, the forecast was obtained analyzing only 2017 data: since the wind data for 2018 was given, by applying the least square method the closeness of wind speed values could be found. After having done this, it was determined that 2015 and 2017 wind speeds were more representative than the 2016 ones; furthermore, since the 2015 data was seen as comparatively older than the 2017 one with respect to 2018, it was decided to disregard it.

3.4.2. Improvements and creativity

To improve the forecast, instead of applying the regression technique to the whole data set for each year a weekly analysis was performed. For 2015, 2016, and 2017, in fact, a trendline for each week of February was optimally found: power values for each hour of each week were then calculated by substituting the 2018 wind-speed at 100m into the obtained equations. Taking an average of the four weeks for each year then yielded the improved forecast. Figure 4 shows the curves for both the strategies (Approach and Improved) with respect to the real observation, it could be observed that the curve for the improved forecast follows the real observation more closely in certain areas, although it rises too above both the curves in specific points which draws the performance downwards. It has to be noted that in order to show a more detailed description of the performance figure 4 only presents the curve for initial days. Please refer to the figure 6 in the appendix for the curve representing the whole month.

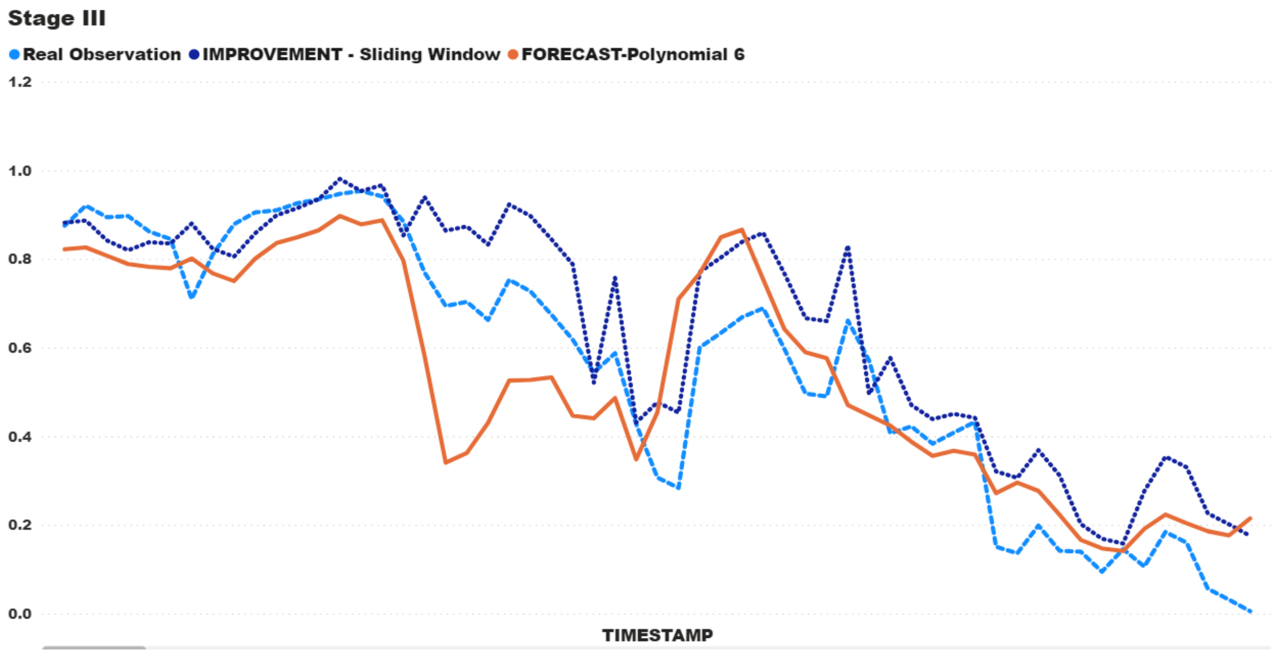

Figure 4: Curve for initial days of February

3.4.3. Calculating RMSE and bias

A drop in the RMSE score is observed for both the cases (Forecast & Improved Forecast) compared to the stage 2 which forecasted for a similar size of data. Splitting of data into small fractions proves to be beneficial and the newly observed RMSE is considerably lower.

Tool	Forecast	Improved Forecast
RMSE	13.72	13.06
BIAS	0.0015	0.096

Also for this stage, similarly as in stage 2, the BIAS score for the improved case is lower than the one obtained with the submitted forecast: the reasons for this can be attributed once more to the nature of this tool's equation.

3.5. Stage 4

The objective for the final stage was predicting March's forecast. It was clear from the previous stages that combining data from different years proved to be optimum. A combination of utilizing the wind speed at a height of 70 meters and the sliding window method was used, 3 different sliding windows for wind speed were created. The winds were grouped in three ranges: low (0 to 5 m/s), medium (5 to 10 m/s) and high (10 m/s to the highest values). This splitting of data was considered in order to minimize the errors since when kept other a limit the scattering of power values got reduced therefore the obtained curve fitted more accurately.

3.5.1. Approach

The "Least Square" method was implemented in this analysis: for each year, namely 2015, 2016 and 2017, After having divided the data, the least square was calculated for each day by taking the difference between the 2018 and the analyzed year's wind values and the square of the difference is taken. For the different wind speed ranges additionally, a representative trendline equation was obtained using 3rd-degree polynomial regression, so in total, for the three years, 9 different equations were found. Finally based on the wind speed value of 2018, one of the equations was chosen and calculated for that wind speed, giving as output the desired normalized power value.

3.5.2. Improvements and creativity

Further improvements could be done by splitting the data depending on other criteria. For instance, the density of any fluid relies legitimately upon the temperature, hence the air density will be lower throughout the summer, and subsequently, a similar breeze speed will have the option to produce more power throughout the winter. Other external factors that could have a significant effect on the power produced are not considered for this study. For example, it could happen that a turbine was broken down and its generator had to be changed at one point in time, therefore changing the production pattern from this point on. Point to be noted the real observation was not provided at the time of writing this report hence the RMSE calculation has not been performed for this section.

4. Discussion

Overall, the outcome of the forecasting process was a success: for each stage an improvement was obtained both in terms of accuracy of the forecast and in development of utilized techniques and tools. The improvement in each stage is further shown from the different RMSE scores obtained, which reduce for each submitted forecast.

Despite having reached decent power production predictions, better performance could have been obtained. Certainly, more precise results could have been obtained if the same techniques were applied to even smaller data sets than those taken in the improvement results. Additionally, sliding windows and least square methods applied to the last stage could have helped in the earlier stages to get improved forecasts. Furthermore, many other different modelling approaches could have been investigated to refine the predicted power productions: among these some machine and learning techniques, more sophisticated statistical tools, different sliding window sizes based on the data etc.

As definite decision, a great amount of knowledge was obtained from working with large measures of genuine information permitting to know about how troublesome it is to develop a conjecture dependent on climate factors. The general fulfillment with the improvement of the underlying model is acceptable, in spite of the fact that as referenced some aspects of the modelling could have been improved by investigating more complex modelling methods.

5. Appendix

- Root Mean Square Error (RMSE) represents the deviation of the forecasted quantities from the real observation ones: therefore, the larger the RMSE, the worst the forecast will be. This quantity is calculated as shown in the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (fc_i - ob_i)^2}{N}} \tag{1}$$

N = Number of values in the forecast, fc = Forecasted value, ob = Observed value

- The BIAS represents the over or under estimation of a value or set of values when compared to a base one. The equation for this is shown below:

$$BIAS = \frac{\sum_{i=1}^{N} (fc_i - ob_i)}{N} \tag{2}$$

N = Number of values in the forecast, fc = Forecasted value, ob = Observed value

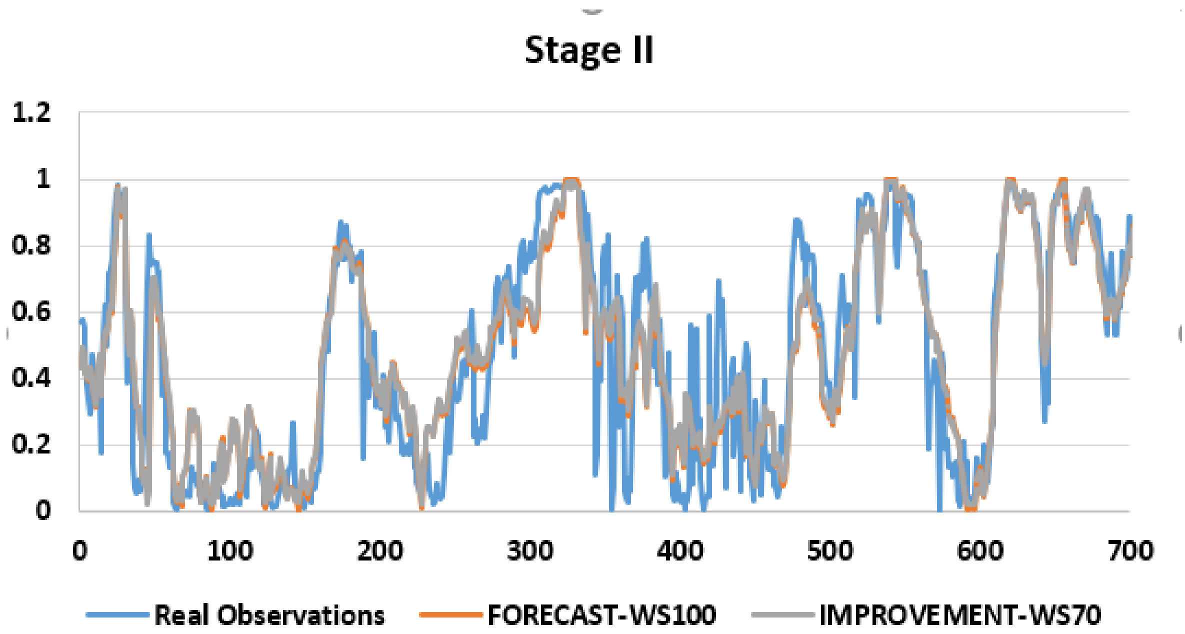

Figure 5: Complete curve for January

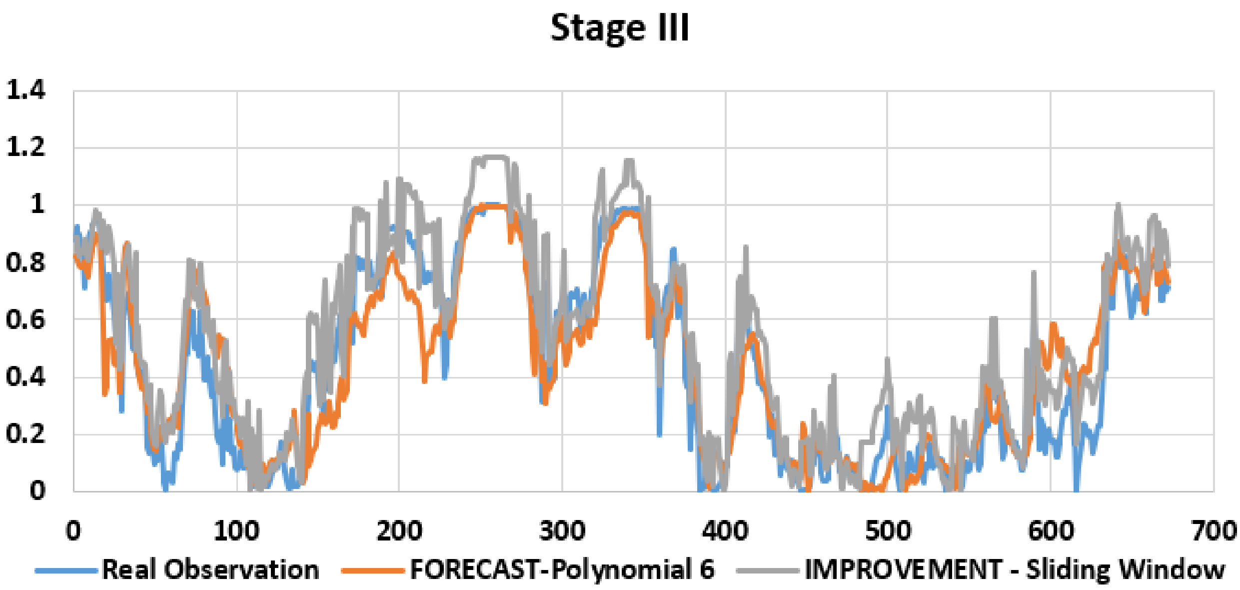

Figure 6: Complete curve for February