PROOF

PoC : Preliminary results

October 2024

Version: 1

Public

With the support of the Energy Transition Fund

Introduction:

The PROOF project focuses on offshore-related grid maintenance scheduling, with specific constraints tailored to this problem. At the core of PROOF lies a wind power forecaster and a scheduling engine. The forecaster provides reliable long-term (days to week ahead) wind power predictions. This longer horizon distinguishes the PROOF forecasters from other available forecasters typically used by system operators (which usually focus on higher accuracy in shorter horizons, e.g., day-ahead to close to real-time). The forecaster's longer-term horizon is crucial for this project as it serves as a key input to the scheduling engine and supports the strategic planning of maintenance activities. The scheduling engine will be detailed in the next article.

Variations in wind power forecasts can have a substantial impact on maintenance schedules. For example, when wind generation is high, maintenance activities are often postponed to capitalize on these favorable conditions. As a result, wind farms optimize energy production and the system operator maximizes wind power utilization.

Forecasting Approach:

This study uses an academic test case that represents a segment of the offshore grid. In this case study, we train and evaluate several forecasting models to predict wind power for 1 to 7 days in advance. We employ a rolling horizon approach, which allows us to update the forecasts daily while covering the entire week-ahead horizon. Figure 1 shows how forecasts are revised every day with this rolling horizon approach.

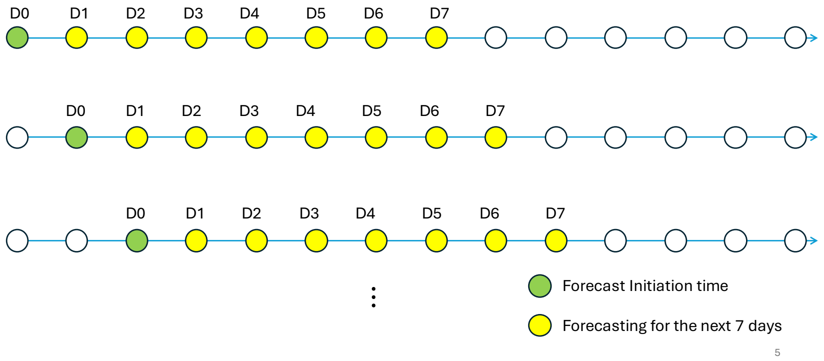

Figure 1) Every day, the model updates predictions for the next 7 days.

We developed different forecasting methods, each tailored to address distinct goals and perspectives. Initially, these methods are compared to each other and subsequently with other available forecasting services, which we refer to as Baseline 1 and Baseline 2. These external forecasters also serve as inputs to our forecasting models, which enables our models to learn from their prediction errors.

- *Time Series Forecaster (Model 1)*

The Time Series Forecaster aims to predict wind power generation at 15-minute intervals with a 7-day horizon. This model leverages temporal dependencies and patterns in the data to make accurate predictions. Figure 2 illustrates the output shape of the Time Series Forecaster, which predicts wind variations with a 15-minute resolution for the next 7 days.

PROOF : PoC Preliminary results (Oct 24)

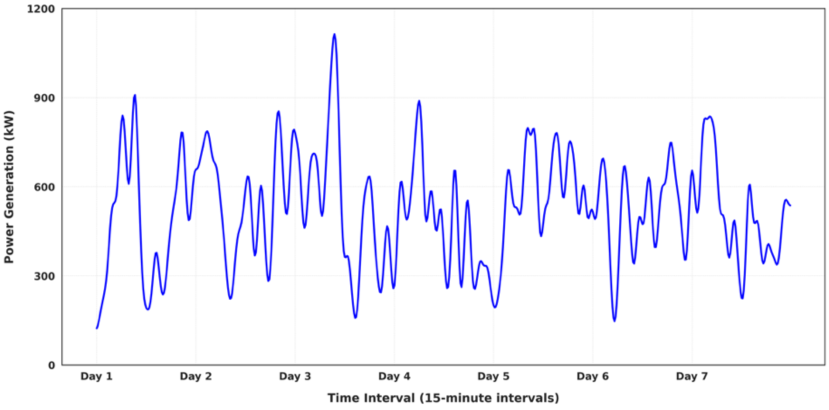

Figure 2) The output shape of the Time Series Forecaster for 7 days ahead with 15-minute granularity

- *Temporal Distribution Forecaster (Model 2)*

The Temporal Distribution Forecaster aims to predict the distribution of wind power generation within 8-hour windows. The predicted distributions provide insights into variability over these blocks. Figure 3 shows the output representation of this forecaster.

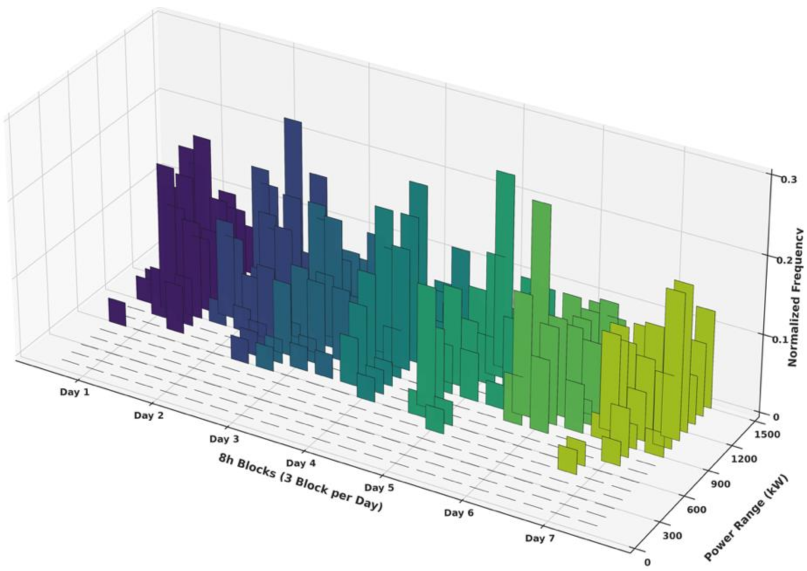

Figure 3) The output representation of the Distribution forecaster, 7 days ahead

- *Multi-Threshold Classification Forecaster (Model 3)*

The Multi-Threshold Classification Forecaster predicts whether wind power will exceed predefined thresholds within average 8-hour windows for the next 7 days. These thresholds are typically defined per asset in the existing process. Figure 4 shows the output representation of this forecaster.

PROOF : PoC Preliminary results (Oct 24)

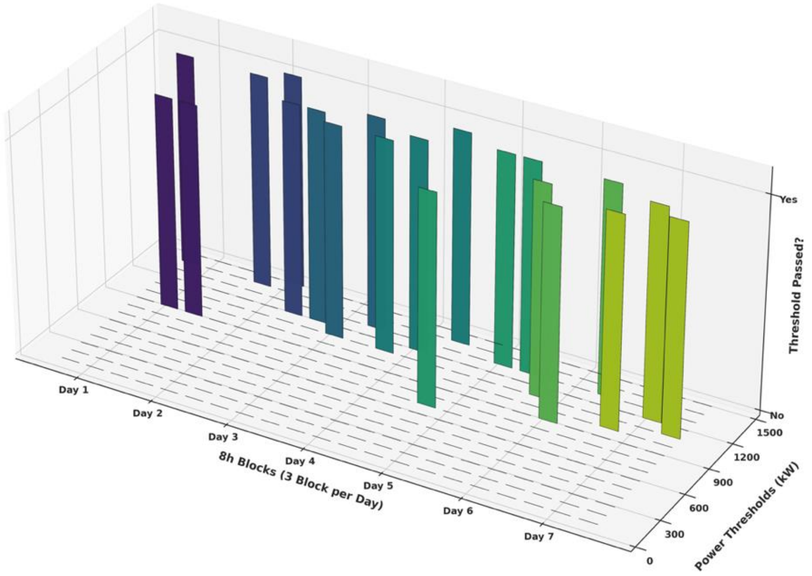

Figure 4) The output representation of the Multi-threshold forecaster, 7 days ahead.

Evaluation Metrics:

To evaluate the performance of our models and baselines, we incorporated traditional metrics like MAE and WD, as well as new problem-oriented metrics that are explained in the following.

- *Mean Absolute Error (MAE):*

MAE is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of observations.

- *Wasserstein Distance (WD):*

WD measures the effort required to transform one distribution into another. It is defined as:

$$W(p,q) = \inf_{\gamma\in\Pi(p,q)} \int_{\mathbb{R}\times\mathbb{R}} |x-y| d\gamma(x,y)$$

where p and q are the distributions being compared, and $\Pi(p,q)$ represents the set of all possible joint distributions with marginals p and q.

- *Problem-Specific Metrics*:

We adapted classification metrics such as True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) to align with the specific requirements of the maintenance scheduling problem. This adaptation provides more intuitive insights into when maintenance should be planned or avoided.

We define outcomes in two categories based on operational relevance:

Relevant Outcome: When actual wind power is low (opportunity to schedule maintenance activity)

Irrelevant Outcome: When actual wind power is high (maintenance activity should be avoided)

Based on these definitions, we identify four possible scenarios, illustrated in Figure 5:

Case 1 (TP): Actual wind power is low, and the predicted wind power is also low. This is an ideal scenario, as we correctly identify a suitable opportunity for maintenance.

Case 2 (FN): Actual wind power is low, but the predicted power is high. This leads to a missed maintenance opportunity due to an overestimated wind power.

Case 3 (FP): Actual wind power is high, but the predicted power is low. This results in scheduling maintenance when conditions are not suitable, which can lead to wasted resources and cancellation costs.

Case 4 (TN): Actual wind power is high, and the predicted power is also high. In this scenario, maintenance is appropriately avoided, saving resources and avoiding disruptions.

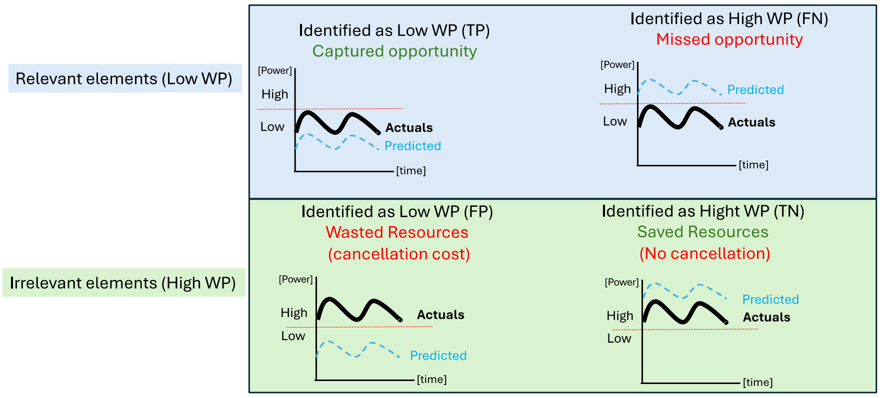

Figure 5) Problem-Specific metrics for evaluating model performance

Case Study:

In this section, we present and analyze the results obtained from the three models, as well as two baselines. For meaningful comparisons, the results have been normalized on a scale from 0 to 1 relative to the maximum capacity of the wind farm. We conducted forecasting over a one-month test set, where, for each forecast initiation time, predictions were made for the next seven days.

- Model 1:

For illustration, seven days ahead time-series forecast results initiated on May 4, 2023, are shown in Figure 6. The actual values are displayed in black, and our time series forecaster predictions (Model 1) are depicted in Blue, Baseline 1 and Baseline 2 results are, respectively, shown in red and orange.

For this forecast period, the MAE (normalized by farm capacity) for Model 1 is 0.10, which is significantly lower than both Baseline 1's MAE of 0.24 and Baseline 2's MAE of 0.17. These results indicate that Model 1 outperforms both Baseline 1 and Baseline 2 in terms of MAE for this specific forecast initiation time.

In particular, between May 5, to May 8, 2023, the predictions from Model 1 align well with the actual generation during this period, while Baseline 1 and Baseline 2 show larger deviations. Another instance occurred around May 9, 2023. Baseline 1's forecast significantly overestimates the generation, while Model 1 provides a prediction much closer to the actual value. This discrepancy highlights Model 1's ability to more effectively capture the variability in wind power generation compared to Baseline 1 and Baseline 2 during certain periods.

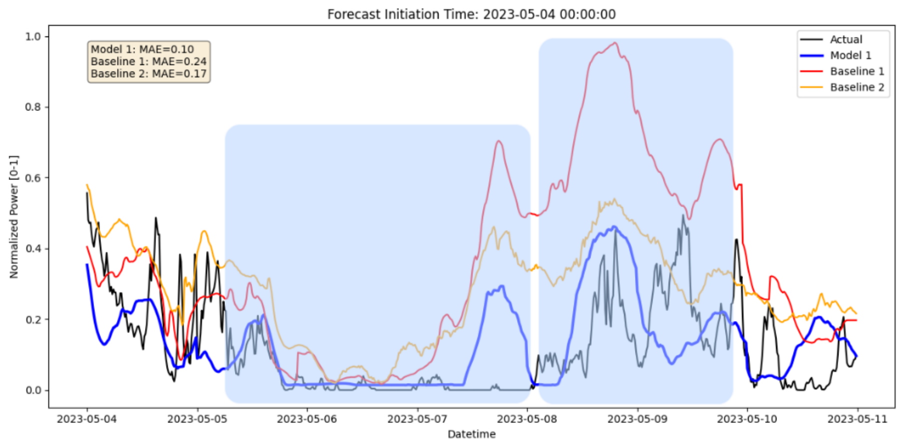

Figure 6: Time series Forecast with forecast initiation time: 2023-05-04

Figure 7 summarizes the overall performance of the forecasting models over the entire one-month test period. The MAE values (scaled by farm capacity) over this period are 0.15 for Model 1, 0.18 for Baseline 1, and 0.16 for Baseline 2. These results demonstrate that, on average, our model outperforms both Baseline models.

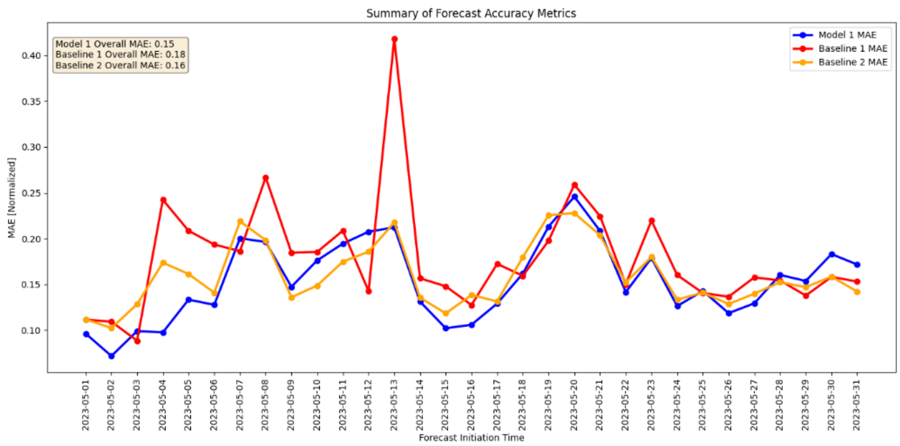

Figure 7: Overall performance of the time series forecaster Model 1 and Baselines 1-2

- Model 2:

The Temporal Distribution Forecaster aims to predict the distribution of 15-minute wind power variations within each 8-hour block over a seven-day horizon. This approach not only provides insights into the variability of wind power generation but also directly captures temporary exceedances of thresholds that might impact assets like transmission lines.

For each forecast initiation time, we generated forecasts spanning the subsequent seven days. Given that a day consists of three 8-hour blocks, this process results in a total of 21 (7×3=21) distributions per initiation time. For a fair comparison, the actual observations, along with the predictions of Model 1, Baselines 1 and 2, were transformed into the output format required by Model 2. The employed evaluation metric is the WD.

For illustration, we present part of the 7-day-ahead forecasts, initiated on May 11, 2023, as shown in Figure 8. Model 2, which is specifically designed for distribution forecasting, demonstrates a WD of 0.1 over the entire week, which is lower than that of Model 1 (0.18) and Baselines 1 and 2, which have WD of 0.18 and 0.16, respectively. When zooming at specific 8-hour blocks, such as the one from 2023-05-16 08:00:00 to 16:00:00 (highlighted in Figure 8), the predicted distribution of Model 2 (M2) is more accurate in terms of both relative frequency (normalized occurrence) and proximity to the actual bin compared to the other models. In this case, the WD of Model 2 is 0.09, which is lower than that of Model 1 (0.29), Baseline 1 (0.78), and Baseline 2 (0.25).

For a more comprehensive view, Figure 9 displays the WD values for all models over one month of the test set. It can be seen that Model 2, which is designed for distribution forecasting, still outperforms the other models.

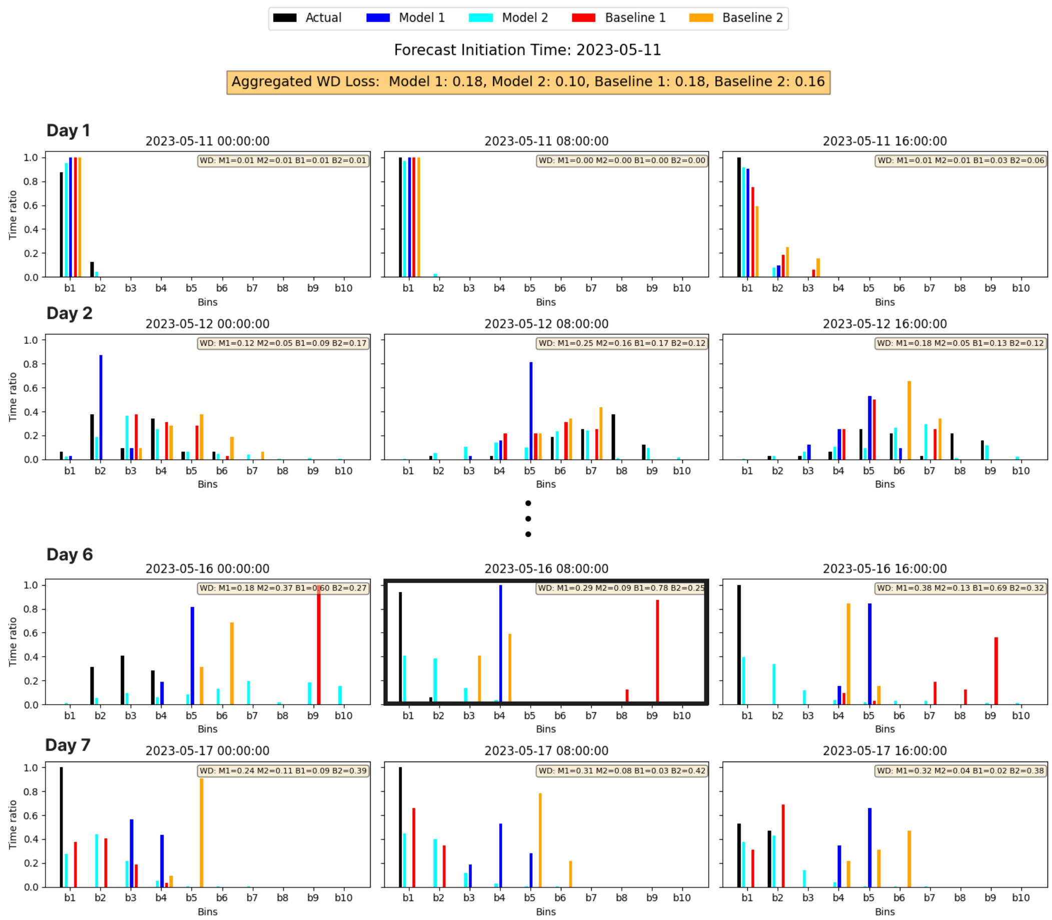

Figure 8) 7-day-ahead wind power distribution forecasts initiated on May 11, 2023.

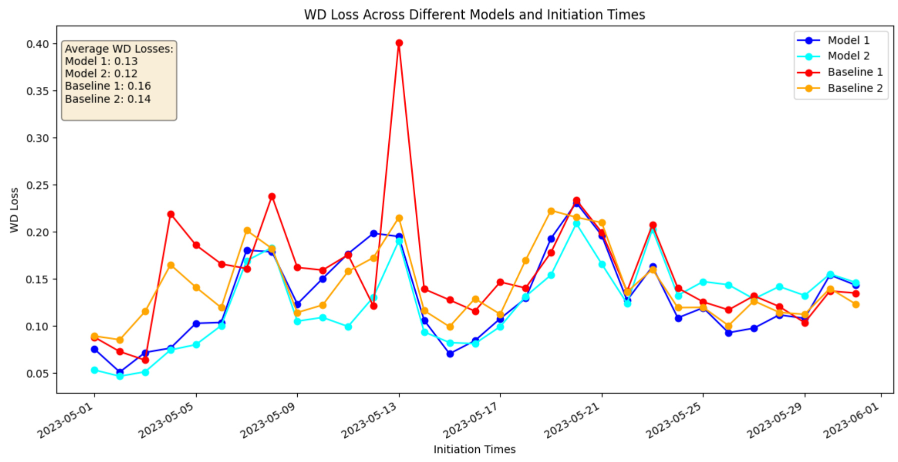

Figure 9: Performance of the distribution forecaster and other models in terms of WD over one month.

- Model 3

This model aims to predict whether the average wind power exceeds or falls below predefined thresholds for each 8-hour interval over the next seven days.

PROOF : PoC Preliminary results (Oct 24)

To compare the results of Model 3 with other models and baselines, we apply the problem-specific metrics, which were discussed earlier. These metrics allow us to assess the models' effectiveness in a domain-relevant manner. These metrics include captured opportunities, missed opportunities, wasted resources, and saved resources. Depending on the specifics of the use cases, these metrics guide model evaluation. For instance, if many maintenance tasks must be scheduled within a short time frame, maximizing captured opportunities and minimizing missed opportunities are important. In another case, if the cost of canceling resources is high and should be seriously avoided, minimizing wasted resources becomes critical. Notably, among these metrics, the saved resources metric is the least significant for this problem. This metric refers to instances when high wind power is correctly predicted. In these instances, no activities were scheduled, nor should any have been. A model that excels solely on this metric does not provide much relevance for our specific needs.

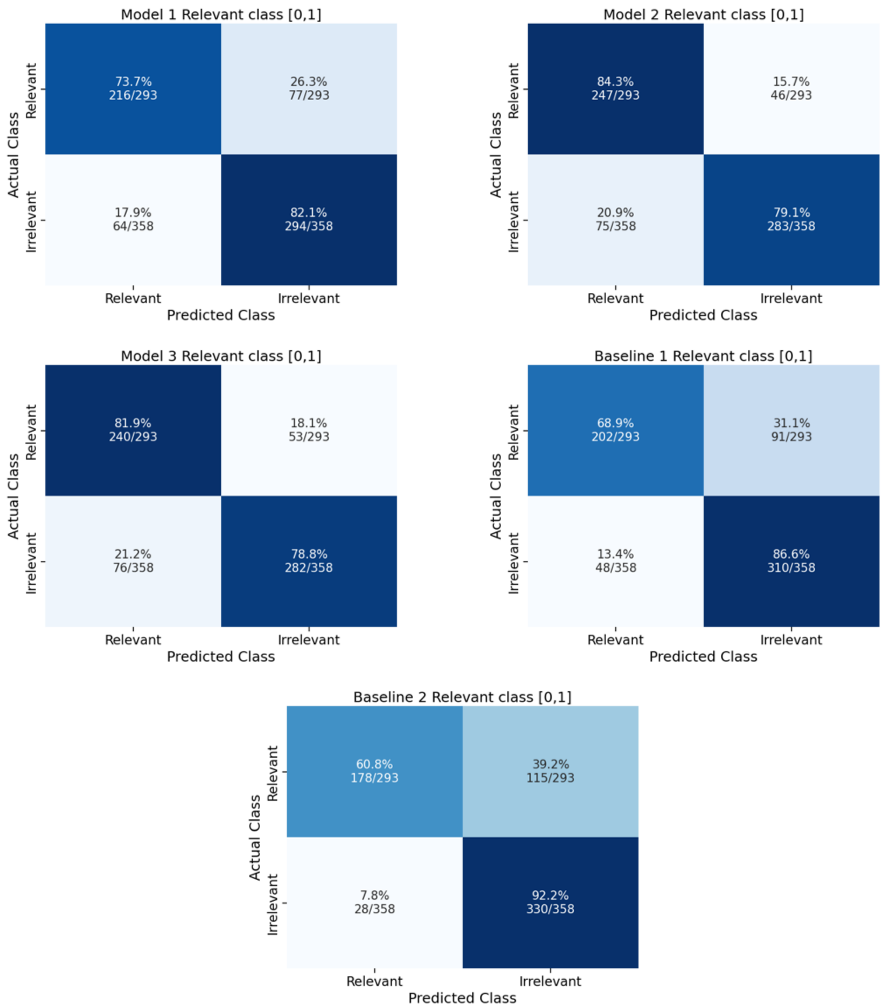

Figure 10: Performance Comparison of Models with Domain-related Metrics

The results of the same test month are shown in Figure 10. The relevant class is wind power below 60 MW. It can be observed that the baseline models, Models 1 and 2, focus less on capturing opportunities (68.9% and 60.8%, respectively) and more on saving resources, which is less relevant for this problem (86.6% and 92.2%). The time series forecaster (Model 1) places greater emphasis on capturing opportunities (73.7%) compared to the baseline models. Model 2 (the distribution forecaster) and Model 3 (the threshold forecaster) both achieve higher scores in capturing opportunities and have fewer instances of missed opportunities. However, it should be noted that, in terms of minimizing wasted resources, Baseline Models 1 and 2 perform better.

The development of this framework for analyzing the performance of various models, without integrating the forecast results into the downstream optimization problem, will facilitate model tuning for this specific use case.

Conclusion

The PROOF project successfully developed three specialized week-ahead forecasting models to enhance offshore wind power maintenance scheduling. These models perform better than the baseline forecasters currently available to the system operator. The first model, the Time Series Forecaster (Model 1), predicts wind power at 15-minute intervals and outperforms traditional baselines in accuracy. The second model, the Temporal Distribution Forecaster (Model 2), predicts the distribution of wind power within 8-hour windows. This model offers valuable insights into power variability. The third model, the Multi-Threshold Classification Forecaster (Model 3), predicts whether wind power will exceed predefined thresholds. We developed problem-specific metrics such as captured opportunities, missed opportunities, wasted resources, and saved resources to evaluate the models not only for statistical accuracy but also for domain-related metrics. This evaluation approach will guide and simplify model selection and tuning without disrupting the forecasting pipeline, as it eliminates the need to integrate predictions directly into the downstream scheduling optimization problem. In the second phase of the project, we will further refine and optimize these models to address the specific complexities of the problem at hand.