AGGREGATION STRATEGIES FOR GRID-BASED NUMERICAL WEATHER PREDICTIONS (NWP) TO IMPROVE POWER CURVE MODELS WITH META-LEARNING EXTENSION

by

Elif Konyar

B.S., Industrial Engineering, Boğaziçi University, 2018

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Industrial Engineering
Boğaziçi University
2021

ACKNOWLEDGEMENTS

Above all, I would like to express my deepest gratitude to my thesis advisor Assist. Prof. Mustafa Gökçe Baydoğan for his endless support, invaluable guidance, constant patience and instantaneous responses to my innumerable questions. He has been a great and motivating mentor to me which I feel extremely lucky for, and I would also like to thank him for introducing me to the field of data mining when I was an undergraduate.

I would also like to thank Assoc. Prof. Wolfgang Hörmann and Assist. Prof. Burak Barutçu for taking part in my thesis committee and their very valuable comments.

I want to thank my fellow colleagues Tarkan, Buğra, Feyyaz, Gizem, Çiğdem and Orkun for their friendship and our memories. I already miss our coffee breaks. I feel very lucky to work and share those moments with them.

I also want to thank my closest friends Zeynep, Tuğçe and İlke for their support and our countless Zoom meetings during the pandemic. I feel very fortunate to have them since high school, and without our long talks, this year would be much more difficult.

I must say that I owe very much to my parents Sevil and Sadettin, my two brothers Yasin and Emre, and Gizem for their invaluable support and readiness to help me. My parents were always supportive in all the decisions I have made in my life. I also feel very lucky to have two very encouraging brothers.

Finally, I am very grateful to my dearest Erdem for his constant support and understanding. He was always there for me when I needed him. I could not complete this study without feeling him by my side.

ABSTRACT

AGGREGATION STRATEGIES FOR GRID-BASED NUMERICAL WEATHER PREDICTIONS (NWP) TO IMPROVE POWER CURVE MODELS WITH META-LEARNING EXTENSION

In the first part of this thesis, alternative strategies to estimate the wind power curve by utilizing grid-based weather forecasts from Numerical Weather Prediction (NWP) models are proposed. Traditional power curve estimation such as Weibull Cumulative Distribution Function and Five Parameters Logistic Function uses actual wind speed information at a single location, and they are known to provide strong results. On the other hand, forecasts from multiple grid locations are used in a short-term wind power forecasting scenario which brings additional difficulties in mapping NWP model forecasts to actual speed values. In order to resolve these, we propose a simple optimization framework which aggregates grid-based NWP predictions to estimate the power curve directly. Due to the problems with the cyclic nature of the direction, we propose alternative aggregation strategies based on generalized additive models. Our experiments on six wind farms show that parameter estimation with Quasi-binomial distribution assumption on the response provides superior performance compared to popular Gaussian likelihood assumption used in the power curve estimation approaches.

In the second part, meta-learning approaches with a dynamic model ranking mechanism is applied. Three mostly used pointwise, pairwise and listwise ranking approaches are utilized with decision trees and tree-based ensembles as learners. Our experiments on the base-level model pool show that pairwise and listwise approaches improve the performance of both pointwise approach and the individual models.

ÖZET

META-ÖĞRENME UZANTISI İLE RÜZGAR GÜÇ EĞRİSİ MODELLERİNİ GELİŞTİRMEK İÇİN SAYISAL HAVA TAHMİNLERİNİ (NWP) BİRLEŞTİRME STRATEJİLERİ

Bu çalışmanın ilk bölümünde, Sayısal Hava Tahmini (NWP) modellerinden grid tabanlı hava tahminlerini kullanarak rüzgar güç eğrisini tahmin etmek için alternatif stratejiler önerilmiştir. Güçlü sonuçlar sağladığı bilinen Weibull Kümülatif Dağılım Fonksiyon ve 5 Parametreli Lojistik Fonksiyonu gibi geleneksel güç eğrisi tahminleme modelleri tek bir konumdan hız bilgisini kullanır. Öte yandan, kısa vadeli güç tahminlemede çoklu grid noktalarından tahminler kullanmak NWP tahminlerini gerçek hız değerlerine eşlemede ek zorluklar getirmektedir. Bunları çözmek için, NWP tahminlerini birleştirip rüzgar gücünü doğrudan tahminleyen basit bir optimizasyon çerçevesi önerilmiştir. Rüzgar yönünün döngüsel yapısıyla alakalı problemler nedeniyle, genelleştirilmiş eklemeli modellere dayanan alternatif birleştirme stratejileri de öneriyoruz. Altı santral üzerinde yapılan deneyler hedef üzerinde Yarı-binom dağılımı varsayımıyla yapılan parametre tahmininin güç eğrisi tahminleme yaklaşımlarında kullanılan popüler Gauss olabilirlik varsayımına kıyasla daha iyi performans sağladığını göstermektedir.

İkinci bölümde ise dinamik model sıralama mekanizmasına sahip alternatif meta-öğrenme yaklaşımları uygulanmaktadır. Karar ağaçları ve ağaç tabanlı topluluk öğrenme modelleri kullanılarak en çok kullanılan noktasal, ikili ve listesel sıralama yaklaşımlarından faydalanılmıştır. Bu modeller ilk aşamadaki model havuzu üzerinde test edilmiştir. Deneyler ikili ve listesel yaklaşımların hem noktasal yaklaşımı hem de modellerin bireysel performanslarını iyileştirdiğini göstermektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . iii
ABSTRACT . . . iv
ÖZET . . . v
LIST OF FIGURES . . . ix
LIST OF TABLES . . . xiii
LIST OF SYMBOLS . . . xv
LIST OF ACRONYMS/ABBREVIATIONS . . . xvii
1. INTRODUCTION . . . 1
2. LITERATURE REVIEW . . . 10
    2.1. Wind Power Curve Modeling . . . 10
    2.2. Algorithm Ranking . . . 12
3. BACKGROUND . . . 16
    3.1. Numerical Weather Prediction (NWP) Models . . . 16
    3.2. Base-level Modeling Methods . . . 16
        3.2.1. Weibull CDF . . . 17
        3.2.2. Five Parameters Logistic Function . . . 17
        3.2.3. Generalized Additive Models . . . 17
    3.3. Meta-level Modeling Methods . . . 18
        3.3.1. Decision Trees . . . 18
        3.3.2. XGBoosting . . . 19
    3.4. Ranking Approaches . . . 19
        3.4.1. Pointwise Approach . . . 19
        3.4.2. Pairwise Approach . . . 20
        3.4.3. Listwise Approach . . . 20
4. METHODOLOGY . . . 21
    4.1. Base-level Modeling Methods . . . 21
        4.1.1. Weibull CDF . . . 21
        4.1.2. Five Parameters Logistic Function . . . 22

4.1.3. Generalized Additive Models (GAMs) 22
4.2. Base-level Aggregation Methods 24
4.2.1. Point-based Power Curve Modeling 24
4.2.2. Region-based Power Curve Modeling 25
4.2.2.1. Modeling with Power Curve Models Based on Shape . 26
4.2.2.2. Modeling with Generalized Additive Models 27
4.3. Meta-level Features 28
4.4. Meta-learning Approaches 33
4.4.1. Pointwise Approach 33
4.4.1.1. Classification Setting - Rank 1 Class Labeling Approach 33
4.4.1.2. Classification Setting - Rank 1 vs Other Approach . . 35
4.4.1.3. Regression Setting 37
4.4.1.4. Ordinal Regression Setting 38
4.4.2. Pairwise Approach 40
4.4.3. Listwise Approach 41
5. EXPERIMENTS ON REAL DATA 42
5.1. Data Description 42
5.2. Performance Metrics 43
5.2.1. Base-level 43
5.2.2. Meta-level 43
5.3. Experimental Settings 44
5.3.1. Base-level 45
5.3.2. Meta-level 46
5.4. Results 46
5.4.1. Point-based Power Curve Modeling 46
5.4.2. Region-based Power Curve Modeling 48
5.4.2.1. Modeling with Power Curve Models Based on Shape . 48
5.4.2.2. Modeling with Generalized Additive Models 48
5.4.3. Meta-learning Algorithm Ranking 53
5.4.3.1. Base-level Model Pool 53
5.4.3.2. Feature Selection 54

5.4.3.3. Pointwise Approach 56
5.4.3.4. Pairwise Approach 58
5.4.3.5. Listwise Approach 60
6. CONCLUSION 64
REFERENCES 66
APPENDIX A: OVERALL RESULTS OF THE OTHER WIND FARMS . . . 71
APPENDIX B: MODEL PARAMETERS 79

LIST OF FIGURES

Figure 1.1. Typical theoretical wind power curve. 2

Figure 1.2. Wind rose (a) and estimated power curve with categorized wind direction (b) of Dinar WF. 4

Figure 1.3. Histogram of normalized production values (July 2018 - June 2019) of Dinar WF. 5

Figure 1.4. Block diagram of the modeling mechanism. 9

Figure 4.1. Wind speed with cubic spline (a), wind direction with cubic vs cyclic splines (b), interaction of wind speed and wind direction (c) and interaction for u and v vectors for Dinar WF. 23

Figure 4.2. U and V components of wind vector in the nearby region of Dinar WF. 26

Figure 4.3. Rank 1 class labeling approach. 33

Figure 4.4. A simple decision tree with rank 1 class labeling approach built with a subset of both prediction-based and weather-based meta-level features for Dinar WF. 34

Figure 4.5. Rank 1 vs other approach. 36

Figure 4.6. A simple decision tree with rank 1 vs other approach built with the same subset from prediction-based and weather-based meta-level features and model information for Dinar WF. 36

Figure 4.7.	Regression error-based approach.	37
Figure 4.8.	A simple tree with regression setting built with a subset of prediction-based and weather-based meta-level features and model information for Dinar WF.	38
Figure 4.9.	Ordinal regression setting.	39
Figure 4.10.	A simple tree with ordinal regression setting built with a subset of prediction-based and weather-based meta-level features and model information for Dinar WF.	39
Figure 4.11.	Pairwise and listwise ranking settings.	40
Figure 5.1.	Wind farms used in experiments.	42
Figure 5.2.	Interaction effect of wind speed and wind direction in one nearby coordinate (a), hourly seasonality plot (b) and monthly seasonality plot (c) for Dinar WF.	50
Figure 5.3.	Comparison of best models from each category in terms of NMAPE for Dinar WF.	50
Figure 5.4.	Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of Dinar WF.	52
Figure 5.5.	The correlation of wind power outputs among base-level model pool for Dinar WF.	54
Figure 5.6.	Selected feature set for Dinar WF.	57

Figure 5.7. Box plots of hourly NDCG values of regression setting and rank 1 vs other approach with Xgboosting from pointwise approach, pairwise approach and listwise approach. 61

Figure 5.8. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Dinar WF. 62

Figure 5.9. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Dinar WF. 63

Figure A.1. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Aliaga Bergama WF. 71

Figure A.2. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Soma 1-2 WF. 71

Figure A.3. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Bares WF. 72

Figure A.4. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Soke WF. 72

Figure A.5. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Geycek WF. 73

Figure A.6. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Aliaga Bergama WF. 74

Figure A.7. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Soma 1-2 WF. 75

Figure A.8. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Bares WF. 76

Figure A.9. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Soke WF. 77

Figure A.10. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Geycek WF. 78

Figure B.1. Model parameters for base-level and meta-level modeling methods 79

LIST OF TABLES

Table 4.1. Prediction-based meta-level features. 30

Table 4.2. Performance-based Meta-level Features. 31

Table 4.3. Weather-based Meta-level Features. 32

Table 5.1. Comparison of predictions in point-based modeling with smoothed and not smoothed wind speed for Dinar WF. 47

Table 5.2. The overall performance of point-based modeling with Weibull CDF in all wind farms. 47

Table 5.3. The overall performance of point-based modeling with Five Parameters Logistic function in all wind farms. 48

Table 5.4. The overall performance of region-based modeling with shape-based methods in all wind farms. 49

Table 5.5. The overall performance of region-based modeling with GAMs in all wind farms in terms of NMAPE. 51

Table 5.6. The overall performance of region-based modeling with GAMs in all wind farms in terms of RMSE. 51

Table 5.7. The overall performance of base-level model pool in all wind farms in terms of NMAPE. 54

Table 5.8. Feature selection scenarios. 55

Table 5.9. Performance comparison of feature selection scenarios in terms of NMAPE. 56

Table 5.10. Performance comparison of pointwise approach methods in terms of NMAPE. 58

Table 5.11. Performance comparison of pointwise approach methods in terms of RMSE. 58

Table 5.12. Performance comparison of pointwise approach methods in terms of MRR. 59

Table 5.13. Performance comparison of pointwise approach methods in terms of $NDCG_{mean}$. 59

Table 5.14. Overall performance of pairwise approach method in terms of NMAPE, RMSE, MRR and $NDCG_{mean}$. 60

Table 5.15. Overall performance of listwise approach method in terms of NMAPE, RMSE, MRR and $NDCG_{mean}$. 60

LIST OF SYMBOLS

a	Five Parameters Logistic Function parameter
b	Five Parameters Logistic Function parameter
c	Weibull parameter
$capacity(t)$	Capacity of the wind farm at time t
d	Discrete wind direction information
e	Five Parameters Logistic Function parameter
f_h	Smoothing function
g	Five Parameters Logistic Function parameter
G	Number of coordinates from GFS
i	Coordinate information
z	Weibull parameter
k	Knots in generalized additive models
$P(t)$	Wind power production at time t
$\widehat{P_i(t)}$	Wind power prediction for i at time t
$\widehat{P_{agg}(t)}$	Aggregate wind power prediction for time t
$p_{m_i}(t)$	Importance of model m_i at time t
R	Number of classes in classification setting
$r_{m_i}(t)$	Rank of model m_i at time t
$\widehat{r_{1t}}$	Predicted rank of base-level algorithm with actual rank 1
s	Five Parameters Logistic Function parameter
T	Number of observations in the training set
$u(t)$	u component of wind vector at time t
$\widehat{u(t)}$	Adjusted u at time t
w_i	Weight for each $\widehat{P_i(t)}$
w_i^u	Weight for each $u_i(t)$
w_i^v	Weight for each $v_i(t)$
$x(t)$	Wind speed at time t
$\widehat{x(t)}$	Adjusted wind speed at time t

x_r	Predictors
$v(t)$	v component of wind vector at time t
$\widehat{v(t)}$	Adjusted v at time t
β_i	Set of parameters of either Weibull CDF or Logistic function
β_r	r^{th} coefficient in generalized additive models
$\theta_i(t)$	Wind direction at i and t

LIST OF ACRONYMS/ABBREVIATIONS

CDF	Cumulative Distribution Function
DCG	Discounted Cumulative Gain
HIRLAM	High Resolution Limited Area Model
GAMs	Generalized Additive Models
GFS	Global Forecasting System
MRR	Mean Reciprocal Rank
NDCG	Normalized Discounted Cumulative Gain
NCEP	National Centers for Environment Prediction
NMAPE	Normalized Mean Absolute Percentage Error
NWP	Numerical Weather Predictions
RankBoosting	Ranking Boosting Algorithm
RankSVM	Ranking Support Vector Machine
RMSE	Root Mean Squared Error
WF	Wind farm

1. INTRODUCTION

The demand for renewable energy sources such as wind and solar has boosted lately in the world due to environmental concerns and increased demand for energy. By 2018, the total net installed capacity of wind energy is 562.9 GW where Turkey constitutes 1.6% of the global wind energy production [1]. By 2020, the total installed capacity of wind energy in Turkey is 8.8 GW where it constitutes 9.2% of Turkey's total installed capacity [2]. Until 2025, 95% of the increment in global installed capacity are planned to be come from renewable energy where wind energy composes the 30% of the planned capacity raise in renewable energy [3]. As wind energy is beginning to play a bigger role in world energy supply, reliable forecasts for wind power production become more significant. The stochastic character of the wind makes wind energy a non-dispatchable source having challenges related to uncertainty [4]. Hence, accurate forecasting of wind power production is critical in incorporating energy supply from wind in the power grid.

Wind power production of a wind turbine depends highly on wind speed which can be shown with a wind power curve [5]. In wind power curves, cut-in speed is the starting point for the wind turbine to produce wind power, whereas rated wind speed is the point where the maximum power is produced [5]. When the speed reaches to a certain point, also known as cut-out speed, the wind turbine is closed down due to the design and safety issues [5]. Theoretical power curves provided by the manufacturers present the ideal wind power production given a wind speed. An example is shown in Figure 1.1. However, in some cases, there is no access to the theoretical power curve. As an alternative, wind speed data can be used to model the empirical power curve.

Power curve estimation approaches make use of two alternative sources of information. One group of studies utilizes actual weather data including wind speed [6, 7], whereas the other group uses forecasts from Numerical Weather Prediction (NWP) models, which are known as physical models [8, 9]. Use of NWP for power curve es-

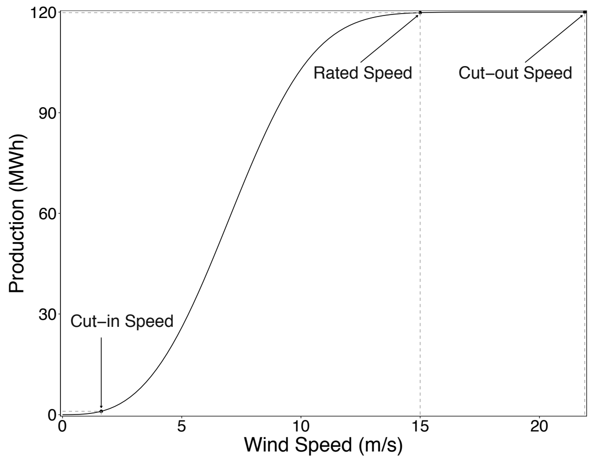


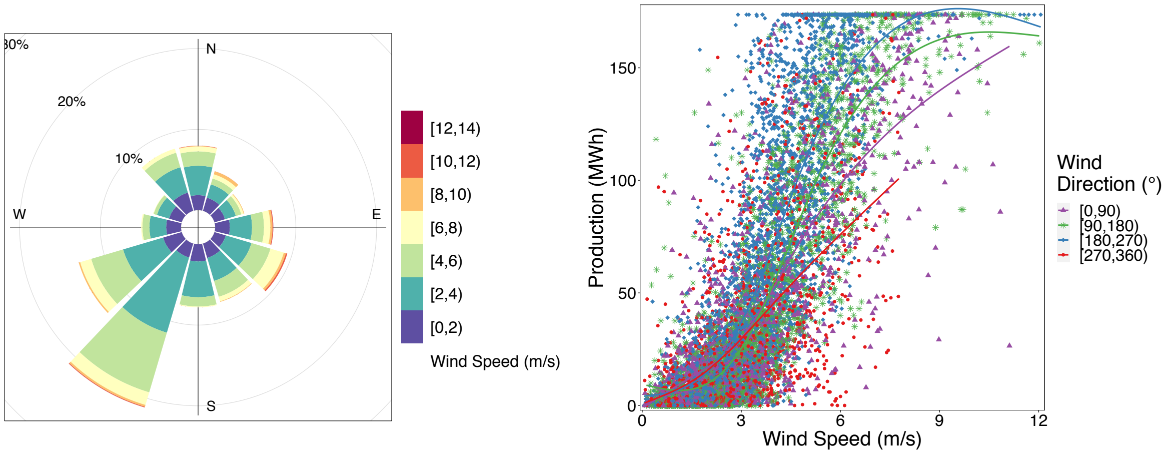
Figure 1.1. Typical theoretical wind power curve.

timation is mainly driven by the prediction requirements in a short-term forecasting scenario in which the weather-related turbine measurements are not available. NWP forecasts consist of weather information such as wind speed, wind direction, atmospheric pressure, temperature, humidity and so on.

NWP models provide weather forecasts on a grid basis [10], which brings about the alternatives of either using information from single coordinate or from multiple coordinates. Use of forecasts from multiple points is known to enhance the predictive performance of the wind power prediction models [11]. It is also shown in the literature that the prediction error diminishes if predictions from different NWP models such as Global Forecasting System (GFS), High Resolution Limited Area Model (HIRLAM) and etc., are combined [8]. Some NWP models are known to be advanced at some specific weather occurrences [8]. Hence, in that case, choosing a set of coordinates along with alternative NWP models becomes more meaningful. However, this poses a challenge in deciding how to fuse the information from the multiple coordinates of alternative NWP models for power curve estimation tasks.

The focus of this thesis is the estimation of power curve at a power plant level. At power plant level, an aggregate version of the turbine power curves are observed. In this case, instead of modeling the power curve of a single wind turbine, production at the power plant in relation to wind speed is modeled. Due to maintenance activities and the aggregation, a noisy version of the power curve is observed. From the methodological point-of-view, there are various frameworks used in wind power prediction to model the wind power curve. These can be classified as parametric approaches and nonparametric approaches [6, 12, 13]. Recent studies show that several parametric distributional forms based on the shape of wind power curves can be used to model power curve [12, 13]. Among a variety of distribution functions, Weibull Cumulative Distribution Function (CDF) and Logistic function are common due to their small number of parameters, their continuity and the resemblance of their shapes to wind power curves [13–16]. In wind power curve modeling, in addition to the wind speed, wind direction plays an important role as well. As shown in Figure 1.2, the speed distribution differs depending on the direction, and there is a requirement in modeling the power curves with direction consideration when NWP models are utilized. To best of our knowledge, except in study [17], wind direction is not utilized in power curve modeling. In [17], authors discretize the wind direction into four bins, and build separate power curve models for each category as in Figure 1.2. Using discretized wind direction is one way to capture the direction-dependent power curve behaviors. However, in this approach, there is loss of information due to the continuous angular nature of the wind direction. Thus, these distribution-based approaches cannot preserve continuous and angular behaviour of direction.

To resolve aforementioned problems with the existing power curve models, we propose two aggregation strategies to utilize multiple coordinates in the distribution-based approaches. In the first proposal, the estimation of the power curve is performed for each coordinate in the first step, and then predictions are aggregated to obtain a final prediction. As this approach has potential to lose the spatial information between the coordinates, we also propose a simple optimization framework which aggregates grid-based NWP forecasts to estimate the power curve directly. Our experiments show that both aggregation strategies show performance improvements compared to the models focusing on the use of NWP predictions at a single location. Unfortunately, these approaches still require the discretization of the wind direction. In order to account for



(a) Wind rose

(b) Power curve

Figure 1.2. Wind rose (a) and estimated power curve with categorized wind direction (b) of Dinar WF.

the continuous nature of the wind direction together with grid-based NWP forecasts from multiple locations, we utilize Generalized Additive Models (GAMs), which are formed by sum of smoothing functions, also called as smoothing splines [18]. In the literature, spline regression methods that utilize wind speed information are widely used to model wind power curves [6], but these models do not consider the interaction between the wind speed and wind direction in a continuous manner. Benefiting from the cyclic splines [18], we enhance the existing strategies by incorporating continuous direction information in building power curves. Since these models are regression-based, they are flexible in adding other variables that can help explain the wind power unlike in Weibull CDF and Logistic function approaches. As a further step, we improve our method by adding variables related to seasonality.

GAMs provide additional flexibility in learning parameters with alternative distributional assumptions [18]. In many regression-based approaches, wind production is assumed to follow a Gaussian distribution, and maximum likelihood estimation is performed for Gaussian parameters. In Figure 1.3, problematic nature of the use of Gaussian distribution is illustrated using the histogram of normalized production values of Dinar WF. Normalized production is basically the ratio of the production to the total installed capacity. These values can be thought as utilization [4, 17], and they are in [0,1] range. To address this issue, we enhance our proposed strategy by modeling utilization values with Quasi-binomial distribution, which is the generalized version of binomial distribution [19], assumption.

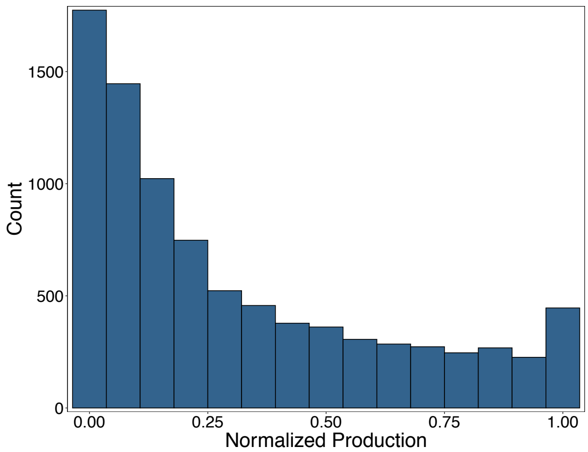


Figure 1.3. Histogram of normalized production values (July 2018 - June 2019) of Dinar WF.

Among the three alternative strategies to model wind power curve in this thesis, the first approach, in which power curve is modeled for each coordinate separately with distribution-based approaches and aggregated in a second step, serves as a benchmark model. The second approach which differs from the first approach in terms of the aggregation strategy brings an improvement. This modeling strategy uses multiple observation points directly in power curve modeling. The third strategy, where GAMs are utilized, brings a further performance enhancement with cyclic splines to model continuous and angular behaviour of wind direction, parameter estimation with Quasi-binomial distribution assumption and flexibility in adding seasonality-related variables.

In almost all problems as in wind power curve modeling task, there are many alternative models to utilize in order to predict a target. However, the sources e.g., time which can be used to try a set of algorithms are limited. Furthermore, many real-life applications necessitate fast decision-making which makes it hard to compare all the possible algorithms. Moreover, as No Free Lunch Theorem suggests, there is no model that provides the best performance in all tasks, and when all the tasks are taken into account, no model provides better or worse performance than random guessing on average [20]. Thus, without any information on a task, it is not possible to suggest using one model over another. However, there can be different sources of information related to a task. These can be domain knowledge and expert knowledge on a specific task as well as information extracted from the task itself. This information which is obtained by the data used in a task can be referred as "meta-knowledge" [21]. By exploiting this knowledge, one model can be in favor of another model in a certain task. However, this means choosing a model over another for the whole task as in some studies [20–22].

On the other hand, in time series applications as wind power curve modeling, some specific conditions change over time due to the nature of the problem. For instance, wind speed and wind direction change in very small time intervals which affect the wind power output as already mentioned. Hence, when there are changes in specific conditions, the performance of the algorithms usually change, too. Although the algorithm with overall superior performance perform better in many intervals, in some other intervals another algorithm can outperform regardless of its overall performance. In this case, selecting the former algorithm over the latter due to its overall superior performance may result in bad performance in the intervals that the latter outperforms. To elaborate the situation more, with dynamically selecting the algorithms by ranking them, an even better performance than of Generalized Additive Models (GAMs) with Quasi-binomial distribution assumption can be achieved. This arouses the need to fuse predictions of different model sources in order to obtain a better performance.

There are different approaches in exploiting predictions of different algorithms in prediction problems. One group of studies focuses on ensembling approaches [23]. In ensembling approaches, predictions of different algorithms are combined with different strategies such as by taking mean, median etc. or by giving different weights to each algorithm [23]. The ensembling approaches are generally known to be performing well in time series applications as well [23]. However, there are several issues related to this approach. The first is how to combine the predictions since there are many alternatives as already mentioned. The second issue is linked to interpretability. Interpretability in machine learning problems is very fundamental since being able to explain why an algorithm is selected or why an algorithm performs better is required in many tasks. In real-life applications such as wind power prediction, machine learning tasks are accompanied with decision-making processes. Hence, it becomes necessary to justify why selecting an algorithm is a better choice to the decision-makers.

Another approach to fuse predictions of different algorithms in dynamic environments such as time series applications is to model it as a ranking problem. This problem can also be thought as multi-target learning problem [22]. In previous studies, there are several multi-target learning approaches that are applied in order to find the best performing model in a whole dataset [20–22]. This is useful when the performance of the models does not change considerable over time and when there is an abundance of algorithms and only a subset of them can be tested due to some limitations such as time. However, in this thesis, instead of finding the best performer for the whole dataset, the aim is to find the best performer in each hour dynamically. Hence, the objective is to rank the algorithm pool for each prediction period t and then to select the algorithm with the first rank for t.

In the literature, ranking problems e.g, document retrieval tasks, personalized rankings, have been formulated with pointwise, pairwise and listwise approaches [24, 25]. Adopting an approach over another makes a change in the loss function calculation related to the number of instances being taken into account at one step [24]. In pointwise approaches, instances are regarded in calculation of the loss function separately, and the ranking problems can be solved as a standard classification or regression task within this approach [24]. On the other hand, pairwise approaches take into account pairs of instances while optimizing the corresponding loss function [24]. Thus, in this approach, relative orders of the pairs are used as inputs within the loss function and are predicted for the test set [24]. Differently, in listwise approaches, the complete set of instances are taken into account in loss function calculation [24]. In this thesis, aforementioned pointwise, pairwise and listwise approaches are adopted and compared in algorithm ranking task in wind power prediction problem. Depending on the formulation type, this ranking problem is modeled as classification, regression, pairwise ranking and listwise ranking tasks from these approaches. Among these, considering pairwise relations and relative orderings or considering whole set of instances improves the performance over pointwise approaches where each algorithm from the algorithm pool is taken as a separate instance.

In order to learn the ranks of the algorithm pool in the second part after obtaining the predictions, features that are extracted from the data can be utilized. These features are known as meta-features [26]. In the literature, previous studies mention some features which are known to be meaningful for this learning to rank task [20, 21, 26]. However, they are generally limited to the statistical and information theoretic characteristics of the whole dataset and to the diversity in the algorithm pool [20]. In [27], where the focus is on dynamically selecting an ensemble algorithm from best models, features related to error and diversity in predictions are extracted and proposed. This thesis extends this feature set based on the predictions to capture different behaviours among the algorithm pool and over time. Additionally, apart from the prediction-based features, performance-based, seasonality-based and weather-based features for the learning to rank approach in wind power setting are proposed. Using additional features in addition to prediction-based features, improves the overall performance. The summary of the modeling mechanism of base-level and meta-level parts is illustrated in Figure 1.4.

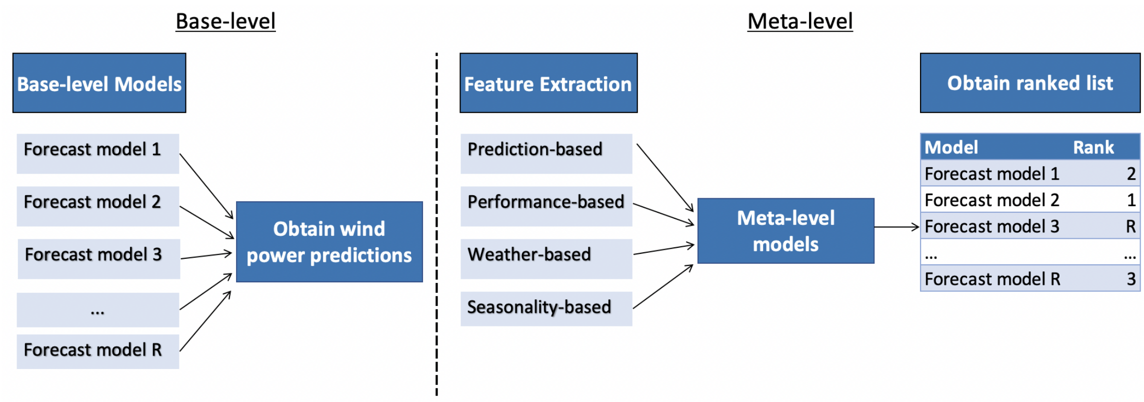


Figure 1.4. Block diagram of the modeling mechanism.

The rest of the thesis is organized as follows: In Section 2, a detailed literature review on wind power curve modeling and algorithm ranking tasks is provided. Section 3 provides background information on numerical weather prediction models and the methods used in this thesis. In Section 4, a detailed explanation for the modeling approaches is presented. Section 5 introduces the data used in this thesis, the experiments and compares the methods, and Section 6 provides the conclusion.

2. LITERATURE REVIEW

This section consists of two parts. First, a detailed literature review on wind power curve modeling is presented. Second, some previous studies on meta-learning, ranking and algorithm selection tasks are summarized.

2.1. Wind Power Curve Modeling

Wind power curve modeling is often classified as parametric approaches and nonparametric approaches [6, 12, 13]. To begin with the parametric approaches from the literature, in [15], the focus is on different logistic functions to model the wind power curve. The advantages of using logistic functions come from their small number of parameters, their continuity and the resemblance of their shapes to wind power curves. This thesis provides the comparison of a variety of logistic functions and shows the superiority of five parameter logistic function with exponential form to the others in terms of its performance. In [13], an analysis on several parametric and nonparametric approaches such as linearized segmented model, four and five parameter logistic functions, neural networks, fuzzy c-means clustering and some data mining algorithms is made. The parameters in the parametric methods are estimated with several different optimization algorithms such as differential evolution, particle swarm optimization. Logistic function in power form with five parameters estimated with differential evolution algorithm outperforms the other methods in this thesis. Another parametric approach which is widely used in wind power curve modeling based on Weibull distribution is demonstrated in [14], which is a study with a primary focus on detecting deviations in power curves. The cumulative distribution function (CDF) of Weibull distribution is also a good alternative to model wind power curves because of its closeness of shape to power curves and uncomplicated form with two parameters. The study shows in the power curve modeling part that Weibull CDF with its parameters estimated with least squares method performs better than linearized segmented model. Another study [16] also shows the advantages of using Weibull distribution in wind power, and it delves more into the several parameter estimation techniques of Weibull distribution. In none of these power curve models based on shape, wind direction is utilized in power curve modeling. Furthermore, the challenge in deciding how to fuse the information from the multiple coordinates is not addressed in these studies.

In [6], four statistical parametric and nonparametric approaches which are polynomial regression, locally weighted polynomial regression, spline regression and penalized spline regression methods are implemented to model wind power curve using wind speed information. Polynomial regression which is subject to many studies brings along overfitting problem, and it is not robust to outliers. Locally weighted polynomial regression is proposed in order to overcome these problems. Besides that, spline regression brings another advantage since it gives more flexibility on the shape of estimated wind power curve by building a piece-wise polynomial regression model. Thus, different local regions are modeled independently without violating the continuity of the shape. This method arouses another problem which is how to select the number and the regions of these polynomials. Penalized spline regression is proposed to overcome this issue, and it provides better results compared to first three statistical approaches in that study. The authors do not utilize the wind direction information in this study as well, and the challenge on fusing the information from multiple coordinates remains. Further detailed review of these parametric and nonparametric approaches in wind power curve modeling can be found in [12]. In addition to the methods mentioned in this section, they analyze some other forms of polynomial power curves, maximum principle method, dynamical power curve, ideal power curve and probabilistic model as parametric methods, and copula power curve, neural networks, fuzzy methods and data mining algorithms such as k-NN algorithm, random forest and boosting as non-parametric methods.

In wind power curve literature, some studies focus on ramp detection problem. Sometimes the wind power production shows sudden changes withing short ranges of time, which is known as ramps [28]. In ramp detection, they try to predict whether a ramp will occur or not within a specified time interval [28]. When the objective is to predict wind power output rather than whether a ramp will occur or not, it is harder to predict these sudden changes, which makes ramp prediction another research area. Thus, this issue is not in the scope of this thesis.

In the works described in this section, weather information from single point is used in wind power curve modeling. In [17], they use multiple points to predict an aggregated wind speed and wind direction. In this thesis, a strategy, where multiple observation points are directly used to predict wind power, is proposed, so that spatial relations are incorporated as well. Moreover, in the studies mentioned here, wind direction information is not utilized. As mentioned in the previous section, in [17], they use discretized wind direction information in four bins. Since wind power output is dependent on wind direction in addition to wind speed, this information with both discrete characteristic and continuous angular characteristic is included into the modeling framework. In [6], where wind power curve is modeled with statistical methods, parameters are estimated with Gaussian distribution assumption on production values. On the other hand, in this thesis, the proposed strategy with GAMs is enhanced by modeling normalized production values with Quasi-binomial distribution assumption. Furthermore, in none of these studies, seasonal variables are used to explain further variation in wind power production. Hence, the proposed strategy improves by adding variables related to seasonality.

2.2. Algorithm Ranking

One group of studies in meta-learning focuses on predicting the best performing algorithm on a whole dataset [20–22]. In [20], with meta-learning techniques, the authors try to find the best performing algorithm on overall in a whole time series data by approaching this problem in classification setting. The pool of the base algorithms include simple models such as moving average, exponential smoothing, polynomial regression, arima models, neural network models, and their ensembles etc. For the meta-learning task, several features considering the whole data set based on general statistics, autocorrelations, frequency domain and diversity among the base algorithms are extracted and used in clustering the datasets. Neural networks, decision trees and support vector machines are applied in a classification setting. Their proposed ranking approach where individual models are weighted outperformed the approaches where single algorithm is selected. In [22], in addition to the current enormously used features such as statistical, information-theoretic and landmarking-based features, the authors suggest using features that show pairwise performances of the algorithms. Moreover, they propose approximate ranking tree forests, which is a tree-based ranker including a combination of approximate ranking trees. Their results show that the pairwise features improve the performance of the rankers, and their proposed ranking algorithm is competitive among the well-known rankers such as k-NN ranker, AdaRank, label ranking trees, predictive clustering trees.

Furthermore, [21] provides a comprehensive review on state-of-art meta-learning algorithms and the meta-knowledge used in these algorithms. According to the survey, the concept of meta-learning has been applied in several application settings to which ensembling several base algorithms, algorithm recommendation, dynamic bias detection and inductive transferring can be given as examples. In algorithm recommendation setting, several machine learning algorithms such as meta-decision trees, support vector machines, neural networks and k-Nearest Neighbours etc. are utilized mostly in classification setting. However, in all these studies, the focal interest is on obtaining the best performing algorithm in a whole dataset.

Another group of studies focus on dynamically selecting best performing algorithms for regression, classification and specifically time series forecasting problems [27, 29, 30]. The authors of [27] compare dynamically selecting best regressor, ensembling all the regressors by dynamic weights and dynamically ensembling a set of best performing regressors to the state-of-art dynamic regression algorithm selection methods as well as common ensembling approaches such as mean, median etc. The authors propose several alternative measures such as variance of predictions, squared error related measures etc. to compute a so-called “region of competence”, which is the neighbouring instances. Their experiments show that dynamically ensembling regressors outperform the other approaches. Similarly, in [29], a review on dynamic classification algorithm selection is provided. Their focus is on different measures while calculating the so-called "region of competence" on which the selection of the set of classifiers is based. These measures are classified as individual-based and group-based. The former including accuracy-based, ranking-based features etc. focuses on individual performances of the classifiers, whereas the latter consisting of diversity, complexity and ambiguity accounts for the relationships between the classifiers as well. They also compare dynamic selection strategy with several other methodologies such as best classifier selection, ensembling all classifiers etc., and dynamic selection strategy shows a good performance. On the other hand, in [30], the focus is on dynamic selection of algorithms in time-series data. In the meta-level, some measures, which belong to statistical and information-theoretic feature categories, such as correlation between numerical predictors, dispersion gain, average of continuous predictors etc. are obtained and used as input to several classification algorithms such as random forests, k-NN and Naive Bayes. Several different strategies such as predicting the best algorithm in a pair or predicting whether to use single best algorithm or an ensemble are analyzed in meta-level. Their results show that their meta-learning approach enhances the performance compared to selecting a single algorithm based on the majority in the past and ensembling by averaging all the regressors. In these studies, only prediction-based meta-features are utilized in several pointwise meta-learning approaches.

Moreover, some studies focus on improving the ranking performance of several state-of-art ranking algorithms such as RankSVM and RankBoosting [31,32]. The motivation in these studies is to enhance the performance of the ranker by using active learning methods in information retrieval setting. The underlying idea is to choose the most explanatory unlabeled examples to feed the ranking models. In [31], the authors achieves this by manipulating the objective function of the mentioned ranking algorithms such that it includes an additional term, which accounts for the error introduced by adding this new example. In the end, the objective tries to select the instance that maximizes this newly introduced error, which is believed to increase the generalization capacity of the model. Similarly in [32], the instances with the highest expected hinge loss are chosen from the unlabeled set in order to improve the ranking performance of RankSVM. These studies are important in the sense that they give an idea on selecting informative instances in order to obtain a better ranking performance.

Furthermore, another group of studies focuses on improving personalized ranking task with adversarial learning [33, 34]. The authors in [33] give an idea on how to manipulate the objective function of a pairwise ranking algorithm in order to increase the robustness along with its performance. They implement the idea of adversarial learning with adversarial perturbations on model parameters. Similarly in [34], the similar idea is implemented with adversarial perturbations in the features to improve the robustness and ranking performance. This formulation framework can also give an idea on how to improve the ranking performance of RankSVM, with adversarial perturbations, and can be applied in algorithm ranking setting.

In this thesis, instead of predicting a best performing algorithm for the whole set as in [20–22], a best performing algorithm is dynamically selected for each hour. In addition to prediction-based meta-features [27, 29], performance-based, weather-based and seasonality-based meta-features are also utilized in the wind power curve algorithm ranking task. Instead of computing "region of competence" with k-NN-based approach as in [27,29], a tree-based clustering is performed for the meta-learning task. In addition to pointwise approaches i.e., modeling as classification or regression problems, pairwise and listwise ranking approaches are also adopted.

3. BACKGROUND

In this section, some important background information on NWP models and the methods used in this thesis is provided.

3.1. Numerical Weather Prediction (NWP) Models

NWP models provide weather forecasts on a grid basis [10]. These forecasts consist of weather information such as wind speed, wind direction, atmospheric pressure, temperature, humidity and so on, and they are provided for different atmospheric pressure levels such as 10 meters above ground, 80 meters above ground, 100 meters above ground etc. and different coordinates on a grid basis [10, 35, 36]. There are two types of NWP models which are global and local models, and they differ based on the region they provide forecasts for [35, 36]. The NWP model provided by Global Forecasting System (GFS) can be an example of a global model, and it is developed by the National Centers for Environment Prediction (NCEP) [35]. GFS model provides weather forecasts with horizontal resolution of 28 kilometers [35]. On the other hand, the NWP model provided by High Resolution Limited Area Model (HIRLAM) program is an example of a local model [36], and it provides weather forecasts with horizontal resolution of 7.5 kilometers [36]. Both models provide forecasts with 3-hour temporal resoulution [35, 36].

3.2. Base-level Modeling Methods

This section provides the fundamental background information related to base-level modeling methods.

3.2.1. Weibull CDF

The Weibull CDF function which is used to model wind power curve can be given as:

$$P(t) = capacity(t) \times (1 - e^{-(\frac{x(t)}{c})^z}) \tag{3.1}$$

where $P(t)$ is the wind power production, $capacity(t)$ is the capacity of the wind farm, $x(t)$ is the wind speed at time t, and c and z are the two Weibull parameters.

3.2.2. Five Parameters Logistic Function

From the logistic functions family, Five Parameters Logistic Function with the following formula is also used to model wind power curve:

$$P(t) = s + \frac{a - s}{(1 + (\frac{x(t)}{e})^b)^g} \tag{3.2}$$

where a, b, e, g and s are the logistic function parameters.

3.2.3. Generalized Additive Models

A generalized additive model is defined as a generalized linear model which is formed by the sum of smoothing splines of the predictors [18]. The simplified structure is:

$$g(E[y(t)]) = \beta_0 + \sum_{\forall r} \beta_r x_r(t) + f_1(x_1(t)) + f_2(x_2(t), x_3(t)) + ... \tag{3.3}$$

where $E[y(t)]$ is the expected value of the response at time t, β_r are the r^{th} coefficient, f_h are the smoothing functions, and x_r are the predictors. Different local regions of the predictors are modeled independently without violating the continuity of the modeled functions and their first two derivatives [18]. The number of these local regions are determined by knots, k. In this approach, the target can be modeled with a wide range of smoothing splines such as cubic regression splines and cyclic cubic regression splines [18]. A cubic regression spline is the combination of cubic polynomials from different local regions [18]. These local regions are divided by knots where the consecutive cubic polynomials have the same value at knots [18]. On the other hand, cyclic cubic regression splines can be used in models when some variables have cyclic nature [18]. Their difference is that the starting and the ending knots have the same value since the smoother is cyclic [18]. Furthermore, with tensor products, the interactions between predictors as well as their main effects can also be modeled [18].

3.3. Meta-level Modeling Methods

This section provides the related background information of the meta-level learners used in this thesis. These meta-level learners are decision trees and XGBoosting.

3.3.1. Decision Trees

Decision trees are widely used models which are not complex and highly interpretable. Depending on the problem setting, trees can handle both categorical and numerical targets [37]. Another advantage is that categorical predictors can directly be used in decision trees without one-hot encoding [37]. Decision trees simply split the input space linearly and learn decision rules from those subregions [37]. However, they can handle nonlinear relationships between the target and the predictors with multiple splits [37]. The predictions are made based on the leaf nodes resulting from these splits. In decision trees, in order to obtain the leaf nodes, the splitting procedure is performed according to different splitting criteria such as Gini index [37], which is the splitting criteria used in decision tree scenarios in this thesis. Gini index can be calculated as:

$$GI = \sum_{r=1}^{R} \hat{p_{nr}}(1 - \hat{p_{nr}}) \tag{3.4}$$

where n is the leaf node on which a training instance ends up, r is the class, $\hat{p_{nr}}$ is the fraction of training instances from r in n, and R is the number of classes in the classification setting. Thus, Gini index is an estimation on variance in leaf nodes [37]. In this thesis, decision trees are used in pointwise approach, which will be explained in detail in Section 4.4.1.

3.3.2. XGBoosting

XGBoosting is an algorithm which implements tree boosting [38]. The idea of tree boosting comes from building consecutive trees such that the next tree learns from the residuals of the previous tree [37]. Hence, the regions where decision trees work badly are focused more in the next tree due to using residuals from the previous tree [37]. Thus, tree boosting is a slow adapting mechanism [37]. On the other hand, XGBoosting is an efficient algorithm which adapts the tree boosting system, and it is known to be performing well on classification, regression and also ranking problems [38]. In this thesis, XGBoosting is used in all the meta-learning approaches, which will be explained in detail in Sections 4.4.1, 4.4.2 and 4.4.3.

3.4. Ranking Approaches

In this section, background information for pointwise, pairwise and listwise ranking approaches are presented.

3.4.1. Pointwise Approach

In pointwise approach, the instances are taken into account separately in optimizing the loss function within the algorithm. Thus, the relative rank orderings of the instances are not considered in this approach. In this approach, the target is modeled as either a categorical variable or a numeric variable. Thus, any classifier or regressor can be used in this approach.

3.4.2. Pairwise Approach

In this approach, meta-learning problem can be modeled as a pairwise ranking problem. XGBoosting is a good option to solve pairwise ranking problems since it uses LambdaMART algorithm, which can optimize a pairwise loss function specific to ranking problems [38]. LambdaMART is known to be a good tree boosting ranking algorithm [39]. The objective in ranking functions is to learn the relative ordering of the pairs [39]. Thus, if there is a pair with $r_{m_i}(t) < r_{m_j}(t)$ where $r_{m_i}(t)$ is the rank of model m_i for time t, or similarly with $p_{m_i}(t) > p_{m_j}(t)$ where $p_{m_i}(t)$ is the importance of model m_i for time t, then the objective is to have a function f such that it gives relative scores as $f(m_i(t)) > f(m_j(t))$ [39, 40].

In a pairwise ranking task, the objective can be to sort documents [39], or as in this thesis, to sort the models. Hence, the probability of the importance of a model m_i being greater than the importance of another model m_j, which is denoted as $P(p_{m_i}(t) > p_{m_j}(t))$, is fundamental in this pairwise task [40]. Thus, LambdaMART uses a set of gradient boosting regression trees and minimizes a cross-entropy loss, which is associated with the above mentioned probability [40].

3.4.3. Listwise Approach

Similar to pairwise approach, XGBoosting is a strong alternative to solve listwise ranking problems. It again uses LambaMART algorithm which maximizes Normalized Discounted Cumulative Gain (NDCG) in listwise setting [41]. NDCG is a widely used ranking measure in previous studies such as in information retrieval tasks [39]. This measure is useful and strong in the sense that it gives more weights to the instances with higher importance [39]. Hence, while maximizing NDCG in this setting, the models with ranks close to 1 are given higher weights.

4. METHODOLOGY

In this section, first, the framework for wind power curve modeling is described. Then, the meta-level modeling strategies are explained. The methodology has three aspects. First, the focus is on types of wind power curve modeling methods which are Weibull CDF, Five Parameters Logistic function and Generalized Additive Models (GAMs). Secondly, the focus is on aggregation methods of weather data from multiple coordinates. As the first alternative, the aggregation is in predictions level, whereas as the second alternative, the aggregation is in weather level. Finally, pointwise, pairwise and listwise meta-learning approaches on wind power curve modeling problem are explained in detail.

4.1. Base-level Modeling Methods

In order to model the wind power curve at power plant level, three base models are used:

(i) Weibull CDF
(ii) Five Parameters Logistic function
(iii) Generalized Additive Models

4.1.1. Weibull CDF

Weibull distribution is a widely used distributional form in power curve modeling [14, 16]. The shape of Weibull CDF is very similar to wind power curves, and it has few parameters to estimate, which makes it a good alternative to be used in base-level methods. With Equation 3.1, wind power predictions, $\widehat{P(t)}$, are obtained for each prediction period t.

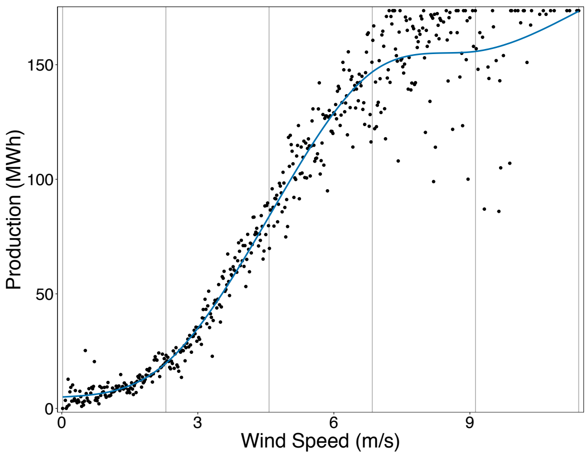
4.1.2. Five Parameters Logistic Function

As an alternative to Weibull CDF, Five Parameters Logistic function is used in power curve modeling. Logistic functions are extensively used in studies due to their small number of parameters, their continuity and the resemblance of their shapes to wind power curves [13, 15]. From the logistic functions family, one of the alternative forms, which is known to be performing well in power curve modeling, is Five Parameters Logistic function, and with Equation 3.2, wind power predictions, $\widehat{P(t)}$, are obtained for each prediction period t.

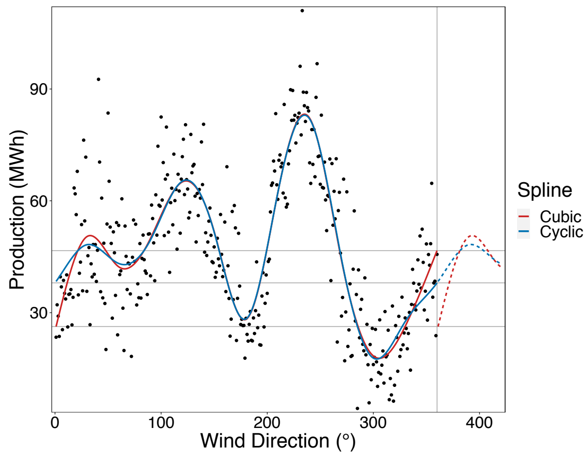
4.1.3. Generalized Additive Models (GAMs)

Generalized additive models are a good alternative to model the nonlinear relationship between the wind power production and variables such as wind speed and wind direction with a variety of splines such as cubic regression spline and cyclic cubic regression spline. To illustrate, Figure 4.1 depicts the cubic regression smoothing spline with an example where averaged power production is modeled with speed information. Furthermore, in wind power prediction, direction information has a continuous angular nature. Figure 4.1 illustrates two example models where power production averaged for rounded direction values is modeled with direction information using cubic regression spline and cyclic cubic regression spline, respectively. This demonstration shows the advantage of using cyclic cubic regression splines in case of having angular variables since the cyclic behaviour is preserved only in the cyclic spline case.

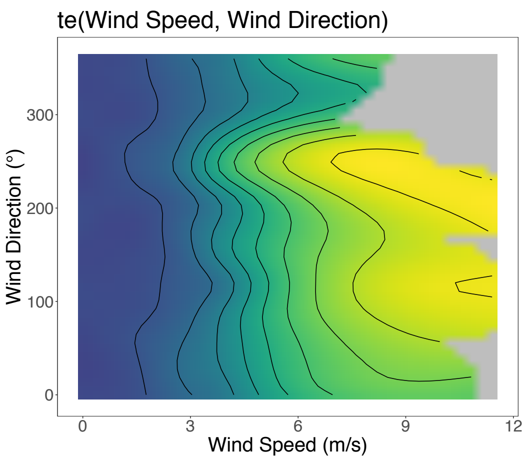
Moreover, the tensor products of the predictors are used to include the interactions among predictors as well as their main effects [18]. In some cases, the relation between a predictor and the target is affected by another predictor. To illustrate, Figure 4.1 shows the interaction between speed and direction where the effect on the wind power is shown with different colors. In the figure, the black contours demonstrate the wind power prediction level, and gray areas indicate that there is no observation in that particular region. Depending on the direction, the impact of speed on wind power



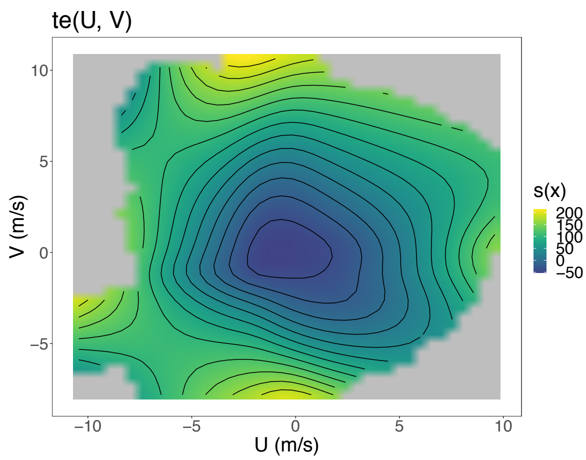
(a) Speed - Cubic



(b) Direction - Cubic vs Cyclic



(c) Speed & Direction



(d) U & V

Figure 4.1. Wind speed with cubic spline (a), wind direction with cubic vs cyclic splines (b), interaction of wind speed and wind direction (c) and interaction for u and v vectors for Dinar WF.

output changes. For instance, when the direction is in 230°-260° range, the power output reaches its maximum given the same wind speed. Similarly, Figure 4.1 shows the interaction between u and v vectors. Depending on the v vector values, the effect of u vector values on the output changes and vice versa.

Additionally, these models are flexible in adding other predictors. Hence, hour and month information are included to capture seasonal behaviour in wind power production. The assumption is that these variables may give us information about temperature, which affects the air density [42]. Studies show that air density also plays a role in wind power output [42]. Thus, adding those seasonal variables into the model may help incorporating the effect of air density in the output. Furthermore, the parameter estimation in GAMs can be made with different distributional assumptions. First, Gaussian assumption and second, Quasi-binomial distribution assumption is made as two alternatives.

4.2. Base-level Aggregation Methods

There are several issues while using grid-based NWP data in power curve modeling. Studies show that using multiple coordinates from multiple NWP models enhances the performance [8, 11]. Another issue is the level of aggregation of data from these multiple coordinates. In this thesis, there are two main approaches to address this problem, which are point-based power curve modeling and region-based power curve modeling.

4.2.1. Point-based Power Curve Modeling

In point-based modeling, data from single coordinate is used at the power curve modeling part. In order to obtain point-based predictions, first two base models, which are Weibull CDF and Five Parameters Logistic function, are used. The parameters in these functions are determined by the method of least squares:

$$\beta_i = \underset{\beta_i}{\operatorname{argmin}} \sum_{t=1}^{T} (P(t) - f(x_i(t), \beta_i))^2 \tag{4.1}$$

where f is either Weibull CDF of Logistic function, T is the number of observations and β_i is the set of parameters depending on f.

In this approach, the direction at coordinate i, $i \epsilon \{1, 2, ..., K\}$, which may involve coordinates from multiple NWP models, is discretized. For each discrete direction d, a different set of parameters are obtained by building separate power curve models. For each coordinate i, the wind power prediction for time t, $\widehat{P_i(t)}$, is obtained. As the final step, in order to obtain an aggregate forecast, the method of non-negative least squares is applied to obtain a weight w_i for each $\widehat{P_i(t)}$.

The aggregate forecast can be written as:

$$\widehat{P_{agg}(t)} = \sum_{\forall i} w_i \times \widehat{P_i(t)} \tag{4.2}$$

where $w_i \geq 0\ \forall i$ and $\sum_{\forall i} w_i = 1$.

4.2.2. Region-based Power Curve Modeling

In region-based power curve modeling, a single aggregate prediction is obtained in one step using weather data from each i. This approach enables modeling spatial relations directly with weather data from multiple coordinates in grids. Weather data used in this thesis consists of u and v vectors which can be expressed as:

$$u_i(t) = x_i(t)cos\theta_i(t) \tag{4.3a}$$

$$v_i(t) = x_i(t)sin\theta_i(t) \tag{4.3b}$$

where $u_i(t)$ is the u vector value at coordinate i at time t, $v_i(t)$ is the corresponding v vector value, and $\theta_i(t)$ is the corresponding wind direction.

Figure 4.2 shows an example of NWP grid points in the nearby region of Dinar WF, where the available wind vectors consisting of u and v components are illustrated. In this type of region-based modeling, spatial interpolation techniques have been used in order to estimate the wind field over the region [9, 43].

In this thesis, two alternative approaches are taken in region-based modeling. Firstly, we propose a modeling approach where the base models, Weibull CDF and

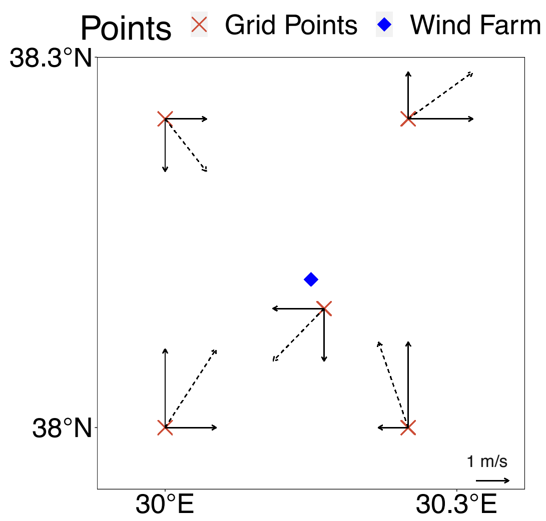


Figure 4.2. U and V components of wind vector in the nearby region of Dinar WF.

Logistic function, are utilized. The second approach is based on GAMs.

4.2.2.1. Modeling with Power Curve Models Based on Shape. In this approach, u and v vectors are directly used as inputs to Weibull CDF and Five Parameters Logistic function. First, weighted speed for a time period t is estimated by weighting u and v vectors separately. This strategy results in weighted u and v values, which are named as adjusted u and adjusted v respectively. These adjusted vectors can be shown as:

$$\widehat{u(t)} = \sum_{\forall i} w_i^u u_i(t) \tag{4.4a}$$

$$\widehat{v(t)} = \sum_{\forall i} w_i^v v_i(t) \tag{4.4b}$$

where w_i^u is the weight for u vector from i and w_i^v is the corresponding weight for v vector. With these adjusted u and v, the adjusted wind speed, $\widehat{x(t)}$, is calculated as:

$$\widehat{x(t)} = \sqrt{(\widehat{u(t)}^2 + \widehat{v(t)}^2)}. \tag{4.5}$$

The wind power predictions for time t, $\widehat{P(t)}$, are obtained using this adjusted speed, $\widehat{x(t)}$, in Weibull CDF and Logistic function in (3.1) and (3.2) separately. The weights for u and v are estimated by minimizing Sum of Squared Errors (SSE) of wind power predictions:

$$SSE = \sum_{t=1}^{T} (P(t) - \widehat{P(t)})^2. \tag{4.6}$$

Similar to point-based modeling, discretized wind direction is utilized in this formulation. For each direction d, different set of weights are obtained.

4.2.2.2. Modeling with Generalized Additive Models. In this approach, speed and direction information are directly used as inputs. Since the wind direction has an angular nature, cyclic cubic regression splines are used as smoothing functions for direction. Therefore, both continuity and cyclic behaviour are preserved. In order to incorporate the interaction between speed and direction, tensor products of speed with cubic regression spline basis and direction with cyclic cubic regression spline basis are utilized. To make the comparison of models accurately, the location of knots for direction is provided such that the intervals in both discretized and continuous cases are equivalent.

Alternatively, instead of using speed and direction directly, u and v vectors are also utilized to model the wind power. Since the direction is determined by the u and v vectors, including the interaction between these two variables adds information related to the wind direction into the model. Therefore, tensor products of u and v vectors with penalized cubic regression spline basis are used. Unlike the categorization of direction in the previous methods, the interaction between u and v has a continuous characteristic due to the nature of GAMs. Furthermore, u and v vectors are not given weights explicitly as in the previous method, but they are aggregated with smoothing splines.

Furthermore, when raw wind vectors, u and v, are utilized in this modeling strategy, the knots in GAMs divide u and v variables such that each region covers a rectangular area in the u-v space. Thus, some instances having the same direction but different speeds can belong to different regional categories. On the other hand, when the model is built with speed and direction, the knots divide these variables such that each instance with the same direction belongs to the same regional category in the speed-direction space.

The model in this approach can be simplified as:

$$P(t) = \sum_{i=1}^{K} f_i(x_i(t), \theta_i(t)) \tag{4.7}$$

where $f_i(t)$ is the tensor product of speed, x, and direction, θ, from coordinate i at time t. In the second case, $f_i(u_i(t), v_i(t))$ is used as the tensor product of $u_i(t)$ and $v_i(t)$ from coordinate i at time t.

In this method, as the first alternative, wind power production values are modeled with Gaussian distribution assumption. As the second alternative, as shown in Figure 1.3, parameter estimation with Quasi-binomial distribution assumption on normalized response is made. As a further step, variables related to seasonality, e.g. hour and month, are added into the model to capture seasonal behaviour in wind power output. Similar to the direction variable, cyclic cubic regression splines are used since both hour and month variables have cyclic behaviours.

4.3. Meta-level Features

This section provides the detailed description of meta-level features extracted from the available data. These features can be divided into four categories: Prediction-based meta-level features, performance-based meta-level features, seasonality-based meta-level features, and weather-based meta-level features. The purpose is to make use of available information from predictions and weather data, which may help predicting the ranks better.

Table 4.1 provides the detailed description of prediction-based meta-level features. Here, the aim is to exploit information from both the predictions and the relationships among the outputs of different models. In addition to predictions, $P_m(t)$ for each model m, and their next and previous neighbours, $Pred\_lead_m$ and $Pred\_lag_m$ respectively, the changes in consecutive predictions may give information about the relative performance of the models, which is denoted as $pdiff$. Furthermore, including the mean and the standard deviation of $pdiff$ from a specified window in a rolling basis can help capturing if there is variability in the outputs of m in that time interval. Those statistics are calculated in Equations 4.8 and 4.9 as:

$$pdiff\_m(t) = \frac{\sum_{i=t-w}^{t+w}(P_m(i) - P_m(i-1))}{2w+1} \tag{4.8}$$

$$pdiff\_sd(t) = \sqrt{\frac{\sum_{i=t-w}^{t+w}[(P_m(i) - P_m(i-1)) - pdiff\_m(i)]^2}{2w}}. \tag{4.9}$$

In order to account for the relationships between the outputs of the models, $corr\_m\_m'(t)$, where m' denotes the other model to which the correlation is calculated from a specified rolling window as shown in Equation 4.10, $hourly\_mean(t)$, which is hourly mean of all the model outputs, $hourly\_sd(t)$, which is hourly standard deviation of all the model outputs, and $dev(t)$, which is deviation of $P_m(t)$ from hourly mean of model outputs, are added:

$$corr\_m\_m'(t) = \frac{\sum_{i=t-w}^{t+w}(P_m(i) - \overline{P_m})(P_{m'}(i) - \overline{P_{m'}})}{\sqrt{(\sum_{i=t-w}^{t+w}(P_m(i) - \overline{P_m})^2)(\sum_{i=t-w}^{t+w}(P_{m'}(i) - \overline{P_{m'}})^2)}} \tag{4.10}$$

where window length is $2w+1$, and $\overline{P_m}$ is calculated as $\overline{P_m} = \frac{\sum_{i=t-w}^{t+w} P_m(i)}{2w+1}$ in a rolling basis.

Table 4.1. Prediction-based meta-level features.

Features	Details
Predicted	Predictions obtained with base-level learners, $P_m(t)$ for each model m
pdiff	Difference between two consecutive predictions
pdiff_m	Mean of difference between two consecutive predictions from a rolling window
pdiff_sd	Standard deviation of difference between two consecutive predictions from a rolling window
Pred_lead	Predictions for the future periods from a specified window
Pred_lag	Predictions from the past periods from a specified window
corr	Correlation between a predictor with each other predictor from a rolling window
hourly_mean	Hourly mean of the predictions
hourly_sd	Hourly standard deviation of the predictions
dev	Hourly deviation of the prediction from hourly mean of the predictions

Moreover, Table 4.2 gives the summary of performance-based meta-level features. The aim is to include information related to the past performance of the models. The underlying motivation is that if a model is recently at rank 1 most of the time, this model may be more likely to perform better in the next period. Additionally, any other performance-related information such as a model performing very bad recently may be important in predicting the ranks for the next period. Thus, the number of times a model ended up at rank 1 on the last day from two days before, which is the last available day in wind energy domain, num_rank1, and daily rank average of last day from 2 days before, $rank\_avg$, are included into the feature set. Moreover, model information is added because a model may outperform the other models in a specific set of conditions, which is related to performance-based meta-level features.

Table 4.2. Performance-based Meta-level Features.

Features	Details
num_rank1	Number of times each model has the first rank on the last day from 2 days before
rank_avg	Daily rank average of last day from 2 days before
model	Model information

Moreover, in wind power setting, including weather-related variables, which are summarized in Table 4.3, is meaningful. Since wind power output highly depends on wind speed and wind direction, variables that would account for either the levels of or changes in wind speed and wind direction are introduced. First, wind speed, $x_i(t)$, and wind direction, $\theta_i(t)$ are added to the feature set. These two variables show changes in very small intervals. Hence, their smoothed versions, mWs and mWd respectively, are calculated in a rolling basis by averaging from a window. Additionally, to include their variability into the model, their standard deviations from a window, $sdWs$ and $sdWd$ respectively, are added in a rolling basis. The reason is that some models may be good at reacting to those sudden changes, and some not to which these two variables would be linked. Furthermore, air turbulence is added to include the variability in wind speed at different levels of atmosphere, and it is estimated by $x_i^{100}(t)/x_i^{10}(t)$, where the $x_i^{100}(t)$ is the wind speed at coordinate i at 100 m above ground, and $x_i^{10}(t)$ is the corresponding speed at 10 m above ground. As mentioned in Section 4.1.3 section, air density is found to be affecting the wind power output, and temperature has a big impact in air density [42]. Thus, temperature information is also added. The variability with respect to NWP models may also be meaningful since some NWP models are known to be good at some certain weather occurrences [8]. Thus, the difference between wind speed from GFS and from HIR, also denoted as ws_gh, is added. Moreover, since information from multiple coordinates can be fused into the model, regional hourly standard deviations of wind speed, and u and v components of wind speed from GFS are included. The regional standard deviation of wind speed is

Table 4.3. Weather-based Meta-level Features.

Features	Details
ws	Hourly speed for each i
wd	Hourly direction for each i
mWs	Mean of hourly speed from a rolling window for each i
sdWs	Standard deviation of hourly speed from a rolling window for each i
mWd	Mean of hourly direction from a rolling window for each i
sdWd	Standard deviation of hourly speed from a rolling window for each i
airt	Air turbulence estimated by (speed at 100 m above ground)/(speed at 10 m above ground) for each i from GFS
ws_gh	Difference in speed between 80 m above ground from GFS and 10 m above ground from HIR fro each i from GFS
Wsgfs_sd	Regional hourly standard deviation of speed including each i from GFS
Ugfs_sd	Regional hourly standard deviation of u component of speed including each i from GFS
Vgfs_sd	Regional hourly standard deviation of v component of speed including each i from GFS
temp	Hourly temperature for each i

calculated in the Equation 4.11, and the other are calculated similarly:

$$Wsgfs\_sd(t) = \sqrt{\frac{\sum_{i=1}^{G}[x_i(t) - \frac{\sum_{j=1}^{G} x_j(t)}{G}]^2}{G-1}} \tag{4.11}$$

where G is the number of coordinates from GFS.

Furthermore, seasonality-based features i.e., hour and month are also included. Since temperature and pressure have impact on air density [42], hour and month information is included because these two values can also give information related to both temperature and pressure, which in turn may affect the performance of a model.

4.4. Meta-learning Approaches

In this section, the alternative frameworks for meta-learning problem are discussed. In this thesis, different approaches are incorporated in order to solve meta-learning problem: pointwise approach, pairwise approach and listwise approach. In pointwise approach, the formulation settings are: rank 1 class labeling approach and rank 1 vs other approach in classification setting, regression setting and ordinal regression setting. According to the problem setting, the target is converted and treated accordingly, which will be explained in detail in the respective sections below.

4.4.1. Pointwise Approach

In this section, different formulation settings for pointwise approach are explained in detail.

4.4.1.1. Classification Setting - Rank 1 Class Labeling Approach. As the starting setting, the meta-learning problem is approached as a classification problem. Thus, the target in this setting is a class variable. In this scenario, rank 1 class labels which show the best performer for each hour are taken as the target. Figure 4.3 illustrates this scenario.

day	hour	feature 1	feature 2	...	target
day 1	0				**model 1**
day 1	1				**model 3**
day 1	2				**model 4**
day 1	...				
day 2	0				**model 3**
...	...				**...**

Figure 4.3. Rank 1 class labeling approach.

In this approach, the features can be selected from the feature set explained in

4.3. The feature set is extended such that for each model m, the corresponding feature is added as a new predictor to the input space. The target, which is the model at rank 1 for that hour, is a categorical variable which allows using any classifier in this setting. Thus, decision trees and XGBoosting can be used to model this dynamic meta-learning problem in rank 1 class labeling approach.

An example tree with the features from Section 4.3 is shown in Figure 4.4. For illustration, six models from base-level model pool are selected. These are aggregated Weibull CDF (P_W) and aggregated Five Parameters Logistic function (P_F) from point-based modeling, and from region-based modeling, Weibull CDF (R_W), Five Parameters Logistic function (R_F), GAMs with gaussian distribution assumption ($GAM1$) and GAMs with quasi-binomial distribution assumption with seasonality ($GAM3$) where speed and direction are used as input with the following tree:

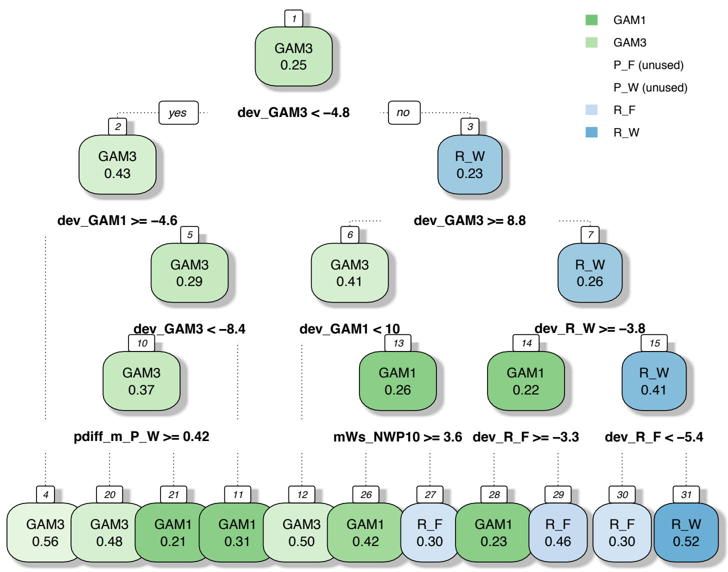


Figure 4.4. A simple decision tree with rank 1 class labeling approach built with a subset of both prediction-based and weather-based meta-level features for Dinar WF.

The figure shows a simple decision tree with a subset of features from prediction-based and weather-based meta-level features, which are dev, denoted as dev_model, $pdiff\_mean$, denoted as $pdiff\_m\_model$ for each model m, and mWs, denoted as $mWs\_coordinate$ for each coordinate i. The leaf nodes show the rank 1 performing model with the corresponding set of decision rules. In this tree, the most important feature, showing the capability of decreasing the variance in the input the most, is *dev* of $GAM3$ model with being in the first split. For instance, the tree shows that if *dev* of $GAM3$ is greater than or equal to 8.8 and *dev* of $GAM1$ is less than 10, then the best performer will be $GAM3$ with 0.5 probability.

The drawback of this modeling strategy is that it ignores the whole set of rankings of the base-level models, but only considers the rank 1 performer. Furthermore, this strategy ignores the pairwise rank relations among the base-level models. Although approaching the problem with rank 1 class labeling approach is very simple, it has these aforementioned drawbacks.

4.4.1.2. Classification Setting - Rank 1 vs Other Approach. In the second approach from classification setting, the input data is expanded such that each day-hour pair has R, which is the total number of base-level models, instances. Hence, the features are not expanded as in the previous rank 1 class labeling approach. Similar to the previous approach, the features are selected from the feature set described in Section 4.3. Unlike in rank 1 class labeling approach, *model* information can be utilized in rank 1 vs other approach. Hence, if in the training data a model has either apparently good or bad performance, which means there is a correlation between model type and rank, this information can be meaningful in splitting procedure. The target, which is whether a base-model is at rank 1 or not, is again a categorical variable. Hence, if the base-model is in rank 1 in that pair, then the target is 1, and if not, the target is 0, which requires modeling a binary classification problem. Hence, decision trees and XGBoosting can be used to model this dynamic meta-learning problem in rank 1 vs other approach. Figure 4.5 illustrates this scenario.

An example tree with the same subset of features from Figure 4.4 and additional *model* information with the same set of base-models is shown in Figure 4.6. From this set, *dev* turned out to be the most important feature with being in the first split. This tree shows that *model* information is also significant in explaining the variance in the

	day	hour	feature 1	feature 2	...	target
model 1	day 1	0				**1**
model 2	day 1	0				**0**
model 3	day 1	0				**0**
...	day 1	...				**0**
model M	day 1	0				**0**
model 1	day 1	1				**0**
model 2	day 1	1				**0**
model 3	day 1	1				**1**
...	day 1	...				**0**
model M	day 1	1				**0**
...	...	...				**...**

Figure 4.5. Rank 1 vs other approach.

input data. Furthermore, the leaf nodes show the class probabilities of either class 0 or class 1. For instance, if *dev* of a model, m, is less than or equal to -3.9 and if mWs of $NWP11$ is greater than or equal to 2.7, then it has 0.77 probability of not being in rank 1.

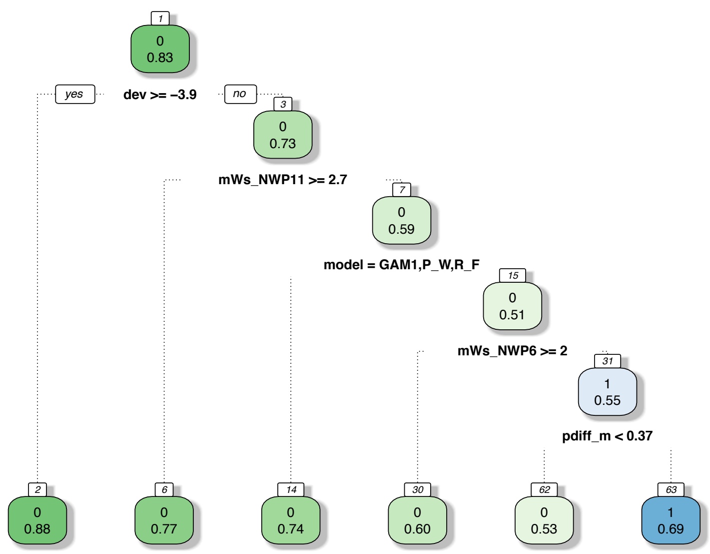


Figure 4.6. A simple decision tree with rank 1 vs other approach built with the same subset from prediction-based and weather-based meta-level features and model information for Dinar WF.

However, this modeling approach has some drawbacks. Similar to rank 1 class labeling approach, this setting ignores the whole set of rankings, thus pairwise relations. Moreover, with the number of base-level models, R, increases, the class imbalance problem becomes more critical. For example, in the scenario where there are six base-level models in the poll, the proportion of class 1 to class 0 is 0.2. Hence, the model will bias towards predicting class 0 due to the class imbalance.

<u>4.4.1.3. Regression Setting.</u> As the third alternative, the meta-learning problem is modeled as a regression problem. In this setting, the error, $E_m(t)$, which is calculated by $P(t) - \widehat{P_m(t)}$, where $\widehat{P_m(t)}$ is the wind power prediction of model m for time t, is used as the target. In this approach, for each time t, there are R instances with the error being the target. Since the target is error, the predictions need a simple postprocessing step such that the model with the least absolute predicted error is chosen as the best performer at time t by $\text{argmin}_m |\widehat{E_m(t)}|$, where $\widehat{E_m(t)}$ is the predicted error of model m at time t. The input data of this approach is depicted in Figure 4.7. In this setting, any regressor can be used to obtain the error predictions for the test set. Hence, regression trees and XGBoosting are utilized to model this task in regression setting.

	day	**hour**	**feature 1**	**feature 2**	**...**	**target**
model 1	day 1	0				**0.031**
model 2	day 1	0				**0.682**
model 3	day 1	0				**-2.381**
...	day 1	...				**12.326**
model M	day 1	0				**3.459**
model 1	day 1	1				**-5.495**
model 2	day 1	1				**18.264**
model 3	day 1	1				**-1.194**
...	day 1	...				**2.387**
model M	day 1	1				**1.593**
...	...	...				**...**

Figure 4.7. Regression error-based approach.

Figure 4.8 shows an example of regression tree built with the same feature set of the tree in Figure 4.6. In this setting, the outputs in the leaf nodes show the mean of the error of the instances, which belong to the corresponding leaf node, with the percentages to the total number of instances. For instance, 6 % of the instances have dev between -8 and 3.1 and mWs_NWP10 greater than or equal to 4.8 with an average error of -9.2. As a result, the mean of the corresponding leaf nodes becomes the predicted errors for the test set.

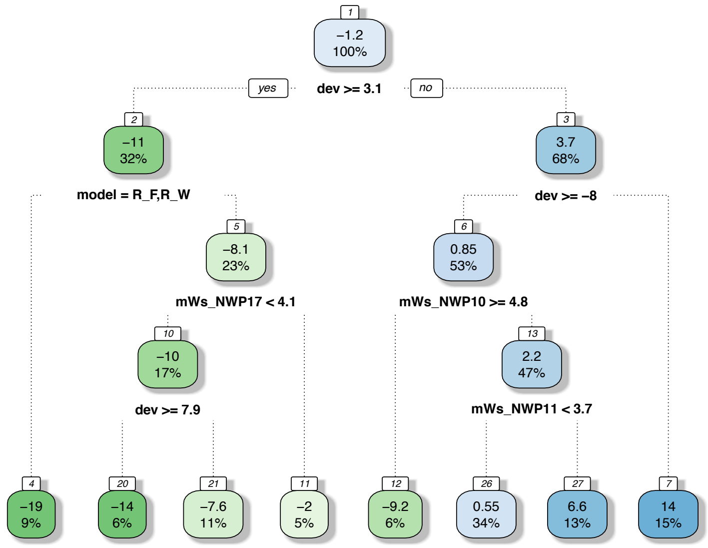


Figure 4.8. A simple tree with regression setting built with a subset of prediction-based and weather-based meta-level features and model information for Dinar WF.

In this setting, there is no class imbalance problem as in rank 1 vs other approach, but this approach still ignores the pairwise relations due to the nature of the regression setting. Furthermore, one drawback specific to regression trees is that multiple models can have the minimum absolute predicted error, which would result in multiple models with rank 1 prediction, again due to the nature of the regression trees. However, this drawback does not exist if XGBoosting is used as the learner since multiple trees are build in XGBoosting algorithm.

4.4.1.4. Ordinal Regression Setting. As the fourth alternative in pointwise approach, the dynamic meta-learning problem is modeled in an ordinal regression setting. In this setting, the target is the rank of each model m as an ordinal variable. Thus, in these approaches, the order in the variable is also taken into account. In this formulation, for each time t, there are R instances with their ranks being the target. This setting is shown in Figure 4.9. In order to solve the meta-learning problem in this setting, decision trees are used in this thesis. When the target is an ordinal variable, the order in the target is also considered unlike with categorical variables.

	day	hour	feature 1	feature 2	...	target
model 1	day 1	0				1
model 2	day 1	0				2
model 3	day 1	0				3
...	day 1	...				5
model M	day 1	0				4
model 1	day 1	1				4
model 2	day 1	1				5
model 3	day 1	1				1
...	day 1	...				3
model M	day 1	1				2
...	...	...				...

Figure 4.9. Ordinal regression setting.

Figure 4.10 shows a simple example of ordinal regression tree built with the same feature set of the trees in Figures 4.6 and 4.8. In this setting, the outputs of the leaf nodes show the mode of the rank of the instances which belong to that leaf node. Hence, the predictions for the test set is made based on the mode of the rank in the corresponding leaf node. For instance, if dev is greater than or equal to 3.4 and mWs of $NWP17$ is greater than or equal to 4.1, then the model instance has 0.23 probability of being in rank 1 with less probabilities for the other ranks.

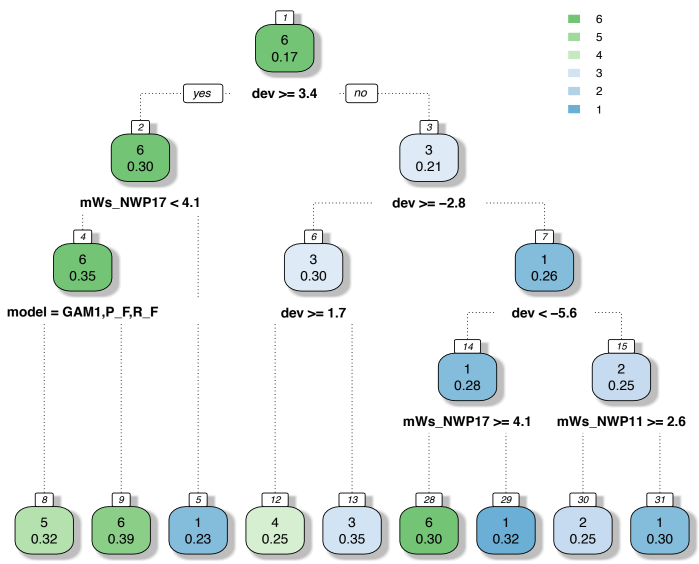


Figure 4.10. A simple tree with ordinal regression setting built with a subset of prediction-based and weather-based meta-level features and model information for Dinar WF.

The advantage of this modeling strategy is that it considers orders in the target variable. However, with this strategy, there may be multiple models with rank 1 prediction or no model with rank 1 prediction. In order to solve the former case, taking the mean of the outputs of the models with rank 1 predictions is a good option. On the other hand, to solve the latter problem, models with the minimum rank prediction can be assumed to be the best performer for that time t.

4.4.2. Pairwise Approach

In this setting, meta-learning problem is approached as a pairwise ranking problem. XGBoosting is a good option to solve pairwise ranking problems since it uses LambdaMART algorithm [38] as explained in Section 3.4.2. The input data used for this setting is shown in Figure 4.11. The difference of this input setting to the previous ordinal regression setting is that now the observations are sorted with respect to their ranks and are fed to the model with group information, which is shown in *group* column. A group shows the day-hour pair in this thesis, and pairs of base-level models are created from each group within the model. This modeling strategy overcomes the drawbacks of the previous settings such as class imbalance problem, ignoring pairwise relations, having the same rank prediction for multiple models etc., hence making this strategy very advantageous.

	day	hour	feature 1	feature 2	...	target	group
model 1	day 1	0				1	1
model 2	day 1	0				2	1
model 3	day 1	0				3	1
model M	day 1	0				4	1
...	day 1	...				5	1
model 3	day 1	1				1	2
model M	day 1	1				2	2
...	day 1	...				3	2
model 1	day 1	1				4	2
model 2	day 1	1				5	2
...	...	...				...	...

Figure 4.11. Pairwise and listwise ranking settings.

4.4.3. Listwise Approach

In this setting, the dynamic meta-learning problem is modeled with the listwise approach. Similar to pairwise approach, XGBoosting is a strong alternative to solve listwise ranking problems as mentioned in 3.4.3. It again uses LambaMART algorithm which maximizes Normalized Discounted Cumulative Gain (NDCG) in listwise setting [41]. In this setting, the complete set of instances are taken into account in loss function calculation [24]. The input data used for this setting is the same with pairwise setting which is shown in Figure 4.11. Similar to pairwise approach, this approach also overcomes the limitations of the pointwise setting e.g., class imbalance problem, having the same rank prediction for multiple models etc., which also makes this strategy strong.

5. EXPERIMENTS ON REAL DATA

In this section, the description of the data, the performance metrics and the experimental settings are explained in detail. As the final step, the performances of the modeling approaches in both base-level and meta-level are compared.

5.1. Data Description

The experiments are made on six wind farms, which are Aliaga Bergama, Bares, Dinar, Geycek, Soke and Soma 1-2 Wind Farms, in Turkey. They are selected such that the spatial conditions are not the same around the wind farms. The locations of the wind farms can be seen in Figure 5.1.

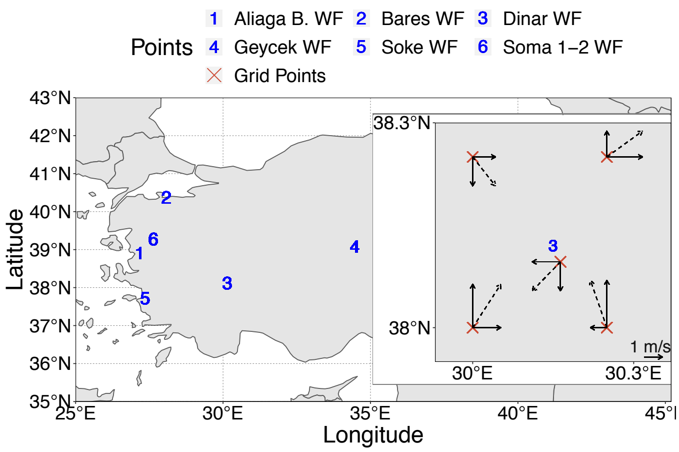


Figure 5.1. Wind farms used in experiments.

Two sets of weather forecasts are used in the experiments. One is provided by Global Forecasting System (GFS) [35], and the other one is provided by High Resolution Limited Area Model (HIRLAM) program [36]. In this thesis, forecasts for u and v components at 80 meters above ground level from GFS and at 10 meters above ground level from HIRLAM with 3-hour temporal resolution are utilized. On the other hand, power production data has 1-hour temporal resolution. Therefore, the weather datasets are interpolated to bring the datasets into the same resolution. In the experiments in both base-level and meta-level, weather data from neighbouring four coordinates from GFS model and the closest coordinate from HIRLAM is utilized.

5.2. Performance Metrics

In this section, the metrics used for comparison of different modeling strategies in both base-level and meta-level are explained.

5.2.1. Base-level

In base-level, the performances of the models are compared based on two performance metrics. Among the metrics extensively used in wind power models [12], normalized mean absolute percentage error (NMAPE) and root mean squared error (RMSE) are used. These metrics can be calculated with the following expressions:

$$NMAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|\widehat{P(i)} - P(i)|}{max_{i=1}^{N}(P(i))} \times 100, \tag{5.1a}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{P(i)} - P(i))^2}. \tag{5.1b}$$

5.2.2. Meta-level

In meta-level, the performances of the models are first compared with respect to NMAPE and RMSE of the models which are predicted as rank 1. This comparison allows to see how much improvement is gained with respect to the model from the base-level with overall superior performance.

Additionally, there are metrics which are widely used in information retrieval tasks. Since algorithm ranking and selecting is a ranking task, those well-known metrics such as mean reciprocal rank (MRR) [44] and normalized discounted cumulative gain

(NDCG) [39] are applicable and useful in this thesis as well. Thus, these two measures are adapted to algorithm ranking task. In this thesis, MRR is calculated as:

$$MRR = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{\widehat{r_{1t}}} \tag{5.2}$$

where $\widehat{r_{1t}}$ is the predicted rank of the base-level algorithm with actual rank 1 for time t. As it can be seen from the equation, this metric measures the performance in predicting algorithms at rank 1.

On the other hand, NDCG@R is computed for top R algorithms, which is the number of the base-level algorithms, in algorithm ranking setting. Hence, this metric measures the overall ranking performance. First, NDCG@R for each time t is computed as:

$$DCG_t@R = \sum_{i=1}^{R}\frac{\widehat{n_{it}}}{log_2(i+1)} \tag{5.3a}$$

$$NDCG_t@R = \frac{DCG_t@R}{max(DCG_t@R)} \tag{5.3b}$$

where $\widehat{n_{it}} = R - \widehat{r_{it}} + 1$, $\widehat{r_{it}}$ is the predicted rank of the base-level algorithm with actual rank i for time t, and max(DCG$_t$@R) is the highest achievable DCG@R. $\widehat{n_{it}}$ is calculated in order to link the ranks in this thesis to the relevance scores in information retrieval tasks. Since this computation calculates NDCG@R for single time period t, the overall NDCG$_{mean}$ is calculated by averaging NDCG@R values as:

$$NDCG_{mean} = \frac{\sum_{t=1}^{N} NDCG_t@R}{N}. \tag{5.4}$$

5.3. Experimental Settings

In this section, the experimental settings for both base-level and meta-level are described in detail.

5.3.1. Base-level

In the experiments, the data is split into the training and test sets such that the training set expands in each iteration, i.e. in each month, in a rolling basis. For the first test month, the training data is from April 2017 to June 2019 for all selected wind farms. The test data is from July 2019 to June 2020 for all wind farms.

In this thesis, as an alternative to using hourly speed and direction data directly, they are smoothed by calculating rolling averages from a specified time window. This approach can harm the forecasts at specific time points when some extreme events occur. However, the results of Weibull CDF and 5 Parameters Logistic function in each coordinate i, $i \epsilon \{GFS_1, GFS_2, GFS_3, GFS_4, HIR_1\}$ for Dinar WF in Table 5.1 show that the models with smoothed speed and direction outperform the models with nonsmoothed hourly wind data in terms of NMAPE. Thus, smoothing is performed in the remaining experiments, and the time window is chosen as 7, which worked better compared to other neighboring time windows. The smoothed speed is calculated by first calculating the smoothed u, $\overline{u}_i(t)$, as $\overline{u}_i(t) = \frac{1}{7}\sum_{w=-3}^{w=3} u_i(t+w)$ and smoothed v, $\overline{v}_i(t)$, as $\overline{v}_i(t) = \frac{1}{7}\sum_{w=-3}^{w=3} v_i(t+w)$.

As the next step, the smoothed speed, $\overline{x}_i(t)$, and the smoothed direction, $\overline{\theta}_i(t)$, are calculated as:

$$\overline{x}_i(t) = \sqrt{\overline{u}_i(t)^2 + \overline{v}_i(t)^2}, \tag{5.5a}$$

$$\overline{\theta}_i(t) = \arctan \overline{v}_i(t)/\overline{u}_i(t). \tag{5.5b}$$

Moreover, in the approaches where discrete wind direction information is used, the number of bins is chosen to be 12 resulting in 30° intervals.

5.3.2. Meta-level

In the meta-level experiments, the data is split into the training and test sets such that the training set expands in each iteration, i.e. in each month, in a rolling basis similar to the base-level. For the first test month, the training data is from April 2019 to June 2019 for all selected wind farms. The test data is from July 2019 to June 2020 for all wind farms. The test set is the same with base-level, so that a comparison with base-level algorithms can also be made.

5.4. Results

In this part, the performances of the methods are provided and compared for the selected six wind farms.

5.4.1. Point-based Power Curve Modeling

In the first method, power curve is modeled with either Weibull CDF or Five Parameters Logistic function. The worst and the best performances from the neighbouring coordinates along with the aggregated results can be found in Tables 5.2 and 5.3. The range of the performances differ in different wind farms, which may arouse from different weather and spatial conditions and various turbine characteristics. Nevertheless, in almost all wind farms, the performance resulting from aggregating the coordinates is better than the best coordinate for both Weibull CDF and Logistic function in terms of both NMAPE and RMSE. Furthermore, Weibull CDF outperforms the Logistic function in the aggregate case in all wind farms in terms of NMAPE.

Additionally, Figure 5.3 shows the monthly superiority of the aggregate case for Weibull CDF ((P) Wei. CDF) in terms of NMAPE for Dinar WF. C_{best} is the individual coordinate with the superior overall performance, and C_{worst} is the coordinate with the worst overall performance. We also observed that the best performing coordinate changes over months, which supports our motivation to aggregate information from

Table 5.1. Comparison of predictions in point-based modeling with smoothed and not smoothed wind speed for Dinar WF.

Coordinates	NMAPE (%)			
	Weibull CDF		5 Parameters Logistic Function	
	Smoothed (Overall)	Not Smoothed (Overall)	Smoothed (Overall)	Not Smoothed (Overall)
GFS_1	14.295*	14.935	14.369*	14.981
GFS_2	13.797*	14.635	13.881*	14.718
GFS_3	15.250*	15.970	15.244*	15.917
GFS_4	14.007*	14.631	13.950*	14.470
HIR_1	13.978*	15.477	14.015*	15.492
Agg.	12.028*	12.872	12.206*	13.007

Table 5.2. The overall performance of point-based modeling with Weibull CDF in all wind farms.

WF	NMAPE (%)			RMSE		
	C_{worst}	C_{best}	Agg.	C_{worst}	C_{best}	Agg.
Aliaga	17.395	10.639	9.732*	30.489	19.080	17.082 *
Soma 1-2	14.879	11.320	10.891*	24.212	19.041	18.170*
Dinar	15.250	13.797	12.028*	37.529	32.253	28.224*
Bares	17.241	12.724	11.769*	12.238	9.621	8.795*
Soke	20.635	16.751	16.732*	11.733	9.891	9.492*
Geycek	11.942	10.662	10.208*	29.572	26.244	24.353*

different coordinates.

Table 5.3. The overall performance of point-based modeling with Five Parameters Logistic function in all wind farms.

WF	NMAPE (%)			RMSE		
	C_{worst}	C_{best}	Agg.	C_{worst}	C_{best}	Agg.
Aliaga	17.419	10.745	9.906*	30.537	18.873	16.865*
Soma 1-2	14.441	11.529	10.992*	23.599	18.833	17.941*
Dinar	15.244	13.881	12.206*	37.351	31.991	28.047*
Bares	17.546	13.016	12.251*	12.232	9.541	8.753*
Soke	20.799	17.015*	17.103	11.621	9.861	9.501*
Geycek	12.168	10.906	10.473*	29.586	26.387	24.376*

5.4.2. Region-based Power Curve Modeling

In this section, the results for region-based power curve modeling are provided.

5.4.2.1. Modeling with Power Curve Models Based on Shape. In the first approach, where power production is modeled with shape-based models with u and v vectors directly, Weibull CDF outperforms Logistic function in all wind farms in terms of NMAPE as shown in Table 5.4. On the other hand, Logistic Function works slightly better in three wind farms in terms of RMSE. However, the performances of these two models are very close.

5.4.2.2. Modeling with Generalized Additive Models. In the second approach, first, tensor products of speed and direction are used to model production with Gaussian assumption (Case 1). Second, the same basis is utilized to model utilization values with Quasi-binomial distribution assumption, which improves the performance (Case 2). Third, variables related to seasonality such as hour and month are added to the second model, which brings a further improvement (Case 3). Additionally, instead of wind speed and wind direction, u and v vectors are directly utilized with tensor

Table 5.4. The overall performance of region-based modeling with shape-based methods in all wind farms.

WF	NMAPE (%)		RMSE	
	Weibull CDF	5 Parameters Log. Function	Weibull CDF	5 Parameters Log. Function
Aliaga	8.803*	8.905	16.265	16.133*
Soma 1-2	10.170*	10.306	17.701	17.381*
Dinar	11.454*	11.615	28.199*	28.317
Bares	11.245*	11.591	8.852	8.794*
Soke	15.913*	16.290	9.676*	9.761
Geycek	10.203*	10.567	25.325*	25.527

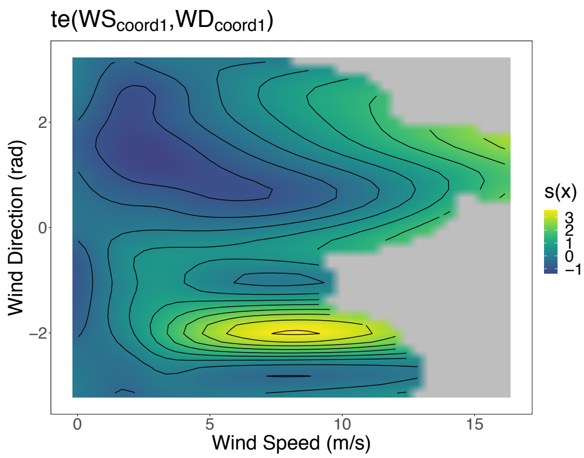
products of penalized cubic regression spline basis in Cases 4 and 5. Case 5 includes seasonality related variables similar to Case 3.

Some of the model outputs of Case 3 is demonstrated in Figure 5.2. The spline and interaction plots are obtained from the model with the following expression:

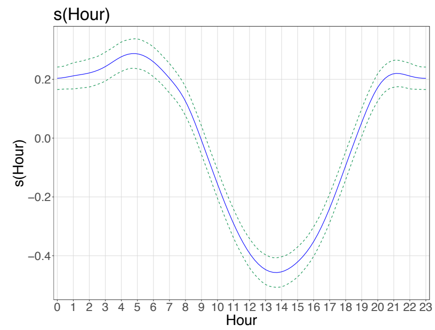
$$P(t) = \sum_{i=1}^{5} f_i(WS_i(t), WD_i(t)) + f_6(hour, month) \tag{5.6}$$

where $f_i(WS_i(t), WD_i(t))$ is the tensor product of speed and direction. On the other hand, $f_6(hour, month)$ can be either only the main effects of hour and month or with their interaction. The figure depicts the interaction effect of speed and direction for one single coordinate. It can be observed that the effect of speed on the level of output is different depending on the direction. Moreover, the effects of hour and month on the output, which show cyclic behaviour, can also be observed.

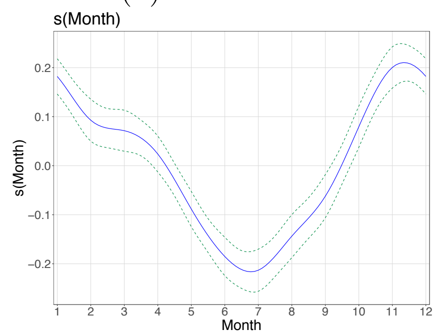
The overall comparison of these cases is provided in Tables 5.5 and 5.6. In terms of both NMAPE and RMSE, parameter estimation with Quasi-binomial distribution assumption enhances the performance in all wind farms (Cases 1 and 2). Moreover,



(a) Speed & Direction effect



(b) Hour effect



(c) Month effect

Figure 5.2. Interaction effect of wind speed and wind direction in one nearby coordinate (a), hourly seasonality plot (b) and monthly seasonality plot (c) for Dinar WF.

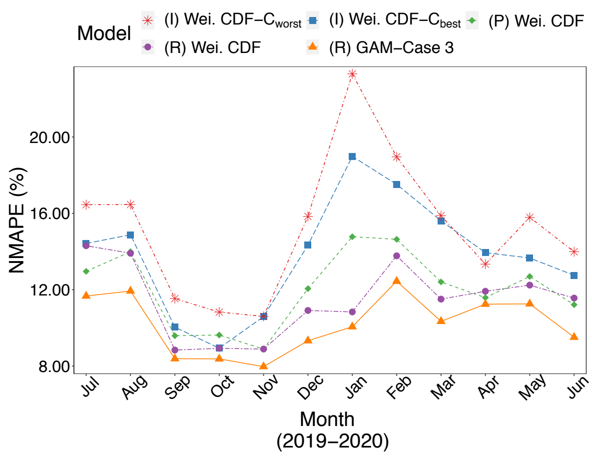


Figure 5.3. Comparison of best models from each category in terms of NMAPE for Dinar WF.

adding seasonality related variables improves the performance further in terms of both NMAPE and RMSE (Cases 3 and 5). These results also show that using wind speed and wind direction information as the input provides a superior performance than using u and v vectors as inputs.

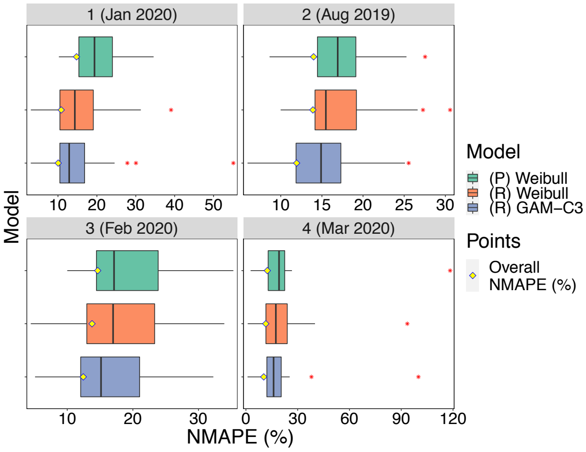
Table 5.5. The overall performance of region-based modeling with GAMs in all wind farms in terms of NMAPE.

WF	NMAPE (%)				
	Case 1	Case 2	Case 3	Case 4	Case 5
Aliaga	8.879	8.491	8.324*	8.678	8.489
Soma 1-2	10.285	10.176	9.744	10.170	9.683*
Dinar	11.160	10.815	10.191*	10.988	10.268
Bares	11.017	10.904	10.582*	10.971	10.649
Soke	16.029	15.952	15.135*	16.132	15.236
Geycek	10.511	10.456	9.908*	10.563	9.980

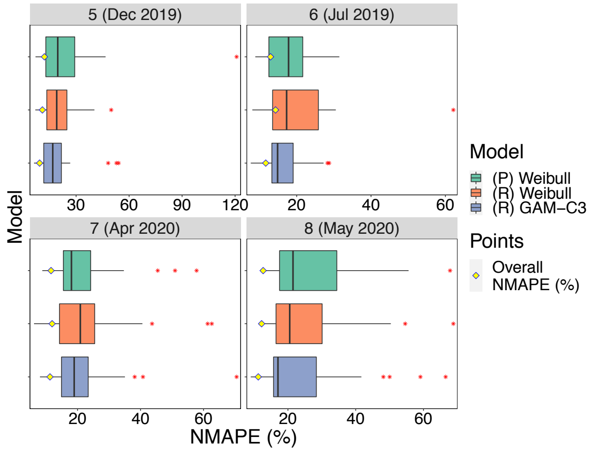
Table 5.6. The overall performance of region-based modeling with GAMs in all wind farms in terms of RMSE.

WF	RMSE				
	Case 1	Case 2	Case 3	Case 4	Case 5
Aliaga	15.772	15.874	15.582*	16.248	15.927
Soma 1-2	17.133	17.256	16.769*	17.427	16.843
Dinar	26.733	26.537	24.756*	26.950	24.896
Bares	8.177	8.340	8.101*	8.411	8.168
Soke	9.434	9.485	9.121*	9.644	9.207
Geycek	25.733	26.241	25.292*	26.460	25.384

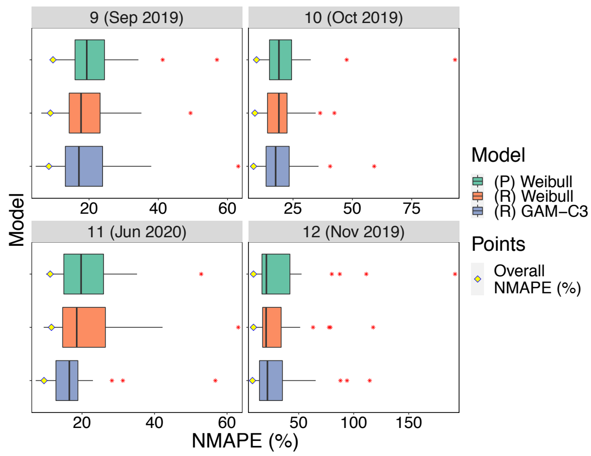
Additionally, the detailed performance comparison of three best scenarios from each aggregation method in Dinar WF in terms of NMAPE is provided in Figures 5.3 and 5.4. Figure 5.3 shows that the third case from region-based approach with GAMs



(a) High Volume



(b) Mid Volume



(c) Low Volume

Figure 5.4. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of Dinar WF.

shows a superior performance in all months. In wind power energy domain, it is also meaningful to analyze the daily performances of the models in order to demonstrate the distributional behavior of the modeling strategies. The box plots in Figure 5.4 are provided for each month separately where months are grouped and ordered based on average power production value. The first four months, which are January, August, February and March, are in high volume category, and the median NMAPE values of these months are mainly lower than the medium and low volume categories. Moreover, except in November and April, the median of daily NMAPE values of region-based approach with GAMs is lower compared to the Weibull CDF case from both point-based and region-based aggregation, whereas the overall NMAPE of GAMs is still better than the other two approaches in those two months. Hence, these results support using region-based modeling strategy with GAMs with Quasi-binomial distribution assumption.

5.4.3. Meta-learning Algorithm Ranking

5.4.3.1. Base-level Model Pool. This section gives the summary of the performances of the methods from meta-learning part. As discussed in the previous sections, three base models with different scenarios are applied in order to predict wind power. These models form the base-level learner pool of the meta-learning task. Hence, the outputs of the following six scenarios are utilized in the meta-learning task:

(i) P_W: Point-based modeling with Weibull CDF (Aggregated case)
(ii) P_F: Point-based modeling with Five Parameters Logistic function (Aggregated case)
(iii) R_W: Region-based modeling with Weibull CDF
(iv) R_F: Region-based modeling with Five Parameters Logistic function
(v) GAM1: Region-based modeling with Generalized Additive Models (Case 1)
(vi) GAM3: Region-based modeling with Generalized Additive Models (Case 3)

These models are selected from the model-aggregation type pair pool such that they are not very similar but strong learners. Thus, the predictions from the individual coordinates obtained before the point-based aggregation are not included. Furthermore, from the GAMs scenarios, one scenario with Gaussian distribution assumption (Case 1) and one scenario with Quasi-binomial distribution assumption on the target (Case 3) as detailed in Section 5.4.2.2 are included to the pool. Figure 5.5 shows the correlation matrix of the base-level model pool for Dinar WF. It can be seen that the correlation among the outputs of the base-level models is still very high. The overall performances of the base-learners for all wind farms, which are used in meta-learning

Table 5.7. The overall performance of base-level model pool in all wind farms in terms of NMAPE.

WF	NMAPE (%)					
	Point-Based		Region-Based			
	Weibull CDF	5 Param. Log. Func.	Weibull CDF	5 Param. Log. Func.	GAMs	
					Case 1	Case 3
Aliaga	9.732	9.906	8.803	8.905	8.879	8.324*
Soma 1-2	10.891	10.992	10.170	10.306	10.285	9.744*
Dinar	12.028	12.206	11.454	11.615	11.160	10.191*
Bares	11.769	12.251	11.245	11.591	11.017	10.582*
Soke	16.732	17.103	15.913	16.290	16.029	15.135*
Geycek	10.208	10.473	10.203	10.567	10.511	9.908*

part, are summarized in Table 5.7.

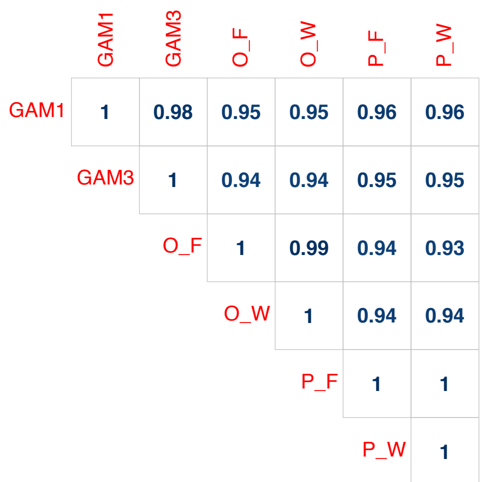


Figure 5.5. The correlation of wind power outputs among base-level model pool for Dinar WF.

5.4.3.2. Feature Selection. In this section, feature selection procedure is described. As explained in Section 4.3, there are 78 features extracted in total from four categories. These features include prediction-based meta-level features, weather-based meta-level features, performance-based meta-level features and seasonality-based features. Weather-based meta-level features are expanded for each coordinate i. These features are summarized in Tables 4.1, 4.2 and 4.3.

In feature selection step, XGBoosting algorithm is used since it has a feature importance property [45]. The algorithm calculates the relative importance of each feature by measuring the information gain with that feature [45]. In this part, first, the combination of four categories from Section 4.3 are compared. The summary of these scenarios are given in Table 5.8. Then, based on XGBoosting feature importance matrix, three different additional scenarios are compared: first 40 features, first 50 features and first 60 features from the whole feature set. In this step, the features are selected from the feature importance matrix of the training data that consists of observations from first three months for meta-learning part of each wind farm.

Table 5.8. Feature selection scenarios.

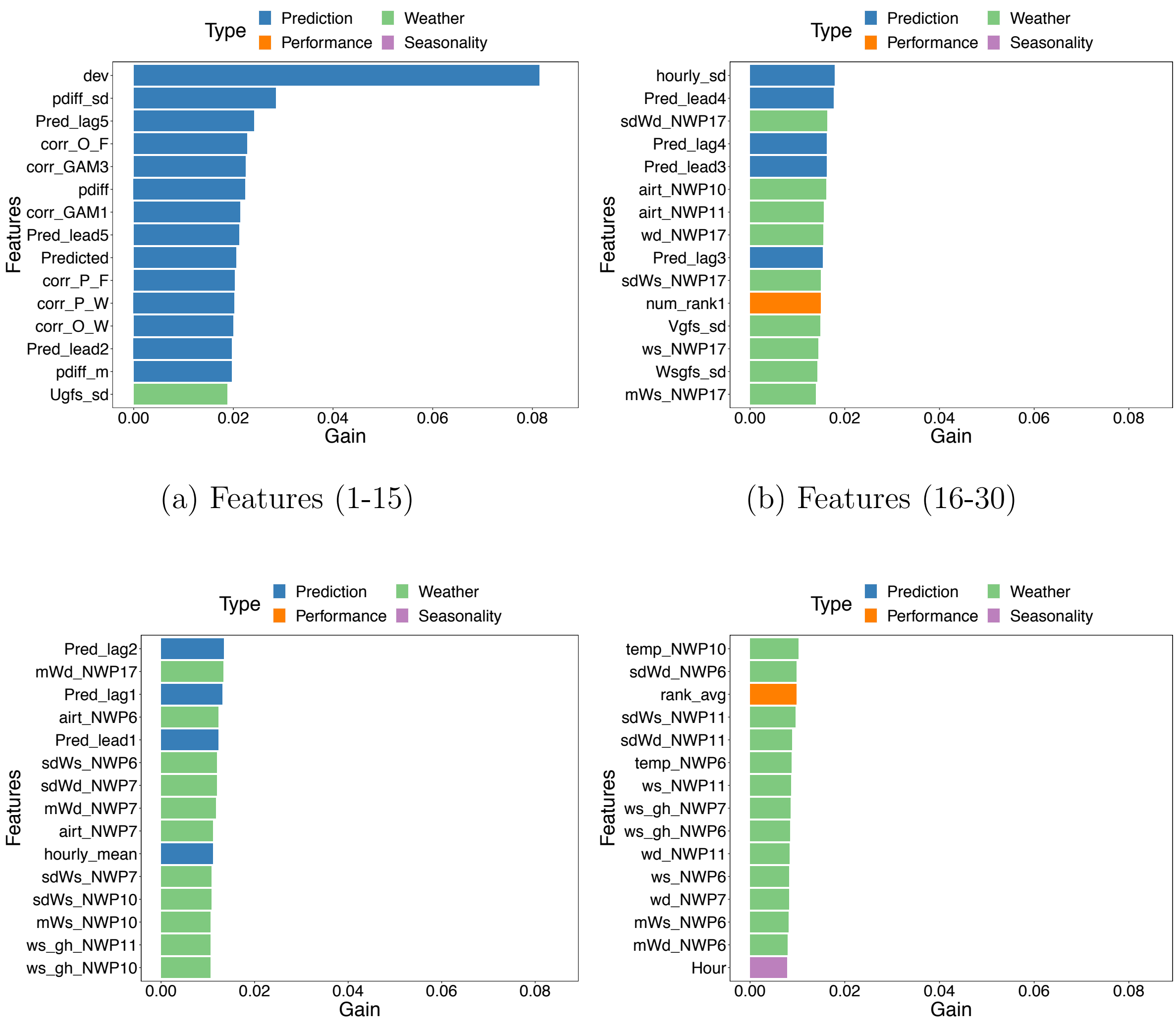
Scenarios	Description
Sce. 1	Prediction-based features
Sce. 2	Prediction-based and weather-based features
Sce. 3	Prediction-based, weather-based and performance-based features
Sce. 4	Prediction-based, weather-based, performance-based and seasonality-based features

The performance comparison of all seven scenarios are given in Table 5.9. It can be clearly seen that weather-based features improve the performance of prediction-based feature set. The table suggests that a feature set combination does not perform the best for each wind farm. However, when the first 60 features from the feature importance matrix are used, it either gives the best performance or a very similar performance to the best scenario for that wind farm. Hence, to be consistent in the performance comparison of all meta-learning approaches, the input set is determined to be the first 60 features from the feature importance matrix of each wind farm. Moreover, Figure 5.6 shows the set of selected features along with their gain values calculated with XGBoosting feature importance property for Dinar WF. These features are color-coded with respect to the feature set category. It can be observed that prediction-based features have generally higher importance compared to the other categories with hourly deviation from hour mean, denoted as dev, being the most significant feature followed by the rolling standard deviation of consecutive prediction differences, $pdiff\_sd$, and predictions from lag 5, $Pred\_lag5$. Furthermore, weather-based feature category has also highly significant features such as regional standard deviation of U, $Ugfs\_sd$, the rolling standard deviation of wind direction, $sdWd$, of $NWP17$ etc. Among the performance-based features, number of times a model is in rank 1 in the last day from previous two days, num_rank1 is more significant than the daily average of ranks in the last day from previous two days, $rank\_avg$.

Table 5.9. Performance comparison of feature selection scenarios in terms of NMAPE.

WF	All Features				Number of Features Selected		
	Sce. 1	Sce. 2	Sce. 3	Sce. 4	40	50	60
Aliaga	8.213	7.924*	7.993	7.966	7.927	7.939	7.937
Soma 1-2	9.282	8.936	8.899	8.951	8.963	8.972	8.878*
Dinar	10.130	9.814	9.879	9.891	9.894	9.942	9.807*
Bares	10.669	10.384	10.362	10.417	10.479	10.353*	10.416
Soke	14.735	14.313	14.309	14.275*	14.362	14.407	14.343
Geycek	9.634	9.423	9.449	9.341*	9.426	9.401	9.373

5.4.3.3. Pointwise Approach. In this section, the performance comparison of pointwise approaches is provided. Table 5.10 shows the performance comparison in terms of NMAPE. The results show that XGBoosting enhances the performance of decision trees in all rank 1 class labeling, rank 1 vs other and regression approaches in all wind farms. Furthermore, rank 1 vs other approach with XGBoosting gives either the best performance or very close performance to the best approach in terms of NMAPE in all wind farms.



(a) Features (1-15)

(b) Features (16-30)

(c) Features (31-45)

(d) Features (46-60)

Figure 5.6. Selected feature set for Dinar WF.

On the other hand, Table 5.11 provides performance comparison in terms of RMSE. The results depict that XGBoosting enhances the performance in terms of RMSE. The comparison also shows that regression-based meta-learning approach gives either the best performance or very similar performance to the best approach in terms of RMSE in all wind farms. Moreover, Tables 5.12 and 5.13 provide performance comparison in terms of selected ranking measures, MRR and NDGG$_{mean}$, respectively. Rank 1 vs other approach with XGBoosting gives a superior performance among pointwise approaches in terms of MRR and NDGG$_{mean}$ in all wind farms.

Table 5.10. Performance comparison of pointwise approach methods in terms of NMAPE.

WF	Rank 1 Class Labeling Approach		Rank 1 vs Other Approach		Regression		Ordinal Regression
	Tree	XGB	Tree	XGB	Tree	XGB	Tree
Aliaga	8.604	8.334	8.515	8.254*	8.454	8.384	8.35
Soma 1-2	9.404	9.179	9.760	9.127*	9.636	9.345	9.361
Dinar	10.478	10.149	10.297	10.089*	10.453	10.118	10.248
Bares	10.847	10.557	10.557	10.545	10.899	10.736	10.483*
Soke	15.211	14.962	15.240	14.893	15.463	14.846*	14.966
Geycek	9.801	9.604	9.923	9.623	9.844	9.599*	9.831

Table 5.11. Performance comparison of pointwise approach methods in terms of RMSE.

WF	Rank 1 Class Labeling Approach		Rank 1 vs Other Approach		Regression		Ordinal Regression
	Tree	XGB	Tree	XGB	Tree	XGB	Tree
Aliaga	16.102	15.822	15.920	15.621	15.464*	15.470	15.836
Soma 1-2	16.564	16.387	17.283	16.311	16.501	16.133*	16.691
Dinar	26.012	25.447	25.919	25.409	25.590	24.608*	26.156
Bares	8.521	8.363	8.274	8.280	8.204	7.993*	8.320
Soke	9.395	9.272	9.356	9.206	9.186	8.888*	9.158
Geycek	25.128	24.598	25.619	24.840	24.348	23.715*	25.352

5.4.3.4. Pairwise Approach. In this section, the overall performance of pairwise approach with XGBoosting is provided in terms of NMAPE, RMSE, MRR and NDCG$_{mean}$ in Table 5.14. Pairwise approach enhances the performance of the best method in

Table 5.12. Performance comparison of pointwise approach methods in terms of MRR.

WF	Rank 1 Class Labeling Approach		Rank 1 vs Other Approach		Regression		Ordinal Regression
	Tree	XGB	Tree	XGB	Tree	XGB	Tree
Aliaga	0.551	0.596	0.598	0.626*	0.433	0.469	0.602
Soma 1-2	0.555	0.601	0.588	0.636*	0.446	0.496	0.602
Dinar	0.565	0.604	0.619	0.629*	0.470	0.500	0.610
Bares	0.575	0.633	0.621	0.646*	0.413	0.457	0.622
Soke	0.557	0.598	0.595	0.616*	0.416	0.478	0.592
Geycek	0.581	0.614	0.617	0.641*	0.443	0.478	0.614

Table 5.13. Performance comparison of pointwise approach methods in terms of $NDCG_{mean}$.

WF	Rank 1 Class Labeling Approach		Rank 1 vs Other Approach		Regression		Ordinal Regression
	Tree	XGB	Tree	XGB	Tree	XGB	Tree
Aliaga	0.878	0.890	0.887	0.894*	0.868	0.871	0.887
Soma 1-2	0.882	0.892	0.888	0.898*	0.869	0.879	0.888
Dinar	0.878	0.889	0.888	0.894*	0.876	0.880	0.886
Bares	0.879	0.891	0.890	0.895*	0.859	0.863	0.890
Soke	0.880	0.889	0.889	0.893*	0.858	0.872	0.887
Geycek	0.883	0.892	0.889	0.896*	0.865	0.870	0.888

pointwise approach in terms of overall NMAPE and RMSE.

Table 5.14. Overall performance of pairwise approach method in terms of NMAPE, RMSE, MRR and NDCG_{mean}.

WF	XGB - Pairwise			
	NMAPE	RMSE	MRR	NDCG_{mean}
Aliaga	7.937	15.130	0.521	0.891
Soma 1-2	8.878	15.675	0.516	0.894
Dinar	9.807	24.527	0.507	0.889
Bares	10.416	8.128	0.500	0.883
Soke	14.343	8.844	0.495	0.884
Geycek	9.373	23.804	0.505	0.887

5.4.3.5. Listwise Approach. This section provides the overall performance of listwise approach with XGBoosting, and the results are summarized in Table 5.15. In terms of NMAPE and RMSE, pairwise approach outperforms the listwise approach, whereas listwise approach gives a better performance than pointwise approach. On the other hand, in terms of MRR and NDCG_{mean}, listwise approach provides a superior performance compared to pairwise approach.

Table 5.15. Overall performance of listwise approach method in terms of NMAPE, RMSE, MRR and NDCG_{mean}.

WF	XGB - Listwise			
	NMAPE	RMSE	MRR	NDCG_{mean}
Aliaga	8.026	15.340	0.603	0.899
Soma 1-2	8.874	15.821	0.605	0.901
Dinar	9.844	24.733	0.608	0.900
Bares	10.522	8.293	0.615	0.896
Soke	14.451	8.972	0.597	0.897
Geycek	9.392	23.957	0.611	0.898

In above calculations, NDCG_{mean} is computed by averaging NDCG values by time t for the whole test horizon. However, it is also important to analyze and compare the distributional behaviour of hourly NDCG values of different scenarios. The box plots of the worst method, which is the regression setting, denoted as Reg. (Tree), and the best method, which is the rank 1 vs other approach with Xgboosting in classification setting, denoted as Rank 1 vs Other (Xgb), from pointwise approaches in terms of NDCG_{mean}, pairwise approach and listwise approach are given in Figure 5.7. The box plots show that the median of hourly NDCG values from listwise approach outperforms the others. Furthermore, pairwise approach also enhances the performance of pointwise approach. Although the NDCG_{mean} values are close, the listwise approach still gives a better overall NDCG perfomance.

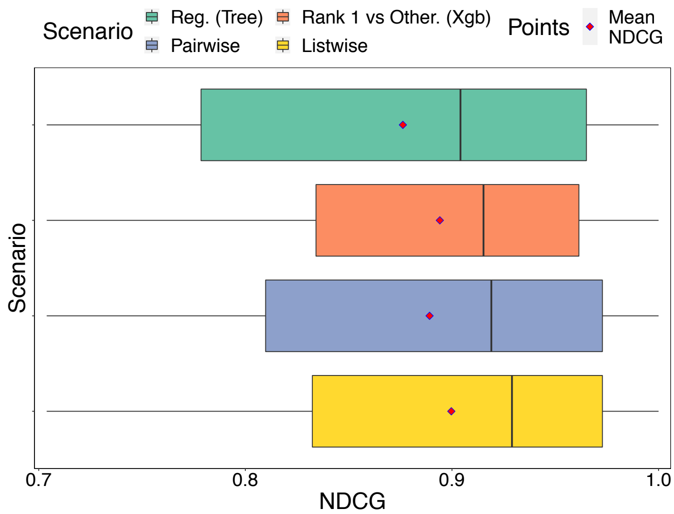


Figure 5.7. Box plots of hourly NDCG values of regression setting and rank 1 vs other approach with Xgboosting from pointwise approach, pairwise approach and listwise approach.

Furthermore, the detailed monthly performance comparison of six models from base-level pool and the pairwise approach, which is the best approach in terms of NMAPE, in Dinar WF in terms of NMAPE is provided in Figure 5.8. It can be observed that the pairwise approach in meta-learning part improves the performance of the best model from the base-level pool, which is Generalized Additive Models with Quasi-binomial distribution, in almost all months.

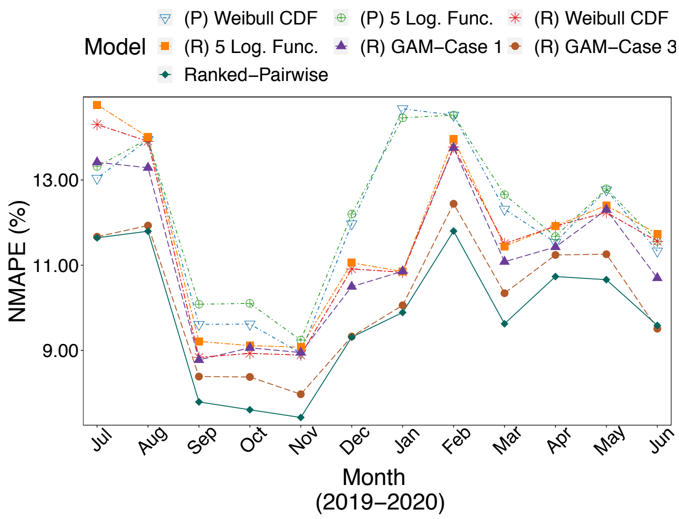
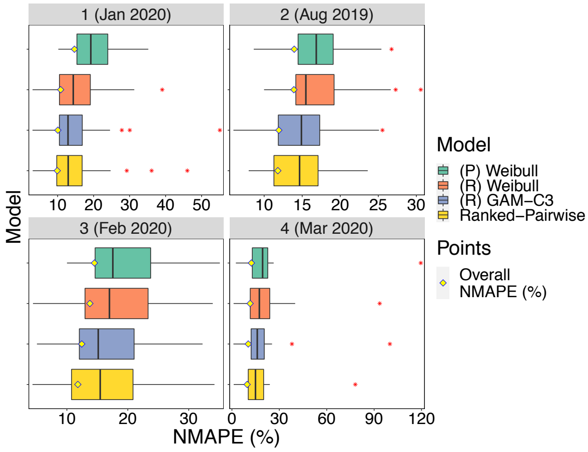
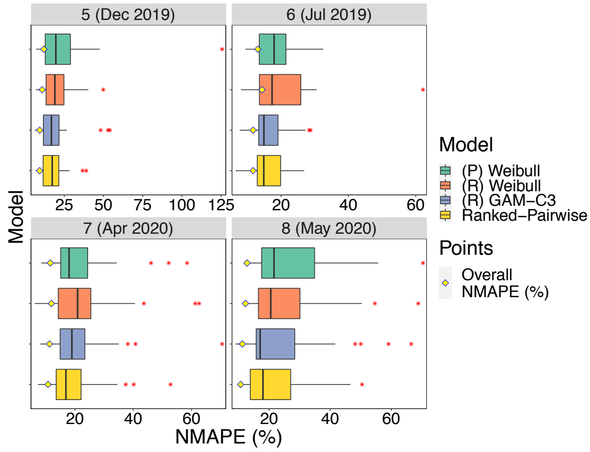


Figure 5.8. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Dinar WF.

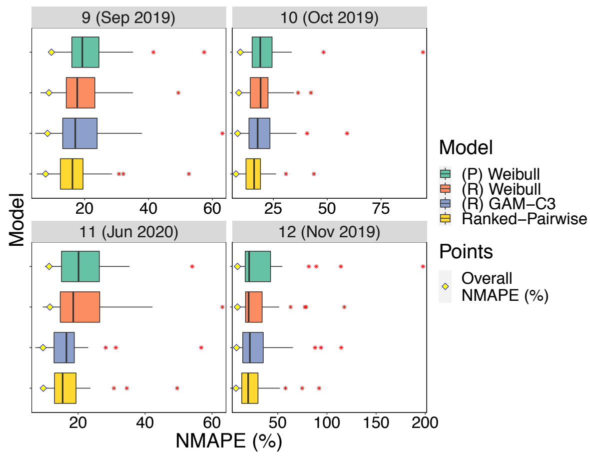
Moreover, as discussed in the base-level results, in wind power energy domain, it is meaningful to analyze the daily performances of the models in order to demonstrate the distributional behavior of the modeling strategies. The box plots of three best scenarios from each aggregation strategy from base-level pool and the pairwise approach from meta-learning part are provided for each month separately, where months are grouped and ordered based on average power production value, for Dinar WF in Figure 5.9 similar to Figure 5.4. In almost all months, the median of daily NMAPE values of pairwise approach is better than the best approach from base-level model pool, which is GAMs with Quasi-binomial distribution, denoted as (R) GAM-C3 in the figure. Additionally, the overall NMAPE of pairwise approach is still better in the months where the median of NMAPE of GAMs is slightly better. Thus, these results show that utilizing a meta-learning step e.g., pairwise approach, enhances the performance of the best base-level model.



(a) High Volume



(b) Mid Volume



(c) Low Volume

Figure 5.9. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Dinar WF.

6. CONCLUSION

In this thesis, in the first part, which is the base-level modeling part, three base models and two different aggregation approaches are implemented to model empirical and aggregate wind power curve. It is observed that using information from multiple coordinates enhances the performance. In point-based modeling, the first drawback is that discretization of direction causes loss of information related to continuity and angularity. Another drawback is that the predictions are obtained in two steps, and the spatial relations are included at prediction level. In order to address the second drawback, a simple optimization framework is proposed. In this approach, the predictions are obtained in one step by utilizing u and v vectors, and the spatial relations are incorporated at weather level. This methodology improves the performance of the first methodology.

Moreover, another approach based on GAMs is proposed to overcome the aforementioned drawbacks. An important capability is that they can model continuous and cyclic variables e.g., direction, with cyclic cubic regression splines. Moreover, with tensor products, the interaction between speed and direction is modeled. Furthermore, parameter estimation with Quasi-binomial distribution assumption in modeling utilization values improves the performance. Including additional variables e.g., seasonality-related variables, which is another flexibility in GAMs unlike in first two approaches, results in an improvement in the performance. All in all, the experiments show that GAMs where parameters are estimated with Quasi-binomial distribution assumption on the response provides superior performance compared to the other proposed alternative strategies.

In the second part, meta-learning approach is implemented in hourly wind power prediction problem. Hence, a dynamic model selection mechanism is built for this problem. Three mostly used meta-learning approaches, which are pointwise ranking approach, pairwise ranking approach and listwise ranking approach, are utilized where decision trees and XGBoosting are used as meta-level learners. Pointwise approach considers the hourly observations as independent instances, whereas pairwise approach takes into account the pairwise performances, which results in an improved performance compared to both pointwise approach and individual performances of models from base-level modeling strategies. On the other hand, listwise approach considers the whole set of ranks, where it gives a comparable performance to the pairwise approach. Furthermore, multiple sets of features are extracted for this meta-learning task, which can be summarized in different categories as prediction-based, weather-based, performance-based and seasonality-based features. Finally, selecting a subset of features from these categories based on XGBoosting feature importance provides either a superior performance than the whole set or an equivalent performance with the whole set, where using a subset is still useful in terms of the computational time.

A future direction based on this thesis for the first part can be on ramp detection strategies since predicting sudden changes in wind power production plays an important role in accurate forecasting of wind power. Another extension to the proposed strategy can be based on vector field interpolation techniques where wind vector fields are directly utilized in the aggregation of grid-based NWP. Moreover, for the meta-learning part, an extension can be based on utilizing active learning strategies on another ranking method, RankSVM, in order to select a meaningful and informative set of observations to improve the ranking performance.

REFERENCES

1. "Key World Energy Statistics 2020 – Analysis - IEA", https://www.iea.org/reports/key-world-energy-statistics-2020, (Accessed on 01/17/2021).

2. "Yük Tevzi Dairesi Başkanlığı - Kurulu Güç Raporu - Aralık 2020", https://webapi.teias.gov.tr/file/2bd6a902-7b75-4226-9c54-55ae9afea22f?download, (Accessed on 01/17/2021).

3. "Renewables 2020 – Analysis - IEA", https://www.iea.org/reports/renewables-2020, (Accessed on 01/17/2021).

4. De Giorgi, M. G., A. Ficarella and M. Tarantino, "Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods", *Energy*, Vol. 36, No. 7, pp. 3968–3978, 2011.

5. Manwell, J. F., J. G. McGowan and A. L. Rogers, *Wind energy explained: theory, design and application*, John Wiley & Sons, 2010.

6. Shokrzadeh, S., M. J. Jozani and E. Bibeau, "Wind turbine power curve modeling using advanced parametric and nonparametric methods", *IEEE Transactions on Sustainable Energy*, Vol. 5, No. 4, pp. 1262–1269, 2014.

7. Thapar, V., G. Agnihotri and V. K. Sethi, "Critical analysis of methods for mathematical modelling of wind turbines", *Renewable Energy*, Vol. 36, No. 11, pp. 3166–3177, 2011.

8. Foley, A. M., P. G. Leahy, A. Marvuglia and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation", *Renewable Energy*, Vol. 37, No. 1, pp. 1–8, 2012.

9. Kim, Y. and J. Hur, "An ensemble forecasting model of wind power outputs based

on improved statistical approaches", *Energies*, Vol. 13, No. 5, p. 1071, 2020.

10. Bauer, P., A. Thorpe and G. Brunet, "The quiet revolution of numerical weather prediction", *Nature*, Vol. 525, No. 7567, pp. 47–55, 2015.

11. Andrade, J. R. and R. J. Bessa, "Improving renewable energy forecasting with a grid of numerical weather predictions", *IEEE Transactions on Sustainable Energy*, Vol. 8, No. 4, pp. 1571–1580, 2017.

12. Lydia, M., S. S. Kumar, A. I. Selvakumar and G. E. P. Kumar, "A comprehensive review on wind turbine power curve modeling techniques", *Renewable and Sustainable Energy Reviews*, Vol. 30, pp. 452–460, 2014.

13. Lydia, M., A. I. Selvakumar, S. S. Kumar and G. E. P. Kumar, "Advanced algorithms for wind turbine power curve modeling", *IEEE Transactions on Sustainable Energy*, Vol. 4, No. 3, pp. 827–835, 2013.

14. Long, H., L. Wang, Z. Zhang, Z. Song and J. Xu, "Data-driven wind turbine power generation performance monitoring", *IEEE Transactions on Industrial Electronics*, Vol. 62, No. 10, pp. 6627–6635, 2015.

15. Villanueva, D. and A. Feijóo, "Comparison of logistic functions for modeling wind turbine power curves", *Electric Power Systems Research*, Vol. 155, pp. 281–288, 2018.

16. Mohammadi, K., O. Alavi, A. Mostafaeipour, N. Goudarzi and M. Jalilvand, "Assessing different parameters estimation methods of Weibull distribution to compute wind power density", *Energy Conversion and Management*, Vol. 108, pp. 322–335, 2016.

17. Khalid, M. and A. V. Savkin, "A method for short-term wind power prediction with multiple observation points", *IEEE Transactions on Power Systems*, Vol. 27, No. 2, pp. 579–586, 2012.

18. Wood, S. N., *Generalized additive models: an introduction with R*, CRC press, 2017.

19. Consul, P. C. and F. Famoye, *Lagrangian probability distributions*, Springer, 2006.

20. Lemke, C. and B. Gabrys, "Meta-learning for time series forecasting and forecast combination", *Neurocomputing*, Vol. 73, No. 10-12, pp. 2006–2016, 2010.

21. Lemke, C., M. Budka and B. Gabrys, "Metalearning: a survey of trends and technologies", *Artificial Intelligence Review*, Vol. 44, No. 1, pp. 117–130, 2015.

22. Sun, Q. and B. Pfahringer, "Pairwise meta-rules for better meta-learning-based algorithm ranking", *Machine Learning*, Vol. 93, No. 1, pp. 141–161, 2013.

23. Adhikari, R., "A neural network based linear ensemble framework for time series forecasting", *Neurocomputing*, Vol. 157, pp. 231–242, 2015.

24. Qomariyah, N. N., D. Kazakov and A. N. Fajar, "Predicting User Preferences with XGBoost Learning to Rank Method", *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 123–128, IEEE, 2020.

25. Cao, Z., T. Qin, T.-Y. Liu, M.-F. Tsai and H. Li, "Learning to rank: from pairwise approach to listwise approach", *Proceedings of the 24th International Conference on Machine Learning*, pp. 129–136, 2007.

26. Vanschoren, J., "Meta-learning: A survey", *arXiv preprint arXiv:1810.03548*, 2018.

27. Moura, T. J., G. D. Cavalcanti and L. S. Oliveira, "MINE: A framework for dynamic regressor selection", *Information Sciences*, Vol. 543, pp. 157–179, 2021.

28. Bossavy, A., R. Girard and G. Kariniotakis, "Forecasting ramps of wind power

production with numerical weather prediction ensembles", *Wind Energy*, Vol. 16, No. 1, pp. 51–63, 2013.

29. Britto Jr, A. S., R. Sabourin and L. E. Oliveira, "Dynamic selection of classifiers—a comprehensive review", *Pattern Recognition*, Vol. 47, No. 11, pp. 3665–3680, 2014.

30. Rossi, A. L. D., A. C. P. de Leon Ferreira, C. Soares, B. F. De Souza *et al.*, "MetaStream: A meta-learning based method for periodic algorithm selection in time-changing data", *Neurocomputing*, Vol. 127, pp. 52–64, 2014.

31. Donmez, P. and J. G. Carbonell, "Optimizing estimated loss reduction for active sampling in rank learning", *Proceedings of the 25th International Conference on Machine Learning*, pp. 248–255, 2008.

32. Donmez, P. and J. G. Carbonell, "Active sampling for rank learning via optimizing the area under the ROC curve", *European Conference on Information Retrieval*, pp. 78–89, Springer, 2009.

33. He, X., Z. He, X. Du and T.-S. Chua, "Adversarial personalized ranking for recommendation", *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 355–364, 2018.

34. Tang, J., X. Du, X. He, F. Yuan, Q. Tian and T.-S. Chua, "Adversarial training towards robust multimedia recommender system", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 32, No. 5, pp. 855–867, 2019.

35. "Global Forecast System (GFS) [0.5 Deg.] - Data.gov", `https://catalog.data.gov/dataset/global-forecast-system-gfs-0-5-deg`, (Accessed on 12/23/2020).

36. "General model description", `http://hirlam.org/index.php/hirlam-programme-53/general-model-description`, (Accessed on 07/07/2021).

37. James, G., D. Witten, T. Hastie and R. Tibshirani, *An introduction to statistical learning*, Vol. 112, Springer, 2013.

38. Chen, T. and C. Guestrin, "Xgboost: A scalable tree boosting system", *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

39. Burges, C. J., "From ranknet to lambdarank to lambdamart: An overview", *Learning*, Vol. 11, No. 23-581, p. 81, 2010.

40. "LambdaMART Demystified", https://staff.fnwi.uva.nl/e.kanoulas/wp-content/uploads/Lecture-8-1-LambdaMart-Demystified.pdf, (Accessed on 05/15/2021).

41. "XGBoost Parameters", https://xgboost.readthedocs.io/en/latest/parameter.html, (Accessed on 06/08/2021).

42. Ulazia, A., J. Sáenz, G. Ibarra-Berastegi, S. J. González-Rojí and S. Carreno-Madinabeitia, "Global estimations of wind energy potential considering seasonal air density changes", *Energy*, Vol. 187, p. 115938, 2019.

43. González-Longatt, F., H. Medina and J. S. González, "Spatial interpolation and orographic correction to estimate wind energy resource in Venezuela", *Renewable and Sustainable Energy Reviews*, Vol. 48, pp. 1–16, 2015.

44. Wu, Y., M. Mukunoki, T. Funatomi, M. Minoh and S. Lao, "Optimizing mean reciprocal rank for person re-identification", *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 408–413, IEEE, 2011.

45. "Python API Reference", https://xgboost.readthedocs.io/en/latest/python/python_api.html, (Accessed on 05/29/2021).

APPENDIX A: OVERALL RESULTS OF THE OTHER WIND FARMS

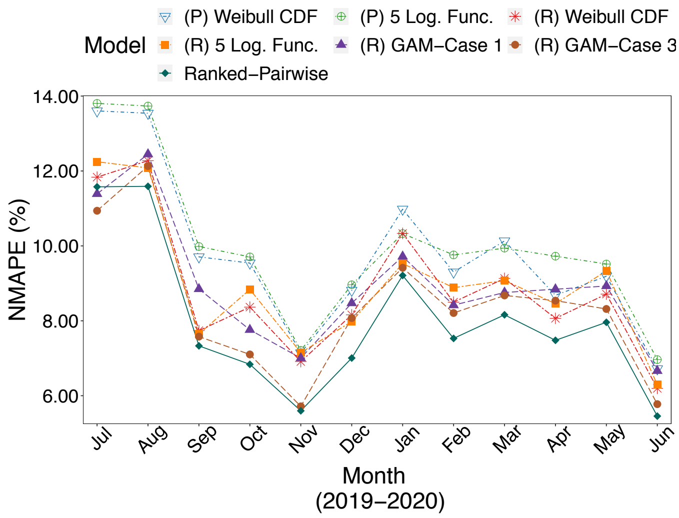


Figure A.1. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Aliaga Bergama WF.

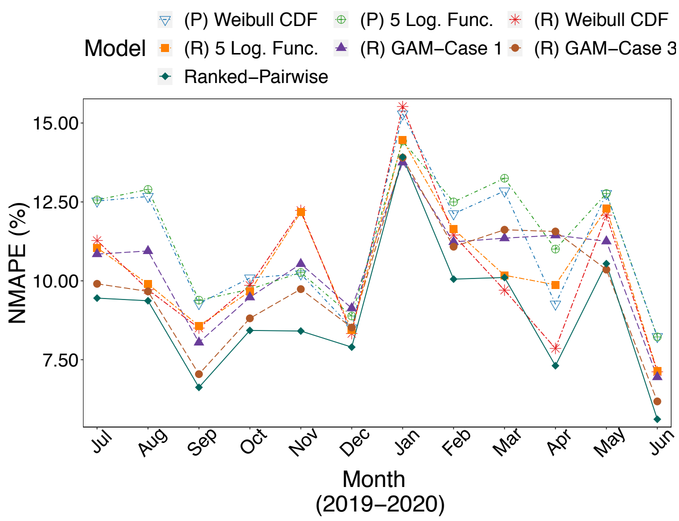


Figure A.2. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Soma 1-2 WF.

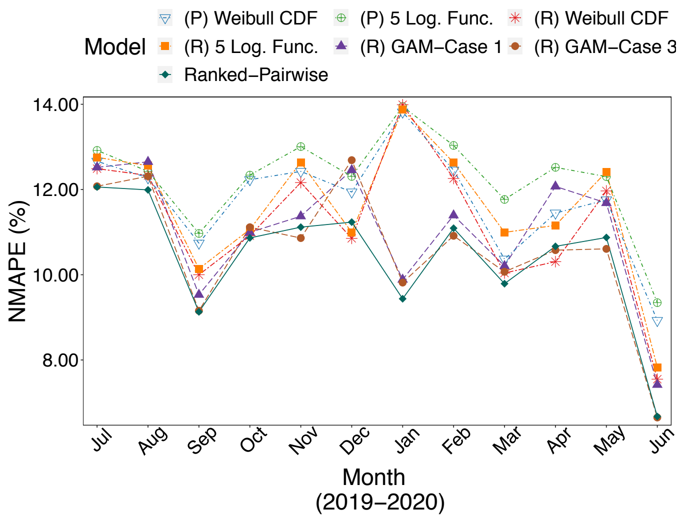


Figure A.3. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Bares WF.

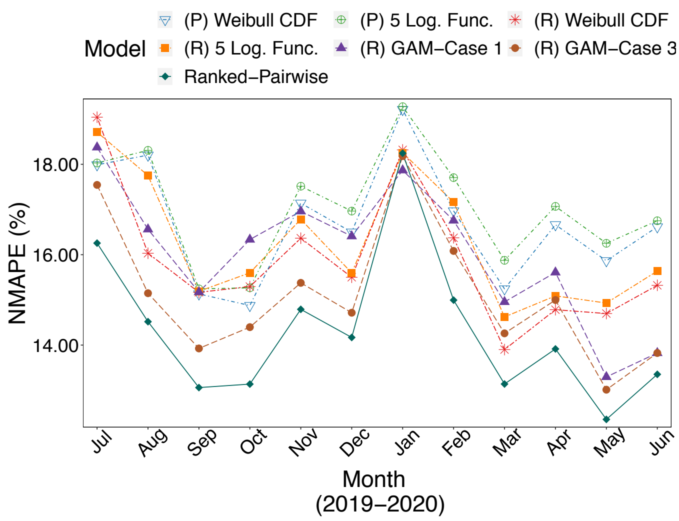


Figure A.4. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Soke WF.

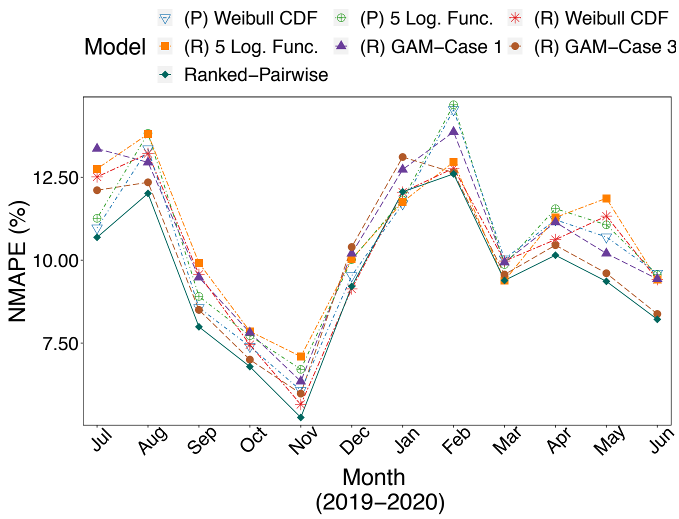
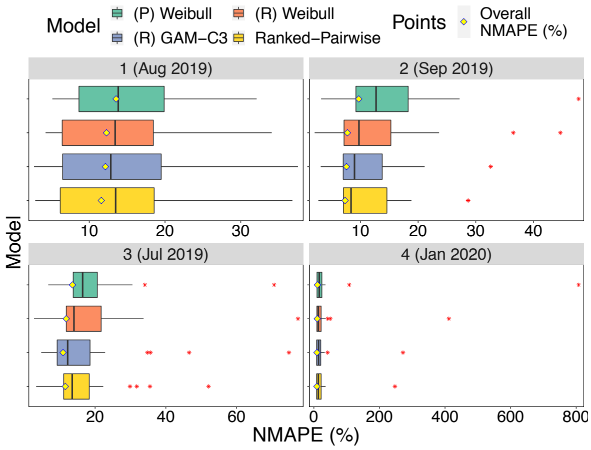
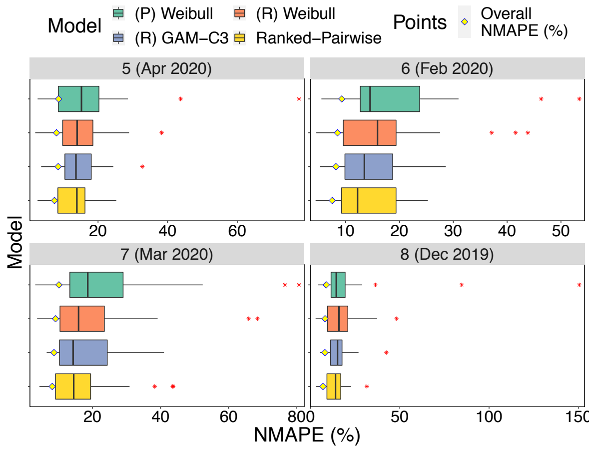


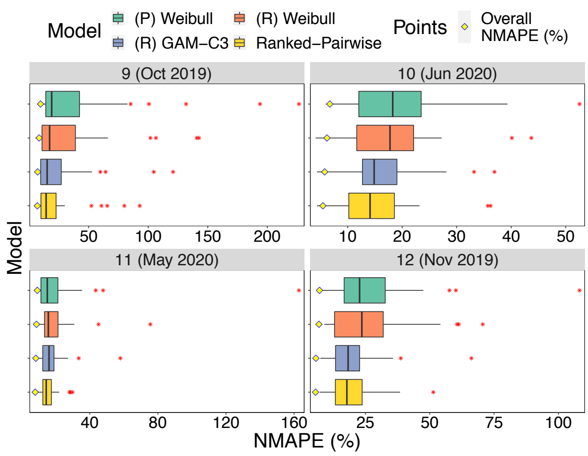
Figure A.5. Comparison of base-level model pool and pairwise approach in terms of NMAPE for Geycek WF.



(a) High Volume

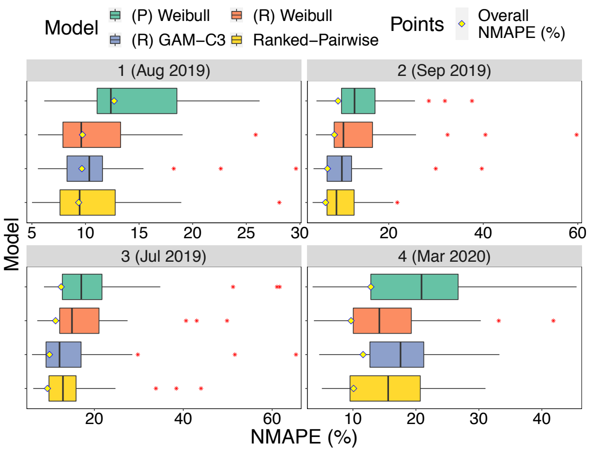


(b) Mid Volume

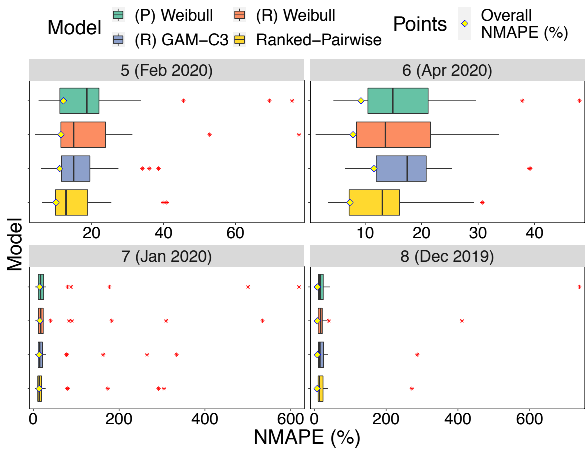


(c) Low Volume

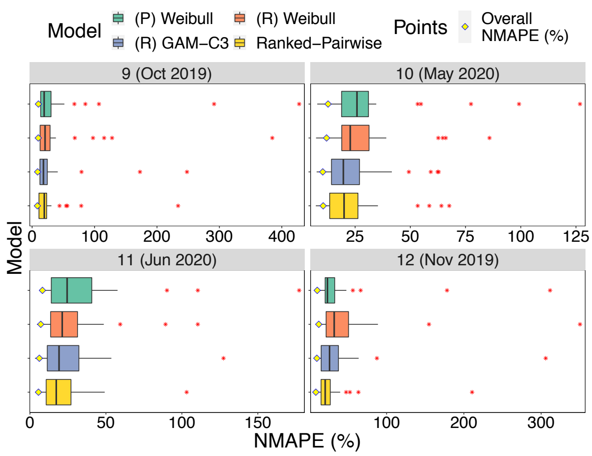
Figure A.6. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Aliaga Bergama WF.



(a) High Volume

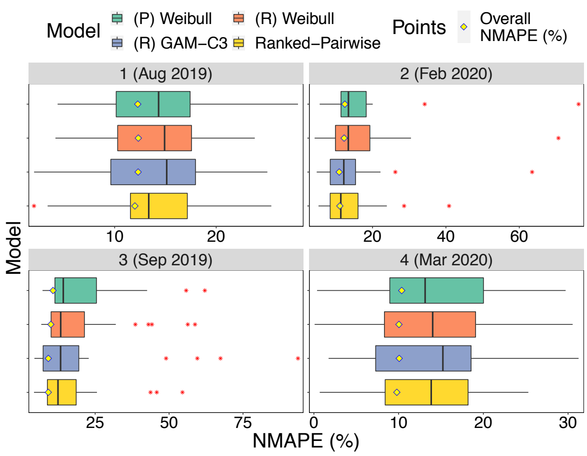


(b) Mid Volume

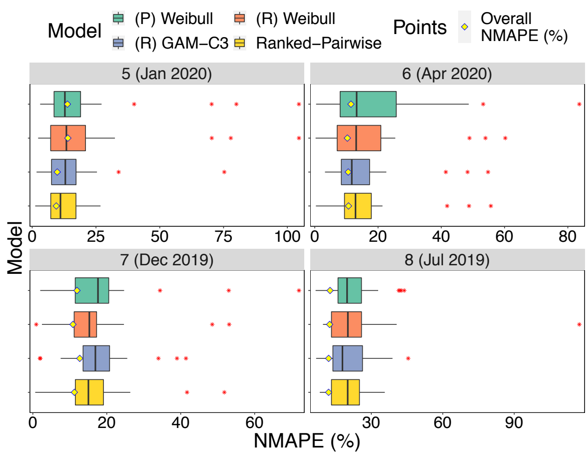


(c) Low Volume

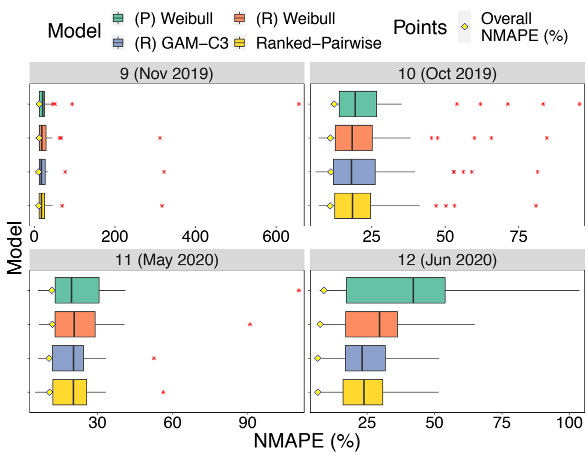
Figure A.7. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Soma 1-2 WF.



(a) High Volume

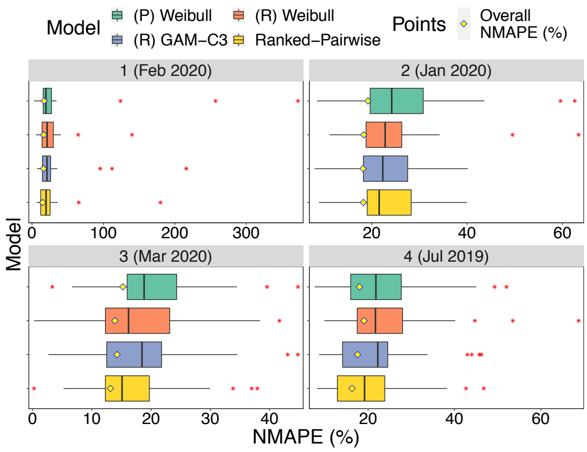


(b) Mid Volume

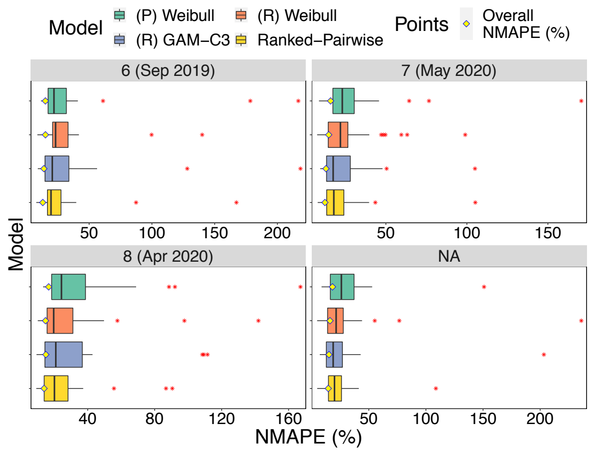


(c) Low Volume

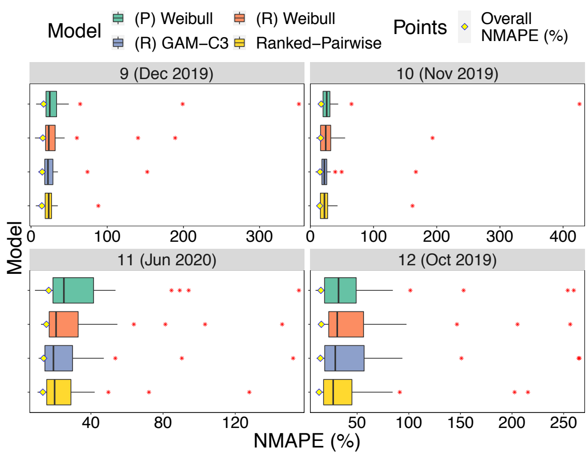
Figure A.8. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Bares WF.



(a) High Volume

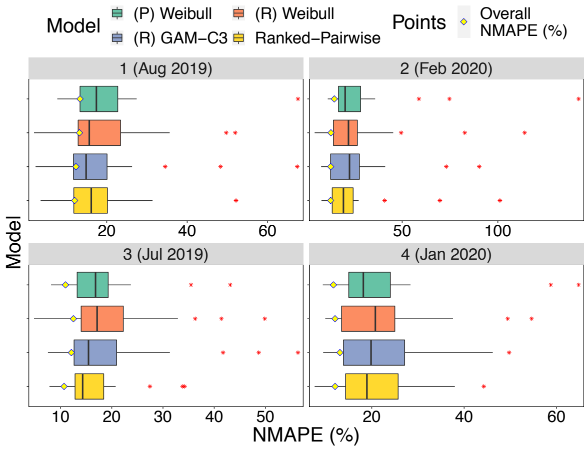


(b) Mid Volume

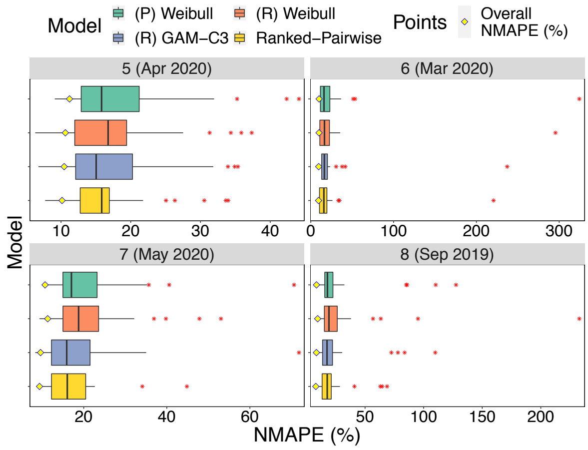


(c) Low Volume

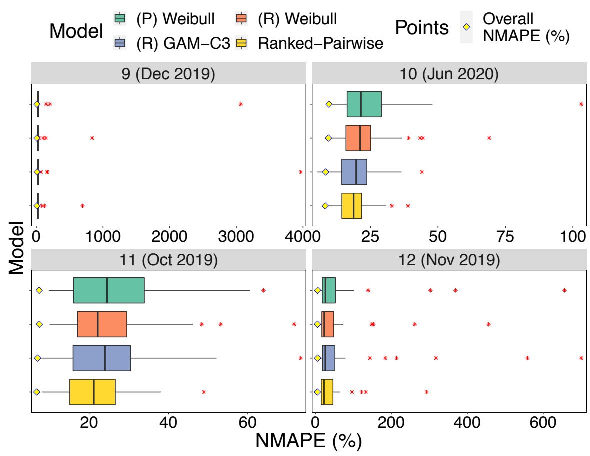
Figure A.9. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Soke WF.



(a) High Volume



(b) Mid Volume



(c) Low Volume

Figure A.10. Box plots of daily NMAPE values of high volume months (a), medium volume months (b) and low volume months (c) of three best approaches from each aggregation strategy from base-level pool and pairwise approach from meta-learning part for Geycek WF.

APPENDIX B: MODEL PARAMETERS

Wind Farm	Region-based Power Curve Modeling Parameters			
	Shape-based Forms	Generalized Additive Models		
	Direction	U & V knots	WS & WD knots	Seasonality
Aliaga	From HIRLAM	(10,10)	(5,13)	te(hour,month)
Soma 1-2	From HIRLAM	(7,7)	(6,13)	hour + ti(hour,month)
Dinar	From HIRLAM	(5,5)	(5,13)	te(hour,month)
Bares	From HIRLAM	(7,7)	(8,13)	hour + month
Soke	From GFS	(6,6)	(6,13)	hour + ti(hour,month)
Geycek	From GFS	(6,6)	Gaussian:(5,13) Quasi-binomial: (3,13)	hour + ti(hour,month)

	Meta-learning Decision Tree Parameters		
	minbucket	max depth	cp
All Wind Farms	(number of training instances)*0.01	7	0

	Meta-learning XGBoosting Parameters		
	nrounds	max depth	eta
All Wind Farms	300	6	0.3

Figure B.1. Model parameters for base-level and meta-level modeling methods