IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 5, 2019, accepted November 20, 2019, date of publication November 27, 2019, date of current version December 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956203

Data Mining-Based Upscaling Approach for Regional Wind Power Forecasting: Regional Statistical Hybrid Wind Power Forecast Technique (RegionalSHWIP)

MEHMET BARIS OZKAN[1,2] AND PINAR KARAGOZ[2], (Member, IEEE)
[1]Energy Institute, Scientific and Technological Research Council of Turkey (TUBITAK), 06000 Ankara, Turkey
[2]Department of Computer Engineering, Middle East Technical University, 06000 Ankara, Turkey

Corresponding author: Mehmet Baris Ozkan (mehmet.ozkan@tubitak.gov.tr)

This work was supported by the RITM Project directed by the TUBITAK Marmara Research Center for Renewable Energy General Directorate of Turkey, Energy Institute, under Grant 5122807.

ABSTRACT With the increasing need for the energy, the importance of renewable energy sources has also been increasing. In order to include the power produced by the wind into electricity grid in a controlled manner, power prediction has an important role. To produce a reliable wind power forecast, obtaining Wind Power Plants' (WPP) power generation data in real time and constructing the power forecast model with historical production values is a desirable action plan. However, this situation may not be applicable for all WPPs in the country due to difficulties in obtaining such data from WPP in real time. Therefore, there is a need for upscaling algorithm for generating the power forecast of such WPPs and producing a regional power forecast for a given region or the whole country. In this work it is aimed to construct an upscaling wind power forecast model to answer these needs. Many models in the literature propose techniques for the estimation of the entire zone rather than an *offline* plant. *Offline* plants are the plants such that their production data is not available in the system. In this work, we propose a method for generating power forecasting for *offline* plants, and for a region at the same time. The technique is based on firstly power forecasting for *offline* plants, and then upscaling to region by using forecasts for both *online* and *offline* wind power plants. The performance of the method is experimentally evaluated with baseline and previous techniques and it is shown to provide higher accuracy for power prediction.

INDEX TERMS Data mining, numerical weather production, regional wind power forecasting.

NOMENCLATURE

RITM	Wind Power Monitoring and Forecast Center
WPP	Wind Power Plant
TSO	Transmission System Operator
PCA	Principal Component Analysis
CFD	Computational Fluid Dynamics
ANN	Artificial Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression
NWP	Numerical Weather Prediction
MLP	Multilayer Perceptron
GFS	Global Forecast System
MGM	Turkish Met Office
RBF	Radial Basis Function
SHWIP	Statistical Hybrid Wind Power Forecast Technique
GPRS	General Packet Radio Service
YTM	Load Distribution Center
SCADA	Supervisory Control and Data Acquisition
TUBITAK	The Scientific and Technological Research Council of Turkey
YEGM	General Directorate of Renewable Energy of Turkey
TEIAS	Turkish Electricity Transmission Company
NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Square Error

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan Bu.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

RV	Random Variable
MW	Megawatt
GW	Gigawatt
WPA	Wind Power Analyser
ECMWF	European Centre for Medium-Range Weather Forecasts

I. INTRODUCTION

Climate change is one of the most important problems that the world has been facing in recent years [1], [2]. The main reason behind the climate change is reported as excessive usage of fossil fuels such as coal, oil and natural gas, which release high carbon dioxide to the atmosphere [1], [2].

For a better future, human beings need to reduce energy consumption and greenhouse gas emissions. In the last century, this situation was also realised by the world leaders and Kyoto Protocol was signed in 1997. The main aim of this agreement is to reduce the proportion of the methane and carbon dioxide release to the atmosphere by using alternative energy sources [3].

Renewable energy sources such as wind, solar, geothermal, hydroelectric and biomass are the main alternative energy sources with their substantial benefits. Use of these sources is not only important for our climate and health but also for our economy.

There are inherent conveniences and difficulties in obtaining energy from each of these alternative energy sources. Wind energy has an important place among these sources with its high potential. However, due to its stochastic nature, unlike other renewable energy sources, it is so unstable and volatile. Therefore, in order to control this unstable energy source, a reliable forecast system is inevitable for the system operator who manages the energy flow in the country [4], [5].

In recent years, as in other developed countries in the world, investments on wind energy are increasing fastly in Turkey. At present, there are nearly 160 WPPs in the country with an installed power of 6.5 GW. This wind power is equivalent to about 8% percent of the installed capacity of Turkey. Investments are increasing with the aim of upgrading this rate to 20% in the near future [6]. In order to manage wind power in the country, RITM project was started in 2010 by TUBITAK for YEGM and first power forecasts in the system were produced in 2011 [7]–[9]. Two main aspects of this project are monitoring the wind power in the country in real time and producing trusty wind power forecasts for improved energy management. In the center, various power forecast models are in operation for solving the different problems such as short term wind power forecasting (up to 48 hours), very short term wind power forecasting (up to 6 hours), probabilistic wind power forecasting and ramp forecasting. For the regional wind power forecasting, the method described in this paper has been recently integrated as the upscaling module of the RITM project.

In this work, we aim to provide regional wind power forecasting capability to RITM. *Regional Wind Power Forecasting* (*i.e., upscaling*) problem deals with the prediction of the aggregated power output of wind farms located within a defined region [4], [5], [10]–[12]. Although this problem has become a very popular research topic in the recent years, there is no standard definition for the *region*. Some researchers define region for areas of a few hundreds of kilometers in diameter, while for some other researchers, regions are defined according to the wind power plants' density. The main problem in the regional wind power forecasting is that dynamic information of all wind power plants such as weather forecasts and production data is not usually available. Currently, in RITM project, a hybrid forecasting method namely *Statistical Hybrid Wind Power Forecast Technique (SHWIP)* [13], [14] is in use as the point forecasting model. However, this method needs past power data of the wind power plant for forecast model construction. So this model is applicable for *online* plants whose production data is monitored by the system in real time. Therefore, for power prediction of *offline* plants, and hence, of a given region, there is a need for an alternative model.

The contributions of this work can be summarized as follows:

- With this method, power estimation can be done for each wind power plant in the country even if the power generation history is not available for some of them.
- Proposed model uses a bottom-up approach (different from the methods in the literature) by first producing power forecasts of all WPPs. Therefore, regional forecast results are more accurate and stable compared to the other direct scaling models.
- This model can be used on a hypothetical plant area to estimate the wind power potential of the location. This problem is named as *Wind Resource Assessment* and by applying this method the wind power potential of a wind farm location under consideration can be estimated. In large scale, a wind power potential atlas for the country can be produced by using the proposed method.

The rest of this paper is organized as follows: In Section II, the upscaling models and related work in the literature are presented. General structures of these models with real world examples are also given in this section. The general architecture of the RITM project and the details of the data used in the forecast application are described in Section III. In Section IV, the proposed models are presented in detail. The computational details of the models are given as well in this section. The evaluation results of the models are presented and discussed in Section V. Finally, the paper is concluded with further remarks and possible future work in Section VI.

II. RELATED WORK

Regional wind power forecast models deal with predicting the wind power production of the region or whole country by using the measurements of some reference *online* wind power plants. Getting the power measurement data of all plants in the country is expensive and sometimes impossible if the number of wind power plants is too high. Therefore there is a need for a model that can generate consistent estimates for the entire region using the existing data in the system [12]. This kind of power forecasts are crucial for the management of the electricity grid and planning of the energy sources more effectively.

In [15], Siebert and Kariniotakis focus on the selection of the reference wind farms in the regional forecasting model. For the selection of the reference wind plant they propose k-means clustering and they evaluate the impact of the number of reference wind farms selection on the accuracy of regional forecasting. According to their results, increasing the number of reference wind farms up to optimal number reduces the error rate.

Lueder *et al.* [16] propose a hybrid physical-statistical upscaling approach based on PCA. They show that spatial decomposition of wind power production can be performed with PCA in order to extract the pattern of variability. With this approach they produce forecast maps for different regions of Germany.

In their work, Drew *et al.* [17] investigate the importance of the regional wind power forecasting for determination of the ramps in the power grid. They discuss three different high resolution weather forecast models for understanding their impact on forecasting. They conclude that an accurate and reliable regional wind power forecast model minimizes the costs of the power grid management.

Lobo and Sanchez propose a regional wind power forecasting model based on smoothing techniques [18]. They apply their model to the Spanish peninsular system. Their approach focuses on searching for similarities between the current wind speed forecasts in the region and historical wind speed forecasts. Power production forecast is generated from the historical power data on the basis of the wind speed similarity.

In [19], [20], regional forecasting model integrated to the Previento forecast tool is described. This forecast tool uses physical characteristics of the plants such as turbine hub height, roughness in the plant area etc. Finally, on the basis of physical characteristics, several reference wind plants are chosen, and from these plants, power forecasts of the whole region are calculated.

Rohrig *et al.* [21] propose neural network based regional forecasting approach. They use two networks in the model. The first network is used for computing the power output of some reference wind plants. Then these forecasts are given as input to the second network to compute the forecasts of the whole region.

Yan *et al.* [22] introduce a multiscale regional wind power forecast model for controlling the high penetration of the wind power on the electricity grid. Their model is constructed on multi-to-multi mapping network and the use of stacked denoising autoencoder. They produce power forecasts for the region with 24 to 72 hours time horizon. According to the reported results, their model provides better performance compared to the other individual regional power forecast models.

A post-processing technique with PCA is presented by Davo *et al.* [23]. They apply the model not only for regional wind power forecasting but also on solar irradiance forecasting. For the numerical weather predictions, they use ECMWF and GFS as the source and their forecast horizon is 0-72 hours. According to their experimental results, combining PCA with these post-processing techniques leads to better performance in comparison to the results of the technique without the PCA reduction.

RegioPred is developed by Ignacio *et al.* [24] by using the point forecasts of *online* wind plants, which are generated by their statistical point forecasting model, namely LocaPred. The model uses Computational Fluid Dynamics (CFD) and power curve modelling for power forecasts. Selection of the reference wind farms is performed by clustering.

In [25], Alvaro *et al.* compare performance of MLPs and SVR for reference wind farm selection in their regional wind farm forecasting model. It is reported that SVR based model offers better performance as the training time is much less than MLP.

The proposed model described in this paper is based on statistical models like most of the methods described above. In the proposed model, reference *online* plants are chosen according to geographical relations and neighboring, historical wind speed and wind direction similarities as in the models described in [15], [18]–[20]. However, unlike these models, in the proposed model, the reference *online* plants are determined separately for each *offline* plant rather than for the region. Therefore, differently from the models in the literature, the proposed model estimates the power forecast for each *offline* plant in the system before estimating the regional power forecasts. As a result, the main advantage of the proposed method is that it is possible to generate power forecasts for every plant in the system while generating power forecasts to the region even if for some of the WPP historical power data is not available. Additionally, another advantage of this method is that it can be applied to a hypothetical plant area in order to investigate the wind power potential. Therefore, the proposed method can be used in wind power assessment problem.

III. GENERAL ARCHITECTURE OF THE RITM SYSTEM

This section summarizes the general architecture of the RITM system and the structure of the data that is used in the proposed method. The structure of the RITM project is shown in Figure 1. The production data of the power plants is measured by wind power analyzers and sent to servers in real time. Apart from the wind power data, from the plant SCADA system and wind masts, wind speed and wind direction values on the turbine levels are sent to the center. The most important data source in the architecture is the NWP. In the absence of the weather forecasts it is impossible to produce power forecasts obtained by the models described in this work. Due

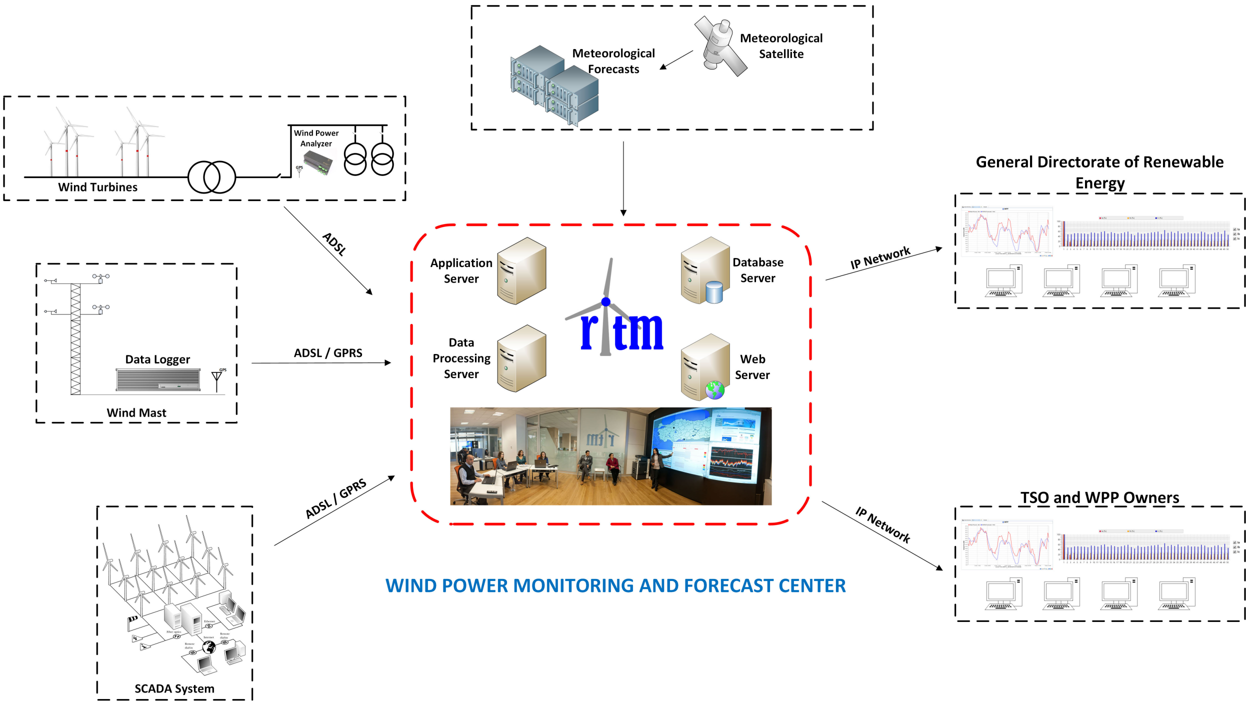

FIGURE 1. The general architecture of the RITM system.

to their importance in power forecasting, these weather forecasts are obtained in hourly resolution for 48 hours (48 hour tuples for the current day and the next day) to the system from three different sources, namely GFS [26], ECMWF [27] and MGM [28]. Since weather forecast inputs are in 48-hour resolution, power forecasts generated in the system are also in 48-hour resolution. These weather forecasts are updated in every 24 hours.

By using the weather forecasts and past power data 48-hour day ahead power forecasts are produced in the system by SHWIP model for *online* WPPs [13] for every day. SHWIP point forecasting model is based on dynamic clustering of weather events according to the most important weather parameters such as wind speed, wind direction and pressure. As of the end of 2017, there are about 160 WPPs in Turkey and 110 of them are monitored online in this system.

As shown in the rightmost modules in Figure 1, produced power forecasts are sent to the WPP owners and they enter the energy market by using these power forecasts. Additionally, by using the monitoring and forecast software, TSO can view the production data in real time and visualize the power forecasts.

IV. PROPOSED TECHNIQUE:REGIONAL STATISTICAL HYBRID WIND POWER FORECAST (RegionalSHWIP)

Regional wind power forecasting is important especially for system planning. Most of the models in the literature make predictions of the wind power of the region by scaling up the power estimates of several reference *online* wind farms in the region. Unlike this approach, in the proposed method, first of all, the wind power forecast is produced for the *offline* power plants in the region and then the power forecast of the region is obtained from the sum of all the *online* and *offline* power plants in the region. With this method, for all the power plants in the country, it is possible to produce power estimates even if they are not monitored in the system, and these forecasts can be used for improving regional forecasts. Unlike previous similar systems, the most important factor in the implementation of this method is the availability of weather forecast data for all the coordinates in the country, not just *online* plants in the system.

The general architecture of the proposed work is presented in Figure 2. The inputs of Stage-1 are 48-hour power forecasts of the *online* plants obtained by SHWIP method [13], [14], 48-hour NWP data for the *online* plants and 48 hour NWP data for the *offline* plants. Individual offlineSHWIP power forecasts are produced in Stage-1 for the *offline* plants and these forecasts are combined to obtain the final 48-hour power forecasts for *offline* WPPs. We have proposed 6 alternative models that can be used for this phase, and their performance have been analyzed as given in Section V. The forecasts of the whole region are determined in Stage-3 by summing up the SHWIP power forecasts of the *online* plants and offlineSHWIP power forecasts of the *offline* plants.

A sample region in the country is presented in Figure 3. In this map, the green circles represent the *online* plants in the region and black circles denote the *offline* ones. With the proposed approach, initially, power forecasts for each of the individual *offline* plants are calculated and the power forecasts of the whole region are determined as the summation of all plants in the region.

Stage 1. In this stage, firstly, 48-hour weather forecast data is retrieved for the candidate reference *online* plants and

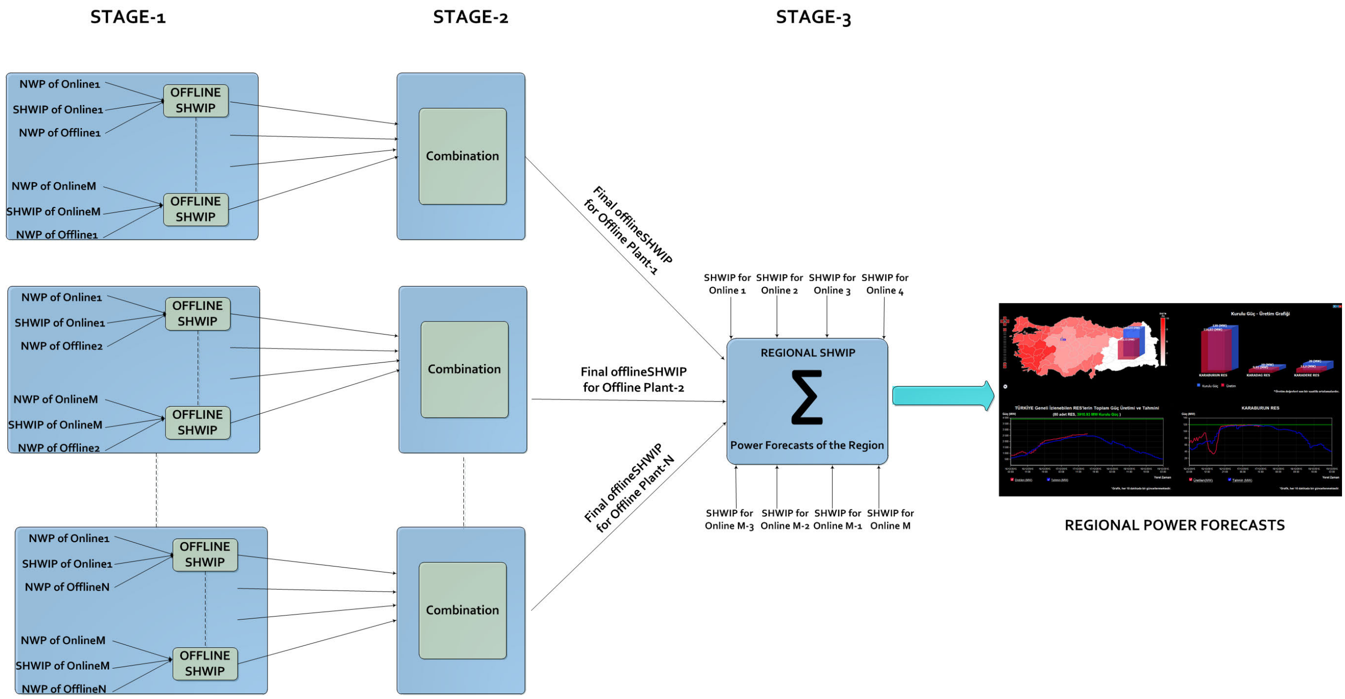

FIGURE 2. General architecture of the proposed RegionalSHWIP method.

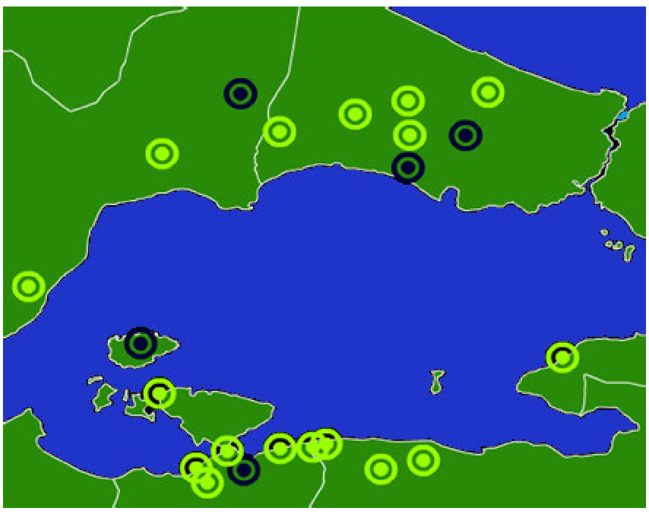

FIGURE 3. A sample region from the marmara sea.

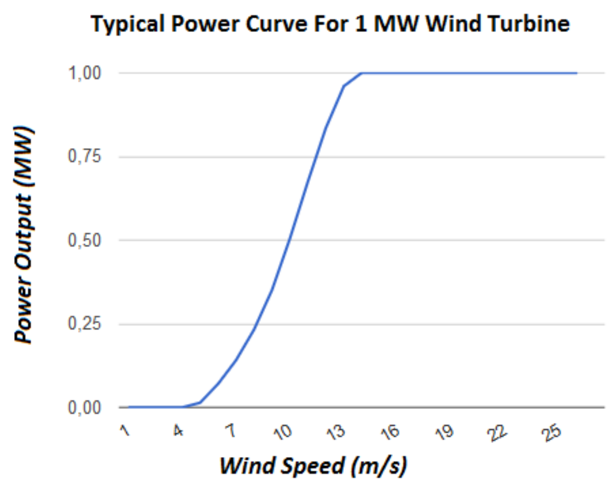

FIGURE 4. Power curve of a typical 1 MW wind turbine.

offline plants. In the constructed model, wind speed values are used in the generation of the power forecasts.

Let $v_{offline}$ and v_{online} are the 48-hour wind speed forecast vector for the time instance t. Equation 1 and Equation 2 represent the 48-hour wind speed input vector for the *offline* and *online* plant, respectively. From these values, capacity factor of each hour is calculated as given in Equation 3, by using a typical power curve of 1 MW wind turbine given in Figure 4. According to power curve in Figure 4, according to the wind speed forecast, expected capacity is determined by using the ranges in Equation 3.

$$v_{offline}[t, t+1, t+2, ..., t+47] \tag{1}$$

$$v_{online}[t, t+1, t+2, ..., t+47] \tag{2}$$

$$C_f(v) = \begin{cases} 0 & v < 5\,m/s \\ 0.014 & v = 5\,m/s \\ 0.07 & v = 6\,m/s \\ 0.1408 & v = 7\,m/s \\ 0.232 & v = 8\,m/s \\ 0.35 & v = 9\,m/s \\ 0.5044 & v = 10\,m/s \\ 0.6744 & v = 11\,m/s \\ 0.8348 & v = 12\,m/s \\ 0.9596 & v = 13\,m/s \\ 1 & v > 13\,m/s \end{cases} \tag{3}$$

$$offlineSHWIP(t) = \frac{C_f(v_{offline(t)})}{C_f(v_{online(t)})} \times \frac{C_{offline}}{C_{online}} \times SHWIP(t) \quad (4)$$

$$distance = \sqrt{(POnline_{lat} - POffline_{lat})^2 + (POnline_{lon} - POffline_{lon})^2} \quad (5)$$

$$weight = e^{\frac{-distance}{100}} \quad (6)$$

From the capacity factor ratio, capacity of the plant and 48-hour SHWIP power forecasts of the *online* plant, 48-hour offlineSHWIP power forecast of the *offline* plant is calculated as given in Equation 4.

Stage 2. The main job of the second stage is selection of the reference *online* WPPs and combination of the individual offlineSHWIP forecasts that were calculated in the previous stage in order to generate the final power forecasts of the *offline* plant. For the selection of the *online* plants, geographical region, neighborhood relation and historical weather data similarities are considered.

To this aim, we define the following 6 alternative models:

- **Model 1: Region Based Exponential Distance Weighting.** Each *online* plant within the *offline* plant's geographical region is selected as reference and individual offlineSHWIP forecasts are weighted in inverse proportion to the distance as shown in Equations 5, and 6.
- **Model 2: Region Based Average Weighting.** Each *online* plant within the *offline* plant's geographical region is selected as reference and individual offlineSHWIP forecasts are equally weighted for each reference *online* WPP.
- **Model 3: Neighbourhood Based Exponential Distance Weighting.** The nearest 15 *online* plants with respect to the *offline* power plant's central coordinates are selected reference and individual offlineSHWIP forecasts are weighted in inverse proportion to the distance given in Equations 5, and 6.
- **Model 4: Neighbourhood Based Average Weighting.** The nearest 15 *online* plants with respect to the *offline* power plant's central coordinates are selected as reference. Individual offlineSHWIP forecasts are equally weighted for each reference *online* WPP.
- **Model 5: Historical Wind Speed Similarity k Nearest Neighborhood based Weighting.**
  For this model initially, for the data retrieval step, the following operations are performed:
  - For every plant, 48-hour daily weather (wind speed) forecast is obtained.
  - For every plant, 30-day historical weather forecast data is obtained. Weather forecasts are obtained as the average forecasts of the 4 (Experimentally shown that 4 is optimal as the nearest neighbor count) nearest grid points to the WPP's central coordinate.
  - For the *online* plants, 48-hour SHWIP power forecasts are obtained.

TABLE 1. r value correlation table.

r value	Relation
+.70 or higher	Very Strong positive relationship
+.40 to +.69	Strong positive relationship
+.30 to +.39	Moderate positive relationship
+.20 to +.29	Weak positive relationship
+.01 to +.19	No or negligible relationship
0	No relationship
-.01 to -.19	No or negligible relationship
-.20 to -.29	Weak negative relationship
-.30 to -.39	Moderate negative relationship
-.40 to -.69	Strong negative relationship
-.70 or higher	Very Strong negative relationship

  After the data retrieval for the selection of the reference *online* WPPs, the following procedure is followed:
  1) For an *offline* plant, from the historical wind speed vectors, a correlation coefficient is calculated for each *online* plant in the system, independent of the regional neighbourhood. For the correlation coefficient *The Pearson Product-Moment Correlation* equation whose formula is given in Equation 7 is used [29].
  $$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}} \quad (7)$$
  In this equation, x value is the 30-day historical wind speed vector of the *offline* wind power plant and y value is the corresponding historical wind speed vector of the *online* WPP.
  2) According to r value correlation in Table 1[1], k *online* WPPs whose correlation coefficient is greater than 0.4 are selected as the reference *online* plants.
  3) Finally from the selected k *online* WPPs final power forecasts of the *offline* plant are calculated as correlation coefficients are assigned as the combination weights.
- **Model 6: Historical 2D Wind Data Similarity Nearest Neighborhood based Weighting.** In Model-5, data similarity is investigated with respect to the most recent 30-day wind speed data of the *offline* plant, and reference *online* plants. While finding the similarity, *the Pearson Product-Moment Correlation* is used. In this model, not only wind speed forecast data but also wind direction forecast data are considered. Additionally, for extracting the seasonal pattern, the training day size is extended to 90 days.
  In this model, the Pearson Product-Moment Correlation is not suitable since the correlation calculation involves

[1] r value correlation table shows the borders of strong, moderate and weak relation between two data sets

TABLE 2. Results of all models (in terms of NMAE %).

WPP	Model-1	Model-2	Model-3	Model-4	Model-5	Model-6	Base Model	Persistent Model
WPP1	13.27 %	13.36 %	12.87 %	13.47 %	12.90 %	**12.71** %	13.43 %	19.32 %
WPP2	16.14 %	16.42 %	16.66 %	16.70 %	15.42 %	**15.09** %	17.65 %	18.95 %
WPP3	**13.92** %	14.07 %	14.34 %	14.65 %	15.10 %	14.79 %	14.40 %	16.40 %
WPP4	13.75 %	13.89 %	13.73 %	13.78 %	**13.11** %	13.17 %	16.37 %	17.15 %
WPP5	17.97 %	18.17 %	18.22 %	19.56 %	17.93 %	**17.55** %	21.79 %	21.01 %
WPP6	15.54 %	16.39 %	15.58 %	16.17 %	15.72 %	**15.43** %	16.52 %	21.70 %
WPP7	15.49 %	15.09 %	16.83 %	16.26 %	14.76 %	**14.00** %	14.99 %	18.14 %
WPP8	16.19 %	16.40 %	17.22 %	18.75 %	16.02 %	**15.84** %	20.32 %	21.12 %
WPP9	17.14 %	17.33 %	17.04 %	17.27 %	**16.59** %	16.70 %	19.43 %	19.21 %
WPP10	17.03 %	17.25 %	16.67 %	17.07 %	14.70 %	**14.38** %	18.83 %	20.78 %

2D matrices that include wind speed and wind direction data. Therefore, instead, we use *RV Coefficient* technique [30]. The RV coefficient is a generalization of the squared Pearson correlation coefficient. As in Pearson correlation, in RV coefficient values range in [0,1], where 1 denotes the highest similarity. As in the Model-5, the *online* plants whose RV correlation coefficients are higher than 0.4 are selected as reference *online* plants and these coefficients are used as the combination weights.

$$RegionalSHWIP = \sum_{n=1}^{N} \begin{cases} SHWIP & n = Online \\ offlineSHWIP & n = Offline \end{cases} \quad (8)$$

Stage 3. In the final stage, the power forecasts of the whole region (RegionalSHWIP) are calculated as the summation of the *online* plants' SHWIP forecasts, and *offline* plants' final offlineSHWIP forecasts within the region, as given in Equation 8.

V. EXPERIMENTAL RESULTS

In this section, experimental evaluation results are presented. Firstly, data used in the experiments is introduced. Then we present the experiments on power forecasts for *offline* plants. Following this, the experiments on regional wind power forecasting are presented in the next section.

A. DATA USED IN THE EXPERIMENTS

In the experiments, we used 10 *offline* WPPs in order to evaluate the forecasting accuracy of the proposed method. Additionally, 9 regions, determined by TSO, are used for evaluating the performance of the regional power forecasts. As the ground truth, we used the power values for these *offline* WPPs and the regions for the period between 1 January 2017 to 31 December 2017, which are obtained from the study in [31].

$$NMAE = \frac{\frac{\sum_{i=1}^{N} |y_i - x_i|}{N}}{C} \times 100\ \% \quad (9)$$

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{N} (y_i - x_i)^2}{N}}}{C} \times 100\ \% \quad (10)$$

Performance results are obtained for the hourly power forecast and power data in this time period. Error results are reported as average NMAE and NRMSE, by using Equation 9 and Equation 10.

In Equation 9 and Equation 10, x_i is the real power, y_i is the power forecast at i^{th} hour, C is the installed capacity of the WPP and N is the total number of hours processed in the test phase which correspond 8760 (365×24) hour data in one year test region.

B. EXPERIMENT 1: ANALYSIS ON POWER FORECAST FOR OFFLINE PLANTS

In the first experiment, performance of the final offlineSHWIP power forecasts (result of Stage-2 in Figure 2) of the 10 *offline* plants for all of the 6 models are investigated in terms of NMAE and NRMSE error rates. These plants are selected from different geographical regions of country. Additionally, these models are compared with two other forecast models whose details are as follows:

- **Base Model:** In this method, the weather forecasts of the *offline* plant are directly used in the power curve in Equation 3 without checking a relation between reference *online* plants.
- **Persistent Model:** In this approach, mean power data of the previous day is assigned as the power forecasts of the next day for each 24-hour tuple.

According to error rates presented in Table 2 and Table 3, Model-6 error rates are lower than Model-5 in 7 plants. Also, for 7 of 10 WPPs the lowest error rates are obtained by Model-6. For three other plants, the error rates of all models are close to each other and these plants have high number of geographical neighbors.

Adding wind direction to selection of the reference *online* WPPs improved the error rates, and compared to other models, Model-6 is more stable especially in the determination of the peak values. An example of this situation is shown in Figure 6. Although WPP1 has 40 MW installed capacity,

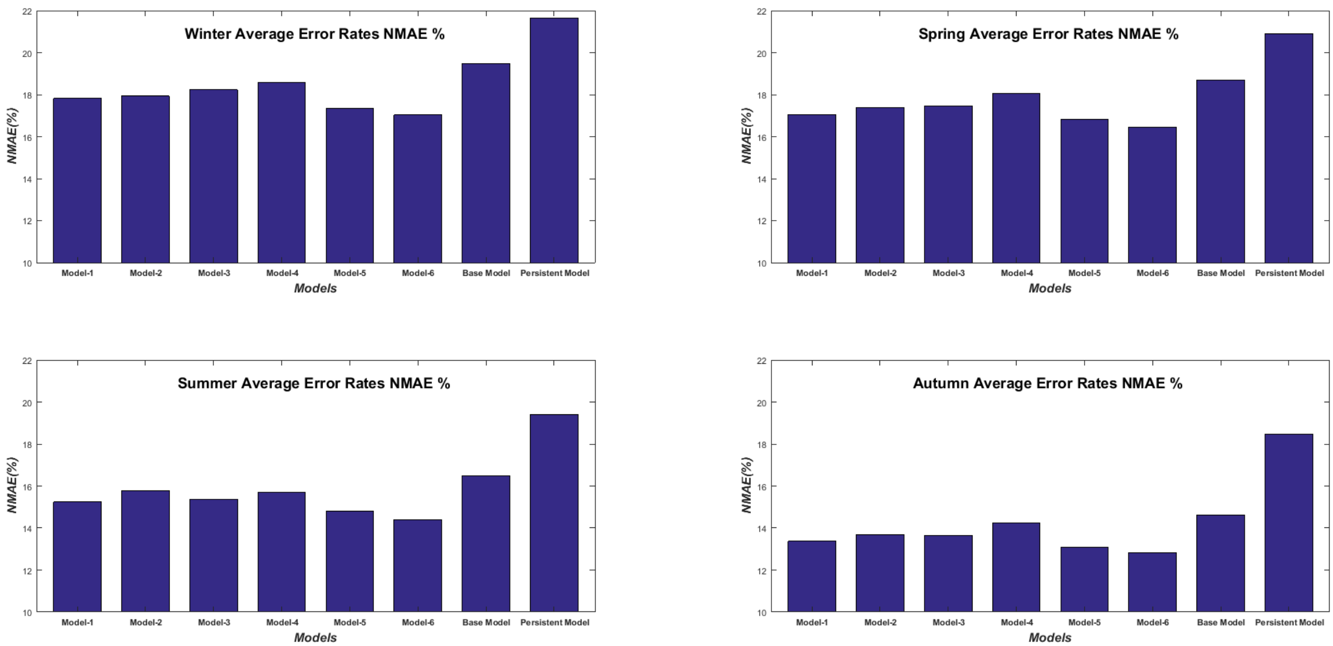

FIGURE 5. Seasonal performance of the all models.

TABLE 3. Results of all models (in terms of NRMSE %).

WPP	Model-1	Model-2	Model-3	Model-4	Model-5	Model-6	Base Model	Persistent Model
WPP1	19.24 %	20.46 %	18.57 %	19.70 %	17.91 %	**17.72** %	19.87 %	27.07 %
WPP2	24.48 %	24.91 %	24.90 %	25.01 %	23.82 %	**23.35** %	26.89 %	25.93 %
WPP3	**20.74** %	20.97 %	21.41 %	21.76 %	22.43 %	22.05 %	22.82 %	24.24 %
WPP4	19.56 %	19.75 %	19.48 %	19.54 %	**18.48** %	18.55 %	23.68 %	24.08 %
WPP5	25.83 %	26.90 %	26.67 %	28.94 %	26.20 %	**26.06** %	31.50 %	30.12 %
WPP6	22.40 %	23.35 %	22.35 %	22.88 %	22.85 %	**22.22** %	24.66 %	29.98 %
WPP7	21.73 %	21.21 %	23.29 %	22.48 %	21.07 %	**20.21** %	22.11 %	26.77 %
WPP8	23.95 %	23.98 %	24.48 %	26.39 %	**23.07** %	23.16 %	28.97 %	31.51 %
WPP9	24.71 %	25.19 %	**24.00** %	24.27 %	24.13 %	24.28 %	28.59 %	26.52 %
WPP10	25.14 %	25.51 %	24.66 %	25.31 %	22.11 %	**21.60** %	28.37 %	29.62 %

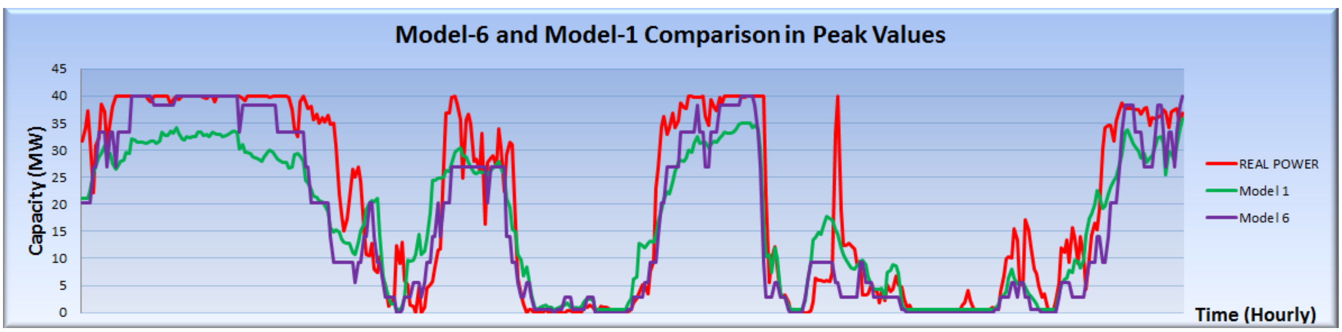

FIGURE 6. Model-6 and model-1 comparison in determination of the peak values.

maximum power forecast value obtained by the Model-1 is at 35 MW levels while Model-6 reaches the plant installed capacity as the power forecast values for the same period.

All of the models are also compared in terms of seasonal performances during the one-year test period as shown in Figure 5. Seasonal error rates are calculated as the average NMAE rates of the 10 *offline* test plants during the seasons. According to the seasonal performances, for all of the seasons, Model-6 shows better performance compared to other models on the average. Similarly, for all of the models, the lowest error rates are obtained in the autumn season while the highest error rates are obtained in the winter

TABLE 4. Results of all regions (in terms of NMAE %).

Region Name	Capacity (MW)	Model-1 Error	Model-2 Error	Model-3 Error	Model-4 Error	Model-5 Error	Model-6 Error	Base Model	ANN	SVM
Trakya YTM	680.35	9.61 %	9.62 %	9.53 %	9.54 %	9.46 %	9.42 %	11.63 %	9.23 %	**9.16** %
North West Anatolia YTM	305	9.79 %	9.82 %	10.17 %	10.04 %	9.92 %	**9.74** %	11.83 %	13.19 %	10.55 %
West Anatolia YTM	3287	8.10 %	8.12 %	7.59 %	7.61 %	7.35 %	**7.11** %	10.10 %	8.30 %	8.40 %
Middle Anatolia YTM	442.5	9.65 %	9.62 %	9.95 %	10.13 %	9.33 %	**9.28** %	12.97 %	11.32 %	11.64 %
West Mediterranean YTM	341.5	10.74 %	11.48 %	11.08 %	10.90 %	10.31 %	**9.54** %	11.00 %	10.77 %	10.22 %
Middle Black Sea YTM	163.8	14.44 %	14.94 %	15.22 %	15.18 %	14.64 %	**14.28** %	16.49 %	15.33 %	15.07 %
Southeastern Anatolia YTM	188.2	9.68 %	9.78 %	9.17 %	9.36 %	9.06 %	8.54 %	10.60 %	**8.15** %	8.34 %
East Mediterranean YTM	804.1	7.55 %	7.61 %	8.10 %	7.39 %	7.16 %	**7.13** %	10.82 %	8.98 %	8.71 %
National YTM (All WPPs)	6212	6.95 %	6.99 %	6.93 %	6.68 %	5.58 %	5.25 %	8.06 %	6.41 %	**5.12** %

season. This situation is merely related with the weather forecasts. Generally, the weather conditions in the winter season are more volatile and the accuracy of the weather forecasts are low compared to other seasons. On the other hand, in summer or autumn, weather conditions are more stable and this issue makes weather forecasts more accurate. Since the weather forecasts are the main and indispensable input to the power forecasts, the seasonal performances show differences.

For all 6 models, the computational complexity for each of 6 models is $O(k\,x\,n)$, where k is the number of *offline* plants in the system and n is the total number of reference *online* plants used in the calculation of the individual offlineSHWIP forecasts. In the first four models n is determined from a specified region with the average number of reference plants are 20-25 for these models. On the other hand, for Model-5 and Model-6, n is the number of all *online* plants in the system (nearly 110 WPPs) and reference plants are selected among them. Since the value for k is expected to be limited and much lower than n, it brings negligible time cost and the growth functions of the models can be formalized as $O(n)$.

C. EXPERIMENT 2: ANALYSIS ON REGIONAL POWER FORECASTING

In the second experiment, performance of the RegionalSHWIP for 9 regions, determined by TSO of Turkey is evaluated. These regions are determined by the transmission system operator according to distribution of the electricity transmission lines in the country. All of the regions have different number of plant density and total installed capacity. In this experiment, proposed models are also compared (in terms of regional bases) with three different approaches described in the literature, whose details are as follows:

- **Base Model:** In this method, for every region, 3 largest (in terms of installed capacity) *online* plants are selected. Then their total of power forecasts is upscaled as the forecast of the whole region with the *upscaling coefficient*. This coefficient is the ratio of total installed capacity of region to total installed capacity of the 3 largest plants in the region.
- **ANN based Model:** In this approach, 1-year of training power forecasts data obtained by SHWIP model (10 months for training, 2 months for validation) of 3 largest *online* plants in the region are given as input to the Feed Forward Back Propagation based ANN [7] with the real production data of the whole region in this 1-year period, which covers the whole 2016 year values. While constructing the network, 3 layers (Input, Hidden, Output) are used with 5 to 10 neurons. The model is implemented in Matlab [32].
- **SVM based Model:** Similar to the ANN based model, for the same reference *online* plants' 1-year of training data is used for constructing the SVM model with the same inputs in ANN model. The model is implemented by using *libsvm* library with the interfaces for Matlab [32], [33]. For the kernel, Radial Bases Function (RBF) is used and optimum values are selected for γ (gamma) and c (penalty) parameters [7], [10]. ($\gamma = 5$, c=2)

Unlike geographical regions, TEIAS identified 9 regions according to the distribution of electricity lines in the country which they named these regions as load distribution center. The results of the all models for these regions are presented in Table 4 and Table 5. Also, total number of plants in the region and total installed capacities of the regions are presented in the result tables. According to the results, the lowest error rate in the regional bases is obtained for the West Anatolia YTM and maximum error rates are obtained for the Middle Black Sea YTM. For the total of all WPPs, namely National YTM, NMAE error rate is calculated as 5.25 % for Model-6. There is a correlation between the error rates and total installed capacities of the region and generally lowest error rates are obtained in the largest regions. The reason behind this result is that, if the number of wind power plants in a region is high, than a forecast error in one plant can be compensated with better performance of another plant in the same region. Hence the error rate drops on the total. In addition to this, SVM and ANN show different performances for the different regions. The reason behind this factor is that, for the regions where these two models show better performance, the 3 largest WPPs in the region

TABLE 5. Results of all regions (in terms of NRMSE %).

Region Name	Capacity (MW)	Model-1 Error	Model-2 Error	Model-3 Error	Model-4 Error	Model-5 Error	Model-6 Error	Base Model	ANN	SVM
Trakya YTM	680.35	12.82 %	12.83 %	12.77 %	12.78 %	12.12 %	12.08 %	14.30 %	12.33 %	**12.00** %
North West Anatolia YTM	305	14.12 %	14.14 %	14.47 %	14.34 %	13.98 %	**13.65** %	16.55 %	17.59 %	14.62 %
West Anatolia YTM	3287	10.50 %	10.51 %	9.95 %	9.96 %	9.27 %	**8.93** %	13.35 %	10.94 %	11.17 %
Middle Anatolia YTM	442.5	13.37 %	13.31 %	13.75 %	13.95 %	12.49 %	**12.39** %	17.74 %	15.33 %	15.94 %
West Mediterranean YTM	341.5	15.41 %	16.39 %	15.87 %	15.63 %	14.83 %	**12.74** %	15.77 %	15.31 %	14.67 %
Middle Black Sea YTM	163.8	22.51 %	23.44 %	24.05 %	24.57 %	**22.90** %	23.35 %	25.28 %	24.72 %	24.71 %
Southeastern Anatolia YTM	188.2	13.76 %	13.76 %	12.88 %	13.22 %	12.49 %	11.08 %	14.99 %	**10.89** %	11.70 %
East Mediterranean YTM	804.1	9.94 %	10.02 %	10.61 %	9.76 %	9.35 %	**9.15** %	13.90 %	11.33 %	11.18 %
National YTM (All WPPs)	6212	8.65 %	8.69 %	9.61 %	8.35 %	7.33 %	6.85 %	10.54 %	8.44 %	**6.79** %

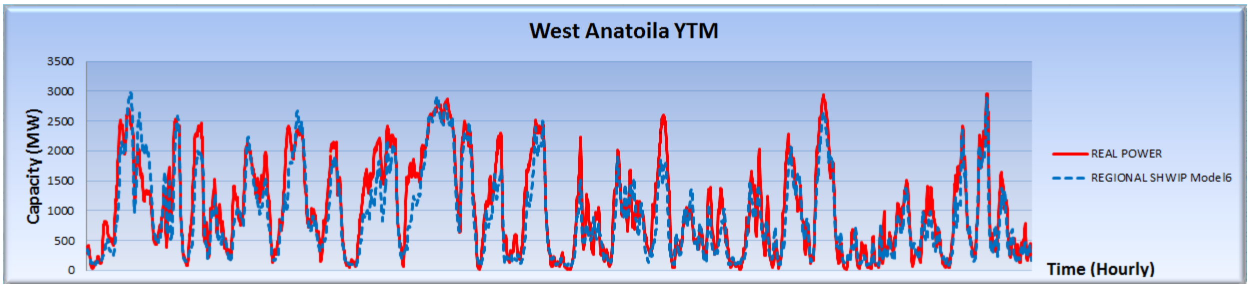

FIGURE 7. Model-6's results for west anatolia YTM.

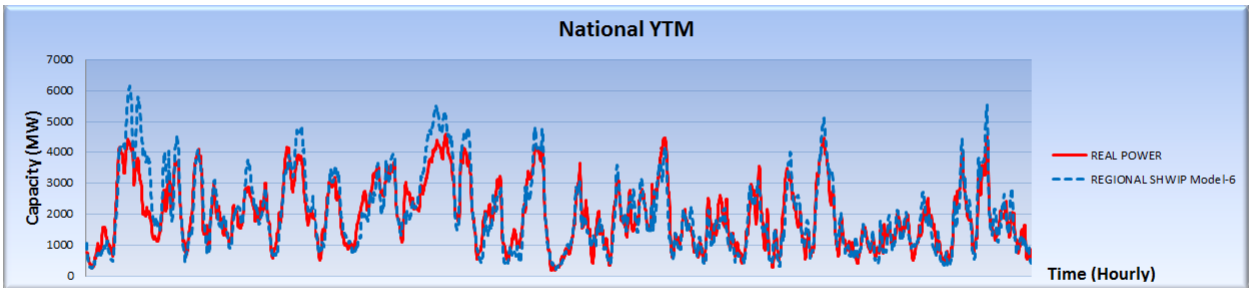

FIGURE 8. Model-6's results for national anatolia YTM.

may represent the whole region better in comparison to the other regions where the models show lower performance. For the West Anatolia YTM and National YTM, the forecast graphics of Model-6 are presented in Figure 7 and Figure 8, respectively. In these graphics, red line represents the actual measured power data in the test region and blue scatter line represents the corresponding power forecasts for the same period. As observed in the figures, the overlap between the forecast and real production value is very high.

VI. CONCLUSION AND FUTURE WORK

Wind energy is an important alternative and clean energy source with high potential. In order to benefit from this renewable energy source effectively there is a need for good planning. The aim of this work is constructing an upscaling wind power forecast model to be integrated to the RITM system. The proposed model is based on statistical methods. For the selection of the reference *online* plants and determination of the combination weights, 6 different models are proposed and evaluated. According the experimental results, performance of Model-6, which is based on historical wind speed and wind direction data similarity performs better than the other models both in plant basis and regional basis. With this approach, by using the weather forecasts and output of the *online* plants point forecasts, 48-hour power forecasts are produced for all *offline* plants in the country. As the final step, all of these power forecasts are accumulated as the forecasts of the whole region. Therefore, by using this method it is possible to generate power forecasts for all the plants in the system and for all the regions at the same time.

For the future work, by using this method wind power potential of all country may be investigated rather than individual forecasts and from that investigation a wind power atlas may be constructed for the country. This atlas can be used for initial investigation for the possible wind plant areas before making investments on the plant site.

ACKNOWLEDGMENT

The authors would like to thank all of the partners of the RITM project.

REFERENCES

[1] J. H. Seinfeld and S. N. Pandis, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Hoboken, NJ, USA: Wiley, 2016.

[2] *Climate Change 2014-Impacts, Adaptation and Vulnerability: Regional Aspects*, Intergovernmental Panel Climate Change, Geneva, Switzerland, 2014.

[3] H. Iwata and O. Keisuke, "Greenhouse gas emissions and the role of the Kyoto protocol," *Environ. Econ. Policy Stud.*, vol. 16, no. 4, pp. 325–342, 2014.

[4] C. Monteiro, "Wind power forecasting: State-of-the-art 2009," Argonne Nat. Lab., Lemont, IL, USA, Tech. Rep. ANL/DIS-10-1, 2009.

[5] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, "The state-of-the-art in short-term prediction of wind power: A literature overview," ANEMOS, Bengaluru, India, Tech. Rep., 2011. [Online]. Available: http://ecolo.porg/documents/documents_in_english/wind-predict-ANEMOS.pdf, doi: 10.11581/dtu:00000017.

[6] C. Ilkiliç and H. Aydin, "Wind power potential and usage in the coastal regions of Turkey," *Renew. Sustain. Energy Rev.*, vol. 44, pp. 78–86, Apr. 2015.

[7] S. Buhan and I. Çadirci, "Multistage wind-electric power forecast by using a combination of advanced statistical methods," *IEEE Trans. Ind. Informat.*, vol. 11, no. 5, pp. 1231–1242, Oct. 2015.

[8] E. Terciyanli, T. Demirci, D. Kucuk, M. Sarac, I. Cadirci, and M. Ermis, "Enhanced nationwide wind-electric power monitoring and forecast system," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1171–1184, May 2014.

[9] *RITM*. Accessed: Oct. 29, 2019. [Online]. Available: http://www.ritm.gov.tr/root/index.php

[10] S. Buhan, Y. Özkazanç, I. Çadırcı, "Wind pattern recognition and reference wind mast data correlations with NWP for improved wind-electric power forecasts," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 991–1004, Jun. 2016.

[11] M. Tan and Z. Zhang, "Wind turbine modeling with data-driven methods and radially uniform designs," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 1261–1269, Jun. 2016.

[12] N. Siebert, "Development of methods for regional wind power forecasting," Ph.D. dissertation, Ecole Nat. Superieure Mines Paris, Paris, France, 2008.

[13] M. Ozkan and P. Karagoz, "A novel wind power forecast model: Statistical hybrid wind power forecast technique (SHWIP)," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 375–387, Apr. 2015.

[14] M. B. Özkan, "Data mining-based power generation forecast at wind power plants," Ph.D. dissertation, Middle Eeast Tech. Univ., Ankara, Turkey, 2014.

[15] N. Siebert and G. Kariniotakis, "Reference wind farm selection for regional wind power prediction models," in *Proc. Eur. Wind energy Conf. (EWEC)*, 2006, pp. 1–11.

[16] L. von Bremen, N. Saleck, and D. Heinemann, "Enhanced regional forecasting considering single wind farm distribution for upscaling," *J. Phys., Conf. Ser.*, vol. 75, no. 1, 2007, Art. no. 012040.

[17] D. R. Drew, D. J. Cannon, J. F. Barlow, P. J. Coker, and T. H. Frame, "The importance of forecasting regional wind power ramping: A case study for the UK," *Renew. Energy*, vol. 114, pp. 1201–1208, Dec. 2017.

[18] M. G. Lobo and I. Sanchez, "Regional wind power forecasting based on smoothing techniques, with application to the Spanish peninsular system," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1990–1997, Nov. 2012.

[19] F. Ulrich, M. Lange, and H. P. Waldl, "Previento—A wind power prediction system with an innovative upscaling algorithm," in *Proc. Eur. Wind Energy Conf.*, Copenhagen, Denmark, vol. 276, 2001, pp. 1–4.

[20] L. Matthias, U. Focken, and D. Heinemann, "Regional wind power prediction with risk control," in *Proc. World Wind Energy Conf.*, 2002, pp. 1–4.

[21] K. Rohrig, "Managing 3000 MW wind power in a transmission system operation Center," in *Proc. Eur. Wind Energy Conf. Exhib.*, Copenhagen, Denmark, 2001, pp. 2–6.

[22] J. Yan, H. Zhang, Y. Liu, S. Han, L. Li, and Z. Lu, "Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3276–3284, May 2018.

[23] F. Davò, S. Alessandrini, S. Sperati, L. D. Monache, D. Airoldi, and M. T. Vespucci, "Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting," *Sol. Energy*, vol. 134, pp. 327–338, Sep. 2016.

[24] M. Ignacio, "LocalPred and RegioPred. Advanced tools for wind energy prediction in complex terrain," in *Proc. Eur. Wind Energy Conf. (EWEC)*, 2003, pp. 16–19.

[25] Á. Barbero, J. López, J. R. Dorronsoro, "Kernel methods for wide area wind power forecasting," in *Proc. Eur. Wind Energy Conf. (EWEC)*, Brussels, Belgium, 2008.

[26] *Global Forecast System*. Accessed: Oct. 29, 2019. [Online]. Available: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs

[27] *European Centre for Medium-RangeWeather Forecasts*. Accessed: Oct. 29, 2019. [Online]. Available: https://www.ecmwf.int/

[28] *Turkish Met Office*. Accessed: Oct. 29, 2019. [Online]. Available: https://mgm.gov.tr/eng/forecast-cities.aspx

[29] T. R. Derrick, B. T. Bates, and J. S. Dufek, "Evaluation of time-series data sets using the Pearson product-moment correlation coefficient," *Med. Sci. Sports Exerc.*, vol. 26, no. 7, pp. 919–928, 1994.

[30] R. Paul and Y. Escoufier, "A unifying tool for linear multivariate statistical methods: The RV-coefficient," *J. Royal Stat. Soc., C (Appl. Statist.)*, vol. 25, no. 3, pp. 257–265, 1976.

[31] S. Eren, D. Küçük, C. ünlüer, M. Demircioğlu, Y. Yanik, Y. Arslan, B. Özsoy, A. H. Güverçinci, I. Elma, Ö. Tanidir, Y. C. Ölmez, and S. Sönmez, "A ubiquitous Web-based dispatcher information system for effective monitoring and analysis of the electricity transmission grid," *Int. J. Elect. Power Energy Syst.*, vol. 86, pp. 93–103, Mar. 2017.

[32] *MATLAB*. Accessed: Oct. 29, 2019. [Online]. Available: https://www.mathworks.com/products/matlab.html

[33] *LIBSVM Library*. Accessed: Oct. 29, 2019. [Online]. Available: https://www.csie.ntu.edu.tw/ cjlin/libsvm/

MEHMET BARIS OZKAN received the B.Sc. and M.Sc. degrees in computer engineering from Middle East Technical University (METU), Ankara, Turkey, in 2011 and 2014, respectively, where he is currently pursuing the Ph.D. degree. He has been a Chief Researcher with the Scientific and Technological Research Council of Turkey (TUBITAK), Marmara Research Center (MRC), Energy Institute, since 2011. His current research interests include power quality systems, renewable energy systems, and wind power forecasting. Some of his articles were published in journals, such as the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS and the IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS.

PINAR KARAGOZ received the Ph.D. degree from the Computer Engineering Department, Middle East Technical University (METU). During the doctoral studies, she was a Researcher with SUNY Stony Brook University, Stony Brook, NY, USA. She had research visits with MIT CSAIL, USA; the Computer Science Department, Ostrava University, Czech Republic; and the Computer Engineering Department, Aalto University, Finland. She is currently a Full Professor with the Computer Engineering Department, METU. She has publications in internationally recognized and indexed journals, including IEEE TKDE, ACM TWEB, IEEE TII, and *The Computer Journal*, and has about 100 articles in international conferences in the below-mentioned areas. Her current research interests include data mining, machine learning algorithms, information retrieval, and social media analysis and mining. In 2016, she received the Best Paper Award from the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. In 2017, her article was nominated for the Wilkes Award of *The Computer Journal*.