A Three-Layer Hybrid Model for Long Term Wind Power Prediction

Jian Gao
Tandon School of Engineering
Computer Science
New York University
New York, NY 10003
Email: jg4631@nyu.edu

TBD
Big Data Laboratory, R&D Division
Hitachi America, Ltd
Santa Clara, California, 95054
Email: tbd@hal.hitachi.com

TBD
Big Data Laboratory, R&D Division
Hitachi America, Ltd
Santa Clara, California, 95054
Telephone: (000) 000–0000
Fax: (000) 000–0000

***Abstract*—Renewable energy such as wind, solar and hydropower has a rapid growth in the past decades. These energy resources give utility companies the opportunity to provide cleaner and cheaper services, which benefits both the customers and our environment. However, the intermittent nature and high variability of wind power generation causes many challenges to power system operators. An accurate wind power prediction system is very essential for economic and stable operation of the electricity markets. In this work, the problem of long-term wind power prediction (WPP) is addressed. We proposed a three-layer WPP model considering the data from historical power measurements and numerical weather prediction (NWP) systems. The first layer is a linear model that learns the wind power generation equation of wind turbines. It maps NWPs wind speed and wind direction forecasts to a basic wind power estimation. The second layer consists of several non-linear models that learn the seasonality and the inertia of wind turbines. The third layer is a stacked regression model that forms linear combinations of the predictors from the previous layer. The proposed approach is tested on a public dataset GEFCom2012, in which the task is to predict 48-hour ahead hourly wind power generation at 7 wind farms. The prediction accuracy is evaluated by RMSE (Root Mean Square Error) and MAE (Mean Absolute Error). We compared the proposed approach against the state-of-the-art algorithm as well as several neural network models. Experiment results show our approach has the best performance.**

I. INTRODUCTION

Under the pressure of global warming and environment pollution, renewable energy systems have a rapid growth in the past decades. Renewable energy resources such as wind farms, solar energy have grown to be a major asset in US electricity market due to their clean form of energy. Many countries have launched projects for large-scale integration of renewable energy resources into their grids. California state is committed to provide 50% of its consumer electricity demand through the renewable energy by 2030.

The increasing integration of wind and solar power generation leads to potential impacts on planning and operations of power systems. The utilities need to maintain the balance of demand and supply to ensure the stability of electric power operation. However, the intermittent nature and high variability of wind power generation causes many challenges to power system operators. Recent adaptations to national grid codes require wind farms to contribute to voltage regulation in the system as the conventional power plants do [1]. They must have the ability to generate or absorb reactive power in order to influence the voltage level at the point of common coupling (PCC) [2]. Energy forecasting can help to integrate wind power generation into conventional power system and satisfy the grid code.

A robust and accurate wind power prediction (WPP) system is very essential for economic and stable operation of the electricity markets. According to a great deal of studies in this area [3], [4], [5], WPP approaches can be classified based on the forecasting horizon into three categories [6]:

- Short-term forecasting (a few minutes to 8 hours-ahead)
- Mid-term forecasting (8 hours to day-ahead)
- Long-term forecasting (multiple-days-ahead)

A long-term wind power forecasting system gives the utilities time to maintain system frequency and operating reserve, as well as the advantage to bid in multiple-days-ahead electricity markets.

Wind power forecasting can be classified into physical approaches, statistical approaches, and hybrid approaches [7], [8], [9]. Physical approaches take into account the physical characteristic of the wind power generation process and establish a mathematical model from wind force to electric energy production, where accurate wind strength measurement is required. These approaches are effective in short-term forecasting and useful in dealing with operation problems. Statistic approaches learn the underlying non-linear relation between NWP wind forecasting and wind power generation through statistical methods. In general, NWP forecasts have major impact on the WPP performance.

Pierre Huyn et al. [10] developed a machine learning model based on support vector machines (SVMs) to forecast day-ahead wind power generation in 15-minute intervals. Shu Fan et al. [11] designed a model using Bayesian clustering and SVM to learn the NWP wind speed forecasting patterns. Bhaskar Kanna and Sri Niwas Singh [12] proposed an adaptive wavelet neural network (AWNN) to learn the mapping from NWPs wind speed and wind direction forecasts to wind power forecasts. Yao Liu et al. [13] provided a short-term wind forecasting model based on discrete wavelet transform and long short-term memory networks (DWT_LSTM). However, due to the large time scale in long-term forecasting and the noisy outputs of weather forecasting systems, short-term WPP approaches cannot be directly used. The industry has not established an effective approach for long-term wind power generation predictions [14].

This work address the problem of long-term wind power prediction (WPP). We proposed a three-layer WPP model considering the data from historical power measurements and numerical weather prediction (NWP) systems. The first layer learns the physical relation between wind force and wind power generation by modeling the wind generation process of wind turbines. It maps NWPs wind speed and wind direction forecasts to a basic wind power estimation based on a linear model. The second layer consists of several non-linear models that learn the seasonality and the inertia of wind turbines. The third layer is a stacked regression model that forms linear combinations of the predictors from the previous layer. The proposed approach is tested on a public dataset GEFCom2012, in which the task is to predict 48-hour ahead hourly wind power generation at 7 wind farms. The prediction accuracy is evaluated by RMSE (Root Mean Square Error) and MAE (Mean Absolute Error). We compared the proposed approach against the state-of-the-art algorithm as well as several neural network models. Experiment results show our approach has the best performance.

This paper is organized as follows: Section II introduces the problem and a brief description of the dataset. Section III provides detailed explanation of the proposed algorithms. Section IV shows the performance evaluation and comparison. Section V concludes the paper.

II. Problem Description

A. The Problem

We focus on the problem of long-term wind power prediction (WPP). The task is to predict 48-hour ahead hourly wind power generation at 7 wind farms. Available information in the training data contains historical power measurements of these wind farms and meteorological forecasts of wind components from the numerical weather prediction (NWP) systems at each farm. The actual wind power generation was measured and recorded hourly and denoted by $wp[t]$. The NWP meteorological forecasts are formulated as vectors containing the predicted wind speed and wind direction (ws_k, wd_k), $k = 1, ..., N$, where N is the total number of records. The NWP forecasts were issued twice a day at time t_k with forecasting horizon of 48 hours ahead.

$$\begin{aligned} ws_k &= \{ws[t] \mid t = t_k + 1, ..., t_k + 48\} \\ wd_k &= \{wd[t] \mid t = t_k + 1, ..., t_k + 48\} \end{aligned} \tag{1}$$

Similarly, the wind power generation is predicted in hourly resolution. The WPP model will learn the mapping from the current NWP wind forecasts and the actual wind power measurement in the past. It can be described by

$$\hat{wp}[t] = f(ws[t], wd[t], wp[t-i], \Theta) \quad i = 1, 2, ... \tag{2}$$

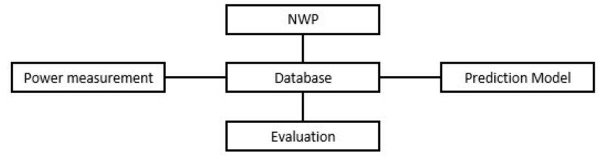

Fig. 1: High-level Architecture

TABLE I. Missing Patterns in the First Period

pos	1-12	13-24	25-36	37-48	49-60	61-72	73-84
horizon	N/A	N/A	N/A	1-12	13-24	25-36	37-48
wp	1	1	1	1	1	1	1
NWP	4	4	4	4	4	4	4

where Θ represents the model parameters learnt from the existing observations. When new NWP forecasts are available, the model is able to forecast the wind power generation at the corresponding hour. Figure 1 shows the high-level architecture.

A big challenge of this problem comes from the intermittent nature and high variability of wind power generation. Another difficult part is to deal with the time-series nature of the dataset. The problem is addressed by extracting features from the following three aspects:

1) Meteorological forecasts from NWP
2) Environmental influence
   a) Seasonal impact
   b) Location impact
3) Historical data of power measurement
   a) The entire historical value (to capture the trend)
   b) Most recent wind power generation (to capture the inertia of windmills)

For long-term wind power prediction ($>$ 24 hours), features from aspect 1) and 2) show more contribution to the prediction performance. The prediction is made by solving a supervised learning problem, and no specific time-series model was used. Therefore, we need to manually incorporate the historical data into our model and keep constantly in mind the time-series property of data streams.

B. The Dataset

The dataset is from Global Energy Forecasting Competition 2012 - Wind Forecasting [15]. It consists of NWP wind speed and direction forecasts and actual wind power measurement. All power values are normalized between 0 and 1 to hide the original characteristics of the wind farms, such as the location and actual weather information. All data is with hourly resolution. The NWP forecast outputs are available twice daily at 00UTC and 12UTC and has forecast horizon of 48 hours ahead. Thus, for each datetime, there are 4 NWP forecasts with different forecast horizons (see Table I). Figure 2 shows the NWP forecast pattern.

There are two parts of available data (in yyyy-mm-dd-hh):

- Series for the period 2009-07-01-00 to 2010-12-30-12
- Series for the period 2010-12-30-13 to 2012-06-28-12

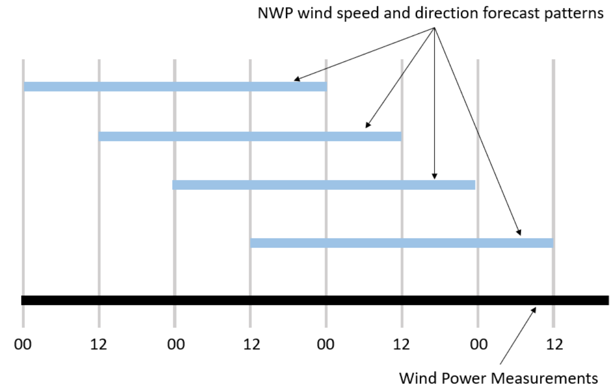

Fig. 2: NWP Forecast Patterns

TABLE II. Missing Patterns in the Second Period

pos	1-12	13-24	25-36	37-48	49-60	61-72	73-84
horizon	N/A	N/A	N/A	1-12	13-24	25-36	37-48
wp	1	1	1	0	0	0	0
NWP	4	4	4	4	3	2	1

In the first part, both actual power and wind forecasting are available at all datetimes. In the second part, a set of 48-hour periods with missing power observations are defined for validation and testing purposes. The first one is from 2011-01-01-01 to 2011-01-03-00, followed by 36-hour available period. The second one is from 2011-01-04-13 to 2011-01-06-12, followed by the second 36-hour available data. The last one is from 2012-06-26-13 to 2012-06-28-12, without data followed. Figure 3 shows the missing pattern in the second period. Both periods can be split into 156 batches, and each has 84 hours of data.

Note that, in the second period, only the meteorological forecasts that were relevant for the periods with missing power data, which would be available in practice, were given [15]. Thus, in the second period, the number of NWP forecasts at each hour various from 1 to 4, which depends on the prediction horizon. Table II listed the missing pattern of wind power and NWP forecasts in the second period.

C. Data Analysis

Figure 4 shows the probability distribution histogram of wind speed and wind direction of all available data. The mean and maximum wind speed is around 4.6 and 17.0 m/s. Since the wind blow is not uniform in all directions, we can include wind direction as a feature.

Figure 5 left shows the correlation between wind speed and wind power generation. It demonstrates strong linear relationship, so we can use linear model to learn the wind

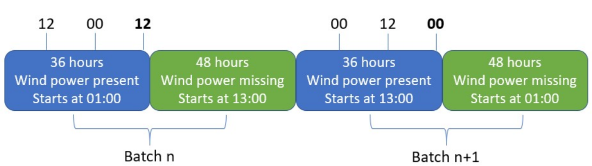

Fig. 3: NWP Forecast Patterns

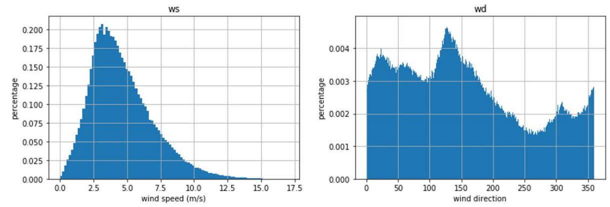

Fig. 4: Distribution of wind speed and wind direction

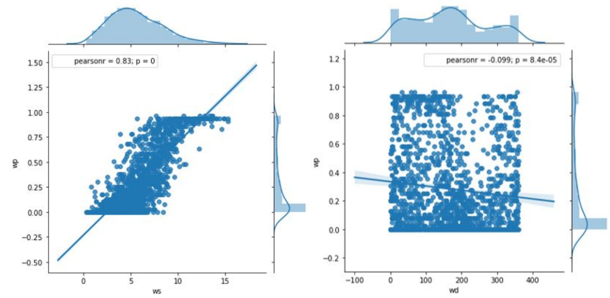

Fig. 5: The correlation with wind power

power generation equation. The correlation between wind direction and wind power generation is displayed in Figure 5 right, which shows strong non-linear relationship. This suggests using categorical variable to represent wind direction.

Figure 6 illustrates the monthly and hourly seasonality of wind power generation. It shows obvious monthly seasonal pattern, so that we can extract month as an important feature. Hourly wind power is almost uniform in one days period. Therefore, we do not include this type of feature. Another finding is that, the wind power has lowest value in July 15 and shows symmetric pattern. So, we create a feature by counting the days away from July 15: $|x - 195|$, where x is the day in a year, and 195 is the day of July 15 in a year.

To analysis the lag correlation of time-series data, we plot the auto-correlation of wind power generation in Figure 7. It shows strong auto-correlation between T-1, T, and T+1.

Although the farms locations are not released, we can still

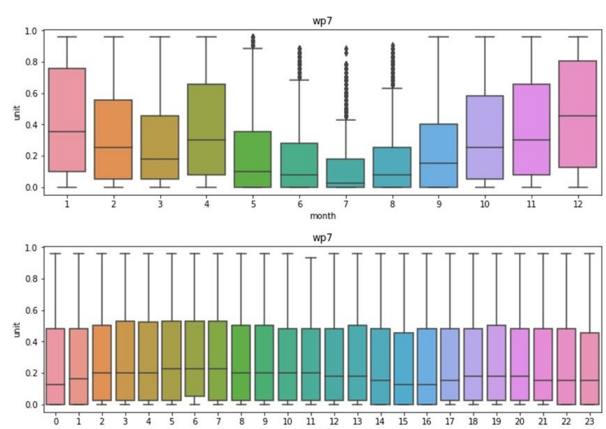

Fig. 6: Seasonality in month (top) and hour (bottom)

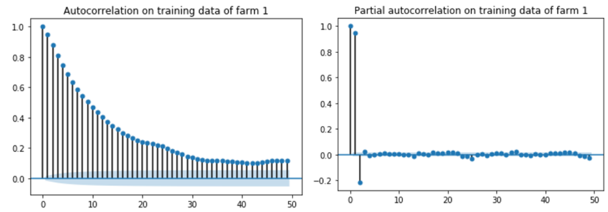

Fig. 7: Auto-correlation of Wind Power Generation

	wp1	wp2	wp3	wp4	wp5	wp6	wp7
wp1	1.000000	0.394313	0.608578	0.784888	0.581040	0.733620	0.723803
wp2	0.394313	1.000000	0.394310	0.502267	0.237813	0.487426	0.512764
wp3	0.608578	0.394310	1.000000	0.655927	0.425660	0.676722	0.703215
wp4	0.784888	0.502267	0.655927	1.000000	0.526760	0.912731	0.881610
wp5	0.581040	0.237813	0.425660	0.526760	1.000000	0.510836	0.490625
wp6	0.733620	0.487426	0.676722	0.912731	0.510836	1.000000	0.927958
wp7	0.723803	0.512764	0.703215	0.881610	0.490625	0.927958	1.000000

Fig. 8: Wind Power Correlation between Farms

get some hint from the dataset. Figure 8 shows the correlation of the wind power generation on each farm pair. We found strong relationship between farm 4, farm 6 and farm 7. This information is useful for designing prediction models.

III. MODEL

Wind power generation is a non-linear and non-stationary time series [16], which is very difficult to forecast due to the intermittent and unstable nature of wind. Since the ability of single model is limited, people use hybrid models to integrate information. The most common combinations are physical and statistical models, short-term and long-term models, and alternative machine learning models [17].

We propose a three-layer hybrid model to predict the wind power generation, in which each layer takes the output of previous layer as training data, and produce its own predictions. The first layer is a linear model that learns the wind power generation equation of wind turbines. It maps NWP wind speed and wind direction forecasts to a basic wind power estimation. The second layer consists of several non-linear models that learn the seasonality and the inertia of wind turbines. The third layer is a stacked regression model that forms a linear combination of the predictors in the previous layer. Figure 9 describes the main functional components of the proposed WPP system. The detailed components of these models are explained in the following subsections.

A. First Layer Prediction

Accurate NWP wind forecasting is decisive to train a good WPP system [18]. The first layer model use NWP data to learn the wind power generation equation of wind turbines.

Wind power is generated from the contact between wind and the blades of wind turbines. The rotating blades slow down the wind and convert wind speed to mechanical energy that drives rotor to generate electricity. The proportion of wind power that can be extracted by wind turbine is described in Figure 10,

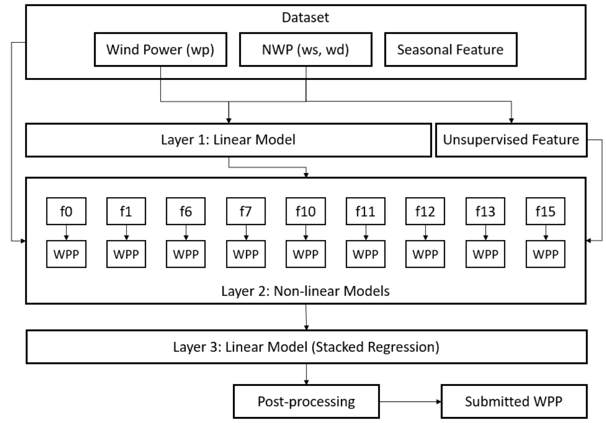

Fig. 9: The Three Layer Wind Power Prediction Framework

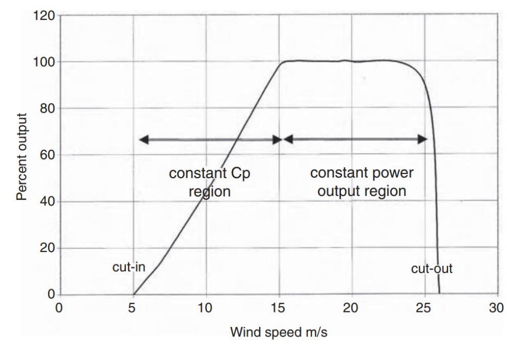

Fig. 10: Five Regions of Turbine speed Control [19]

where the conversion rate C_p is limited to 59% by Betz′s Law. The speed-power curve can be split into three regions [19]:

1) Constant C_p region, power linearly increases with wind
2) Constant power region, power reaches a controlled limit
3) Region of power shutdown, wind exceeds the upper limit

The electric energy comes from the kinetic energy of wind, which is a function of the wind speed (ws) and air mass (m). The kinetic energy of wind is

$$KE = \frac{1}{2} * m * ws^2 \tag{3}$$

and momentum in the wind is $m * ws$, thus

$$power\ per\ unit\ area = KE * momentum \propto m^2 * ws^3 \tag{4}$$

which indicates the power extracted from wind is proportional to cube of wind speed. Taking into account the impact of wind direction (wd), we use the following formula to learn this relation

$$wp \sim wd * (ws + ws^2 + ws^3) \tag{5}$$

Since the relationship between wind direction and wind power is non-linear, we split the 360-degree wind direction into 8 categories and build a categorical representation of wd. $*$ means the interactions in linear regressions. The relation between NWP wind forecasts and wind power generation is learnt by a linear model, whose output will be used in the next layer models.

B. Second Layer Predictions

Except for the meteorological forecasts from NWP, there are many other kinds of information essential for wind power prediction. The second layer takes the first layer predictions as a new feature $wind$ and extracts features from other domains. The information captured by these features include seasonal pattern, historical observation, and environmental influence. We develop multiple non-linear statistical models to learn those kinds of information and combine them into a hybrid model to obtain better predictive performance.

1) Seasonal Pattern: Based on the previous data analysis, we found that the wind has a variation according to the season of the year (Figure 6). The underlying reason is that the season may affect the temperature and humidity, which have impact on the air density. Since the wind power generation equation is based on wind speed, wind direction and air density, seasonal features like month, day, hour can express these factors in a non-explicit way. In this work, we extract four kinds seasonal features $year$, $month$, day, $hour$. All variables are in numerical format and the variable $day = |x - 195|$ represents the gap from July 15.

2) Time-Series Property: Machine learning models assume observations to have independent and identically distributed (i.i.d.) distribution. However, the temporal dependence of time-series data violates this assumption. For each timestamp, we integrate the $wind$ feature with its previous and next m-hours data as a vector $p_1, \cdots, p_m$, $n_1, \cdots, n_m$. It learns the inertial behavior of wind turbines by capturing the temporal dependence in time series observations.

Another issue is that the wind power observations are non-stationary time series. For stationary stochastic process, the mean and variance are constant values. The covariance matrix only depends on the lag and distance between two time stamps, and is independent of the actual time. But the real data of wind power measurement shows a changing mean and variance over time. We use a feature set to learn this tendency. It adds more weight to the recent observations and reduce the impact of the observations with a long distance.

3) Historical Observation: The main purpose of time-series modeling is to forecast the future by studying the past observations. Statistical models use previous wind power observations to generate prediction over the next few hours. The prediction performance falls significantly as the forecasting horizon increases. Statistical models like auto regressive integrated moving average (ARIMA) and neural network models like long short-term memory (LSTM) are only good at small horizons. For larger horizons, the partial auto-correlation of wind power observation is small, as shown in Figure 7.

We extract two types of historical features: one is to learn the long-term trend, the other is to learn the short-term inertial of wind windmills. For each 48-hours missing period, we extract the actual wind power observation before and after it: $wph0$, $wph49$. These values are the nearest available observations to the missing block and are shared by all 48 predictions in the block. In addition, we extract historical features $h1, h2$ for each horizon, representing the two most recent observations. For horizon $t \leq 24$, $h1$ is the observation t hours before, and $h2$ is the observation $t + 1$ hours before; for horizon $t \geq 25$, $h1$ is the observation $49 - t$ hours after, and $h2$ is the observation $50 - t$ hours after. Instead of an overall model for all forecasting horizons, training separate models with different historical features for each horizon can significantly improve the prediction performance.

4) Recursive Forecast: Recursive forecast is another way to use historical data. Ordinary machine learning approaches train independent models for each horizon and perform forecasting in parallel. Recursive forecasting models are trained sequentially so that the predictors at adjacent horizons can help each other. Suppose we want to forecast y using its past observation and feature x, the recursive model would be:

$$y_{t+1} = \alpha_0 + \alpha_1 y_t + \alpha_2 x_{t+1} + \epsilon_{t+1} \tag{6}$$

One step ahead forecast is

$$\hat{y}_{t+1} = \hat{\alpha}_0 + \hat{\alpha}_1 y_t + \hat{\alpha}_2 x_{t+1} \tag{7}$$

Two steps ahead forecast is

$$\hat{y}_{t+2} = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{y}_{t+1} + \hat{\alpha}_2 x_{t+2} \tag{8}$$

The forecasting models are trained recursively at each horizon by including the previous output as the historical feature. Predictions are supposed to converge to the unconditional mean for long horizons. This approach provides performance improvement in the first 12 horizons.

5) Environmental Influence: The location of wind farms have influence on wind power generation. Farms locate close to each other may have similar patterns. We extract unsupervised features to discover the environment influences. Figure 8 shows strong correlation between farm 4, 6, 7, and weak correlations between other farms. First, we extract unsupervised features $pos12$, $start$, $cluster\_all$, $cluster\_farm$ as in [20] to learn the weak correlations. Then, we add a post-processing procedure to smooth the output of the predictions. Denote (y_4, y_6, y_7) as the raw prediction vector at farm 4, 6, 7, then the vector after smoothing is the weighted average $(y_4, y_6, y_7) * (a_1, a_2, a_3)^T$. The smoothing coefficient matrix (a_1, a_2, a_3) is learnt by linear regression on the validation set. Another smoothing is conducted on the time sequence. We use moving average of a 3-hour window size to smooth the predictions. Experiments show the smoothing trick can always reduce the prediction error.

6) Summary: We have designed nine models in the second layer with different features and different type of training data. Some models are trained separately at 7 farms and/or 48 horizons, others are trained on all farms and/or all forecasting horizons. All models are trained by gradient boosting regression [21]. The summery of model variants and the feature used in layer two are listed in Table III, Table IV and Table V.

TABLE III. Prediction Models in the Second Layer

feature / data	7 farms, all hors	7 farms, 48 hors	all farms, 48 hors	all farms, all hors
forecast only	f0	N/A	N/A	f10
forecast + his	f1	f12	f11, f15	N/A
uns only	N/A	f6, f7	N/A	N/A
forecast + his + recursive	N/A	N/A	f13	N/A

TABLE IV. Features Created in the Proposed Algorithm

feature	meaning	type	range
$wind$	layer one prediction	float	[0,1]
$p_1, \cdots, p_m, n_1, \cdots, n_m$	lag of $wind$	float	[0,1]
$month$	month	int	[1,12]
$year$	year	int	[2009,2012]
$hour$	hour in a day	int	[0,23]
day	date difference to July 15	int	[0,195]
$dist$	forecasting horizon	int	[1,48]
set	batch number	float	[1,313]
$h1, h2, wph1, wph49$	historical feature	float	[0,1]
$pos12$	$(dist$-1)%12	categorical	$\{0, 1, \cdots, 11\}$
$start$	start hour of forecast	categorical	{0,12}
$cluster_farm$	cluster in one farm	int	[1,6]
$cluster_all$	cluster in all farms	int	[1,24]
wd_c8	categorical wind direction	categorical	$\{0, 1, \cdots, 8\}$
wd_c12	categorical wind direction	categorical	$\{0, 1, \cdots, 11\}$
$ws2, ws3$	wind speed square & cube	float	[0,17]
$r1$	recursive feature	float	[0,1]

C. Third Layer Prediction

Since no individual forecasting approach can capture all the information, we use hybrid method to combine the knowledge learnt by single models. This layer takes the prediction outputs from the previous layer to learn a hybrid ensemble model. We introduce three varieties of hybrid model to learn the ensemble: choose best, simple average, and stacked regression.

1) Choose Best: Model with different types of feature shows different performance at each forecasting horizon. For example in Figure 11, model $f0$ has lower prediction error in the middle while model $f1$ performs much better in the first and last few hours. The reason is, $f1$ includes historical features before and after the missing hours. This type of feature is benefit to the adjacent hours, but may deteriorate the prediction at larger distance. The ensemble model is obtained by combining experts at different forecasting horizons. Based on the prediction performance on validation set, a hybrid model $f01$ is formed by choosing $f1$ at horizons close to the available data and $f0$ at other forecasting horizons.

2) Simple Average: The mean-squared-error (MSE) of an predictor $\hat{\theta}$ with respect to the real value θ is defined as

$$MSE(\hat{\theta}) = E_{\hat{\theta}}[(\hat{\theta} - \theta)^2] = Var_{\hat{\theta}}(\hat{\theta}) + Bias_{\hat{\theta}}(\hat{\theta}, \theta)^2 \quad (9)$$

The first hybrid approach is aimed at reducing the bias by choosing the best model at different horizons. The variance

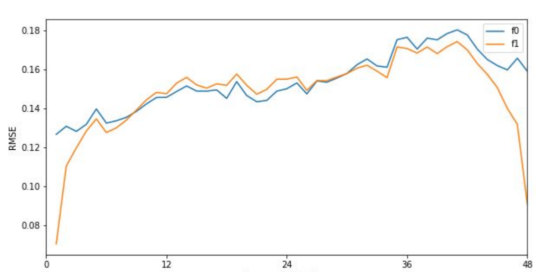

Fig. 11: Prediction error of model $f0$, $f1$ over horizons

TABLE V. Feature used in the Second Layer Models

model	feature
f0	ws, wind, season, lag of wind, dist
f1	ws, wind, season, lag of wind, dist, h1, h2
f6	wind, season, pos12, start, cluster_farm, cluster_all
f7	wind, season, pos12, start, cluster_all
f10	ws, wind, season, lag of wind, dist, farm, wd_c12
f11	ws, wind, season, lag of wind, dist, farm, wd_c12, wph1, set
f12	ws, wind, season, lag of wind, dist, farm, h1, h2, set
f13	ws, wind, season, lag of wind, dist, farm, h1, h2, set, r1
f15	ws, wind, season, lag of wind, dist, farm, wd_c12, wph1, wph49, set

can be simply reduced by averaging the predictions of multiple models. Table VI shows the averaged prediction error 0.14695 is better than the best single model's error 0.14873.

3) Stacked Regression: Suppose we have a set of predictors $f_1(x), \cdots, f_K(x)$, instead of selecting a single one from the set, a more accurate predictor can be obtained by combining them. We restrict attention to linear combinations

$$f(x) = \sum_{k=1}^{K} \alpha_k f_k(x) \quad (10)$$

In the simple average approach, $\alpha_k = \frac{1}{K}$ is a constant. Given samples $\{(x_n, y_n), n = 1, \cdots, N\}$, we can learn the coefficient α_k to minimize

$$\sum_{n=1}^{N} (y_n - \sum_{k=1}^{K} \alpha_k f_k(x))^2$$

Since the single model predictions are highly correlated, we use ridge regression with regularization term $\sum_k \alpha_k^2 = s$. It was pointed out in [22] that for non-negative coefficients α_k, the trained linear combination will be better than the best single model in set $\{f_1(x), \cdots, f_K(x)\}$. The procedure of training stacked regression is as follows :

1) Split the training set into two sets (train & holdout)
2) Train multiple base models on the first set (train)
3) Test these base models on the second set (holdout)
4) Train a high-level meta-model. Use the predictions from step 3 (out-of-folds predictions) as inputs (feature), and the actual values (target) as outputs (see Figure 12)

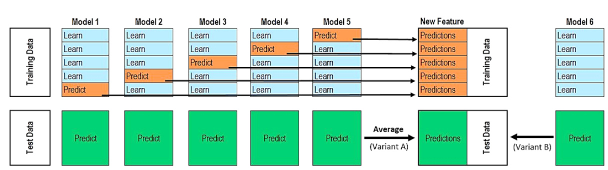

Fig. 12: Stacked Regressions

IV. Experiment Result

The dataset comes from a Kaggle competition[1], in which hundreds of teams have submitted their prediction results. The performance of the proposed approach is compared with the winner's approach [20] as well as two neural network based models NDEKF_RNN [23] and AWNN [12]. The performance

[1]https://www.kaggle.com/c/GEF2012-wind-forecasting

TABLE VI. Prediction Error in RMSE for Comparison.

model	note	RMSE
Linear Regression	layer 1	0.17390
f0	layer 2	0.15435
f1	layer 2	0.14997
f6	layer 2	0.15704
f7	layer 2	0.15777
f10	layer 2	0.15297
f11	layer 2	0.15390
f12	layer 2	0.14914
f13	layer 2	0.14949
f15	layer 2	0.15045
f01	layer 3	0.14873
simple average	layer 3	0.14695
Proposed	layer 3	**0.14508**
Leustagos [20]	winner's approach	0.14567
DuckTile [15]	local linear regression	0.14719
Duehee Lee [15]	neural network & Gaussian process	0.15501
AWNN [12]	wavelet neural network	0.15014
NDEKF_RNN [23]	recurrent neural network	0.15347
PERSIST	persistence model	0.35366
Random	random guess	0.46070

of the prediction models are evaluated by Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

In this work, we have proposed a three-layer hybrid model for wind power prediction. Table VI shows the prediction error for single models in each layer and for the hybrid model. The RMSE of our model is 0.14508, which outperforms the state-of-the-art approach and the neural network models. The performance of persistence model and random guess are also listed for reference.

In this task, the WPP model predicts 48-hour ahead hourly wind power generation at 7 wind farms. Figure 13 plots the prediction error over 48 forecasting horizons, in which $f7$ and $f12$ are the worst and the best model in the second layer. Including historical observations significantly improves the performance at horizons close to the available data, that is, the first and last few hours. Figure 14 plots the prediction error over 7 wind farms. All models show larger error in farm 3 and 5, probably due to the inaccurate NWP forecasts.

Model performance various over forecasting horizons. Figure 15 listed the performance over four horizon periods $[1, 12]$, $[13, 24]$, $[25, 36]$ and $[37, 48]$, in which the first row is the average over four periods. A comparison with Adaptive Wavelet Neural Network (AWNN) [12] and Node Decoupled Extended Kalman Filter trained Recurrent Neural Network

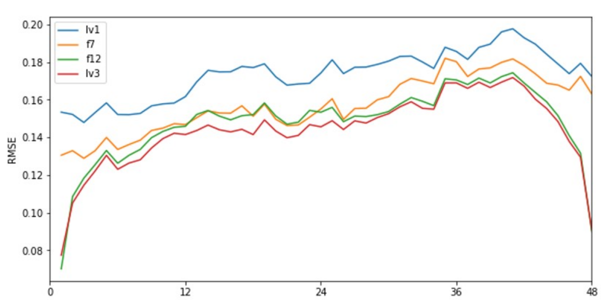

Fig. 13: Prediction error over forecasting horizons

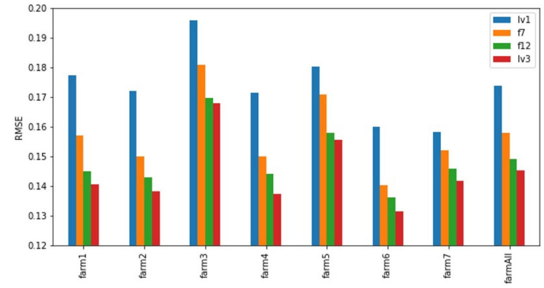

Fig. 14: Prediction error over wind farms

horizon	lv1	f0	f1	f6	f7	f10	f11	f12	f13	f15	lv3
[01,48]	0.17390	0.15435	0.14997	0.15704	0.15777	0.15297	0.15390	0.14914	0.14949	0.15045	0.14508
[01,12]	0.15585	0.13688	0.12976	0.13828	0.13917	0.13478	0.13138	0.12860	0.12964	0.13070	0.12531
[13,24]	0.17234	0.14799	0.15213	0.15220	0.15139	0.14645	0.14797	0.15135	0.15143	0.14771	0.14376
[25,36]	0.17953	0.15834	0.15781	0.16091	0.16336	0.15817	0.15674	0.15609	0.15653	0.15678	0.15287
[37,48]	0.18653	0.17164	0.16207	0.17404	0.17451	0.17028	0.17513	0.16231	0.16228	0.16720	0.15976

Fig. 15: Prediction error over four horizon periods

(NDEKF_RNN) [23] is displayed in Table VII. Note that the horizon period $[13, 36]$ is where the neural network models have the largest improvement over the persistence baseline. Actual and predicted wind power is displayed in Figure 16.

TABLE VII. Prediction error over horizon period $[13, 36]$

model	note	RMSE	MAE
Proposed	layer 3	**0.1496**	**0.1055**
AWNN [12]	wavelet neural network	0.1531	0.1172
NDEKF_RNN [23]	recurrent neural network	0.1690	0.1280
PERSIST	persistence model	0.3710	0.2725

V. CONCLUSION

This paper proposed a three-layer hybrid model for wind power prediction. The useful information comes from two aspects: the NWP wind forecast and the historical wind power measurement. We developed multiple models targeting at different aspects of knowledge. The hybrid model integrates the physical and statistical models specialized for short and long forecasting horizons. Experiment results on a public competition dataset show that the proposed prediction model has the best performance compared with the state-of-the-art approach as well as several neural network models.

Hybrid approach offers significant improvement over the single models. It combines different characteristics in the single models and reduces the bias and variance of the prediction result. The best individual model $f12$ is trained on 7 farms and 48 horizons separately, and takes into account the historical observations. The smoothing trick alleviates the effect of outliers by considering the correlation over time and location. It is found that recurrent neural network models including adaptive wavelet, extended Kalman filter and LSTM do not show advantage on this dataset. The use of recursive feature improves the prediction only on small forecast horizons.

Future work is to incorporate discrete wavelet transform with LSTM network for short-term forecasting. Another work is to develop power prediction models for solar energy and produce probabilistic predictions.

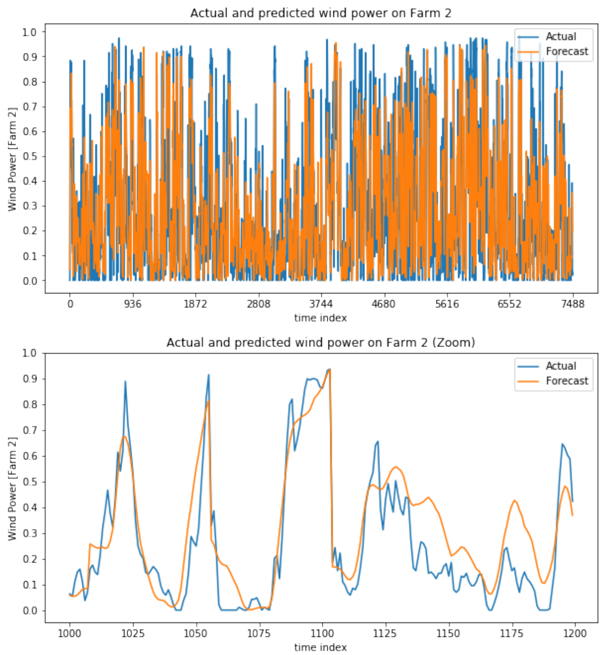

Fig. 16: Actual and Predicted Wind Power Generation

Acknowledgment

The authors would like to thank Pierre Huyn, Kishan Guddanti, and Zehan Xu for their comments and suggestions.

References

[1] H. Karbouj and Z. H. Rather, "A comparative study on the impact of grid code regulations on stability of wind integrated power systems," in *2019 1st Global Power, Energy and Communication Conference (GPECOM)*, June 2019, pp. 342–347.

[2] M. Mohseni and S. M. Islam, "Review of international grid codes for wind power integration: Diversity, technology and a case for global standard," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3876 – 3890, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032112002225

[3] H. B. Azad, S. Mekhilef, and V. G. Ganapathy, "Long-term wind speed forecasting and general pattern recognition using neural networks," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, pp. 546–553, April 2014.

[4] L. Lazi, G. Pejanovi, M. ivkovi, and L. Ili, "Improved wind forecasts for wind power generation using the eta model and mos (model output statistics) method," *Energy*, vol. 73, pp. 567 – 574, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544214007555

[5] Q. Chen and K. A. Folly, "Comparison of three methods for short-term wind power forecasting," in *2018 International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–8.

[6] X. Wang, P. Guo, and X. Huang, "A review of wind power forecasting models," *Energy Procedia*, vol. 12, pp. 770 – 778, 2011, the Proceedings of International Conference on Smart Grid and Clean Energy Technologies (ICSGCE 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1876610211019291

[7] L. LI, Y. qian LIU, Y. ping YANG, S. HAN, and Y. mei WANG, "A physical approach of the short-term wind power prediction based on cfd pre-calculated flow fields," *Journal of Hydrodynamics, Ser. B*, vol. 25, no. 1, pp. 56 – 61, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1001605813603388

[8] J. Gao and H. Tembine, "Distributed mean-field-type filters for big data assimilation," *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pp. 1446–1453, 2016.

[9] D. Alencar, C. de Mattos Affonso, R. Oliveira, J. Moya Rodrguez, J. Leite, and J. C. Reston Filho, "Different models for forecasting wind power generation: Case study," *Energies*, vol. 10, p. 1976, 11 2017.

[10] P. Huyn, P. Chongfuangprinya, S. Raju, and C. Venkatraman, "A case-study of wind turbine power forecasting using machine learning techniques," *Big Data Laboratory, R&D Division, Hitachi America, Ltd., USA*, 2017.

[11] S. Fan, J. R. Liao, R. Yokoyama, L. Chen, and W. Lee, "Forecasting the wind generation using a two-stage network based on meteorological information," *IEEE Transactions on Energy Conversion*, vol. 24, no. 2, pp. 474–482, June 2009.

[12] B. Kanna and S. N. Singh, "Long term wind power forecast using adaptive wavelet neural network," in *2016 IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics Engineering (UPCON)*, Dec 2016, pp. 671–676.

[13] Y. Liu, L. Guan, C. Hou, H. Han, Z. Liu, Y. Sun, and M. Zheng, "Wind power short-term prediction based on lstm and discrete wavelet transform," *Applied Sciences*, vol. 9, no. 6, 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/6/1108

[14] S. Han, Y. hui Qiao, J. Yan, Y. qian Liu, L. Li, and Z. Wang, "Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network," *Applied Energy*, vol. 239, pp. 181 – 191, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261919302065

[15] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 357 – 363, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169207013000745

[16] C. Wang, H. Zhang, W. Fan, and P. Ma, "A new chaotic time series hybrid prediction method of wind power based on eemd-se and full-parameters continued fraction," *Energy*, vol. 138, pp. 977 – 990, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544217312902

[17] W.-Y. Chang, "A literature review of wind forecasting methods," *Journal of Power and Energy Engineering*, vol. 2, pp. 161 – 168, 2014. [Online]. Available: http://m.scirp.org/papers/44881

[18] D. Wang, H. Luo, O. Grunder, and Y. Lin, "Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction," *Renewable Energy*, vol. 113, pp. 1345 – 1358, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148117305992

[19] M. Patel, *Wind and Solar Power Systems: Design, Analysis, and Operation, Second Edition*. CRC Press, 2005. [Online]. Available: https://books.google.com/books?id=8aQDw14fQnkC

[20] L. Silva, "A feature engineering approach to wind power forecasting: Gefcom 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 395 – 401, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169207013000836

[21] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016. [Online]. Available: http://arxiv.org/abs/1603.02754

[22] L. Breiman, "Stacked regressions," *Machine Learning*, vol. 24, no. 1, pp. 49–64, Jul 1996. [Online]. Available: https://doi.org/10.1007/BF00117832

[23] B. Kanna and S. Niwas Singh, "Improved rnn and awnn based wind power forecast using meteorological inputs," in *1st Annual International Conference on Power, Energy and Electrical Engineering (PEEE 2013)*, 08 2013.