Feature Extraction of Numerical Weather Prediction Results Toward Reliable Wind Power Prediction

Kazutoshi Higashiyama, Yu Fujimoto, Yasuhiro Hayashi
Waseda University
Tokyo, Japan
Email: higashiyama@akane.waseda.jp

***Abstract*—Wind power prediction is necessary for stable operation of a power grid under the introduction of significant wind power generation. Wind power prediction approaches based on the results of numerical weather prediction (NWP) have been developed successfully in recent years. However, the high dimensionality of NWP results can be a major obstacle when training models. This paper proposes a feature extraction scheme based on convolutional neural networks to compress high-dimensional NWP results by deriving critically important low-dimensional information for wind power prediction. The experimental results show that the proposed feature extractor can contribute to the improvement of wind power prediction accuracy.**

***Index Terms*--Convolutional neural network, deep learning, feature extraction, numerical weather prediction, wind power prediction.**

I. INTRODUCTION

Renewable energy frameworks have been installed throughout the world in recent years. The spread of renewable energy has considerable advantages, not only for reducing greenhouse gas emissions, but also for building an environmentally sustainable society. Japan has diligently reduced its dependence on nuclear power and has promoted renewable energy since the Fukushima nuclear disaster in 2011. According to the policies of the Japanese government, renewable energy is expected to supply approximately 24.3% of the country's gross electricity consumption by 2030 [1]. Therefore, more renewable energy systems will be widely introduced in the near future. In order to promote the use of renewable energy sources, it is necessary to enhance technological development so that renewable energy frameworks can be introduced at lower cost.

Wind power is one of the most attractive forms of renewable energy. However, the instability of wind power cannot be ignored when attempting to achieve a stable power supply. In particular, wind ramp events [2] are sudden variations in wind patterns. These events have a harmful impact on the frequency stability of a power system. Advanced wind power prediction can help to optimize power operation plans managing such ramp events and to minimize additional costs for strengthening a power grid with wind power.

Wind power prediction methods are mainly classified into three categories: physical, statistical, and hybrid approaches [3]. In physical approaches, numerical weather prediction (NWP) results are obtained by using a physical model set based on the observed weather data as initial values; then, the wind power generation is predicted based on the NWP results and the power curve of a wind power turbine [4]. This approach has been successful for long-term forecasting from one day to one week [2]. Meanwhile, statistical approaches generally utilize historical wind power generation data. The autoregressive integrated moving average (ARIMA) model is a typical linear model [5], and the neural networks (NNs) and support vector machines (SVMs) are well known as nonlinear models [6]. These statistical models are generally suitable for short-term forecasting from 30 min to 6 h [2]. Hybrid approaches, on which this paper focuses, utilize both physical models and statistical models. Several frameworks with hybrid approaches show good performance in medium-term forecasting from 6 h to 1 d [2].

Generally, machine learning frameworks have been adopted in the hybrid approaches, particularly for utilizing NWP results and historical wind power generation data [7], [8]. However, the high dimensionality of NWP results can be an obstacle to improving wind power prediction accuracy because of the difficulty in optimizing the parameters of machine learning models. We address this serious issue through a feature extraction scheme; the effectiveness is verified through a numerical experiment.

This paper is organized as follows. Section II explains the conventional approach to wind power prediction. Section III introduces a feature extraction scheme to tackle the high dimensionality of the NWP results. Section IV discusses the results of numerical experiments with a real-world dataset. Finally, Section V concludes with remarks on the proposed feature extraction scheme.

978-1-5386-1953-7/17/$31.00 ©2017 IEEE

II. WIND POWER PREDICTION

This paper focuses on regional wind power prediction, which targets total output of wind farms (WFs) in a region. In terms of grid operators, regional wind power prediction has more benefits than targeting an individual WF. Regional wind power prediction is still a challenging task although smoothing effects can reduce wind power fluctuation through consideration of several WFs over a wide area [9]. Fig. 1 (a) shows a conventional approach to regional wind power prediction using NWP results. The NWP results are obtained by using a physical model based on observed weather data [10], and are stored in a data server system with the observed regional wind power generation obtained from WFs. In the power prediction phase, hourly forecast regional wind power generation can be obtained by using the power predictor g_h as follows:

$$\hat{p}_{t+h} = g_h(\boldsymbol{W}_{t+h}, \boldsymbol{P}_t), \tag{1}$$

where $\boldsymbol{W}_{t+h}$ is the NWP result for the horizons of h ahead obtained at the current time t, $\boldsymbol{P}_t$ is the subsequence of observed regional wind power generation up to the current time t, and $\hat{p}_{t+h}$ is the forecast regional wind power generation at time $t+h$. The wind power predictor g can be composed of a regression model. Support vector machines (SVMs) [6], random forests (RFs) [11], and neural networks (NNs) [7] are typical implementations of such a wind power predictor. In this paper, we assume that the wind power predictors are independently trained for each horizon $h \in \mathcal{H}$.

The main focus of this paper is the dimensionality of the NWP results $\boldsymbol{W}$. The NWP results contain forecast meteorological variables such as wind speed and direction at every grid point in three-dimensional space. Therefore, the dimensions of the NWP results are relatively high. For example, the NWP results [12] used in this research are composed of three kinds of forecast meteorological variables on 120×120 grid points at a certain height from the ground, so that the data contains 43,200 dimensions for each time horizon. It is well known that statistical models employing such high-dimensional input show poor accuracy in practical tasks. The problem caused by such high dimensional data is called the "curse of dimensionality" [13], which is a common issue in statistical models. Focusing on the limited number of NWP grid points surrounding WFs can be a possible approach to avoid this issue [14]. However, the prediction errors included in the NWP results yield positional displacement of the forecast weather variables at the grid points; thus, this type of arbitrary dimension selection may overlook important information contained in the original NWP results. For this reason, in this paper, a nonlinear feature extractor f is introduced as shown in Fig. 1 (b). First, the NWP results are converted into low-dimensional feature maps $\boldsymbol{M}_{t+h}$ by the feature extractor f as follows:

$$\boldsymbol{M}_{t+h} = f(\boldsymbol{W}_{t+h}). \tag{2}$$

Then, the hourly regional wind power generation at time $t+h$ is obtained using g_h with $\boldsymbol{M}_{t+h}$ as follows:

$$\hat{p}_{t+h} = g_h(\boldsymbol{M}_{t+h}, \boldsymbol{P}_t). \tag{3}$$

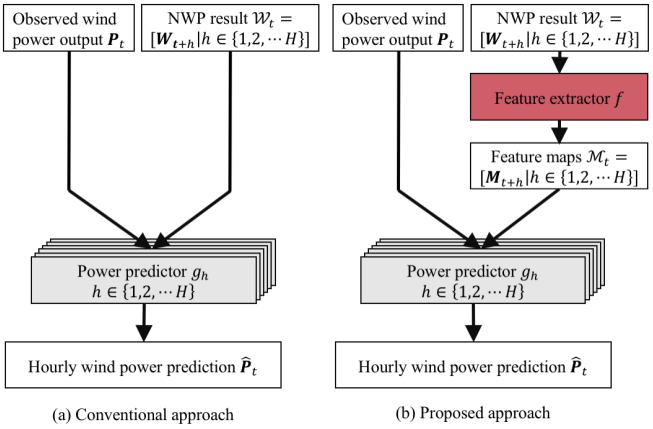

Figure 1. Procedure for prediction

Transforming a high-dimensional feature into a low-dimensional feature is a general and powerful technique for improving the performance of a trained model.

III. FEATURE EXTRACTION METHOD

In this paper, feature extractor scheme based on convolutional neural networks (CNNs) [15] are proposed. CNNs are feed-forward neural networks in which units have local receptive fields, and are mainly utilized as feature extractors for two-dimensional images. CNNs can be applied to NWP results because they have the grid topology. Two specialized layers, the convolution and pooling layers, are repeated in CNNs. They are explained in the following subsections; subsequently, the structure of our feature extraction model is described.

A. Convolution Layer

Let $\boldsymbol{I} = [I_{m,n} | m = \{1,\cdots,M\}, n = \{1,\cdots,N\}]$ be two-dimensional input data. Kernel $\boldsymbol{K} = [K_{p,q} | p = \{1,\cdots,P\}, n = \{1,\cdots,Q\}]$ is also a two-dimensional array of parameters obtained via training. Convolution, which computes a dot product between elements of the input and the kernel, is defined as

$$u_{m,n} = \sum_P \sum_Q I_{m-p,n-q} K_{p,q} + b_{m,n}, \tag{4}$$

where $b_{m,n}$ indicates bias. Then, the result $u_{m,n}$ is converted into the output of convolution layer $z_{m,n}$ as follows:

$$z_{m,n} = \phi(u_{m,n}), \tag{5}$$

where $\phi(\cdot)$ is an activation function. There are several well-known types of activation function. Our implementation adopts a rectified linear unit (ReLU) [16] for the activation function, which is defined as

$$\phi(x) = \max(0, x), \tag{6}$$

where x is the input value. A set of output $z_{m,n}$ is called a feature map, the size of which is smaller than the input because the size of the kernel is smaller than the input. Intuitively, the operation in this layer detects the two-dimensional patterns that are similar to kernels on input, and converts them into the smaller patterns.

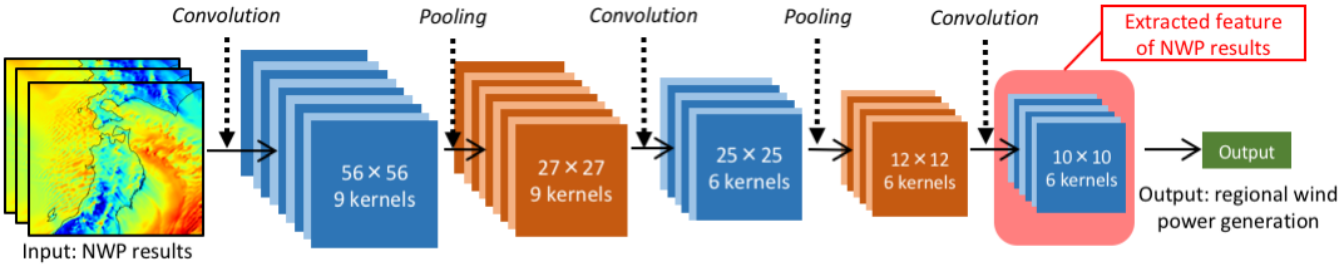

Figure 2. Proposed feature extraction model based on CNN.

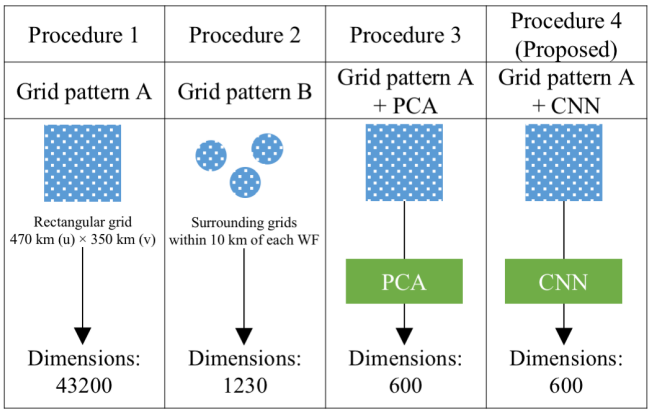

Figure 3. Four types of processing procedure for NWP results.

B. *Pooling Layer*

The pooling layer achieves a nonlinear down-sampling of the convolution layer output. The role of the pooling layer is to reduce and summarize feature maps obtained in the convolution layer. For example, max pooling returns the maximum value within subregion $H_{p,q}$ according to

$$s_{p,q} = \max_{(m,n)\in H_{p,q}} z_{m,n}, \tag{7}$$

where $H_{p,q}$ is a separated region centered around (p, q) in the inputs. Note that the pooling layer provides robustness against a small amount of positional shift by summarizing the feature maps.

C. *Proposed Model*

Fig. 2 shows the proposed feature extractor f^{CNN} constructed with CNN architecture. Although the model originally outputs the regional wind power generation by using the NWP results, the target of this model is the $\boldsymbol{M}$ obtained in the last convolution layer. The parameters of this model are determined by minimizing the mean squared error using a backpropagation algorithm [15]. Note that the proposed feature extraction model $f^{CNN}: \boldsymbol{W}_t \rightarrow \boldsymbol{M}_t$ can reduce the dimensionality of the NWP results by 98.6% (from 43,200 variables to 600 variables).

IV. Numerical Experiment

To verify the effectiveness of the proposed feature extraction scheme, numerical experiments were conducted by using a dataset collected in the Tohoku region in northeastern Japan. The region includes 20 WFs, which generate 440 MW at maximum. The regional wind power generation is normalized into 0 to 1 by dividing by the maximum value. The region experiences seasonal changes in the weather; in particular, there are constantly strong winds in winter (October to December) and spring (January to March). In the experiments, two months (December 2012 and March 2013) are individually evaluated as a validation set.

TABLE I. Specifications for Numerical Experiment

Prediction target	Regional wind power generation of Tohoku area Rated output: 440 MW	
Training period	Apr 2012–Nov 2012 Jan 2013–Mar 2013	Apr 2012–Feb 2013
Validation period	Dec 2012	Mar 2013
Frequency of prediction	Every 60 minutes with horizons up to 48 hours	

The NWP results used in this research are provided by the Numerical Weather Forecasting and Analysis System (NuWFAS) [12]. The data includes forecast meteorological variables with horizons up to 72 hours. It is assumed that the data are operationally uploaded at every 17:00 Japan Standard Time (JST) each day. By contrast, the observed regional wind power generation is uploaded every 30 minutes and recorded to a data server system. Two grid patterns of the NWP results are prepared for comparison. Grid pattern A is a rectangular grid of 120 × 120 points covering all of the targeted WFs. The grid stretches 470 km from east to west, and 350 km from north to south. Although every grid point contains many kinds of forecast meteorological variables, only the absolute wind speed and directions at above 60 m are employed, where the directions represent the sine and cosine of the wind speed. By contrast, grid pattern B includes 410 points of NWP results within a radius of 10 km around each WF. This grid pattern is used to avoid the curse of dimensionality in the conventional approach [14]. Note that grid pattern B cannot be applied to the CNN feature extractor because the input of CNN must have grid topology.

Fig. 3 shows the processing procedures for NWP results. In procedures 1 and 2, grid patterns A and B are directly used, respectively. Meanwhile, in procedures 3 and 4, grid patterns A are converted by feature extractors. The feature extractor in procedure 3 utilizes a principal component analysis (PCA), which is known as unsupervised learning for dimensionality reduction. The PCA model adopts 600 principal components for comparison with procedure 4, which employs the proposed CNN feature extractor. The number of the principal

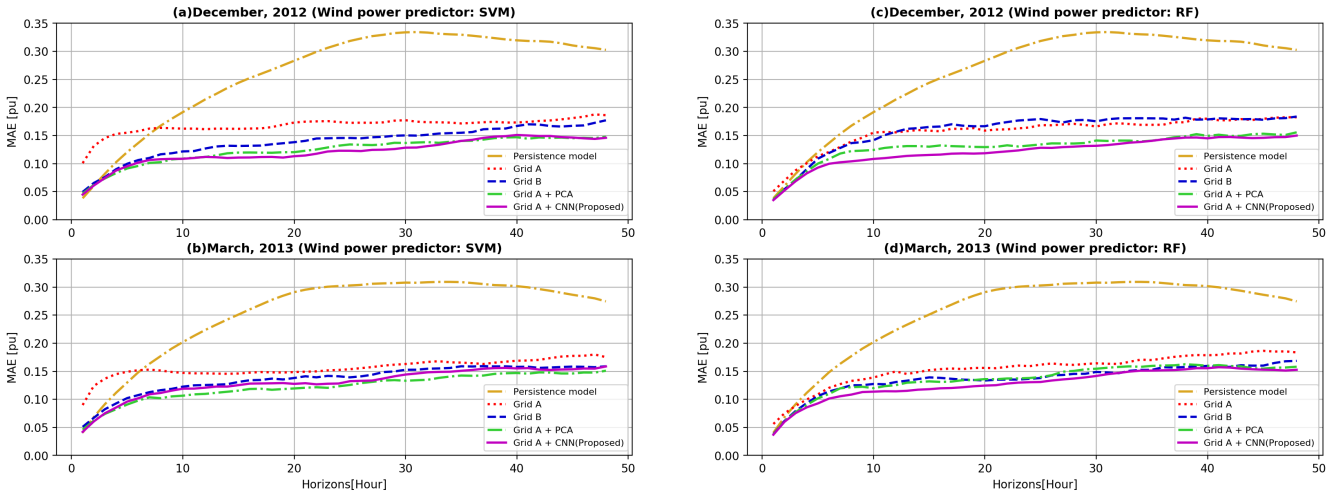

Figure 4. MAE of wind power generation.

TABLE II. MAE (VALIDATION PERIOD: DECEMBER 2012)

Horizons [Hours]	Persistence model	Wind power predictor: SVM				Wind power predictor: RF			
		Grid A	*Grid B*	*Grid A + PCA*	*Grid A + CNN*	*Grid A*	*Grid B*	*Grid A + PCA*	*Grid A + CNN*
6	0.137	0.157	0.106	0.099	0.101	0.120	0.118	0.108	0.097
12	0.215	0.161	0.126	0.111	0.111	0.155	0.156	0.129	0.111
24	0.313	0.174	0.145	0.131	0.122	0.164	0.178	0.132	0.126
36	0.328	0.172	0.156	0.140	0.142	0.169	0.180	0.143	0.143
48	0.303	0.186	0.177	0.147	0.145	0.182	0.183	0.155	0.149

TABLE III. MAE (VALIDATION PERIOD: MARCH 2013)

Horizons [Hours]	Persistence model	Wind power predictor: SVM				Wind power predictor: RF			
		Grid A	*Grid B*	*Grid A + PCA*	*Grid A + CNN*	*Grid A*	*Grid B*	*Grid A + PCA*	*Grid A + CNN*
6	0.148	0.151	0.107	0.097	0.101	0.120	0.114	0.109	0.101
12	0.224	0.146	0.125	0.109	0.120	0.149	0.130	0.125	0.114
24	0.301	0.152	0.141	0.123	0.128	0.159	0.136	0.137	0.130
36	0.308	0.164	0.158	0.141	0.152	0.172	0.157	0.160	0.151
48	0.276	0.175	0.158	0.150	0.158	0.183	0.168	0.158	0.151

components indicates 96.2% of the cumulative proportion of variance.

Although the kind of method adopted for g should be considered, our aim is to construct a feature extraction model that is capable of accepting any kinds of g. Thus, we adopt two popular machine learning methods, SVMs and RFs, which are applied in many fields. All datasets in this experiment were observed for 1 year, from April 1, 2012, to March 31, 2013. Table I lists the specifications for this experiment.

The dataset is split into periods by months. For December 2012 and March 2013 as validation datasets, we fit the models to the other months of the data, and evaluate the fitted models when predicting each of the validation datasets. These models and the persistence model, which is known as a naïve method, are implemented. To evaluate the regional wind power prediction, the mean absolute error (MAE) and root mean squared error (RMSE) are used as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|P(i) - \hat{P}(i)\right|, \tag{8}$$

$$RMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(P(i) - \hat{P}(i)\right)^2}, \tag{9}$$

where N is the number of samples in a validation set.

Figs. 4 and 5 illustrate the MAE and the RMSE, respectively, for each horizon. Tables II–V also present the MAE and RMSE for several horizons. The plots in Figs. 4 and 5 (a), (b) denote the results of SVM-based wind power predictors. By contrast, the plots in Figs. 4 and 5 (c), (d) denote the results of RF-based wind power predictors. The results in Figs. 4 and 5 (a),(b) indicate that the SVM model, which uses grid pattern A, is inferior to the other models owing to the high dimensionality of the NWP results. Meanwhile, the RF model with grid pattern A in procedure 1 is superior to the one with grid pattern B in Figs. 4 and 5 (c) because RF works well when the input data has redundant variables. The proposed procedure 4 employing the CNN feature extractor shows good performance from 10 to 30 horizons ahead, except for the SVM model in March 2013. Although the proposed procedure 4 is partially inferior to procedure 3, the proposed model can prevent significant

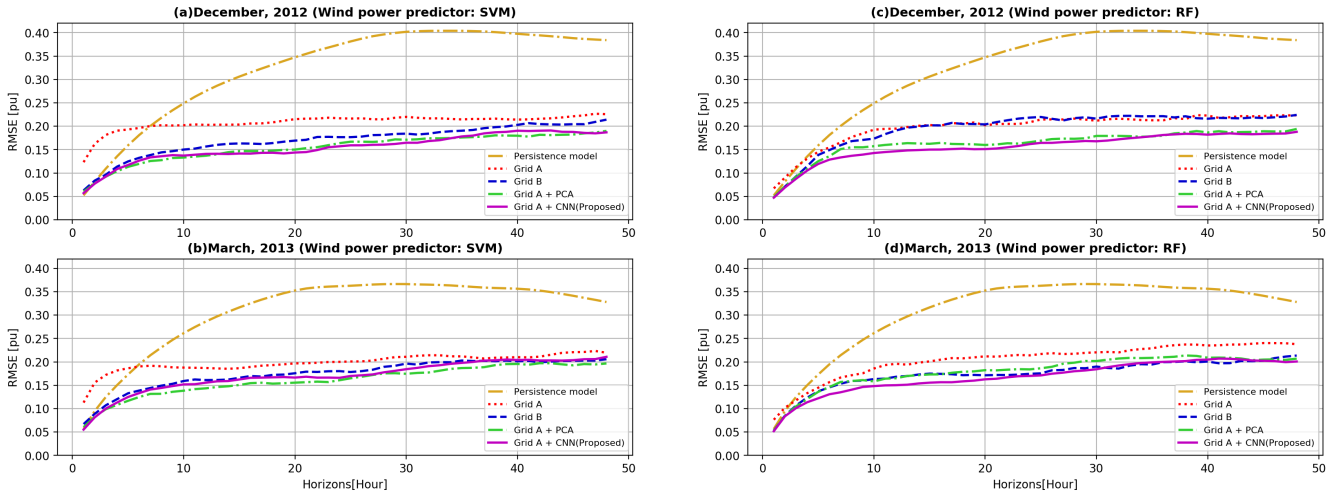

Figure 5. RMSE of wind power generation.

TABLE IV. RMSE (VALIDATION PERIOD: DECEMBER 2012)

Horizons [Hours]	Persistence model	Wind power predictor: SVM				Wind power predictor: RF			
		Grid A	*Grid B*	*Grid A + PCA*	*Grid A + CNN*	*Grid A*	*Grid B*	*Grid A + PCA*	*Grid A + CNN*
6	0.181	0.196	0.132	0.119	0.126	0.154	0.148	0.135	0.128
12	0.275	0.202	0.156	0.136	0.140	0.194	0.190	0.161	0.146
24	0.375	0.216	0.176	0.163	0.156	0.208	0.218	0.165	0.161
36	0.403	0.215	0.191	0.176	0.179	0.212	0.221	0.179	0.181
48	0.384	0.225	0.214	0.189	0.186	0.223	0.224	0.194	0.187

TABLE V. RMSE (VALIDATION PERIOD: MARCH 2013)

Horizons [Hours]	Persistence model	Wind power predictor: SVM				Wind power predictor: RF			
		Grid A	*Grid B*	*Grid A + PCA*	*Grid A + CNN*	*Grid A*	*Grid B*	*Grid A + PCA*	*Grid A + CNN*
6	0.195	0.190	0.138	0.124	0.133	0.156	0.145	0.146	0.130
12	0.287	0.186	0.161	0.143	0.155	0.196	0.168	0.165	0.150
24	0.362	0.200	0.180	0.160	0.165	0.213	0.174	0.183	0.170
36	0.361	0.209	0.201	0.187	0.199	0.233	0.199	0.209	0.200
48	0.329	0.219	0.205	0.196	0.210	0.237	0.213	0.206	0.200

deterioration of prediction accuracy owing to the CNN feature extractor.

Figs. 6 and 7 show the NWP results and feature maps, which were extracted from the NWP results by feature extractor f^{CNN} at a time when the wind power generation was large. Comparing Figs. 6 and 7, the NWP results have different patterns, whereas the feature maps look similar, e.g., (c), (e), and (f). This also holds when the regional wind power generation is low, e.g., (e) and (f) in Figs. 8 and 9. These results imply that the feature extractor f^{CNN} can not only perform mapping from high- to low-dimensional data, and can also extract essential features for wind power prediction from the NWP results. Therefore, even if two patterns of NWP results appear different, feature extractor f^{CNN} can recognize that they are similar patterns from the viewpoint of regional wind power generation. The proposed approach, which utilizes feature maps obtained by f^{CNN} instead of directly utilizing the NWP results, can be applied in a wide range of fields.

V. CONCLUSION

This paper focuses on the problem caused by the high dimensionality of NWP results when utilizing a machine learning framework to predict wind power generation. A feature extraction scheme based on CNNs was proposed to address this issue. In experiments, the proposed approach shows a higher performance than the other comparative approaches. Thus, the proposed feature extractor helps to improve the prediction accuracy for wind power generation and can be employed as a foundational tool for utilizing NWP results.

In our current implementation, the proposed CNN feature extractors was constructed by using all the NWP results without distinguishing forecast horizons, so that the feature extractor ignored even characteristics for each forecast horizon. In future work, we will implement individual feature extractors corresponding to each forecast horizon of NWP results which can reveal the trend of prediction errors in the NWP results.

ACKNOWLEDGMENT

We thank the Central Research Institute of Electric Power Industry (CRIEPI) for providing NWP datasets. This research was supported by the New Energy and Industrial Technology Development Organization (NEDO).

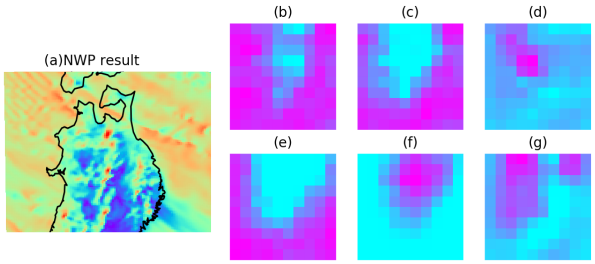

Figure 6. Example of NWP result and feature maps at 22:00 on Dec 19, 2012, when the amount of wind power generation is 0.85 p.u.

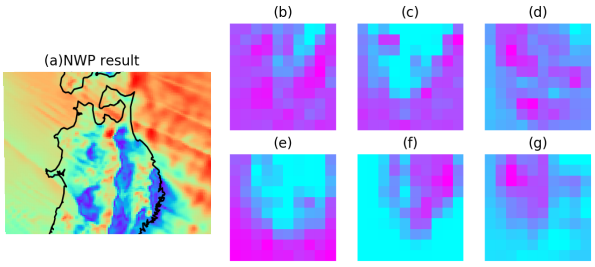

Figure 7. Example of NWP result and feature maps at 19:00 on Dec 26, 2012, when the amount of wind power generation is 0.85 p.u.

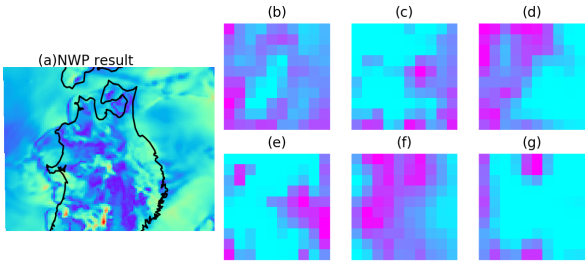

Figure 8. Example of NWP result and feature maps at 22:00 on Dec 2, 2012, when the amount of wind power generation is 0.05 p.u.

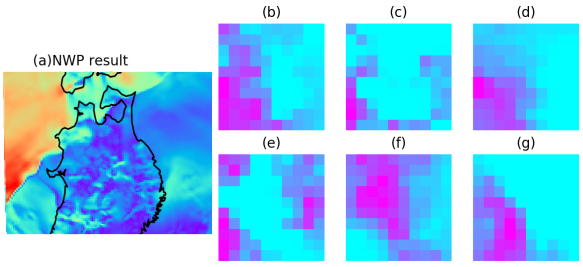

Figure 9. Example of NWP result and feature maps at 5:00 on Dec 22, 2012, when the amount of wind power generation is 0.06 p.u.

REFERENCES

[1] Ministry of Economy, Trade and Industry (METI), "Long-term Energy Supply and Demand Outlook," July 2015. [Online]. Available: http://www.meti.go.jp/english/press/2015/pdf/0716_01a.pdf

[2] R. Sevlian and R. Rajagopal, "Detection and statistics of wind power ramps," in *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3610-3620, Nov. 2013.

[3] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," *North American Power Symposium 2010*, Arlington, TX, 2010, pp. 1-8.

[4] J. Tambke, M. Lange, U. Focken, J. Wolff, and J. Bye, "Forecasting Offshore Wind Speeds above the North Sea," *Wind Energy*, vol. 8, pp. 3 – 16, 2005.

[5] A. K. Nayak, K. C. Sharma, R. Bhakar, and J. Mathur, "ARIMA based statistical approach to predict wind power ramps," *2015 IEEE PES General Meeting*, Denver, CO, 2015, pp. 1-5.

[6] J. Zeng and W. Qiao, "Short-term wind power prediction using a wavelet support vector machine," in *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 255-264, April 2012.

[7] Y. Li, C. Dai, T. Wang, Z. Zhou, S. Zhou, L. Cai, P. Musilek, and E. Lozowski, "Separate wind power and ramp predictions based on meteorological variables and clustering method," *IEEE 6th International Conference on Power Systems (ICPS)*, New Delhi, 2016, pp. 1-6.

[8] G. Sideratos and N. D. Hatziargyriou, "An advanced statistical method for wind power forecasting," in *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 258-265, Feb. 2007.

[9] M. Asari, T. Nanahara, T. Maejima, K. Yamaguchi, and T. Sato, "A study on smoothing effect on output fluctuation of distributed wind power generation," *IEEE/PES T&D Conference and Exhibition*, 2002, pp. 938-943 vol.2.

[10] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X. Huang, W. Wang, and J. G. Powers, "A description of the. advanced research WRF version 3," National Center for Atmospheric Research, Boulder, CO, NCAR TECHNICAL NOTE, NCAR/TN-475+STR, June 2008.

[11] Y. Lin, U. Kruger, J. Zhang, Q. Wang, L. Lamont, and L. E. Chaar, "Seasonal analysis and prediction of wind energy using random forests and ARX model structures," in *IEEE Transactions on Control Systems Technology*, vol. 23, no. 5, pp. 1994-2002, Sept. 2015.

[12] A. Hashimoto, H. Hirakuchi, H. Tamura, Y. Hattori, and S. Matsunashi, "A 53-year reproduction run over Japan using a regional climate model," Central Research Institute of Electric Power Industry, Technical Report N13004, Japan, Oct 2013.

[13] R.E. Bellman, *Adaptive Control Processes*, Princeton University Press, NJ, 1961, p. 94.

[14] Y. Takahashi, Y. Fujimoto, and Y. Hayashi, "Forecast of infrequent wind power ramps based on data sampling strategy," *11th International Renewable Energy Storage Conference (IRES 2017)*, Germany, Mar 2017. (to appear in *Energy Procedia*)

[15] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov 1998.

[16] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. ICML-10*, pp. 807-814, 2010.