

Empirical Mode Decomposition-k Nearest Neighbor Models for Wind Speed Forecasting

Ye Ren, P. N. Suganthan

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore City, Singapore
Email: re0003ye@ntu.edu.sg, epnsugan@ntu.edu.sg

Received December 2013

Abstract

Hybrid model is a popular forecasting model in renewable energy related forecasting applications. Wind speed forecasting, as a common application, requires fast and accurate forecasting models. This paper introduces an Empirical Mode Decomposition (EMD) followed by a k Nearest Neighbor (kNN) hybrid model for wind speed forecasting. Two configurations of EMD-kNN are discussed in details: an EMD-kNN-P that applies kNN on each decomposed intrinsic mode function (IMF) and residue for separate modelling and forecasting followed by summation and an EMD-kNN-M that forms a feature vector set from all IMFs and residue followed by a single kNN modelling and forecasting. These two configurations are compared with the persistent model and the conventional kNN model on a wind speed time series dataset from Singapore. The results show that the two EMD-kNN hybrid models have good performance for longer term forecasting and EMD-kNN-M has better performance than EMD-kNN-P for shorter term forecasting.

Keywords

Wind Speed Forecasting; Empirical Mode Decomposition; k Nearest Neighbor

1. Introduction

Wind energy generated by wind turbines is a potential renewable energy for fossil energy complementary. However, wind energy is intermittent thus more difficult to be integrated to the grid than power generated by conventional generators. In order to utilize the wind energy, accurate forecasting is necessary so that the distribution grid has a controllable demand-supply equilibrium [1].

The empirical relation between wind speed and wind power follows a non-linear function (3rd order) and so it prefers forecasting wind power generated by the wind turbine to the wind speed due to less accumulated error. However, wind speed data is usually easier to obtain even without the presence of the wind turbine, thus this paper will focus on wind speed forecasting.

Wind speed time series (TS) is a one-dimensional vector that each value in the vector consists of a wind speed value and a time-stamp. Some statistical models such as Autoregressive Moving Average (ARMA) can be used for TS modelling and forecasting and was employed for wind speed/power TS forecasting [2]. However, ARMA

is a linear model but the nonlinear wind speed TS is not suitable for linear modelling. Some Computational Intelligence (CI) based models were applied for wind speed/power TS forecasting [3], [4]. These CI based models were evaluated and compared with statistical models and the results showed that they usually outperform statistical models.

Hybrid models are developed by combining more than one model together. There are mainly four kinds of hybridization: (i) a linear forecasting model followed by a non-linear forecasting model such as ARMA-Neural Network (NN) [5]; (ii) an optimization tool to find the optimal parameters for the forecasting model such as Genetic Algorithm (GA)/Partial Swarm optimization (PSO)-Support Vector Regression (SVR) [6]; (iii) a decomposition tool to decompose the TS into several sub-TS and different forecasting models to different sub-TS such as Wavelet-SVR [7] and Empirical Mode Decomposition (EMD)- NN [8] and (iv) a further combination of the previous three kinds such as Wavelet-PSO-Adaptive Neuro-Fuzzy Inference System (ANFIS) [9].

EMD is a non-linear and non-stationary TS decomposition tool [10]. It decomposes a TS into a collection of Intrinsic Mode Functions (IMFs) and one Residue. Each IMF or Residue reflects a certain characteristic of the original TS, which is less complex than the original TS. The details of EMD are described in Section II-A.

In the literature, there are several hybrid models consisting EMD and statistical/CI based models such as EMD-SVR [11], EMD-NN [8] and EMD-MA-Persistent [12]. However, there are simpler models consisting EMD and k-Nearest Neighbor (kNN) for TS forecasting. The advantage of kNN is that it is simple, non-parametric and robust.

In [13], an EMD-kNN model was reported for annual average rainfall forecasting. The EMD-kNN model was used to forecast two rainfall datasets in China and the performance was evaluated by three error measures. The results showed that the EMD-kNN model outperformed kNN model with respect to all three error measures for the two datasets. EMD-kNN model was also employed for financial TS forecasting [14]. Four stock TS were used for evaluation against kNN and ARMA models and the EMD-kNN model was the most accurate one.

This paper applies two configurations of EMD-kNN models for wind speed forecasting: one is to construct feature vectors for each IMF or Residue and the other is to construct a feature vector set from all IMFs and Residue. The detailed configurations of the two EMD-kNN models are discussed in Section III.

The remaining of the article is organized as follows: Section II introduces EMD and kNN. Section III discusses the two configurations of EMD-kNN models. Section IV evaluates the two models with a wind speed TS collected in Singapore and Section V concludes the paper and recommends for future work.

2. Methodology

In this section, EMD and kNN are introduced in details.

2.1. Empirical Mode Decomposition

EMD [10] decomposes non-linear and non-stationary complex TS into a finite number of TS known as IMFs and one residue. The IMF has a well behaved Hilbert transformation so that the instantaneous frequency of the IMF can be calculated and therefore it is able to locate any event on time and frequency domain with IMF [10].

The procedure of EMD (known as sifting) is as follows [10], [15]:

1) Identify all local maxima and local minima in the TS $x(t)$ and interpolate all local maxima with an interpolation method such as cubic spline to form an upper envelope $e_{\max}(t)$ and use the similar interpolation method for all local minima to form a lower envelope $e_{\min}(t)$.

2) Calculate the mean of upper and lower envelopes

$$m(t) = \frac{e_{\max}(t) + e_{\min}(t)}{2}$$

3) Subtract $m(t)$ from the original TS to obtain a detail component $d(t) = x(t) - m(t)$.

4) Iterate on $d(t)$ for the previous three steps until the stopping criteria met.

5) Iterate on $m(t)$ for the previous four steps for another IMF formation.

The stopping criteria defined in [10] are: (i) in the whole TS, the number of zero-crossings and the number of extrema must equal or differ at most by one; and (ii) the mean value of the envelope must approach zero.

However, it is difficult to implement the two above criteria in computing and therefore an additional quantitative criterion is defined by Rilling [15]:

Define a mode amplitude $a(t) = \frac{e_{\max}(t) - e_{\min}(t)}{2}$, and define an evaluation function $\delta(t) = \left| \frac{m(t)}{a(t)} \right|$. The sifting iteration stops when $\delta(t) < \theta_1$ for $(1-\alpha)$ fraction of the total duration and $\delta(t) < \theta_2$ for α fraction of the total duration.

Where θ_1, θ_2 and α are user-defined thresholds.

Finally, the original TS is decomposed as:

$$x(t) = \sum_{i=1}^N \text{IMF}_i + R_N \quad (1)$$

2.2. k Nearest Neighbor

kNN is a method developed for both classification and regression. It is a kind of lazy learning [16]. The advantage of kNN is that it is a non-parametric model and therefore it can be applied to wide range of applications [13].

The procedure of kNN for regression is as follows [13], [17]:

1) Form d -dimensional feature vectors D from the historical data $x: D = \{x_t, x_{t-1}, \dots, x_{t-d+1}\}$; Their corresponding successors are denoted as x^h .

2) Form a n -dimensional distance vector DIST_i for each testing vector D_i by calculating the Euclidean distance between D_i and the remaining D : $\text{DIST}_i = \{\|D_i - D_j\|, j \neq i\}$.

3) Sort DIST_i in ascending order and select the first K entries as the nearest neighbors $D_k, k \in \{1, \dots, K\}$.

4) Form a kernel function $K(j) = \frac{1/j}{\sum_{j=1}^K 1/j}$ as a weighted averaging factor for kNN aggregation.

5) Calculate the final estimation as $\hat{x}^h = \sum_{j=1}^k K(j)x_j^h$. Where x_j is the corresponding successors of D_j .

In order to have a better generalization, kNN undergoes a leave-one-out cross validation during training to select the optimal k and d [17].

3. Two EMD-kNN Models

In the previous section, the details of EMD and kNN are discussed. The advantage of kNN as a regression tool is also discussed. However, when the TS is non-stationary and nonlinear, there will be difficulties in feature vector formation, weighted aggregation and so on [13]. A hybrid model consisting EMD and kNN is then introduced to overcome the drawbacks of kNN.

This section discusses two configurations of EMD-kNN: (i) kNN modelling for each IMF and Residue and (ii) kNN modelling by constructing feature vectors from all IMFs and Residue. We denote the first configuration as EMD-kNN-Parallel (EMD-kNN-P) and the second configuration as EMD-kNN-Multiple-Features (EMD-kNN-M). The flowchart of the two configurations is shown in **Figure 1**. The details of the two configurations are discussed in the following subsections.

3.1. EMD-kNN-P

EMD-kNN-P applies kNN to each IMF and Residue. The mathematical representation is shown:

$$\hat{y}_i = \hat{x}_i^h = \text{kNN}(\text{IMF}_i, K_i, d_i) \quad (1)$$

$$\hat{y}_R = \hat{x}_R^h = \text{kNN}(R, K_R, d_R) \quad (2)$$

$$\hat{y} = \hat{x}^h = \sum_i^N \hat{y}_i + \hat{y}_R \quad (3)$$

where $\text{kNN}(x, k, d)$ is a conventional kNN process described in Section II-B.

The advantage of EMD-kNN-P is that we can use leave-one-out cross validation to select the optimal k and

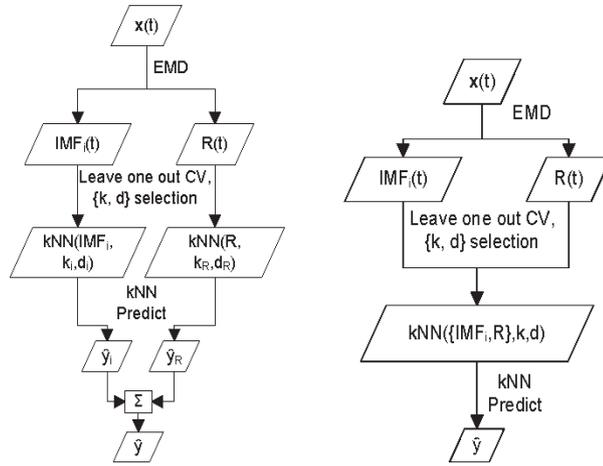


Figure 1. Flow chart of the two EMD-kNN models: (a) EMD-kNN-P and (b) EMD-kNN-M.

d for each IMF and Residue series but the disadvantage is that it requires more computing power as there are Ntimes more kNN executed for EMD-kNN-P than the conventional kNN.

3.2. EMD-kNN-M

EMD-kNN-M forms the feature vector by combining decomposed historical data. It is advantageous over the conventional kNN that the EMD-kNN-M’s feature vector contains more information and thus the distance measure is more resistance to non-linearity and non-stationarity. The mathematical representation is shown:

$$\hat{y} = \hat{x}^h = \text{kNN}(\{\text{IMF}_i, R\}, k, d), i \in \{1, \dots, N\} \tag{5}$$

Compared with EMD-kNN-P, there are less kNN processes in EMD-kNN-M. Compared with the conventional kNN, there is more information in the feature vectors.

4. Results and Discussion

The wind speed was recorded by an anemometer on top of an apartment building located at 78 Marine Dr., Singapore. The data was collected from May to Jun 2013 with 10 minute average. The dataset was split into five weekly datasets (WK17 to WK21) for evaluation. The beginning 70% was used for training and the remaining 30% was used for testing. The 5 datasets were scaled to [0,1] interval. An example of the EMD decomposed TS for WK17 training set is shown in **Figure 2**.

To evaluate the performance of the models, three error measures are used in the experiment: Root Mean Square Error (RMSE), symmetric Mean Absolute Percentage Error (sMAPE) and Mean Absolute Scaled Error (MASE):

$$\text{RMSE} = \sqrt{E[(\hat{y} - y)^2]} \tag{6}$$

$$\text{sMAPE} = E\left[\frac{|\hat{y} - y|}{\hat{y} + y}\right] \times 100\% \tag{7}$$

$$\text{MASE} = \frac{\sum_{j=1}^n |\hat{y}_j - y_j|}{\frac{n}{n-1} \sum_{i=1}^n |y_i - y_{i-1}|} \tag{8}$$

where \hat{y} is the predicted data, y is the desired data and n is the number of data points in the TS.

As shown in **Table 1**, the error measures had an increasing trend with respect to forecasting horizons for the persistent model but the trend is not consistent increasing for kNN, EMD-kNN-P or EMD-kNN-M. The RMSE,

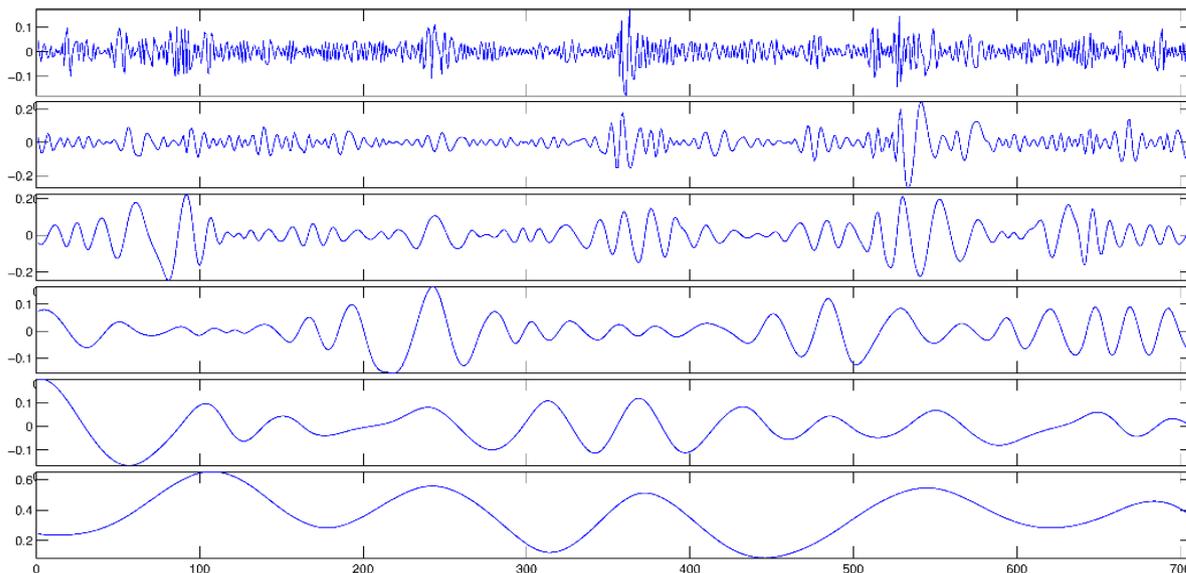


Figure 2. An Example IMFs (first 5 TS) and Residue (last TS) of WK17 Training Set after EMD.

Table 1. Comparison of different models on RMSE, sMAPE and MASE for 1, 3, 5, 7 and 9 steps ahead forecasting.

Dataset	1 Step ahead				3 Step ahead				5 Step ahead				7 Step ahead				9 Step ahead			
	Persistent	kNN	EMD-kNN-M	EMD-kNN-P																
RMSE																				
WK17	0.1015	0.1671	0.1355	0.1592	0.1436	0.1345	0.1169	0.1362	0.1618	0.1178	0.1290	0.1228	0.1760	0.1574	0.1441	0.1139	0.1802	0.1756	0.1550	0.1215
WK18	0.0843	0.2153	0.2163	0.2353	0.1551	0.1947	0.2057	0.1949	0.2019	0.2168	0.2265	0.1937	0.2385	0.2729	0.2261	0.1910	0.2652	0.3016	0.2295	0.1907
WK19	0.0575	0.1267	0.1098	0.1283	0.0926	0.0974	0.0850	0.1034	0.1174	0.0853	0.1021	0.0849	0.1320	0.1176	0.1105	0.0790	0.1456	0.1335	0.1232	0.0909
WK20	0.1104	0.1970	0.1747	0.2024	0.1641	0.1480	0.1484	0.1762	0.1915	0.1266	0.1630	0.1563	0.2121	0.1753	0.1546	0.1305	0.2312	0.2103	0.1587	0.1265
WK21	0.0857	0.1415	0.1299	0.1489	0.1289	0.1236	0.1077	0.1178	0.1399	0.1097	0.1091	0.0928	0.1503	0.1312	0.1210	0.0827	0.1740	0.1547	0.1406	0.1045
sMAPE																				
WK17	25.14	39.55	32.37	39.89	33.52	31.37	30.11	34.11	38.97	31.50	32.69	31.03	42.81	39.96	35.57	28.20	45.27	45.62	37.31	30.41
WK18	13.64	33.43	33.30	36.35	24.79	28.36	30.46	29.99	30.36	29.94	34.53	29.76	35.74	40.76	34.58	29.16	40.42	48.20	33.98	27.70
WK19	24.29	43.79	40.47	46.05	35.33	34.77	32.40	37.55	43.78	32.47	37.29	30.85	47.62	41.77	37.45	27.98	49.25	45.01	42.43	33.94
WK20	25.59	43.79	40.05	44.50	35.61	33.81	35.86	39.87	42.42	30.74	39.60	34.71	47.49	38.90	37.77	29.75	49.57	44.38	35.03	29.53
WK21	33.24	48.33	46.00	53.72	42.51	41.48	38.79	43.80	48.20	41.18	41.27	35.74	53.08	46.26	44.87	35.67	61.23	51.90	48.89	38.18
MASE																				
WK17	0.998	1.812	1.414	1.706	1.435	1.388	1.269	1.421	1.652	1.286	1.406	1.286	1.865	1.718	1.597	1.188	1.966	2.029	1.710	1.281
WK18	0.999	2.617	2.608	2.856	1.875	2.273	2.442	2.343	2.390	2.464	2.773	2.354	2.889	3.241	2.776	2.330	3.234	3.636	2.753	2.241
WK19	0.998	2.357	2.130	2.427	1.605	1.755	1.578	1.902	2.137	1.584	1.947	1.531	2.436	2.210	2.024	1.399	2.648	2.512	2.335	1.705
WK20	0.997	1.885	1.664	1.888	1.470	1.397	1.421	1.668	1.747	1.212	1.581	1.437	1.995	1.666	1.495	1.206	2.111	1.992	1.465	1.200
WK21	0.997	1.735	1.603	1.801	1.402	1.452	1.288	1.460	1.600	1.362	1.410	1.144	1.824	1.644	1.559	1.065	2.107	1.933	1.774	1.245

sMAPE and MASE plots with respect to forecasting horizon for the persistent model, the kNN, the EMD-kNN-M and the EMD-kNN-P on the five datasets are shown in Figures 3-5.

Compared with the persistent model, kNN, EMD-kNN-P and EMD-kNN-M had higher error measures for 1 and 3 step-ahead forecasting in most of the cases but the error measure of kNN, EMD-kNN-P and EMD-kNN-M became smaller than the persistent model for 7 and 9 step-ahead forecasting. We can infer that kNN, EMD-kNN-P and EMD-kNN-M have better performance than the persistent model for longer term forecasting.

kNN had smaller error measures than EMD-kNN-P but larger error measures than EMD-kNN-M for 1, 3 and 5 step-ahead forecasting in most of the cases. For 7 and 9 step-ahead forecasting, kNN had larger error measures than EMD-kNN-M and EMD-kNN-P.

The error measures for EMD-kNN-M for 1 and 3 step-ahead forecasting were smaller than that for EMD-kNN-P in most of the cases but the error measures for EMD-kNN-M for 5, 7 and 9 step-ahead forecasting were larger than that for EMD-kNN-P. Therefore we can conclude that EMD-kNN-P is the best model for longer term forecasting among the four models. EMD-kNN-M has advantage in shorter term and mid-term forecasting.

5. Conclusion and Future Work

This paper has described two configurations of EMD based kNN models for wind speed forecasting. The first configuration has applied kNN on each IMF and residue and has aggregated the predicted values to form a final

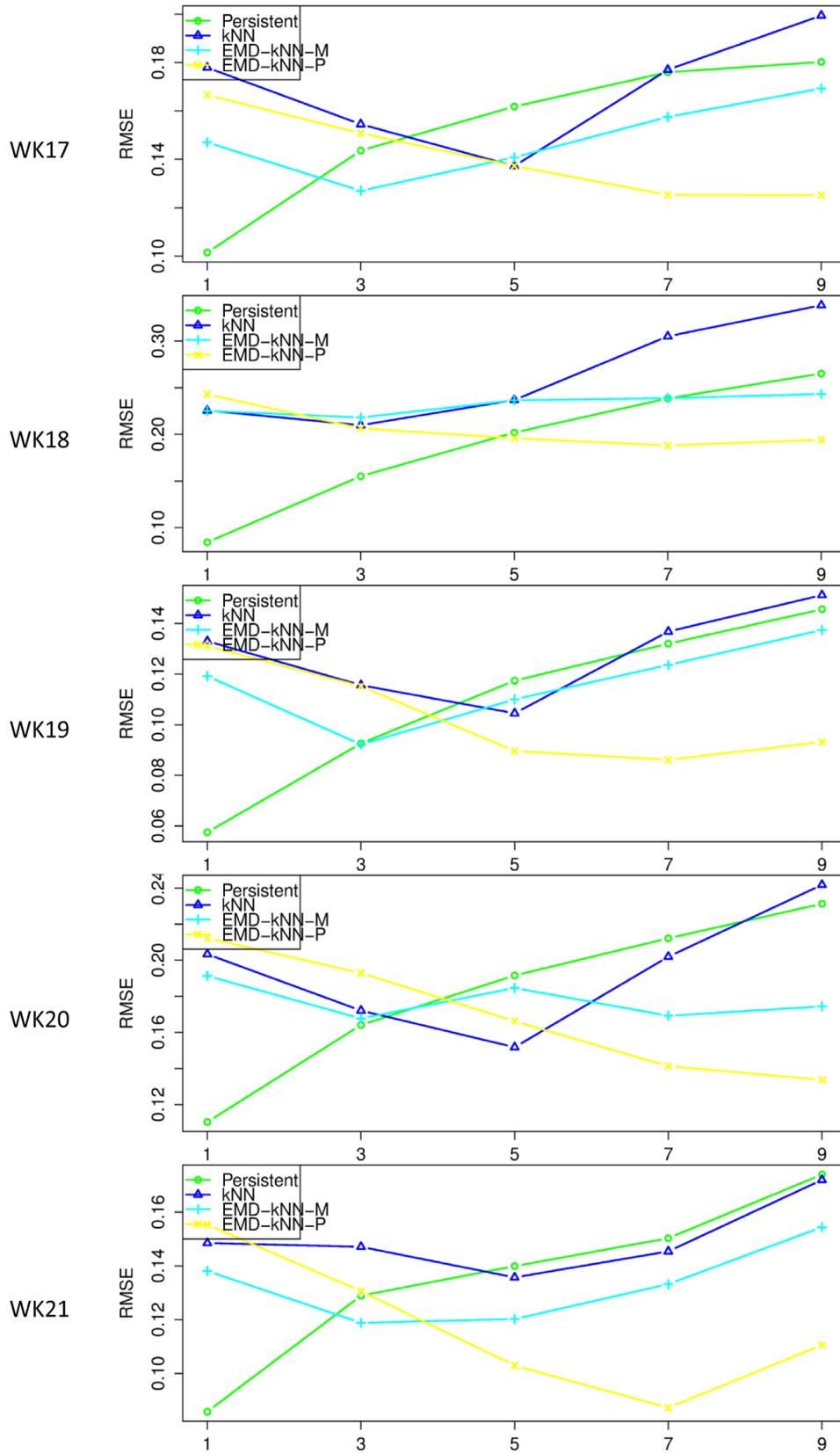


Figure 3. RMSE v.s. Horizon Plots for WK17, WK18, WK19, WK20 and WK21 datasets.

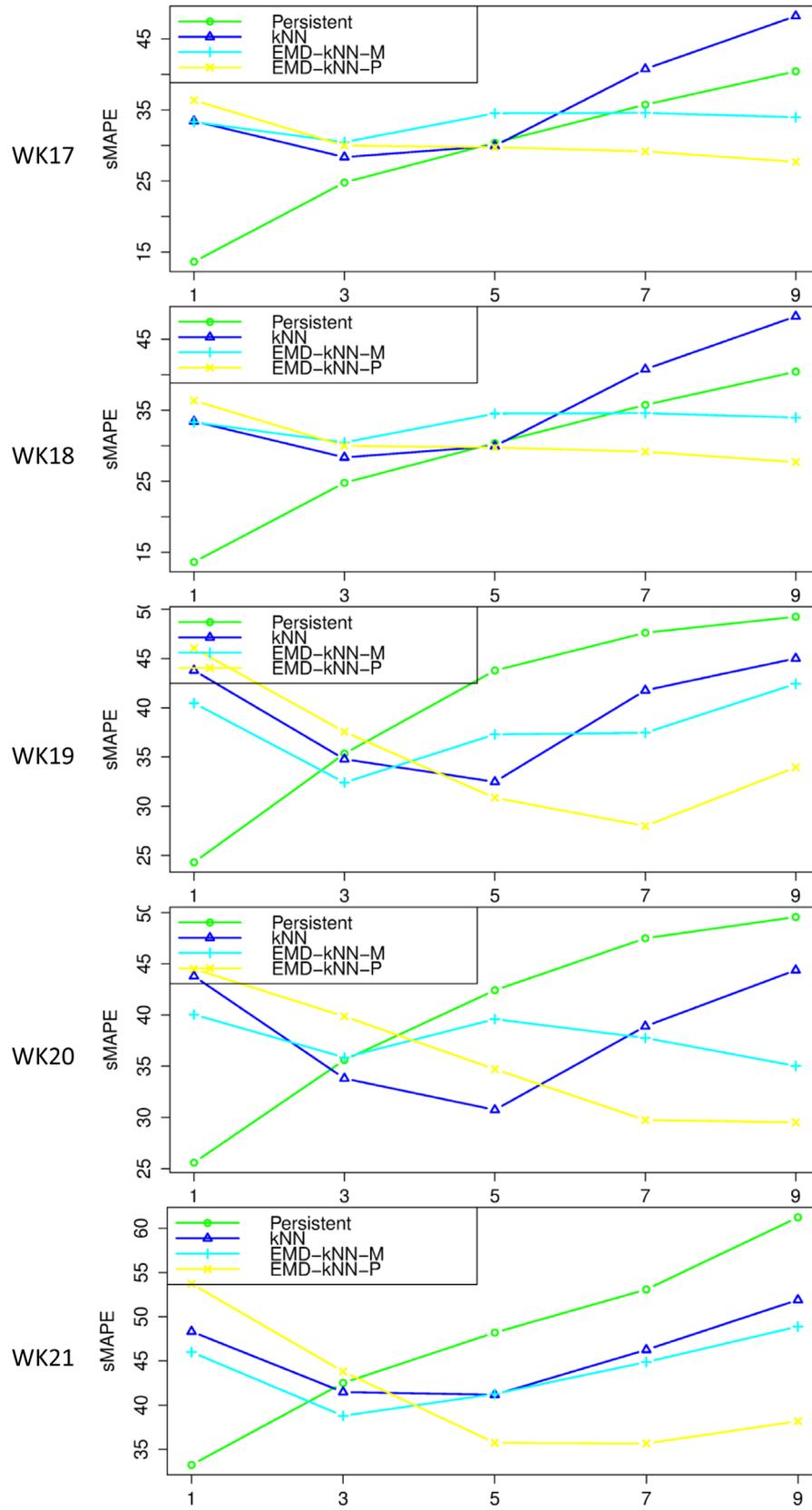


Figure 4. sMAPE v.s. Horizon Plots for WK17, WK18, WK19, WK20 and WK21 datasets.

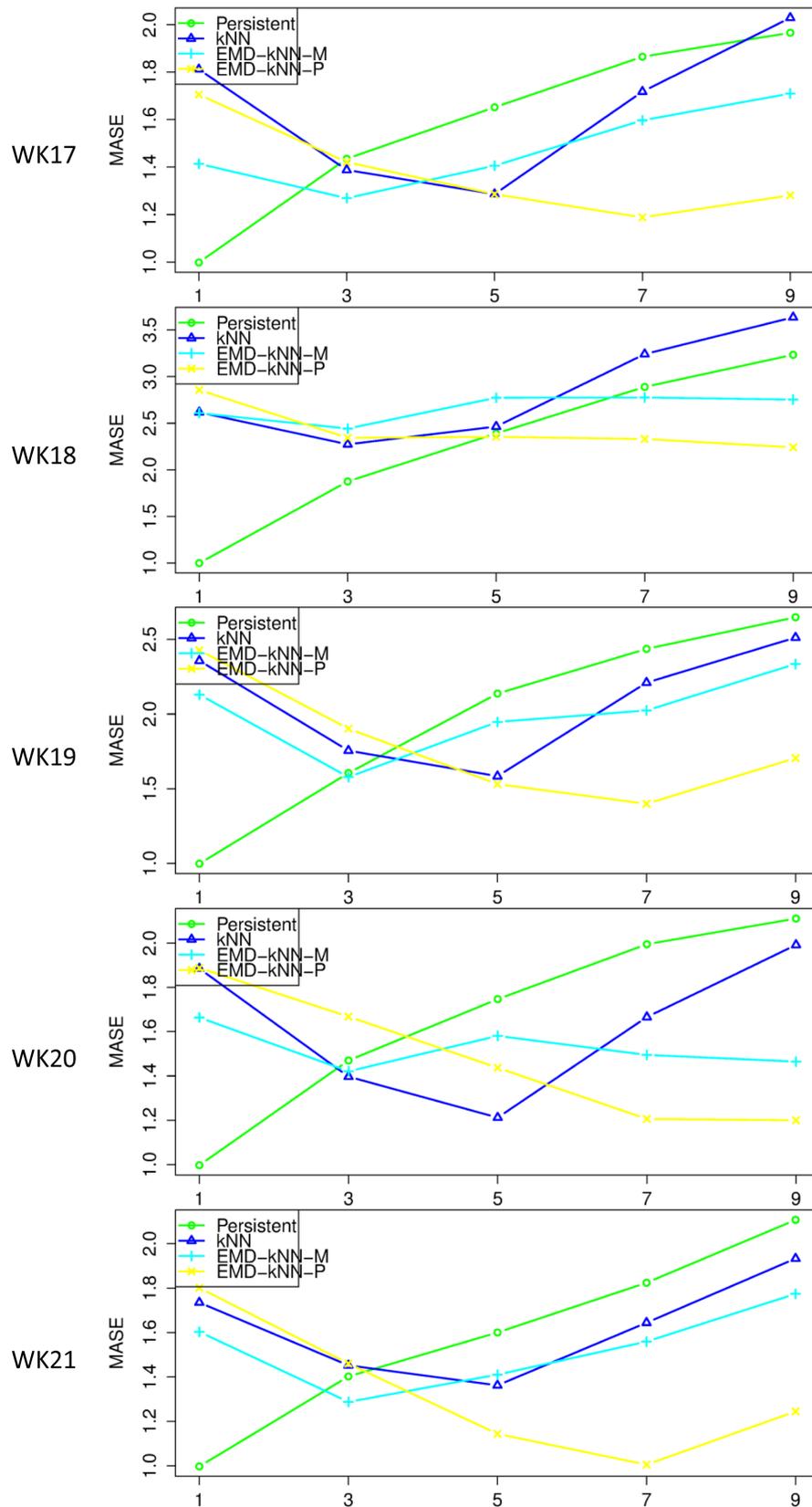


Figure 5. MASE v.s. Horizon Plots for WK17, WK18, WK19, WK20 and WK21 datasets.

prediction. The second configuration has combined all IMFs and residue together to form a feature vector set for computing the distance matrix and the prediction has followed the conventional kNN model. The two configurations have been compared with the persistent model and the kNN model with a wind speed TS recorded in Singapore. The results have shown that the two configurations outperformed the persistent model and kNN for longer term forecasting. The second configuration has outperformed the first configuration for 1 and 3 step-ahead forecasting.

For future work, a possible improvement is on the feature vector selection. Some statistical methods can be applied for the dimension selection instead of user-defined range followed by grid search based on cross validation performances. Another possible future work is to apply different weight scheme w to the distance matrix creation stage. Instead of a uniform w , a linear or exponential decayed w can be used to weigh the distance. This weighted distance may improve the selection of nearest neighbors.

Acknowledgements

The author Ren Ye would like to thank National Research Foundation (NRF) for providing the Clean Energy Program (CEPO) research scholarship.

References

- [1] Wu, Y.K. and Hong, J.S. (2007) A Literature Review of Wind Forecasting Technology in the World. *IEEE Lausanne Power Tech*, Lausanne, 1-5 July 2007, 504-509.
- [2] Hill, D.C., McMillan, D., Bell, K.R.W. and Infield, D. (2012) Application of Auto-Regressive Models to U.K. Wind Speed Data for Power System Impact Studies. *IEEE Transactions on Sustainable Energy*, **3**, 134-141. <http://dx.doi.org/10.1109/TSTE.2011.2163324>
- [3] Damousis, I., Alexiadis, M., Theocharis, J. and Dokopoulos P., (2004) A Fuzzy Model for Wind Speed Prediction and Power Generation in Wind Parks Using Spatial Correlation. *IEEE Transactions on Energy Conversion*, **19**, 352-361. <http://dx.doi.org/10.1109/TEC.2003.821865>
- [4] Salcedo-Sanz, S., Ortiz-Garcia, E.G., Perez-Bellido, A.M., Portilla-Figueras, A. and Prieto, L. (2011) Short Term Wind Speed Prediction Based on Evolutionary Support Vector Regression Algorithms. *Expert Systems with Applications*, **38**, 4052-4057. <http://dx.doi.org/10.1016/j.eswa.2010.09.067>
- [5] Shi, J., Guo, J. and Zheng, S. (2012) Evaluation of Hybrid Forecasting Approaches for Wind Speed and Power Generation Time Series. *Renewable and Sustainable Energy Reviews*, **16**, 3471-3480. <http://dx.doi.org/10.1016/j.rser.2012.02.044>
- [6] Wang, Y., Niu, D. and Ma, X. (2010) Optimizing of SVM with Hybrid PSO and Genetic Algorithm in Power Load Forecasting. *Journal of Networks*, **5**, 1192-1200. <http://dx.doi.org/10.4304/jnw.5.10.1192-1200>
- [7] Zeng, J. and Qiao, W. (2012) Short-Term Wind Power Prediction Using a Wavelet Support Vector Machine. *IEEE Transactions on Sustainable Energy*, **3**, 255-264. <http://dx.doi.org/10.1109/TSTE.2011.2180029>
- [8] Guo, Z., Zhao, W., Lu, H. and Wang, J. (2012) Multi-Step Forecasting for Wind Speed Using a Modified EMD-Based Artificial Neural Network Model. *Renewable Energy, Elsevier*, **37**, 241-249. <http://dx.doi.org/10.1016/j.renene.2011.06.023>
- [9] Catalao, J., Pousinho, H. and Mendes, V. (2011) Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Wind Power Forecasting in Portugal. *IEEE Transactions on Sustainable Energy*, **2**, 50-59.
- [10] Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C. and Liu, H. (1998) The Empirical Mode Decomposition and Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis. *Proceedings of the Royal Society London A*, **454**, 903-995. <http://dx.doi.org/10.1098/rspa.1998.0193>
- [11] Ye, L. and Liu, P. (2011) Combined Model Based on EMD-SVM for Short-Term Wind Power Prediction. *Proceedings of the CSEE*, **31**, 102-108.
- [12] Sun, C., Yuan, Y. and Li, Q. (2012) A New Method for Wind Speed Forecasting Based on Empirical Mode Decomposition and Improved Persistence Approach. *Conference on Power & Energy (IPEC2012)*, Ho Chi Minh City, 659-664.
- [13] Hu, J., Wang, J. and Zeng, G. (2013) A Hybrid Forecasting Approach Applied to Wind Speed Time Series. *Renewable Energy, Elsevier*, **60**, 185-194. <http://dx.doi.org/10.1016/j.renene.2013.05.012>
- [14] Lin, A.J., Shang, P.J., Feng, G.C. and Zhong, B. (2012) Application of Empirical Mode Decomposition Combined with K-Nearest Neighbors Approach in Financial Time Series Forecasting. *Fluctuation and Noise Letters*, **11**. <http://dx.doi.org/10.1142/S0219477512500186>

- [15] Rilling, G., Flandrin, P. and Goncalves, P. (2003) On Empirical Mode Decomposition and Its Algorithms. *IEEE-URASIP Workshop on Nonlinear Signal and Image Processing (NSIP2003)*, **3**, 8-11.
- [16] Solomatine, D.P., Maskey, M. and Shrestha, D.L. (2008) Instance-Based Learning Compared to Other Data-Driven Methods in Hydrological Forecasting. *Hydrological Processes*, **22**, 275-287. <http://dx.doi.org/10.1002/hyp.6592>
- [17] Lall, U. and Sharma, A. (1996) A Nearest Neighbor Bootstrap for Resampling Hydrologic Time Series. *Water Resources Research*, **32**, 679-693. <http://dx.doi.org/10.1029/95WR02966>