International Journal of Fuzzy Logic and Intelligent Systems, vol. 11, no. 4, December 2011, pp. 247−253
http://dx.doi.org/10.5391/IJFIS.2011.11.4.247

A Short-Term Wind Speed Forecasting Through Support Vector Regression Regularized by Particle Swarm Optimization

Seong-Jun Kim*, In-Yong Seo**

***Gangneung-Wonju National University, Gangneung, 210-702, South Korea**
sjkim@gwnu.ac.kr
****KEPCO Research Institute, Daejon, 305-380, South Korea**
iyseo@kepri.re.kr

Abstract

A sustainability of electricity supply has emerged as a critical issue for low carbon green growth in South Korea. Wind power is the fastest growing source of renewable energy. However, due to its own intermittency and volatility, the power supply generated from wind energy has variability in nature. Hence, accurate forecasting of wind speed and power plays a key role in the effective harvesting of wind energy and the integration of wind power into the current electric power grid. This paper presents a short-term wind speed prediction method based on support vector regression. Moreover, particle swarm optimization is adopted to find an optimum setting of hyper-parameters in support vector regression. An illustration is given by real-world data and the effect of model regularization by particle swarm optimization is discussed as well.

Key Words : Support vector regression, Hyper-parameter, Particle swarm optimization, Wind speed forecasting, Root mean square error

I. Introduction

The generation of wind power as renewable energy has been rapidly growing all around the world. Undoubtedly wind energy is unlimited in potential. However, although there are many advances in wind turbine technology and wind resource identification skill, its own intermittency and volatility make a restriction for efficient utilization of the wind power. Hence the accurate wind speed prediction is a primary requirement for efficient large-scale integration of wind generation in power systems[1]. For example, a 10% deviation of the wind speed leads to about a 30% deviation in the wind power generation. This is because the power potential is proportional to the cubic power of the wind speed[2].

Models for the wind power forecasting are broadly classified into physical and statistical ones[3]. It is reported that, compared with physical persistent models, statistical methods are more useful for short-term forecasting of wind speed. Short-term prediction of wind speed is made in the order of several days and also from minutes and hours[2]. Recently artificial intelligence (AI) techniques including fuzzy logic, neural network, support vector regression, and some hybrid methods have been employed for the wind power forecasting[1, 2, 4-8]. It is shown that some AI techniques outperform statistical approaches in terms of predicting accuracy. Among them, support vector regression (SVR) has much attention in the literature and it has been successfully applied to regression problem and function approximation[9]. For example, Mohandes *et al.*[6] applied this method to wind speed prediction and compared its performance with feedforward neural networks (FNN). The results indicate that SVM outperforms FNN in terms of root mean square error. Unlike FNN, SVR has distinctions that it allows the use of kernel theory to increase the quality of modeling and, moreover, that it can be efficiently solved as a convex optimization problem[7].

The purpose of this paper is to present an SVR approach for the short-term wind speed forecasting. One thing to keep in mind is that the performance of SVR heavily depends on the choice of kernel function, training sample, and several hyper-parameters. In particular, a search algorithm should be applied to find the best parameter setting so that SVR can guarantee the maximum accuracy for wind speed forecasting. For this purpose, a particle swarm optimization (PSO) is adopted in this paper. PSO is a population based stochastic optimization technique inspired by social behavior of bird flocking and fish schooling[10]. Many literatures show that PSO has a good optimization performance. Though there have been some variants of PSO, its standard edition is applied in our experiments. The remaining part of this paper is organized as follows. Section 2 outlines the proposed method for the wind speed forecasting and briefly describes SVR and PSO. In Section 3, numerical illustrations are given by using real-world wind farm dataset. In particular, effects of hyper-parameters on the forecasting accuracy are

Manuscript received Sep. 4, 2011; revised Nov. 4, 2011; accepted Nov. 7, 2011;
+Corresponding author
This work was supported by the National Research Foundation Grant funded by the Korean Government (NRF-2009-0069111).

investigated in terms of root mean square error. Finally, our works are concluded with summary in Section 4.

2. Framework of The Proposed Method

Support vector machine (SVM) is a novel machine learning technique for classification developed by Vapnik[11]. In this method, classifications are done by finding an optimum separating hyperplane (OSH) which maximizes the minimum distance between classes. In particular, when OSH cannot be found in the original space, a nonlinear and high-dimensional mapping should be applied as illustrated in the following figure[12].

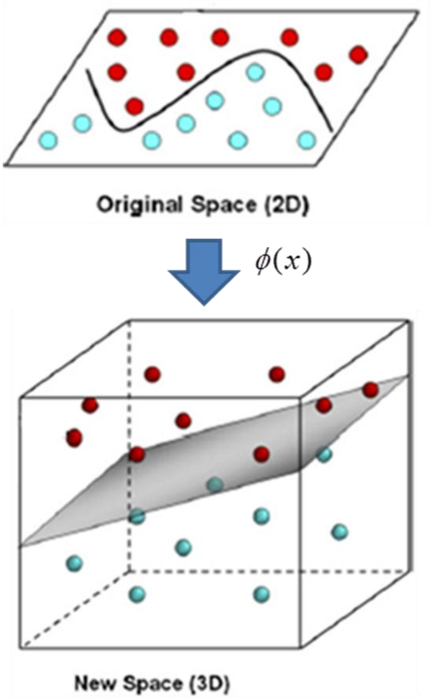


Fig. 1. Illustrating OSH in the high-dimensional feature space[12]

The notion of OSH can be extended to regression problems. The basic idea of support vector regression (SVR) is to map the feature vector x into a high dimensional space by using a nonlinear mapping $\phi(x)$. The resulting regression line $f(x)$ is expressed as:

$$f(x) = \sum w_i \phi(x_i) + b \qquad (1)$$

where w_i and b are regression coefficients obtained by minimizing the following risk function.

$$R(w,\varepsilon) = w^T w + C \cdot \sum L(y,\varepsilon) \qquad (2)$$

In the equation, C is a penalty constant for regularization and $L(y,\varepsilon)$ is an ε-insensitive loss function defined by:

$$L(y,\varepsilon) = \begin{cases} 0, & |f(x)-y| \le \varepsilon \\ |f(x)-y|-\varepsilon, & \text{elsewhere} \end{cases} \qquad (3)$$

By substituting (3) into (2), the risk function is rewritten as:

$$R(w,\varepsilon) = w^T w + C \cdot \sum (\xi + \xi^*) \qquad (4)$$

where ξ and ξ^* are non-negative slack variables. The following figure illustrates the regression line found by minimizing the above risk function.

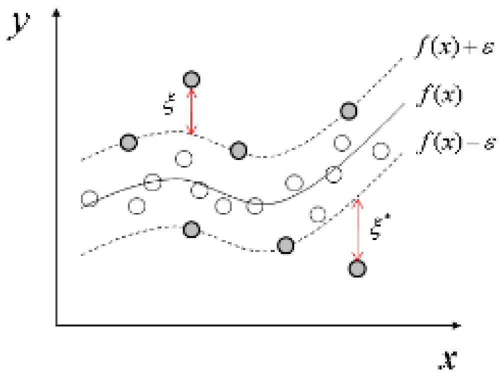


Fig. 2. Illustration of regression line with ε-insensitive zone

A new additional parameter ν can be accommodated in minimizing (4). This is because ν is capable of controlling the number of support vectors and the training errors. In general, $\nu \in (0,1]$ stands for an upper bound on the fraction of training errors and/or a lower bound of the fraction of support vectors[7]. In this case, the optimum regression line and its capability depend on C and ν which should be specified *priori*. Moreover, choice of kernel has to be considered as well. A kernel function is an inner product of two transformed feature vectors and it is written by $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. In the literature, there are several kernel functions, namely linear, polynomial, sigmoidal, and Gaussian kernels. Among them, Gaussian kernel is most commonly used and therefore adopted in this paper. Gaussian kernel is defined by:

$$k(x_i, x_j) = \exp(-\| x_i - x_j \|^2 / \sigma^2) \qquad (5)$$

where $||\cdot||$ and σ denote the 2-norm and the kernel bandwidth respectively. It is noted that modeling accuracy as well as predicting capability is affected by the value of σ. This will be illustrated later by example of wind speed dataset.

Particle swarm optimization (PSO) is a heuristic approach proposed for evolutionary computational optimizations[10]. A particle represents one of potential solutions in PSO. Let d denote the number of decision variables. A swarm of m particles is generated in which each particle is assigned to a random position in the solution space. Let s and t respectively denote the position and the velocity of particles. Each particle moves towards its best previous position p and towards the best previous particle g. The velocity and the position of a particle are updated by using the following equations:

$$t_i(k) = \lambda t_i(k-1) + c_1 r_1 [p_i(k) - s_i(k)] + c_2 r_2 [g(k) - s_i(k)] \quad (6)$$

$$s_i(k) = s_i(k-1) + t_i(k) \quad (7)$$

where λ is an inertia constant to control the impact of the previous history on the current velocity, c_1 and c_2 are learning factors, and r_1 and r_2 are random numbers uniformly distributed in [0,1]. Henceforth, a particle moves towards $p_i(k)$ and $g(k)$ by the stochastic mechanism to escape from local optima. The algorithm of PSO is terminated when maximum number of generations is reached or the best particle position of the entire swarm is little improved.

The proposed approach to wind speed prediction in this paper is based upon the combination of SVR and PSO. Future values of the wind speed are predicted by constructing SVR model of past wind speed data. In the process of modeling and validation, PSO is used to regularize hyper-parameters of SVR. This is essential to improve the accuracy of wind speed forecasting, which is evaluated by root mean square error in this study. Our framework for wind speed forecasting is shown in the following figure.

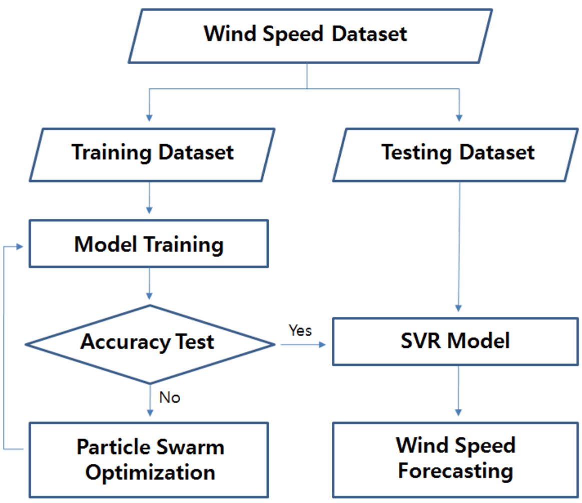


Fig. 3. Procedure of the proposed method for wind speed forecasting

3. Application to Wind Speed Forecasting

The proposed approach for wind speed prediction is illustrated by using real-world dataset shown in the following figure. The wind speed dataset was measured by the meter per second (mps) and obtained at March 2011 from a wind farm site in South Korea. Several hundred cases are depicted in the figure. The dataset is divided for model training and testing. In particular, samples for training are randomly selected from the entire dataset in order to reduce a localization effect. Two hundred sample cases are included in the training dataset.

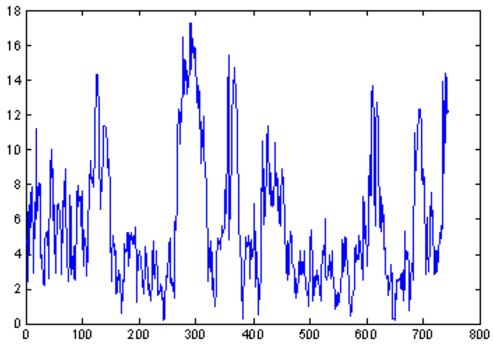


Fig. 4. Illustration of wind speed data

The forecasting accuracy of wind speed is in general evaluated by root mean square error (RMSE) defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / n} \quad (8)$$

where n is the number of samples in the training dataset and y_i and $\hat{y}_i$ respectively denote i th observation and its prediction. RMSE of (8) is a widely used statistic in time series analysis. Model training is conducted so that RMSE can be minimized. Nevertheless, RMSE is normalized by the standard deviation in this study. Normalized RMSE is useful for comparing accuracies over different datasets.

The wind speed prediction system under consideration uses p recent wind speed observations as input to obtain k future predictions of wind speed. This paper sets $p = 3$ and $k = 1$ for illustration of the proposed method. Gaussian radial basis function in (5) is chosen as kernel for SVR modeling. As described in Section 2, the penalty constant C, the fraction of support vectors ν, and the kernel bandwidth σ are specified in SVR training and testing. The best combination of them will be searched by PSO. Two datasets of which starting points were randomly selected are prepared for testing SVR. The model training is conducted by using LIBSVM, a Matlab toolbox provided in [13]. However, according to our preliminary experiments, C tends to be large as iterations go on. This leads to a larger ν and, subsequently, a model overfitting would be incurred. Therefore, we apply a restriction $0.5 \le C \le 3.0$ to the experiment. The figure below shows trends of the normalized RMSEs obtained by PSO-SVR training.

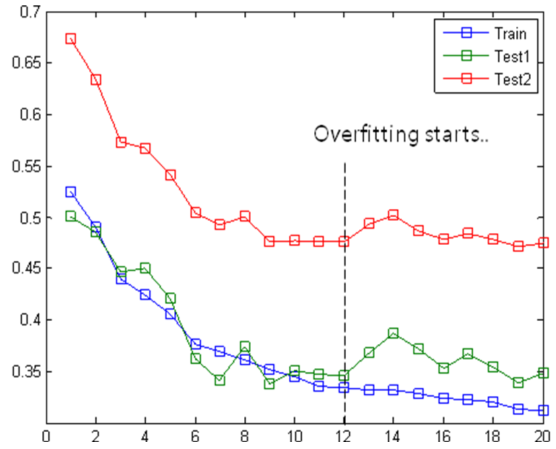


Fig. 5. Normalized RMSE by generation in PSO-SVR training

In this experiment, the numbers of particles and generations for PSO search are given as 10 and 30 respectively. We can see that RMSE of the training dataset is strictly decreasing as the generation goes on. Although the smaller RMSE can be preferred, a caution should be placed against model overfitting. Above two datasets for testing are useful to check this problem up. In the figure, RMSEs of test datasets are not improved but increased after generation 14. This strongly indicates that model overfitting began at that point and the final hyper-parameter setting should be chosen by $(C, \nu, \sigma) = (1.9187, 0.3069, 0.0364)$ from the 13th generation. The corresponding movements of generation-best particles are illustrated by the following figures.

Once a prediction model is configured by the above setting of hyper-parameters, a series of wind speed predictions is easily obtained. The resulting calibration-prediction plot of the training and testing datasets is depicted in Fig. 7.

RMSE of the training dataset is 1.1709. Similarly, RMSEs of two testing datasets are 1.5283 and 1.3004 respectively. Predictions are relatively better in the testing dataset 2. This is because the variation of the testing dataset 2 is smaller than that of the testing dataset 1. Standard deviations of testing datasets are respectively 5.5598 and 3.0304. The corresponding individual predictions are shown in Fig. 8.

So far we explained a PSO-SVR procedure for short term wind speed forecasting and illustrated the usefulness of PSO in the process of training and validation. According to our experiments, RMSEs are also affected by variations of testing dataset. In order to take a closer look at this point, more numerical experiments are conducted. The following table summarizes means, standard deviations, and RMSEs obtained through 20 experimental runs.

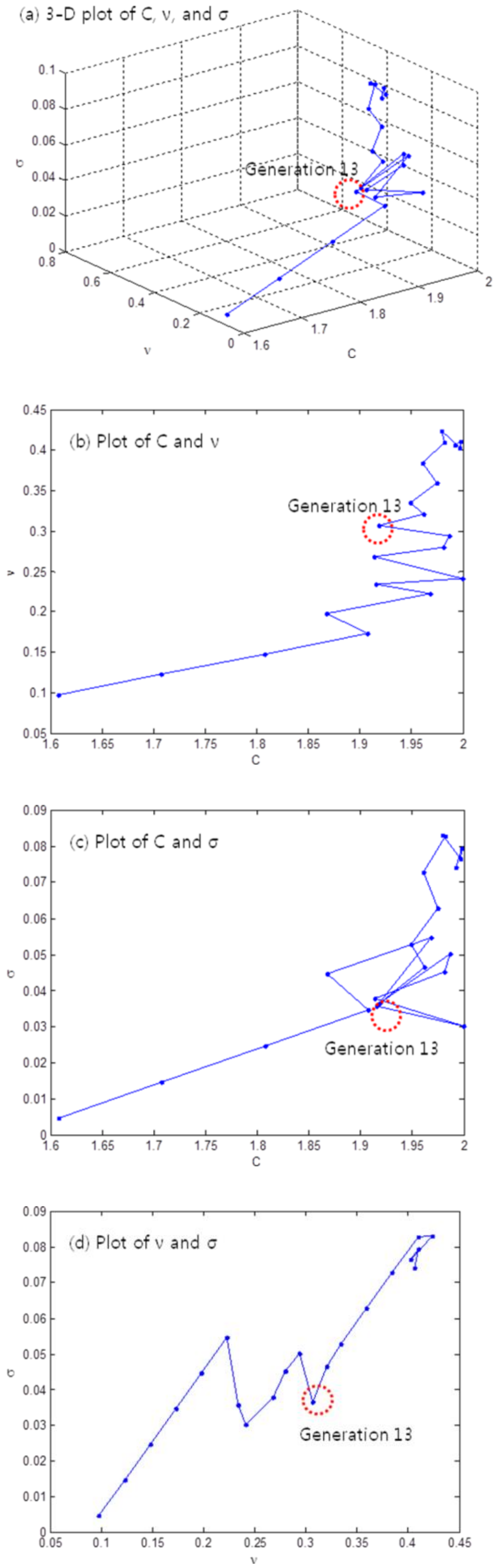


Fig. 6. Generation to generation movement of particles in PSO-SVR training

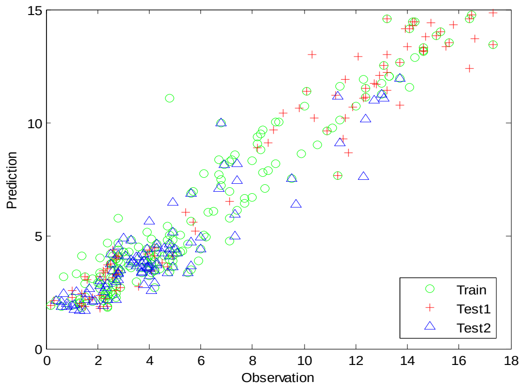
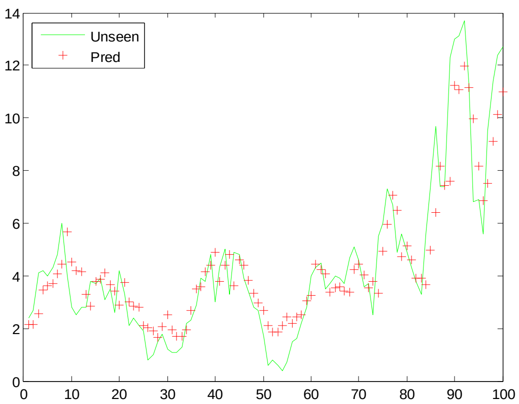


Fig. 7. Calibration-prediction plot of training and testing datasets

(a) Dataset 1



(b) Dataset 2

Fig. 8. Individual predictions for testing datasets

We can see that means and standard deviations of the training dataset are fairly stable in the table. This is caused by randomly selecting the training samples. On the other hand, testing datasets have large variability. By considering practical situations for the wind speed prediction, both variability and locality are unavoidable. This point is undoubtedly recognized by the line plots in Fig. 9.

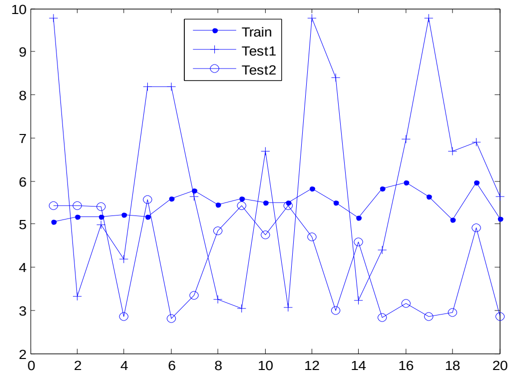
Table 1. Mean, standard deviation, and RMSE according to dataset

Mean			Standard Deviation			RMSE		
Train	Test1	Test2	Train	Test1	Test2	Train	Test1	Test2
5.041	9.767	5.422	3.277	4.843	3.782	1.218	2.427	1.705
5.164	3.333	5.421	3.481	1.798	2.656	1.175	0.873	1.327
5.177	4.974	5.403	3.335	4.231	2.679	1.131	1.715	1.378
5.211	4.193	2.861	3.308	2.630	1.157	1.420	0.897	0.848
5.159	8.183	5.557	3.267	5.539	3.823	1.389	2.016	1.724
5.597	8.176	2.818	3.917	5.549	1.245	1.237	1.708	0.846
5.780	5.644	3.337	3.978	3.530	1.668	1.041	0.898	1.000
5.452	3.255	4.845	3.546	1.592	3.303	1.493	0.805	1.580
5.600	3.043	5.422	3.982	1.260	3.782	1.239	0.851	1.632
5.486	6.687	4.740	3.549	2.856	2.711	1.499	1.571	1.133
5.494	3.068	5.422	3.592	1.266	3.782	1.467	0.751	1.579
5.813	9.772	4.705	4.046	4.846	3.182	1.256	1.806	1.333
5.505	8.391	2.986	3.533	5.336	1.262	1.023	2.216	0.987
5.151	3.229	4.580	3.250	1.599	3.246	1.347	0.929	1.460
5.830	4.391	2.838	4.082	3.544	1.255	1.206	1.126	0.926
5.964	6.974	3.168	3.892	3.046	1.461	1.014	1.358	1.045
5.626	9.772	2.847	3.925	4.846	1.258	1.191	1.997	0.881
5.098	6.691	2.945	3.237	5.625	1.174	1.408	1.964	0.836
5.975	6.896	4.917	3.850	3.107	3.306	0.995	1.381	1.629
5.129	5.644	2.859	3.319	3.530	1.270	1.228	1.018	0.845

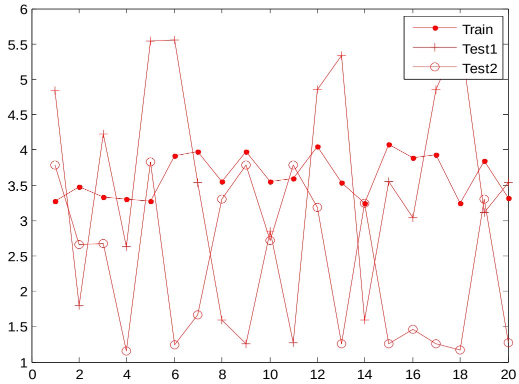
Among testing datasets, the former has much more variability. As mentioned earlier, this implies that the forecasting accuracy of wind speed can be deteriorated in the testing dataset 1. A scatter plot of standard deviation and RMSE is shown in Fig. 10. By this figure, we can see that the larger standard deviation leads to the higher RMSE. In other words, the forecasting accuracy is correlated with the sample variability.

4. Conclusions

SVR is a powerful technique for solving function estimation problems and therefore has a potential for prediction applications. This paper proposes to use SVR for the short term wind speed forecasting. In particular, by incorporating a PSO technique with SVR, we attempt to find the best SVR model and to improve the forecasting accuracy which is measured by RMSE. The presented approach is illustrated by real-world dataset. According to the illustration, PSO is helpful to find the best hyper-parameter setting of SVR. The presented method focuses on the single-step forecasting of wind speed. Therefore, for the purpose of practical applications, it should be extended to the multi-step forecasting.



(a) Means



(b) Standard deviations

Fig. 9. Means and standard deviations of all datasets

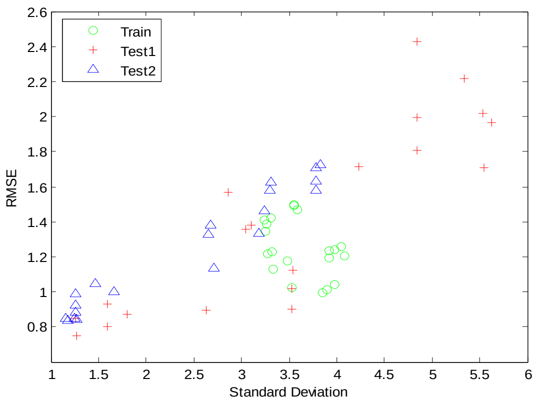


Fig. 10. Dataset variability and forecasting accuracy

By using numerical experiments, we demonstrate that data variability affects the predictive accuracy of wind speed. Such variability should be accommodated in the process of hyper-parameter tuning. This indicates that other additional features should be extracted from dataset and accommodated in the process of hyper-parameter tuning. Although not included in this paper, adopting a wavelet based de-noising as a preprocessing step for wind speed forecasting would be fruitful for future research.

References

[1] J. Catalao, H. Pousinho and V. M. F. Mendes, “Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal,” *IEEE Transactions on Sustainable Energy*, vol. 2, pp. 50-59, 2011.

[2] J. Zhou, J. Shi and G. Li, “Fine tuning support vector machines for short-term wind speed forecasting,” *Energy Conversion and Management*, vol. 52, pp. 1990-1998, 2011.

[3] H. Liu, H. Q. Tian, C. Chen and Y. F. Li, “A hybrid statistical method to predict wind speed and wind power,” *Renewable Energy*, vol. 35, pp. 1857-1861, 2010.

[4] M. C. Mabel and E. Fernandez, "Analysis of wind power generation and prediction using ANN: A case study," *Renewable Energy*, vol. 33, pp. 986-992, 2008.

[5] M. Monfared, H. Rastegar and H. M. Kojabadi, "A new strategy for wind speed forecasting using artificial intelligent methods," *Renewable Energy*, vol. 34, pp. 845-848, 2009.

[6] M. A. Mohandes, T. O. Halawani, S. Rehman and A. A. Hussain, "Support vector machines for wind speed prediction," *Renewable Energy*, vol. 29, pp. 939-947, 2004.

[7] S. Salcedo-Sanz, E. G. Ortiz-Garcia, A. M. Perez-Bellido, A. Portilla-Figueras, and L. Prieto, "Short term wind speed prediction based on evolutionary support vector regression algorithms," *Expert Systems with Applications*, vol. 38, pp. 4052-4057, 2011.

[8] D. S. Moon and S. H. Kim, "A study on wind speed estimation and maximum power point tracking scheme for wind turbine system," *Journal of Korean Institute of Intelligent Systems*, vol. 20, pp. 852-857, 2010.

[9] H. Song, R. D. Dosano, and B. Lee, "Power system voltage stability classification using interior point method based support vector machine," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 9, pp. 238-243, 2009.

[10] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of IEEE Conference on Neural Networks*, vol. 4, pp. 1942-1948, 1995.

[11] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1990.

[12] F. J. Martinez-de-Pison, C. Barreto, A. Pernia and F. Alba, "Modelling of an elastomer profile extrusion process using support vector machines," *Journal of Materials Processing Technology*, vol. 197, pp. 161-169,

2008.
[13] C. C. Chang, and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1-27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Seong-Jun Kim

Professor of Industrial Engineering, Gangneung-Wonju National University

Ph.D. in Industrial Engineering, KAIST, Korea, 1995

Research Area : Computational Intelligence, Soft Sensing, Applied Statistics
Email: sjkim@gwnu.ac.kr

In-Yong Seo

Principal Researcher of KEPCO Research Institute

Ph.D. in Electrical Engineering, Brown University, United States, 2003

Research Area : Smart Power Grid, Distribution Automation, Complex System Modeling
Email: iyseo@kepri.re.kr