

Genetic Driven Multi-Relevance Vector Regression Forecasting of Hourly Wind Speed in Smart Power Systems

Miltiadis Alamaniotis
Applied Intelligent Systems Laboratory
Purdue University
West Lafayette, IN, USA
malamani@ecn.purdue.edu

Georgios Karagiannis
Department of Mathematical Sciences
Durham University
Durham, United Kingdom
georgios.karagiannis@stat.durham.uk

Abstract—Implementation of smart power systems implicitly demands the extensive use of learning. Learning methods may be used for forecasting of several critical parameter values of the power system infrastructure that will lead to efficient management of energy distribution and utilization. This paper places itself in the area of integrating renewable sources in the power grid. In particular the objective is to develop an intelligent data driven method for forecasting the average hourly wind speed. Wind speed is crucial in predicting the power output coming from wind generators. The proposed methodology, implements a two-stage learning method: in the first stage, it employs three relevance vector regressors (RVRs) that learn from observations and make predictions, while in the second stage a genetic algorithm is employed to learn and assimilate the discrepancies of predictions with the most recent observations. The proposed method is applied to a set of real world wind speed hourly data, while benchmarked against learning Gaussian processes. Results exhibit the superiority of the multi-kernel method in forecasting wind speed.

Index Terms—Wind speed forecasting, RVR, genetic algorithm, optimization problem.

I. INTRODUCTION

Smart power systems refer to the integration of advanced information technologies with power systems. This integration aims at ensuring the reliable, safe and continuous flow of power from the generation units to the load centers [1]. The cornerstone of smart power systems is forecasting the values of critical variables prior to any action-taking.

The machine learning library offers the necessary tools for developing intelligent forecasting tools. In particular, intelligent tools are capable of assimilating observed data, learn from observations, and provide accurate forecasts [2]. In addition, such intelligent tools, which are data-driven, are able to capture the latest trends in data and avoid explicitly modeling of the complex physical processes.

Wind speed is a crucial parameter in integration of renewable sources to the power grid [3]. Inevitable, it has been also marked as a crucial parameter in developing smart power systems. Forecasting of hourly wind speed allows planning and the ahead of time scheduling of wind energy generation and utilization. Furthermore, it helps market operators to provide incentives consumers to shift their consumption to time slots where wind energy is available.

Hourly wind speed forecasting has been a topic of study and various methods have been proposed. The majority of proposed methods adopts tools from machine learning and statistics. In [4], three different types of neural networks have been applied for hourly wind speed prediction, while neural networks have been also studied and tested in [5]. In [6], authors utilize support vector machines. Furthermore, a hybrid approach that utilized empirical mode decomposition and neural networks is presented in [7], and the synergism of neural network with autoregressive integrated moving average (ARIMA) in [8]. Other hybrid methods contain the use of Bayesian statistics with neural networks as discussed in [9], neuro-fuzzy approaches [10], and the integration of neural networks with wavelets [11]. In addition, a fuzzy logic approach is proposed in [12], while time series methods such as autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) have been proposed in wind speed forecasting in [13] and [14]. Kernel density estimators, and dynamic regression models are presented in short term wind speed forecasting in [15] and [16] respectively. It is evident that wind speed forecasting gains popularity and more sophisticated methods are under development. To this point, the proposed methods do not explicitly show an adequate degree of sophistication in capturing the wind dynamics.

In this paper, a new method is proposed for hourly wind speed forecasting. The proposed method adopts a set of intelligent tools and more particularly a set of relevance vector regressors (RVR) [17]. Each regressor is equipped with a

different kernel function where a kernel models different data properties. The set of RVRs are put together to form a linear ensemble forecaster allowing each regressor to contribute to the final forecast. The contributions are evaluated using a single objective problem whose solution is identified using genetic algorithms [18]. The above schema allows capturing of the various wind dynamics through modeling of various data properties.

The rest of the paper is organized as follows. Section II briefly presents RVR, while section III describes the presented method. Section IV provides the obtained results on a real-world dataset, and lastly, section V concludes the paper.

II. RELEVANCE VECTOR REGRESSION

The machine learning library contains all those models that learn from data and are able to perform tasks such as prediction and classification [19]. Among the machine learning models, there is the set of “kernel machines”, which are models that employ the “kernel trick” [19]. Kernel trick includes those models who incorporate a kernel function, which is any valid analytical function that is cast into the dual representation given below:

$$k(x_1, x_2) = \phi(x_1)^T \phi(x_2) \quad (1)$$

where $\phi(x)$ is also an analytical function called the basis function. A kernel function (or simply kernel) expresses the properties among the data. Thus, the kernel trick allows the modeler to select the form of the kernel and subsequently control the output of the kernel machine [19].

Relevance vector machines (RVM) is a learning model that belongs to kernel machines and is employed for classification and regression problems [17]. In the latter problems, it is simply called relevance vector regression. RVR follows the basic simple linear regression model [19]:

$$y(x) = \sum_{n=1}^N w_n k(x, x_n) + w_b \quad (2)$$

where, $y(x)$ is the predicted output for input x , w_n are the regression weights, x_n is the training dataset and the regression factors consist of kernel functions, i.e., $k(x, x_n)$. It is assumed that the regression weights in (2) follow a normal distribution defined by [19]:

$$p(\mathbf{w} | \boldsymbol{\alpha}) = \prod_{n=1}^{N+1} N(w_n | 0, \alpha_n) \quad (3)$$

where α_n is the variance of weight w_n , while \mathbf{w} and $\boldsymbol{\alpha}$ consolidate in vector form the weights and the respective variances.

Deriving the RVR model assumes adoption of a posterior probability over the weights \mathbf{w} . The posterior probability is also taken as a normal distribution given by

$$p(\mathbf{w} | \mathbf{t}, \mathbf{X}, \boldsymbol{\alpha}, \sigma^2) = N(\mathbf{w} | \mathbf{m}, \Sigma) \quad (4)$$

with $\boldsymbol{\alpha}$ the vector unknown parameters, \mathbf{m} , Σ are the mean and covariance matrix respectively evaluated by the kernel in the training phase. Next, the logarithmic marginal likelihood function of (4) is formed and the maximum likelihood method is taken using an iterative algorithm such as the Expectation-Maximization (EM) algorithm [19]. The iterative algorithm will provide a solution where some of the regression coefficients are set equal to zero; therefore, the respective kernel contributions drop to zero as well. The kernels of non-zero contribution are called “relevance vectors”, and the RVR is formulated as a predictive distribution formula given by

$$p\left(t | \mathbf{x}, \mathbf{X}, \mathbf{t}, a^*, \frac{1}{(\sigma^2)^*}\right) = N\left(t | \mathbf{m}^T \phi(\mathbf{x}), \sigma^2(\mathbf{x})\right) \quad (5)$$

where the mean and variance values are respectively taken by

$$\mathbf{m}^T \phi(\mathbf{x}) = \left(\frac{1}{(\sigma^2)^*} \boldsymbol{\Sigma} \Phi^T \mathbf{t} \right) \phi(\mathbf{x}) \quad (6)$$

$$\sigma^2(\mathbf{x}) = \left(\frac{1}{(\sigma^2)^*} \right)^{-1} + \phi(\mathbf{x})^T \boldsymbol{\Sigma} \phi(\mathbf{x}) \quad (7)$$

where $(\sigma^2)^*$ are the optimal values computed by EM (or any other iterative algorithm employed for solution seeking).

A detailed derivation of RVR (and RVM as well) is given in [17] and [19]. At this point, it should be noted that in practical problems selection of kernel functions is based upon modelers experience and intuition – there is not a dominant kernel selection method.

III. WIND SPEED FORECASTING METHOD

The underlying idea behind the hourly wind speed forecasting in the current study is the use of an linear ensemble of RVRs, and the subsequent optimization of the ensemble using a genetic algorithm [20,21]. The block diagram of the proposed methodology is depicted in Fig. 1, where its individual steps are shown.

Initially, a set of three RVR models is determined. The three RVR are equipped with different kernel functions and more specifically with the i) Gaussian, ii) Polynomial and iii) Spline kernel respectively. The analytical forms of the three kernels are given below respectively:

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (8)$$

where σ^2 is a hyperparameter evaluated in the training phase.

$$k(x_1, x_2) = \left(ax_1^T x_2 + c \right)^d \quad (9)$$

where the slope α , the constant term c , and the polynomial degree d are parameters determined in the training.

$$k(\mathbf{x}_1, \mathbf{x}_2) = 1 + \mathbf{x}_1 \mathbf{x}_2 + \mathbf{x}_1 \mathbf{x}_2 \min(\mathbf{x}_1, \mathbf{x}_2) - \frac{\mathbf{x}_1 + \mathbf{x}_2}{2} (\min(\mathbf{x}_1, \mathbf{x}_2))^2 + \frac{1}{3} (\min(\mathbf{x}_1, \mathbf{x}_2))^3 \quad (10)$$

that has no adjustable parameters.

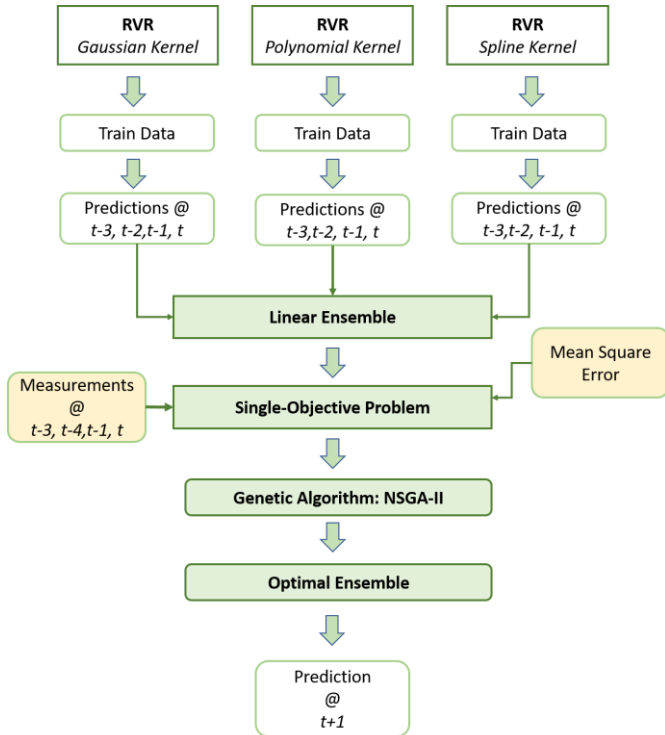


Figure 1. Block diagram of the proposed wind speed forecasting method.

The RVR models are fed with a set of known data, i.e., training datasets, and then are utilized for prediction making. By assuming that the current time is t , then prediction making is performed for time points $t-3$, $t-2$, $t-1$, and t . The rationale of predicting the wind speed at previous points is that we can compare those prediction with the available wind speed measurements (i.e., the real wind speed has been observed until the current time t). In particular, this comparison will allow the assessment of the prediction performance of each RVR model in the most recent measurements. Therefore, by assessing the RVR performance in the most recent measurements, we are able to capture the most recent wind speed dynamics.

In the next step, a linear ensemble of the three RVRs is formed as given below:

$$E(t) = \alpha_G G_t + \alpha_P P_t + \alpha_S S_t \quad (11)$$

where $E(t)$ is the ensemble value at time t , while G_t , P_t , S_t are the individual RVR values and α_G , α_P , α_S the respective ensemble coefficients for Gaussian, Polynomial and Spline kernels. Respectively, the variance associated with the ensemble is obtained by:

$$V_e(t) = \alpha_G^2 G_t + \alpha_P^2 P_t + \alpha_S^2 S_t \quad (12)$$

where the the coefficients and the RVR values are the same as in Eq. (11). It should be noted that the unknown parameters in the ensemble are the coefficients α_G , α_P , α_S . Evaluation of the coefficients is performed by formulating and solving a single-objective optimization problem.

To that end, a single performance evaluation measure is adopted as the problem objective function, namely the mean square error (MSE). The analytical formula of the objective function is given by [22]:

$$MSE = \frac{1}{M} \sum_{m=1}^M (E_m - R_m)^2 \quad (13)$$

where M is the number of samples, E_m is the ensemble value and R_m is the real value at time m . Therefore, by utilizing the individual predictions and the measure function a single-objective problem is formulated:

$$\begin{aligned} & \text{minimize} [MSE] \\ & \text{subject to } 0 \leq \alpha_G, \alpha_P, \alpha_S \leq 1 \\ & \text{where } M = 4 \end{aligned} \quad (14)$$

with the box constraints determining that the ensemble coefficients lay in the interval $[0, 1]$.

A solution to the above single-objective optimization problem is sought by a genetic algorithm. In particular, the algorithm adopted is the Non-Dominated Sorting Genetic Algorithm – II (NSGA-II) [23], which is able to identify a global optimal solution [21]. NSGA-II guarantees an optimal solution in the form of a three-entry vector where each entry is a value for the ensemble coefficient will be identified.

In the last step, the optimal ensemble is utilized for prediction making of the speed value in the next hour. In addition, the wind speed of the variance associated with the predicted value is also computed. To make it clearer, the computed optimal coefficients are input to Eq. (11) and (12) and subsequently the wind speed and its variance are found.

IV. RESULTS

In this section, the proposed method is tested on a set of real world dataset. The speed data have been taken from the National Renewable Energy Laboratory (NREL) Observed

Atmospheric and Solar Information System (OASIS) [24]. The data have been recorded as the hourly wind speed in the form of m/s and span the period of December 31, 2016 to January 16, 2017. The method is benchmarked against the individual RVR models as well as against another kernel machine, and more specifically against the Learning Gaussian Process (GP) model equipped with Gaussian kernel [2,19]. The GP model is also trained on the same datasets.

The initial training dataset is comprised of the 24 values of the day December 31, 2016. Then, every hour the training data shifts by one hour, i.e., by including the most recent measurement and removing the oldest one, hence, keeping the number of measurements at the number of 24 at each instance.

The obtained results are recorded with respect to mean square error and the variance, while are grouped with respect to daily performance. Table I provides the obtained results with respect to prediction accuracy with respect to MSE, while Table II provides the mean predicted variance spanning the whole tested period.

TABLE I. WIND SPEED FORECASTING RESULTS

Day Year 2017	Mean Square Error (MSE)				
	RVR Gaussian	RVR Polynomial	RVR Spline	Ensemble	GP Gaussian
Jan 1	5.3610	3.3705	3.3061	2.96	3.2278
Jan 2	1.4019	1.1178	1.1190	0.9986	0.9930
Jan 3	1.4061	1.4323	1.3723	1.1996	3.4183
Jan 4	1.0629	1.2427	1.2459	0.8185	2.5974
Jan 5	2.3628	3.3515	3.3185	2.2440	2.6757
Jan 6	1.1515	1.2985	1.2295	0.7477	4.4388
Jan 7	2.4088	1.8215	1.6032	1.1158	2.5706
Jan 8	0.8689	1.1724	1.1628	0.7942	3.5819
Jan 9	2.9338	3.0779	3.3819	1.2651	6.3367
Jan 10	4.0200	3.9396	3.9879	0.8625	5.2828
Jan 11	3.2309	3.2301	3.2889	2.7124	7.3023
Jan 12	1.5058	1.2968	1.3257	0.4645	1.5808
Jan 13	0.7776	0.9812	1.0014	0.3945	3.4640
Jan 14	1.9912	2.3138	2.3020	1.9808	3.9875
Jan 15	1.7828	1.2242	1.3827	1.1043	2.8951
Jan 16	4.3299	3.7886	3.9418	1.3081	5.8776

TABLE II. MEAN AVERAGE PREDICTED VARIANCE FOR THE TESTED TIME PERIOD

Time Period	Mean Predicted Variance				
	RVR Gaussian	RVR Polynomial	RVR Spline	Ensemble	GP Gaussian
Jan 1- 16	0.8026	0.0808	0.0759	0.0483	1.225

We observe in Table I that the proposed ensemble method is a very powerful method that provides the lowest error in all the tested days. It does not always provide clearly the lowest MSE that far away from the rest but there are instances that is marginally more accurate than the other methods. Such an example is wind speed prediction for day January 2, 2017, where the presented ensemble provides slightly lower MSE than the GP-Gaussian kernel model. Therefore, the domination of the ensemble in all cases, exhibits the validation of our method in weighting the contributions accordingly and obtain a final forecast using the ensemble. In addition, we observe in Table II that the proposed method also provided the lowest mean variance among all tested methods. This does not come as a surprise given that the variance also is determined by the optimal ensemble, by weighting the individual RVR variances accordingly. For visualization purposes, Fig. 2 and 3 show the forecasted values taken with the optimal ensemble and the individual RVR models against the real values for the days of January 2 and January 8.

Furthermore, we observe that the proposed method clearly outperformed the GP model with respect to variance as well; it provided much lower mean variance than GP. Hence, we conclude that the proposed method is not only more accurate than the rest forecasters but also more precise.

With regard to the rest forecasters, i.e., the individual RVR and the GP, there is no a single model that performs better than the rest in all cases. Though the RVR with Gaussian kernel provides the lowest error in many cases, it does not consistently perform better than the rest. This observation supports the statement that we do not know a priori which kernel will be the best performer in our forecasting. With regard to GP forecaster, with a few exceptions - such as the days of January 1 and 2 - it is slightly worse than the individual RVR forecasters in the majority of the tested cases. Overall, given that the RVR ensemble consistently provides low error, then the RVR ensemble method is clearly superior over the individual RVR forecasters.

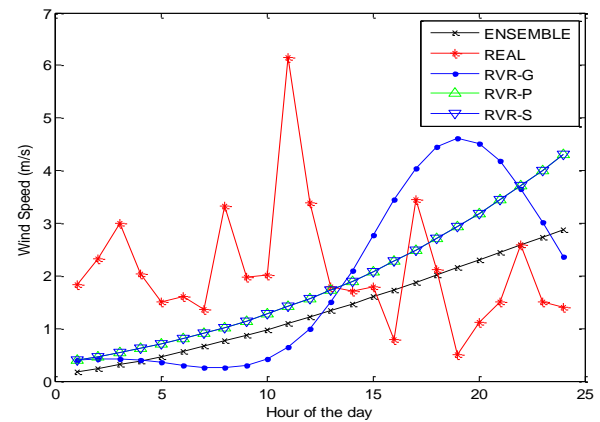


Figure 2. Ensemble and individual RVR forecasted against real wind speed values for the tested day of January 1.

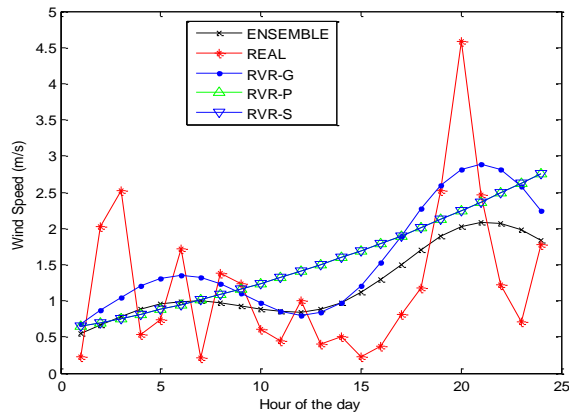


Figure 3. Ensemble individual RVR forecasted against real wind speed values for the tested day of January 8.

V. CONCLUSION

A new methodology for wind speed forecasting that is applicable in developing smart power systems is discussed and in the current manuscript. The presented methodology that integrates a set of three kernel modeled RVR models equipped with three different kernels with genetic algorithm optimization is tested on a set of real wind speed data. Results exhibit the powerfulness of the methodology in predicting the hourly wind speed, while proved that the use of Genetic algorithms in a two-stage process indeed improved the individual predictions. In addition, the presented method outperformed another kernel machine forecasting and in particular the learning GP equipped with Gaussian kernel.

Future work will move to two directions: i) testing of a higher number of kernels beyond the three ones presented in this work, and ii) extensive testing in a larger dataset with more datasets taken along the year to test the performance of the proposed method. Furthermore, comparison with other optimization methods will be also planned.

REFERENCES

- [1] L.H. Tsoukalas, and R. Gao, "From smart grids to an energy internet: Assumptions, architectures and requirements," In *Proc. 2008 IEEE Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT 2008)*, pp. 94-98.
- [2] M. Alamaniotis, A. Ikonopoulou, and L.H. Tsoukalas, "Evolutionary Multiobjective Optimization of Kernel-based Very Short-Term Load Forecasting," *IEEE Transactions on Power Systems*, vol. 27(3), August 2012, pp. 1477-1484.
- [3] M. Lei, L. Shiyang, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews*, vol. 13(4), May 2009, pp. 915-920.
- [4] G. Li, and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Applied Energy*, vol. 87(7), July 2010, pp. 2313-2320.
- [5] A. Daraeepour, and D.P. Echeverri, "Day-ahead wind speed prediction by a Neural Network-based model," In *Proc. 2014 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1-5, February 2014.
- [6] J. Zhou, J. Shi, and G. Li, "Fine tuning support vector machines for short-term wind speed forecasting," *Energy Conversion and Management*, vol. 52(4), 2011, pp. 1990-1998.

- [7] Z. Guo, W. Zhao, H. Lu, and L. Wang, "Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model," *Renewable Energy*, vol. 37(1), 2012, pp. 241-249.
- [8] E. Cadenas, and W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model," *Renewable Energy*, vol. 35(12), 2010, pp. 2732-2738.
- [9] Li G, Shi J, Zhou J. Bayesian adaptive combination of short-term wind speed forecasts from neural network models. *Renewable Energy*, vol. 36(1), pp. 352-359, 2011.
- [10] F. Castellanos, and N. James, "Average hourly wind speed forecasting with ANFIS," In *Proc. 11th American Conference on Wind Engineering*, pp. 1-8, June 2009.
- [11] H. Liu, T.Q. Tian, D.F. Pan, Y.F. Li, "Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks," *Applied Energy*, vol. 107, pp. 191-208, 2013.
- [12] X. Wang, G. Sideratos, N. Hatzigiorgiou, and L.H. Tsoukalas, "Wind speed forecasting for power system operational planning," In *Proc. IEEE International Conference on Probabilistic Methods Applied to Power Systems*, pp. 470-474, September 2004.
- [13] E. Erdem, and J. Shi, "ARMA based approaches for forecasting the tuple of wind speed and direction," *Applied Energy*, vol. 88(4), pp. 1405-1414, 2011.
- [14] Colak, I., Sagioglu, S., Yesilbudak, M., Kabalci, E. and Bulbul, H.I., 2015, November. "Multi-time series and-time scale modeling for wind speed and wind power forecasting part I: Statistical methods, very short-term and short-term applications," In *Proc. 2015 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, pp. 209-214.
- [15] J. Zhou, J. Hongjie, Z. Tian, and L. Hu, "Short-term wind speed forecasting based on non-parametric kernel density estimation," In *Proc. IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, pp. 1-5, May 2012.
- [16] B. Banerjee, D. Jayaweera, and S.M. Islam, 2011, November. "Wind power forecasting with dynamic regression models and ageing considerations," In *Proc. IEEE PES Innovative Smart Grid Technologies Asia (ISGT)*, pp. 1-7, Nov. 2011.
- [17] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The journal of machine learning research*, vol. 1, pp. 211-244, 2001.
- [18] A. Gambier, "MPC and PID Control Based on Multi-Objective Optimization," in *Proc. of American Control Conf.*, pp. 4727-4732, 2008.
- [19] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer, 2006.
- [20] R. T. Marler, and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidisc. Optim.*, vol. 26, pp. 369-395, 2004.
- [21] M. Alamaniotis, N. Bourbakis, and L.H. Tsoukalas, "Very-Short Term Forecasting of Electricity Price Signals Using a Pareto Composition of Kernel Machines in Smart Power Systems," In *Proc. 3rd IEEE Global Conference on Signal and Information Processing*, pp. 780-784, December 2015.
- [22] M. Alamaniotis, D. Bargiotas, N. Bourbakis, and L.H. Tsoukalas, "Genetic Optimal Regression of Relevance Vector Machines for Electricity Price Forecasting in Smart Grids," *IEEE Transactions on Smart Grid*, vol. 6(6), pp. 2997-3005, November 2005.
- [23] K. Deb, A. Pratap, S. Agarwal, and T.A.M.T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation*, vol. 6(2), pp.182-197, 2002.
- [24] A. Andreas, and S. Wilcox, "Observed Atmospheric and Solar Information System (OASIS)," Tucson, Arizona (Data), NREL Report No. DA-5500-56494, <http://dx.doi.org/10.5439/1052226>.