EasyChair Preprint

№ 6094

Super Local Models for Wind Power Detection

María Barroso and Ángela Fernández

July 16, 2021

Super Local Models for Wind Power Detection

María Barroso and Ángela Fernández[0000−0002−6602−3753]

Dept. Ingeniería Informática, Universidad Autónoma de Madrid
maria.barrosoh@estudiante.uam.es, a.fernandez@uam.es

Abstract. Local models can be a useful, necessary tool when dealing with problems with high variance. In particular, wind power forecasting can be benefited from this approach. In this work, we propose a local regression method that defines a particular model for each point of the test set based on its neighborhood. Applying this approach for wind energy prediction, and especially for linear methods, we achieve accurate models that are not dominated by low wind samples, and that implies an improvement also in computational terms. Moreover it will be shown that using linear models allows interpretability, gaining insight on the tackled problem.

Keywords: Local Models · Regression · Wind Power Forecasting.

1 Introduction

Local models are a natural option when facing processes with high variance. This is the case when we are dealing with renewable energy power predictions, and in particular with wind energy forecasting. Wind power presents wide, fast changing fluctuations, specially at farm level, making difficult to obtain accurate predictions specially for medium or long term horizons. Nevertheless, nowadays obtaining a good prediction in this context is crucial due to the importance of this kind of energies in the power market, so it is necessary to know in advance how much wind energy will be injected into the power grid, for minimizing operating system costs and maximizing benefits of the electric market agents.

Analyzing more in detail wind power properties, we will check that, as expected, wind speed has a great influence on it. It is well-known that the frequencies of this variable follow a Weibull distribution, that is, a stretched exponential where large frequencies match the lowest wind values, and hence these low wind values are dominating in some sense the distribution. This fact, added to the sigmoid-like structure of wind turbine power curves that clearly shows different regimes at low, medium, and high wind speeds, made local models an attractive option when predicting wind energy [4, 12].

This general situation can be applied to Spain, one of the main producers of wind energy in the World, with 27GW of installed wind-generation capacity that covered a 21% of the Spanish electric consumption in 2020 [1]. Moreover, it is expected to be grown up to 50GW extra installed capacity in 2030, covering the 42% of energy consumption by just using renewable energies [3].

In this paper we are going to tackle the wind power forecasting problem for a Spanish wind farm since a local perspective, proposing a new, model-agnostic algorithm. We will see how the proposed local approximation together with a liner model, appears as a good competitor for a Support Vector Regression (SVR) model, state of the art for this particular problem. We will see that, for this particular problem, the proposed Super Local Linear Model outperforms SVRs in computational terms and it has the advantage of been easily interpretable.

This paper is organized as follows: in Section 2 we will briefly review the state of the art in local models and wind power prediction, and in Section 3 we will present our proposed local algorithm. In Section 4 we show the different experiments that we have carried out and, finally, in Section 5 we offer some conclusions and point out to lines of further work.

2 Local Models and Wind Power Forecasting

In this section we will briefly review the different methodologies and models applied in the state-of-the-art literature for wind energy prediction. When facing this problem we can find two different basic approximations: using physical or statistical models. Physical models used physics knowledge to define formulas that give us an approximation to the wind energy production at some horizon [10]. By contrast, the statistical approach does not need any knowledge about the system, but it will try to find, in an automatic manner, relations between some features and the predictive target. We will focus our attention on the latter models where, for the problem we are tackling, the features are always numerical weather predictions (NWP) and the target is given by the wind power.

Reviewing the works done in the context of statistical models, we find a wide variety of methods dealing with wind energy forecasting from different points of view. On one side, considering the data as a temporal series, models like ARMA [8], ARIMA [6] or ARX [7] has been defined. This kind of models used to provide good results for short term forecasting, but for horizons longer than 48-72 hours the predictions are weaker [11]. On the other side, traditional machine learning models are also frequent in the state-of-the-art prediction of wind power. In particular, Artificial Neural Networks (ANNs) [15, 17] and Support Vector Machines (SVMs) [13] repeatedly appear in the literature, obtaining very good results.

Finally, as previously mentioned, due to its intrinsic properties, wind power forecasting can be correctly modeled following a local approach. For example, in [16] local models based on ANN, Adaptive Neuro-Fuzzy Inference System (ANFIS), or Least Squares Support Vector Machine (LS-SVM) are applied, improving their global versions. Another example can be found in [4] where a specific model for high wind powers is built, defined in terms of the own productions.

Lastly, it should be mentioned that the traditional metric in the energy context is the Normalized Mean Absolute Error (NMAE), defined as:

$$NMAE = \frac{1}{N} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{pc} ,$$

where y_i represents each real output (the wind energy), $\hat{y}_i$ its corresponding prediction for a sample with N points and pc represents the installed capacity of the wind farm.

3 Super Local Models

As mentioned before, the idea of building local models is to construct a model for a particular region based on the characteristics of that subset of points, i.e., just using similar points. We will proposed in this section a local algorithm following the idea of local regression models [5]. Local regression (LOESS, or LOWESS) is a well-known method in Statistics, that generalizes the concepts of moving averages and polynomial regression. It consists on defining a linear regression function around a point of interest to model the data, but just using "local" interesting information, i.e. its neighborhood. The LOESS model allows to weight the importance of each neighbor using a kernel function centered at the point in question (LOWESS).

Following this reasoning we define a new local algorithm that can be applied together with any machine learning method. The main idea is to train a different model for each new, unseen pattern, so its training will use just points with similar features to the ones of the point that we want to predict. We will call *Super Local Models* to these individual local models. The neighborhood used for training each model will be defined by the K training nearest neighbors of each test data point, in terms of the Euclidean distance. The most challenging part will be to find the optimal hyperparameters for each model. In this work we will consider a global model built over the whole training dataset, and we will hyperparametrized it using cross validation in this global context. We have considered that the optimal hyperparameter values for the global model will be a good approximation for each super local model hyperparameter.

Validation, training and prediction processes of the proposed method are presented in Algorithm 1. Apart from the input dataset, the number K of nearest neighbors considered is needed to train each particular model. The selection of an optimal K is another challenge that will be addressed in further work.

Looking at the steps in the algorithm, firstly the dataset is divided into training and testing sets to evaluate the performance of the super local model. Second, training and test regressor variables are standardized using the standard deviation of the training set, such that the fitted model does not depend on the scale at which the variables were measured. Next, a validation process is launched over the entire training dataset to obtain the parameters that reduce the overall error model. These parameter values will be used to train each particular model on test data over the training selected neighborhood. Predictions of each model are computed and finally the model is evaluated using some error measure.

Algorithm 1: Super Local Models algorithm

input : $\{X, Y\}$ – Dataset
K - Number of nearest neighbors
output: $\{X_{test}, Y_{pred}\}$ – Prediction dataset

$X_{train}, X_{test}, Y_{train}, Y_{test} \leftarrow$ split(X, Y);
standardScaler(X_{train},X_{test});
validation(X_{train}, Y_{train});
for (x_{test}, y_{test}) *in* (X_{test}, Y_{test}) **do**
  $K_x, K_y \leftarrow$KNN(X_{train}, Y_{train}).kneighbors(x_{test}, K);
  fit(K_x, K_y);
  $Y_{pred} \leftarrow$ predict(x_{test});
end
Error $\leftarrow$ evaluation(Y_{test}, Y_{pred});

4 Experiments

This section presents the description of the experimental setup, results, and analysis. First, a brief description of used data and their preprocessing is presented. Next, super local regression models based on a linear model, in this case Ridge Regression (RR, [9]), and on a non-linear model, for instance Support Vector Regression (SVR, [14]), are tested in order to prove their efficiency and viability. Results will be compared with those obtained by their corresponding global models. Also, in the linear model, data descriptive statistics of more correlated regressor variables with the target are shown. Finally, a study about computational capability required by each of the models is presented. *scikit-learn* implementations for RR and SVR methods are used.

4.1 Dataset description

Production data are from Sotavento Wind Farm in Northwestern Spain, collected during 2016, 2017 and 2018, and they consist of hourly mean powers measured in kWh. In order to predict hourly energy production in this farm, NWP variables from European Centre for Medium-Range Weather Forecasts (ECMWF) [2] will be used as predictors. In particular, we will use the next NWP variables:

- Surface pressure (sp).
- 2 m temperature (2t).
- Eastward component of the wind at 10 m (10u).
- Northward component of the wind at 10 m (10v).
- Module of velocity of the wind at 10 m (10Vel).
- Eastward component of the wind at 100 m (100u).
- Northward component of the wind at 100 m (100v).
- Module of velocity of the wind at 100 m (100Vel).

A grid of 0.125° resolution, approximately centered at the farm, is used with southwest and northeast coordinates being (43.25°, −8°) and (43.5°,−7.75°), respectively. This results in a 9 point grid and, using 8 different variables at each point, gives a total of 72 predictive features. We use years 2016 and 2017 for training, and 2018 for testing purposes. In this way, training and test sets dimensions are 17 544 x 72 and 8760 x 72, respectively. Moreover, as previously mentioned, all the regressor variables are standardized using the training set.

4.2 Ridge Regression models

Our first experiment is focused on evaluating the super local regression model based on Ridge Regression (RR) using K different nearest neighbors. Later, a statistical analysis of the selected neighbors will be done for those variables more correlated with the final power.

The validation process of the proposed model was performed through a 5-fold cross validation procedure in the whole training set using a grid search over $\{10^k : -1 \leq k \leq 3\}$ for optimizing the regularization parameter λ. This validation process resulted in an optimal value of $\lambda = 100$; recall that a high value of λ implies a reduction in variance at the cost of an increase in bias. Now, this λ will be used for training each particular local model based on the nearest neighbors of the test data.

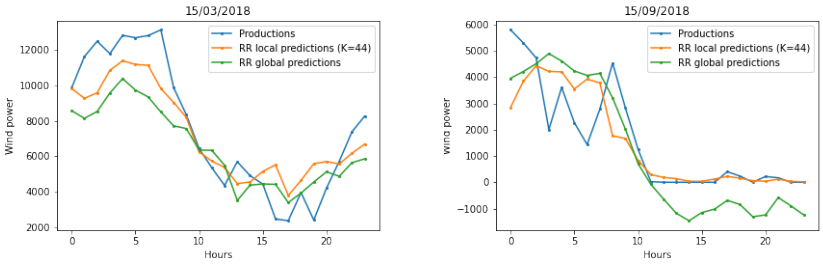
Employing NMAE as metric for evaluating, a global RR model with $\lambda = 100$ on the test set obtains an error equal to 8.44%. This error is far from the state-of-the-art errors for wind power forecasting in Sotavento [13]; thus one of the aims of the experiment is to improve it using the proposed Super Local Models.

As previously said, K-NN is used to select the subset of training data to fit the RR model on. These K values were chosen by rounding different percentages over the 17 544 total training data. Table 1 shows the percentage of testing error made, for each selected K, when predictions of the 2018 data are evaluated using NMAE as indicator.

Table 1: Local RR NMAE for different number of neighbors K.

	0.1%	0.25%	0.5%	1%	5%	10%	25%	50%	75%
K	17	44	87	175	877	1754	4386	8772	13 158
NMAE (%)	6.92	**6.83**	6.88	6.94	7.15	7.26	7.53	7.95	8.22

Looking at NMAE results obtained for each K, certain conclusions can be drawn, such as the fact that a larger number of neighbors implies a larger error in the predictions. Moreover, the bigger the error, the closer to the global model, as expected. This result makes sense since the regularization parameter



(a) 15-03-2018 (b) 15-09-2018

Fig. 1: Real productions and local and global RR predictions.

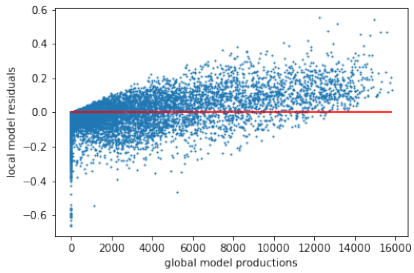


Fig. 2: Local model residual plot.

λ used is fixed. On the other hand, the most accurate local model is obtained by selecting 0.25% of training data, i.e. 44 neighbors, making an error equal to 6.83%. Therefore, we have got that a linear model as RR is a competitive model in wind power forecasting by making it local.

Next, we analyze predictions obtained by the local Ridge Regression model with 44 neighbors and the global model predictions, and we compare them against real productions for some particular days of 2018 with a high or low production profile.

Figure 1 shows, for two particular days, the comparison between local and global forecasts against the real productions. Local predictions seems to be closer to actual powers than the global ones, specially for high wind energy, and both model predictions maintain the trend of the real curve.

Figure 2 shows the residual plot, where it can be appreciated that high productions are infra-estimated by the proposed method and, by contrast, low productions are overestimated. When the power is null, there are a huge number of big errors that are probably due to hours when the machines stopped production for maintenance work. These errors, added to those derived from meteorological forecasting errors, are difficult to avoid in our model and we should have them in mind when analyzing the results.

In order to understand how the model is selecting the K training patterns to define a particular super local model for each test data point, statistics of these neighbors have been also studied for some regressor variables. This analysis is possible because we are working with a linear regression model that allows us to evaluate the effect of each feature, using the correlation matrix. The possibility of this analysis is one of the main advantages of the proposed method.

Modules of velocity of the wind at 10 and 100 m in the 9 coordinates around Sotavento are the predictor variables that have more correlation with the registered power; temperature and pressure are hardly relevant in the global model. By contrast, correlation matrix of individual models did not show any variable more influential than the others. This is because the value of the regularization parameter is high which implies a reduction of the model coefficients.

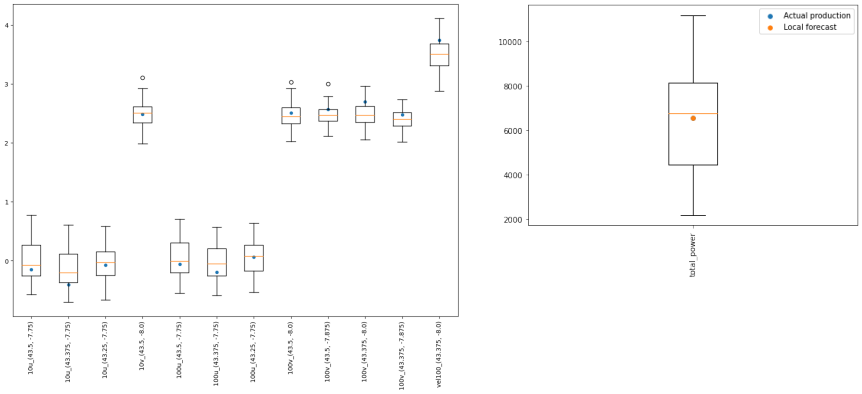
We have selected for this purpose two particular models with high and low residual as specified in Table 2.

Table 2: Residuals, local predictions and productions of the selected models.

Datetime model	**Residuals**	**Prediction**	**Production**
09/01/2018 - 07:00:00	1.41×10^{-6}	6563.40	6563.43
13/02/2018 - 20:00:00	0.55	2573.31	12 255.83

Figure 3a shows, using boxplot diagrams, descriptor statistics of some of the most correlated features with the target for the K neighbors selected. The point to be predicted is represented as a blue point, and it can be seen that it is close to the median values of the sample and that the interquartile range is narrow. The corresponding outputs of these neighbors are represented in Figure 3b, its range goes from 2000 to 12000 kW and it can be seen that the actual production value for this test point is also very near to the median. This model produces, as expected, a good prediction.

In Figure 4a it can be seen the wind speed modules statistics of the selected neighbors for the second model. In this case, the features of the point to be predicted are, in most of the coordinates, also close to the central values of the sample and the corresponding outputs of these neighbors (see Figure 4b) have a narrow interquartile range. In this case, the prediction obtained is right on the median but the actual production reaches a value of 12 255.83 kW, high above the expected value according to its neighbors production. This strange result leads us to think that, for this particular point, there should be an error in the weather forecasts used as features.



(a) Regressor variables (b) Outputs

Fig. 3: Neighbors statistics for 09/01/2018 - 07:00:00 model.

4.3 SVR models

This second experiment is focused on evaluating the performance of an SVR model with an RBF kernel using the proposed super local method.

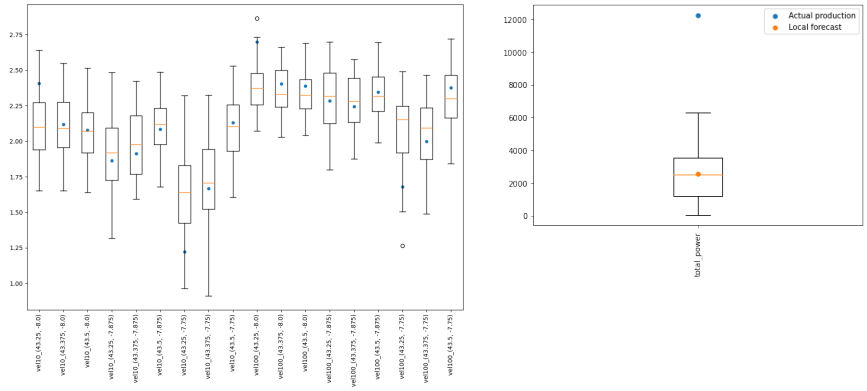
Again, a 5-fold cross validation procedure over the training set using a grid search over the values specified in Table 3 is launched. In the table, d is the number of dimensions of training data and σ is the standard deviation of the target data.

Table 3: SVR hyperparameters and the grid used to find them.

Hyperparam.	**Range**
C	$\{10^k : 3 \leq k \leq 5\}$
ϵ	$\{\sigma/2^k : 2 \leq k \leq 6\}$
γ	$\{4^k/d : -2 \leq k \leq 1\}$

After validating the total training set, the smallest error obtained corresponds to the values $C = 10^4$, $\epsilon = \sigma/2^6$ and $\gamma = 4^{-1}/d$. Notice that the low value of γ indicates that the Gaussian shape in the RBF kernel is very flat.

Evaluating the SVR global model over the 2018 test data, using the hyperparameter values described above, we obtain an NMAE equal to **6.35%**. The evaluation of the super local model approach by selecting different K nearest neighbors over the 17 544 total training data points and applying the same percentage method of the previous experiment, is presented in Table 4.



(a) Regressor variables (b) Outputs

Fig. 4: Neighbors statistics for 13/02/2018 - 20:00:00 model.

Table 4: NMAE for local SVR selecting different number of neighbors K.

	0.1%	0.25%	0.5%	1%	3%	5%	10%	25%	50%
K	17	44	87	175	350	877	1754	4386	8772
NMAE (%)	6.79	6.64	6.60	6.53	6.46	6.37	6.36	6.35	6.34

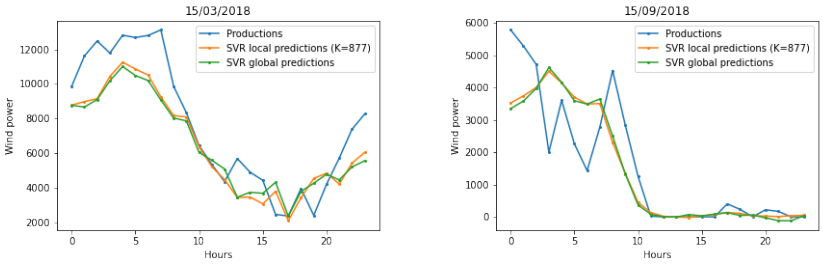
The proposed local model improves as the number of neighbors increases until it stabilizes at $K = 877$, presenting an error similar to the one obtained with the global SVR model. It should be noted that the implicit local character of the SVR model spotlights the neighboring data of each point through the use of the RBF kernel.

Next, the hourly predictions given by the global SVR model, the local SVR model with 877 neighbors and the real productions will be illustrated, for the same days than in the previous experiment. As shown in Figure 5, the forecasts of both models hardly differ and they maintain the same trend as the real curve.

4.4 Computation time

In order to compare the time costs of the aforementioned models, the time taken by each model, either RR or SVR, to predict 8760 hourly powers was calculated. For doing so, we need to measure times to perform the validation of the two models with the 17 544 training data points, as well as to train the models with the optimal hyperparameters.

All the experiments have been executed under the same conditions in the Scientist Computation Center of the Autonomous University of Madrid. Some



(a) 15-03-2018 (b) 15-09-2018

Fig. 5: Real productions and local and global SVR predictions.

of the server characteristics used consist in 40 cores, 700GB RAM, 40 TB hard disk drive and 2 GPUs tesla p100.

For RR validation, we only need 0.6 seconds to find the optimal hyperparameters over the entire training dataset. The SVR validation was very costly and took a total of 12 hours to complete. Training and predicting time of RR and SVR global models were negligible.

On the other hand, the total time to perform the training and the predictions of the local models corresponding to each test data point has been calculated. The training process of each point to predict has been performed using the optimal parameters of the global model using 44 neighbors for RR and 877 neighbors for SVR.

Training and prediction of each individual linear model costs an average time of 0.17 seconds. For non-linear super local models, this average time is 0.24 seconds. Therefore, in a real context of medium-term wind power forecasting, only 5 seconds are needed to validate, train and predict the power of the next 24 hours for the local RR model, compared to more than 12 hours for the global SVR model. By contrast, SVR super local model does not represent an improvement in computational times considering that the validation process cannot be lightened.

5 Conclusions

In this work we have proposed a new, model-agnostic local approximation called Super Local Model that we have applied to a real problem, the wind energy forecasting. This new algorithm is based on point-wise models built just over its neighborhood, i.e. its more similar points.

These individual models have appeared as effective methods specially when applied together with linear models for energy forecasting. In the case studied, the Ridge Regression (RR) super local models were able to compete against Support Vector Machines (SVRs), one of the main state-of-the-art models for wind power prediction. Moreover, RR Super Local Models are very efficient in computational terms, outperforming SVRs. Finally, we have analyzed several

particular cases detecting possible causes of the biggest errors produced, thanks to the interpretability allowed by linear models.

Nevertheless, some work remains to be done. In particular, we would like to define a proper validation method for detecting the best number of neighbors K for the Super Local Models. Also, other metrics for defining similarity should be studied. Regarding the wind power problem, this prediction could be improved using NWP ensembles, so certain uncertainty will be associated to the energy forecasting in terms of the meteorology reliability. Finally, it will be interesting to try the proposed approximation with other real problems, like for example the solar energy forecasting, and also with other datasets far away from the energy field, so a general good performance of the Super Local Models can be tested.

Acknowledgments

The authors acknowledge financial support from the Spanish Ministry of Science and Innovation, project PID2019-106827GB-I00/ AEI/10.13039/501100011033. They also thank the UAM–ADIC Chair for Data Science and Machine Learning and gratefully acknowledge the use of the facilities of Centro de Computación Científica (CCC) at UAM.

References

1. Asociación empresarial eólica. https://www.aeeolica.org/sobre-la-eolica/la-eolica-espana, last access 27-04-2021
2. Centro europeo de previsiones meteorológicas a medio plazo. https://www.ecmwf.int/, last acces 10-03-2021
3. Instituto para la diversificación y ahorro de la energía. https://www.idae.es/informacion-y-publicaciones/plan-nacional-integrado-de-energia-y-clima-pniec-2021-2030, last access 27-04-2021
4. Alaíz, C., Barbero, A., Fernández, A., Dorronsoro, J.: High wind and energy specific models for global production forecast. In: Proceedings of the European Wind Energy Conference and Exhibition - EWEC 2009. EWEA (March 2009)
5. Cleveland, W.S., Devlin, S.J.: Locally weighted regression: An approach to regression analysis by local fitting. Journal of the American Statistical Association **83**(403), 596–610 (1988)
6. De Felice, M., Alessandri, A., Ruti, P.M.: Electricity demand forecasting over italy: Potential benefits using numerical weather prediction models. Electric Power Systems Research **104**, 71–79 (2013). https://doi.org/https://doi.org/10.1016/j.epsr.2013.06.004, https://www.sciencedirect.com/science/article/pii/S0378779613001545
7. Duran, M., Cros, D., Santos, J.: Short-term wind power forecast based on arx models. Journal of Energy Engineering-asce - J ENERG ENG-ASCE **133** (09 2007). https://doi.org/10.1061/(ASCE)0733-9402(2007)133:3(172)
8. Gallego, C., Pinson, P., Madsen, H., Costa, A., Cuerva, A.: Influence of local wind speed and direction on wind power dynamics – application to offshore very short-term forecasting. Applied Energy **88**(11), 4087–4096 (2011). https://doi.org/https://doi.org/10.1016/j.apenergy.2011.04.051, https://www.sciencedirect.com/science/article/pii/S0306261911002868

9. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning: with Applications in R. Springer (2013), https://faculty.marshall.usc.edu/gareth-james/ISL/
10. Jung, J., Broadwater, R.P.: Current status and future advances for wind speed and power forecasting. Renewable and Sustainable Energy Reviews **31**, 762–777 (2014). https://doi.org/https://doi.org/10.1016/j.rser.2013.12.054, https://www.sciencedirect.com/science/article/pii/S1364032114000094
11. Jung, S., Kwon, S.D.: Weighted error functions in artificial neural networks for improved wind energy potential estimation. Applied Energy **111**, 778–790 (2013). https://doi.org/https://doi.org/10.1016/j.apenergy.2013.05.060, https://www.sciencedirect.com/science/article/pii/S030626191300473X
12. Pinson, P., Nielsen, H., Madsen, H., Nielsen, T.: Local linear regression with adaptive orthogonal fitting for the wind power application. Statistics and Computing **18**(1), 59–71 (2009)
13. Ruiz, C., Alaíz, C.M., Dorronsoro, J.R.: Multitask support vector regression for solar and wind energy prediction. Energies **13**(23) (2020). https://doi.org/10.3390/en13236308, https://www.mdpi.com/1996-1073/13/23/6308
14. Smola, A., Schölkopf, B.: A tutorial on support vector regression. Statistics and Computing **14**, 199–222 (08 2004)
15. Wang, J., Yang, W., Du, P., Niu, T.: A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm. Energy Conversion and Management **163**, 134–150 (2018). https://doi.org/https://doi.org/10.1016/j.enconman.2018.02.012, https://www.sciencedirect.com/science/article/pii/S0196890418301079
16. Wu, S.F., Lee, S.J.: Employing local modeling in machine learning based methods for time-series prediction. Expert Systems with Applications **42**(1), 341–354 (2015). https://doi.org/https://doi.org/10.1016/j.eswa.2014.07.032, https://www.sciencedirect.com/science/article/pii/S0957417414004394
17. Zhang, J., Yan, J., Infield, D., Liu, Y., sang Lien, F.: Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and gaussian mixture model. Applied Energy **241**, 229–244 (2019). https://doi.org/https://doi.org/10.1016/j.apenergy.2019.03.044, https://www.sciencedirect.com/science/article/pii/S0306261919304532