Forecasting the Wind Generation Using a Two-Stage Network Based on Meteorological Information

Group Number: 5

Shubhan Mital
Tejasvi Chabbra
Raghav Gupta
Dishank Srivastava

Introduction

<u>Wind and its advantages and disadvantages</u>

- Economical benefits
- Environmental benefits
- Stochastic nature of wind energy

<u>The solution?</u>

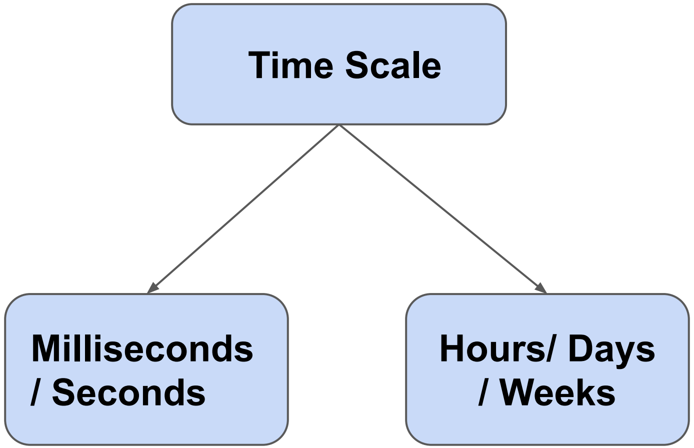
Improve the wind generation forecast

<u>Papers Approach</u>

Time Scale: 1-48 Hrs

Wind speed, Wind generation, numerical weather prediction (NWP) model or meteorological services provide the essential information for the forecast

<u>Approach to forecasting</u>

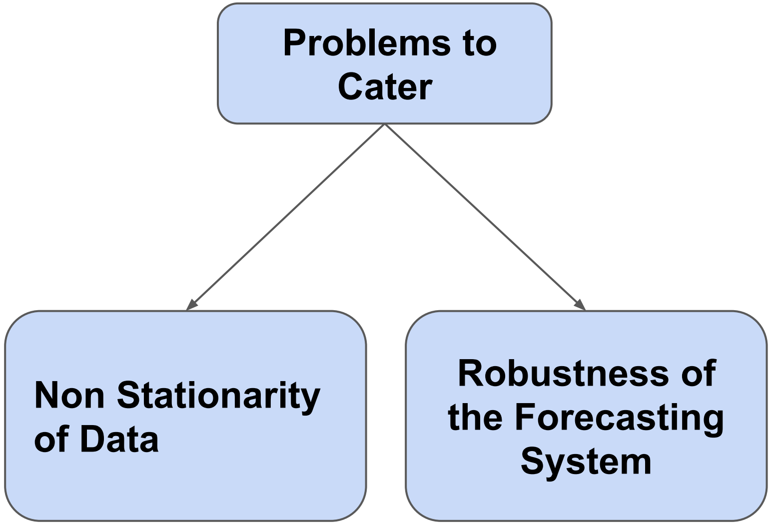


Introduction

Physical vs Statistical Models

- Why Statistical in this case?

Statistical model implementation



Model used

forecasting model with Bayesian clustering by dynamics (BCD) and support vector regression (SVR) is proposed in this paper

It tackles the two problems:

- The proposed model is well suited for capturing the dynamics of WG/WS time series by using hybrid architecture
- it has strong robustness and can be easily modified for different wind farms.

Task Description and Data analysis

Blue Canyon I wind farm

Nameplate capacity of 74 MW

Dots- weather stations within 80 kms

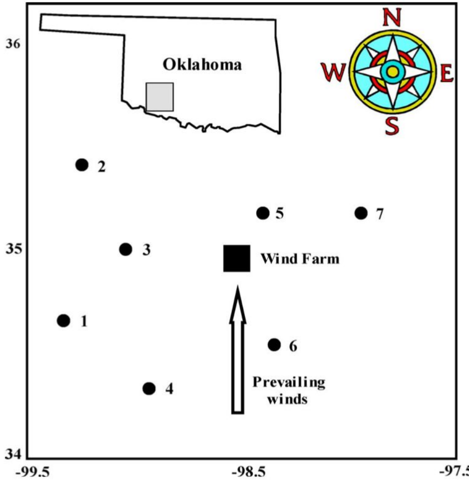


Fig. 1. Geographical location of blue canyon wind farm with altitude and longitude labeled in the map.

Data Collected for model establishment

- 10 min data from wind farm
- Hourly data from surrounding weather stations
- Hourly meteorological forecasts using NWP

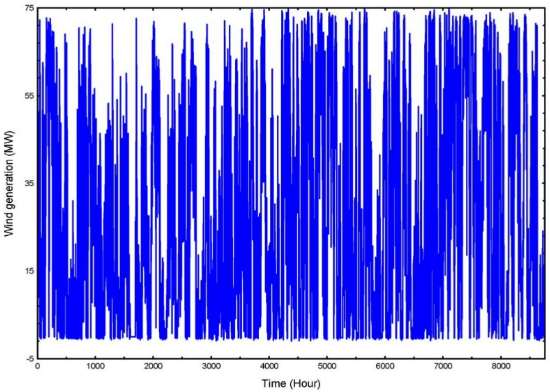


Fig. 2. Ten-minutes wind generation in the wind farm from June 1, 2004 to May 31, 2005.

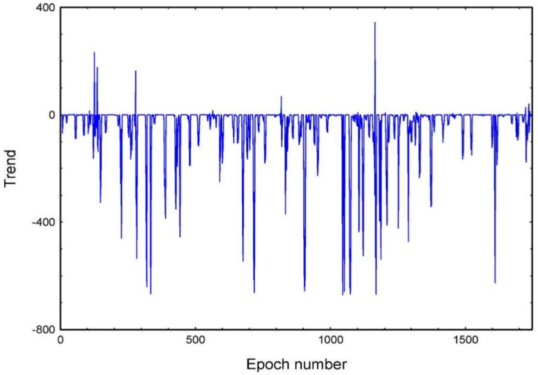


Fig. 3. Trend statistic for ten-minutes wind generation series. RAQ parameters: delay = 6; embedding dimension = 3; radius = 1.5; line definition = 10 points.

Task Description and Data analysis

Correlation analysis

Wind Direction vs Wind Speed

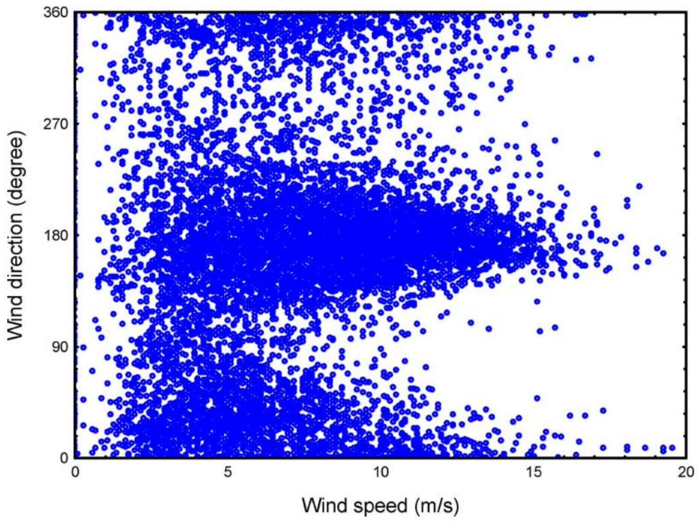


Fig. 4. Correlation between the local wind speed and the direction from June 1, 2004 to May 30, 2005. 0 or 360 is recorded as north, 90 = east, 180 = south, and 270 = west.

Correlation Analysis

Aggregated Generation vs Wind Speed

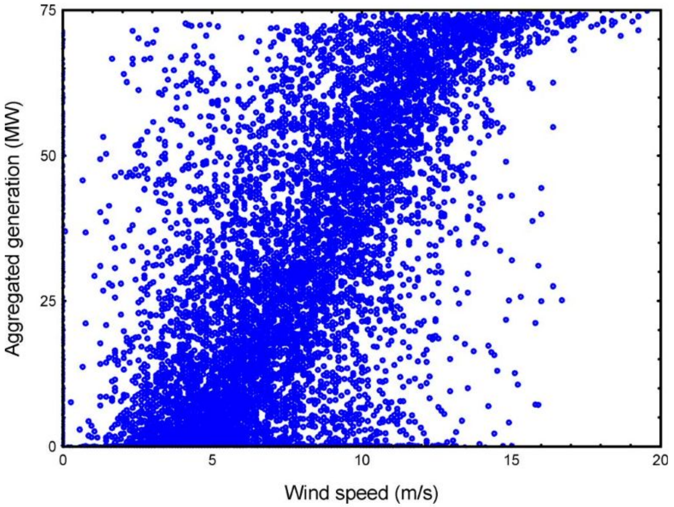


Fig. 5. Correlation between the wind speed and the aggregated wind generation at the wind farm from June 1, 2004 to May 30, 2005.

Method and Learning Algorithm

A.Architecture of Forecasting system

$$y(t+1) = f(y(t), \ldots, y(t-m+1); X)$$

Time series based, nonlinear, discrete time model.

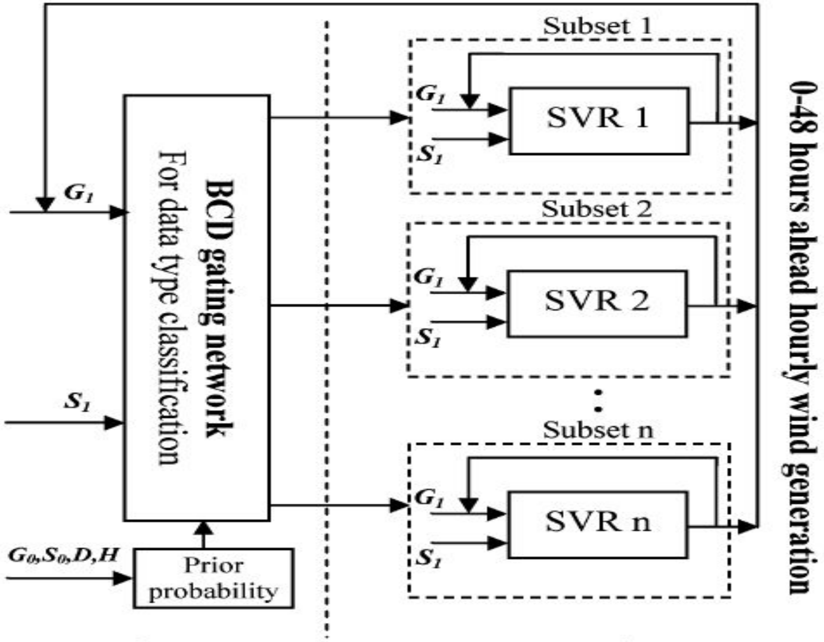
Task: Extrapolate past wind generation behavior.

The model used: Hybrid Architecture with Multi staged Design.

1) BCD classifier: **input dataset** $\xrightarrow{\text{(Unsupervised)}}$ **several subsets**

2) SVR Group: **fit training data** $\xrightarrow{\text{(Supervised)}}$ **each subset**

SVR fits the training data inside the unsupervised subsets formed by the BCD Classifier on the basis of dynamics.



Hybrid model for the wind generation forecasting.

BCD Classifier:

- Partitioning the space of input dataset.
- BCD Vs others(HC and SOM) - risk of overfitting
- Input variables/factors: WG/WS data, wind direction and humidity

SVR Group:

- Advanced machine learning technique based on Structure Risk minimization principle.
- Resistant to overfitting problems ⟶ high generalization performance
- Input variables/factors:Wind generation and speed data

❖ **Wind generation forecasting: voting manner between the BCD and SVR**

❖ **Future estimates derived by using past and forecast values of the input as well as network's output at previous time stage.**

Prediction equation:

$$\hat{y}(t+1) = f(\hat{y}(t), \hat{y}(t-1), \dots, X)$$

B.Selection of input variables

- Input variables have to be selected keeping these 3 aspects in mind:
  - Which exogenous Variables have significant correlation to wind generation
  - Which station's data is suitable for our forecasting model
  - How many lagged hours to be included in the wind generation time series
- Table 1 shows that wind generation has positive cross correlation to 2 exogenous variables namely wind direction and gust speed.
- Table 2 shows the cross correlation between different stations and wind generation.We set a benchmark of 0.60 and select stations which have a correlation (>0.60) i.e. Stations 2 , 3 , 5.

TABLE I

CROSS-CORRELATIONS BETWEEN AGGREGATED WIND GENERATION AND DIFFERENT OBSERVATIONS WITHIN THE WIND FARM

	Wind direction	Barometric pressure	Temperature	Humidity	Gust speed
Aggregated WG	0.14	-0.02	-0.02	-0.07	0.62

TABLE II

CROSS-CORRELATIONS BETWEEN THE WIND SPEEDS IN THE WEATHER STATIONS AND THE AGGREGATED GENERATION OF THE WIND FARM

Stations	1	2	3	4	5	6	7
Cross-correlation With generation	0.56	0.66	0.67	0.54	0.62	0.50	0.51

TABLE IV

LIST OF INPUT DATA OF THE BCD CLASSIFIER

Input	Variable	Detail description
1-6	G_1	Wind power generation series
7-30	S_1	Wind speed series
31-40	Variables to determine prior probability	G_0: Average wind generation
		S_0: Average wind speed
		D: Forecasted wind direction
		H: Average humidity

- Table 4 is used for deciding the input variables for the BCD classifier.
- First 6 input variables - Wind power generation time series.
- Next 24 input variables - Wind speed time series.
- Last 10 input variables - Avg wind generation , Avg wind speed , Forecasted wind direction and average humidity which is used to determine the prior probability of the classifier.

- Table 3 is used for deciding the input variables for the SVR network.
- This table consists of hourly wind generation series of previous 6 hours as well forecasted and observed wind speeds of wind farm and weather stations 2 ,3 5 that were chosen.
- Here lag 1-6 indicates the observed wind generation for the previous 6 hours
- Similarly for wind speeds lag 0 indicated the forecasted wind speed at and lag 1-5 are for the observed wind speeds for previous 5 hrs

TABLE III

LIST OF INPUT DATA OF THE SVR NETWORK

Input	Variable name	Lagged value (hours)
1-6	Hourly wind power generation (G_1)	1,2,3,4,5,6
7-12	Hourly wind speed 0 (S_1)	0,1,2,3,4,5
13-18	Hourly wind speed 2	0,1,2,3,4,5
19-24	Hourly wind speed 3	0,1,2,3,4,5
25-30	Hourly wind speed 5	0,1,2,3,4,5

The wind speed 0 means the speed within the wind farm, and accordingly number 2, 3, or 5 means weather station 2, 3, or 5.

The hour of the predication is assumed at 0, the lag 0 represents the target instant, and the 6 lagged hours means the values that were measured 6 h earlier than the hour of predication.

C.The learning Algorithm - BCD classifier

- Method of clustering is used in the classifier using the principle of similarity of time series
- Step 1 of BCD classifier is to model the input time series using an autoregressive model. The matrix form of the model is : $x_j = X_j \beta_j + \varepsilon_j$ (eqn 4). Using given data and standard bayesian techniques , parameters are calculated.
- Step 2 of the BCD classifier is the forming of initial clusters which is done using a model based bayesian procedure.The matrix form of these clusters is $x_k = X_k \beta_k + \varepsilon_k$ (eqn 5). Here $x_k = \begin{pmatrix} x_{k1} \\ \vdots \\ x_{km_k} \end{pmatrix}, \quad X_k = \begin{pmatrix} X_{k1} \\ \vdots \\ X_{km_k} \end{pmatrix}.$
- Step 3 of the BCD Classifier is ranking these initial clusters using the posterior probability given by $P(M_C \mid x) \propto P(M_C) f(x \mid M_C)$.(eqn 6). To calculate the marginal likelihood we use $f(x \mid M_C) = \int f(x \mid \theta_C) f(\theta_C)\, d\theta_C$ (eqn 7) which is simplified to remove the integral and results in $f(x \mid M_c) = \frac{\Gamma(1)}{\Gamma(1+m)} \times \prod_{k=1}^{c} \frac{\Gamma(m_k/m + m_k)}{\Gamma(m_k/m)} \times \frac{(\mathrm{RSS}_k/2)^{(q-n_k)/2}\, \Gamma(n_k - q)/2}{(2\pi)^{(q-n_k)/2} \det(X_k^T X_k)^{(1/2)}}$ (eqn 8).
- However these clusters grow exponentially and Step 4 of the BCD classifier is to reduce the no of clusters by merging them.There are two ways to do this :-
  - Agglomerative search strategy - Uses marginal likelihood to merge clusters
  - Heuristic strategy applied on agglomerative clustering - Uses a similarity measure namely euclidean distance between 2 times series $S_j = \{x_{i1}, \ldots, x_{in}\}$ and $S_j = \{x_{j1}, \ldots, x_{jn}\}$ given by $D(S_i, S_j) = \sqrt{\sum^{n} (x_{it} - x_{jt})^2}.$ (eqn 9)

D.The learning Algorithm - The SVR algorithm

Consider training data $(x_1,y_1),(x_2,y_2),....,(x_n,y_n)$.
The support vector regression solves an optimization problem:

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{n}(\xi_i+\xi_i^*)$$

$$\text{subject to } y_i - (\omega^T\phi(x_i)+b) \le \varepsilon + \xi_i^*$$

$$(\omega^T\phi(x_i)+b) - y_i \le \varepsilon + \xi_i$$

$$\xi_i, \xi_i^* \ge 0, \quad i = 1,\dots,n$$

x_i **:** mapped to a higher dimension space by the function Φ

ξ_i^* **:** slack variable of the upper training error subject to

ε **-** insensitive tube

$$\boxed{(\omega^T\phi(x_i)+b) - y_i \le \varepsilon}$$

- If x_i is not kept in ε, there is an error of ξ_i^* that has to be minimized.
- SVR: Minimizing the training error $C\sum_{i=1}^{n}(\xi_i+\xi_i^*)$ as well as regularization term $\omega^T\omega/2$

❖ Traditional least square method: ε is always zero and data are not mapped into higher dimensional spaces.

Usage of Lagrange theory:

$\Phi(x_i) \longrightarrow$ High of infinite dimensional space

Solving for ω: two Options available

$$\max_{\alpha_i,\alpha_i^*} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i-\alpha_i^*)^T Q(\alpha_j-\alpha_j^*) - \varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)$$

$$+\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)$$

$$\text{subject to, } \sum_{i=1}^{n}(\alpha_i-\alpha_i^*) = 0$$

$$0 \leq \alpha_i, \quad \alpha_i^* \leq C, \quad i = 1,\ldots,n$$

$$Q_{ij} = \phi(x_i)^T\phi(x_j)$$

α_i **and** α_i^* **are Lagrange multipliers.**

Implicit mapping: Kernel trick

$\phi(x_i)^T\phi(x_j) = (\gamma x_1^T x_2 + c_0)^d$: Polynomial Kernel

$\phi(x_i)^T\phi(x_j) = e^{-\gamma(x_1-x_2)^2}$: RBF Kernel

Can be calculated even without knowing Φ(x)

- For numerical experiments, we use the software LIBSVM, which is a library for support vector machines and efficient implementation.

Numerical Experiments

A. *Data Collection and Implementation*

Time Duration Chosen: Two months, from June 1, 2004 to May 30, 2005.

(Upon the request of the power company running the wind farm, the 48-h forecasts are made in **each morning** for **generation schedule** and **electricity markets** of the **next operation day**)

The learning procedure of the proposed architecture is as follows:

1. Classify the entire training dataset into two groups: a training set and a verification set
2. Determine prior probability of the BCD classifier according to the control variable
3. Classify the input data type using BCD according to the dynamics of WG/WS time series
4. In each subset of the input space, train the SVR to fit the data subset
5. Forecast the wind generation using the dataset for verification and calculate the MAPE
6. Tune the parameters of the model and repeat steps 1)–5) until satisfactory results are obtained
7. Select the network parameters at the minimum of the MAPE as the final ones.

Numerical Experiments

A. Data Collection and Implementation

After the training procedure is finished, the following test process is applied to verify the proposed model:

1. Identify the type of the input test data according to the information of the test hour and previous hours, using the BCD classifier
2. Use the corresponding SVR network to get the output of the next hour wind generation
3. Recurrently derive the 2–48 h estimates using past values of the inputs, as well as the network's outputs at previous time steps

Numerical Experiments

B. Bayesian Clustering Analysis

Now, we will go over the following steps:

1. Describe the clustering results of BCD
2. Carry out RQA analysis to the WG series in each cluster

This analysis not only **assists the parameter adjustment** in the training procedure, but also explains the effectiveness of Bayesian clustering.

The Analysis

The whole training dataset has been partitioned into **nine clusters** by the BCD classifier.
The **test data** was assigned to **four** of them
For these four subsets, the **RQA analysis** is used to investigate the **nonstationarity of the WG series**, keeping the RQA parameters same as before

Numerical Experiments

B. <u>Bayesian Clustering Analysis</u>

The trend statistics for the WG series in the subsets are **closer to zero** than that in the whole dataset, indicating the effect of clustering for reducing the movable nature of the WG series.

(For simplicity, we did not plot the trend statistic in every cluster.)

Numerical Experiments

C. Numerical Results

Criteria to compare performance:

a) **Normalized Mean Absolute Error (MAE)**
MAE is defined as

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}(|y_{ai} - y_{fi}|)/P_N \times 100\%$$

b) **Root Mean Square Error (RMSE)**
RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(y_{ai} - y_{fi})}{P_N}\right)^2} \times 100\%$$

$\mathbf{y_{ai}}$ is the actual value,
$\mathbf{Y_{fi}}$ is the forecast value,
P□ is the nameplate capacity,
and **n** is the total number of value predicted.

Numerical Experiments

C. Numerical Results

For comparison, simulations which are comparing with Persistent forecast, which uses the most recent information available are also conducted.

The benefit gained by using the proposed model is measured as the accuracy improvement over the persistent model using the following formula:

$$\mathrm{Im}\,p = (\mathrm{Err}_P - \mathrm{Err}_m)/\mathrm{Err}_p \times 100\%$$

where, Err□ is the evaluation criterion (i.e., MAE or RMSE) of the persistence
Err□ is the evaluation criterion of the proposed model.

Numerical Experiments

C. Numerical Results

The tables below show the numerical results of the proposed model and the improvements over the persistent model.

(b) 24-hour ahead predication

Persistence error: MAE=21.24 $\%P_n$ RMSE=29.84 $\%P_n$

SVR Inputs (stations)				Errors(% of P_n)		Improvement (%)	
0	3	2	5	MAE	RMSE	MAE	RMSE
O	×	×	×	15.13	21.06	28.77	29.42
O	O	×	×	14.88	20.55	29.94	31.13
O	O	O	×	14.56	20.08	31.45	32.71
O	O	O	O	14.38	19.74	32.28	33.84

(c) 48-hour ahead predication

Persistence error: MAE=25.42 $\%P_n$ RMSE=34.81 $\%P_n$

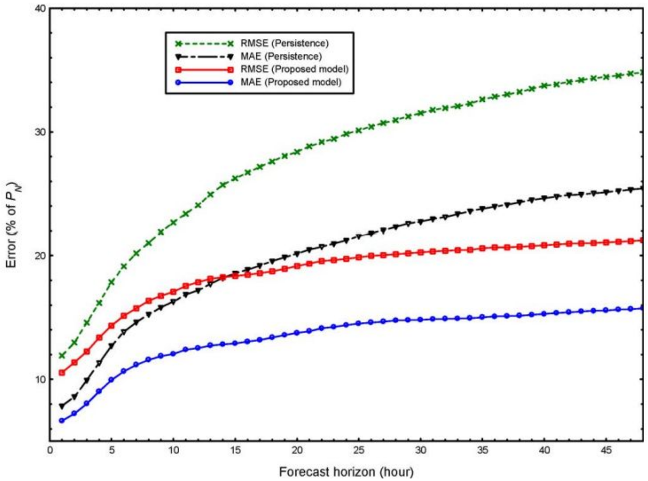
SVR Inputs (stations)				Errors(% of P_n)		Improvement (%)	
0	3	2	5	MAE	RMSE	MAE	RMSE
O	×	×	×	16.33	22.17	35.76	36.31
O	O	×	×	16.09	21.84	36.70	37.26
O	O	O	×	15.89	21.56	37.49	38.06
O	O	O	O	15.73	21.24	38.12	38.98

The following conclusions can be derived:

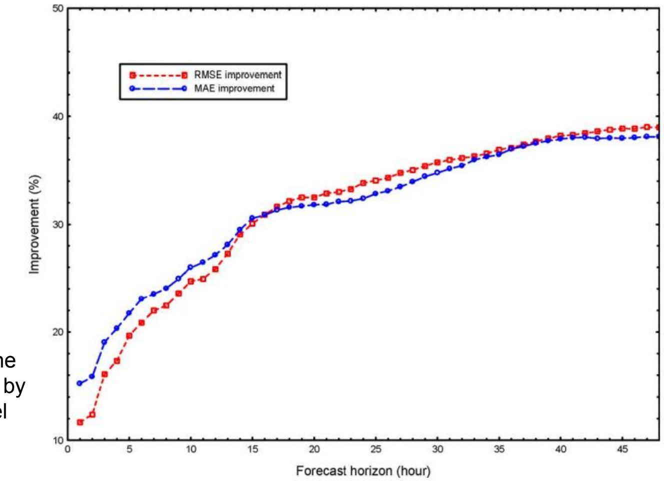
1. The proposed forecasting model **outperforms the persistence forecast** in all the situations, even in the first time step
2. The addition of data from remote stations **improves forecasting accuracy**, and their importance grows as the predication window extends to the future

Numerical Experiments

C. <u>*Numerical Results*</u> *(These are done for a horizon of 48 h)*



The MAE and RMSE performance obtained by the proposed model compared with the performance obtained by persistence.



The percentage improvement over the persistence obtained by the proposed model

As shown, the proposed model can provide **much more accurate forecast** in the whole forecast horizon. It is able to produce **robust multistep ahead** estimations compared with persistent forecast, and considerable improvement up to **more than 40%** over the persistence is achieved.

Conclusion

- Developed a model which gives us a clear understanding of the projections on wind power generation for any wind farm for the next 48 hours.
- Used Machine Learning in the form of a BCD classifier coupled with a SVR network
- Benefits include:
  - Easy implementation
  - Cost efficiency
  - Handles non stationary data due to robustness
- Model gives lesser RMSE and MAE as compared to the basic persistence model
- As the time duration of the forecast increases , accuracy increases and therefore we settle at 48 hour prediction