Performing Experimentation with Physics Model to Predict Statistical Weather Condition

A.Jayanthiladevi¹,², V.Nagaraju², L.Devi³, N. Padmavathi⁴, P.Sailaja⁵, M.S Minu⁶ and Tapas Bapu B R

¹Professor-Computer Science & Information Science, Srinivas University, Karnataka, India
²Professor / Department of ECE, Rajalakshmi institute of technology
³Assistant professor in computer application, Muthayammal College of arts and science, Rasipuram, Tamilnadu, India
⁴Assistant professor in computer application, Muthayammal College of arts and science, Rasipuram, Tamilnadu, India
⁵Assistant Professor/ Department Of Technology, Velammal Institute Of Technology
⁶Assistant professor, Dept of CSE, SRM Institute of Science and Technology, Ramapuram Campus
⁷Professor / Dept of ECE, S. A. engineering college

*Corresponding author e-mail: drjayanthila@gmail.com

Abstract. Ice in wind turbines may cause a tremendous reduction in energy conservation. As, ice over turbines are not considered to be a traditional weather prediction data, prediction towards power can leads to higher error. This work anticipates a statistical approach dependent on Naive bayes regression to identify production loss has to be analyzed. It measures input of regional weather condition and various other conditions, and identify power production loss for 48 hours to enhance prediction of next generation energy loss. This can be trained with various prediction measurements and drastically enhances other conventional approaches for longer period. It may diminish absolute production error by ~100kW and it computes its skill with other models. Prediction of weather data is considered to be one of the effectual data for diverse statistical prediction and some calculations are not so absolute. This method can be computational less cost and may be trained again for next prediction.

Key words. Weather prediction, power loss, ice over turbines, naive bayes, regression

1. Introduction

Appropriate recognizing of wind power prediction is essential to handle power demand and energy production. During cold climates, ice over wind turbines may leads to various crises [1]. Owing to variation in aero-dynamic balance, vibration generation, load improvement, ice may outcomes in essential production losses. During end of 2016, 25% of complete installation of wind energy capacity was in cold environment, which makes ice an essential factor in producing energy [2]. Ice modelling over structures like wind turbines or power lines have been performed with physics based ice model.
generally depends on equation specified as Makkonen equation [3]. With complex terrain, small structural scale and complexity of measuring ice, modelling ice and ice development are more confronts. Author in [4], demonstrated the usage of various NWP models or NWP ensemble prediction as input to ice system that may cause an extremely complex load prediction [5]. By initiating least variations of certain parts of input factors to ice models also determined small variation of ice accumulation that uses applied ice model. Therefore, icing prediction in physics-based ice models are uncertain in common [6]. Physics-based weather prediction have been utilized to model ice based production losses. It shows certain skill in estimating production loss [7]. Moreover, due to uncertain weather prediction, these predictions are not perfectly appropriate.

Machine learning has prior been utilized to generate ice based production loss model, however to certain investigators knowledge, everything will be originated from physics based ice model outcomes merged with certain power curves [8]. As records and historical weather prediction of on-site calculations, ice prediction can be performed in statistical manner [9]. However it may use weather prediction data integrated with machine learning/statistical approach that identifies production losses. There are huge applications of machine learning approaches in geo-scientific model by author in [10] and also certain specific in isolation with weather predictions.

One of the general way is there are certain progression towards entire replacing conventional numerical weather prediction approaches, however there should be a simplified realities [11]. In addition, there is huge progression with the use of neural network approaches to modify arguments in context to climate model [12].

Here, this work uses a machine learning based approaches termed naive bayes regression is trained with previous weather prediction [13]. Upper range for all 48 hours was selected as it has maximal time to lead regional NWP model that was used. Initializing prediction with UTC for 48 hours is most crucial; however it is specified in 'next few days'.

In this technical sections are utilized by input data that are describes, the selection of machine learning approaches will be elaborated and used as a method described. In these outcomes, this work will discuss power loss prediction skill and it is translated to appropriate power predictions [14]. As well, it discusses the significance of weather prediction data verses on-site measurements and individual input variables [15].

2. Data
For this analysis, various weather predictions have been analyzed with online available weather prediction data of certain data. Here, weather data from online is considered as an input for further process. This is explained in next section.

3. Observations
This work uses anonymized production data and with meteorological observations of wind data which is located with range of 58-60° N at terrain or hill as various surroundings. With these sites, there are numerous wind turbines may be observed Turbines form various stations may have de-icing operation system to eliminate ice-growth. Moreover, this work may be incorrectly and it may still experience production loss periods. Henceforth, this work included in this investigation for de-icing system. Only 15 minutes value of every hour is utilized in verification. To identify production loss from various measured production and wind speed, it has to recognize production that may expect provided wind speed (potential production). With this, this model constructs easier production model that may be associated with wind speed at site with production ability, therefore production is encountered with power curve. As this does not hold ice occurrence relation, only threshold values are set to all these days.
4. Data splitting
With diverse machine learning tasks, algorithms are prone to ‘over-fitting’ that it may work for all data with algorithm that is seen during training and testing data, however it may not work effectively on newer data (it may not be generalized). Henceforth, it should be evaluated with separate testing scenario. As this data may use non-zero auto-correlation, simpler randomized splitting of individual prediction into testing and training sets is not suitable, as it may cause to over-confident outcomes (owing to learning algorithms with correlation). As an essential approach, this may splits the data into chunks and these chunks are randomly distributes for training and testing, i.e. (20% and 80%). Training set generated in this dataset is used for selecting and tuning machine learning approaches.

As well, this work may compute the use of randomized data with operational practice: for prediction initialization at 5 days can be done with all available prediction data for sample validation—therefore every predictions are initialized with certain time period. As this method may be computationally expensive with an operational context it may be re-trained easily for all new predictions. This model is termed as operational validation all through the work.

5. Naive Bayesian
It is extremely hard to predict in previous that some machine learning approaches may work effectively for some crisis, even if they are already known to certain works on related constraints. Henceforth, section of appropriate techniques possesses numerous ways known as hyper-parameters. There parameters have to be provided in prior for training purpose. In recent time, there is a necessity of using neural networks with various neuron layers (termed as deep learning) for extensive application range, deep learning are generally based on extremely huge amount of training data to use its maximum result.

This work does not consider deep learning approaches here. Indeed, this method is tested with various approaches like neural networks with hidden layers, random forest, support vector machine and multi-layer regression. Every test is done with online available dataset of various hyper-parameters over training set (expect multi-linear regression, this may not have hyper-parameters), and naive bayes is considered to perform training set. For performing this training, partitioned the data into 3 equal parts. Every technique is trained with 2/3 of training data and computed with remaining 1/3. This procedure may be repeated for 3 times, then average value is computed. This approach shows the superior average skill that was chosen. Appropriate configurations of all these techniques are validated with hyper-parameters merged with corresponding codes.

6. Numerical results
The simulation is carried out in MATLAB environment. Here, comparison is made of power prediction with naive bayes method with other physics ice model. Here, RMSE is measured as metric for performing this computation. The time will be of 18-20h. Unbiased RMSE is depicted as in Eq. (1):

$$Unbiased_{RMSE} = \sqrt{RMSE^2 - bias^2}. \quad (1)$$

Here, two methods are used for prediction, they are: deterministic and ensemble prediction. Mean over entire period is generally studies. As, naive bayes is used as a classification factor it is essential to partition data into testing and training. The evaluation is performed for subset functionality. However, it may facilitate reasonable evaluation. Here, NB based prediction is slightly higher than deterministic physics model for prediction as in Fig 1. Comparison with mean prediction is slightly superior to other station, and slightly worse with various stations as in Table 1.

**Table 1. Station based prediction rate**

| Station | A   | B   | C   | D   |
|---------|-----|-----|-----|-----|
| Naive   | 0.29| 0.33| 0.30| 0.40|
| Method      | Wind speed | Wind direction | Temperature | Velocity | Dew point | Pressure | Cloud water | humidity | Graupel | Cloud ice | Snow | Ice | Water vapour | Kinetic energy |
|------------|------------|----------------|-------------|----------|-----------|----------|-------------|----------|---------|-----------|------|-----|--------------|----------------|
| Bayes Random forest | 0.31       | 0.35           | 0.33        | 0.42     |
| Physical model    | 0.47       | 0.42           | 0.36        | 0.42     |
| Ensemble          | 0.39       | 0.37           | 0.28        | 0.36     |

| Table 2. NWP and observation rate |
|-----------------------------------|
| NWP                               | Wind speed | Wind direction | Temperature | Velocity | Dew point | Pressure | Cloud water | humidity | Graupel | Cloud ice | Snow | Ice | Water vapour | Kinetic energy |
| Observations                      | Wind direction | Wind speed | Power produced | Potential power | Temperature | Absolute production loss | Relative production loss |

**Fig 1. Production loss**

Based on available dataset, 14 prediction variables and 7 observation variables are considered with total 21 input variables as in Table 2. The input variables are extremely based on available data. Moreover, with this variables prediction loss can be computed.
Feature significance of input variables are considered with standardized procedure to construct naive bayes algorithm. The essential feature considered here is snow. Next, empirical models are considered to train the variables as in Fig 2. All possible combinations with these variables are left out.

Therefore, $2^{21} - 1 = 2097150$ possible combinations are identified. There is no essential impact when eliminating single variables for any sort of inputs. This specifies that variables are replaced with significant features. When eliminating snow, cloud ice turns to be a most essential variable after wind speed as in Fig 3. When eliminating both snow and ice, there is no significant loss in skill of naive bayes predictions. This step may be trained for variables with NWP forecast. Henceforth, when no NWP information is attained, naive bayes is considered to be simply persistence. Thus, NWP forecast is most useful, this may increase on-site measurements which are not statistically essential.
7. Conclusion
This work tested and developed a machine learning model to predict weather even with ice in wind turbines. It also handles production loss in wind turbines for every 48 hours. This method utilizes prediction from conventional NWP model integrated with production and meteorological observations at wind power as input to naive bayes regression. This model is trained with past observations and predictions and to identify relative production loss in every stages of time. This method is considered to be effectual and simpler that may persists over observed production time loss.

REFERENCES
[1] A. G. Kraj and E. L. Bibeau, “Phases of icing on wind turbine blades characterized by ice accumulation,” Renewable Energy, vol. 35, no. 5, pp. 966–972, 2010.
[2] L. Makkonen, “Models for the growth of rime, glaze icicles and wet snow on structures,” Philosophical Transactions of The Royal Society Lond., vol. 358, pp. 2913–2039, 2000.
[3] N. N. Davis, P. Pinson, A. N. Hahmann, N.-e. Clausen, and M. Žagar, “Identifying and characterizing the impact of turbine icing on wind farm power generation,” Wind Energy, vol. 16, no. September 2015, pp. 1503–1518, 2016.
[4] B. E. K. Nygaard, H. Àgústsson, and K. Somfalvi-Tóth, “Modeling Wet Snow Accretion on Power Lines : Improvements to Previous Methods Using 50 Years of Observations,” Journal of Applied Meteorology and Climatology, vol. 52, pp. 2189–2203, 2013.
[5] P. Thorsson, S. Söderberg, and H. Bergström, “Modelling atmospheric icing: A comparison between icing calculated with measured meteorological data and NWP data,” Cold Regions Science and Technology, vol. 119, pp.124–131, 2015.
[6] H. Bergström, E. Olsson, S. Söderberg, P. Thorsson, and P. Undén, “Wind power in cold climates, Ice mapping methods,” Stockholm, 2013.
[7] N. Davis, A. N. Hahmann, N. E. Clausen, and M. Žagar, “Forecast of icing events at a wind farm in Sweden,” Journal of Applied Meteorology and Climatology, vol. 53, no. 2, pp. 262–281, 2014.
[8] P. D. Dueben and P. Bauer, “Challenges and design choices for global weather and climate models based on machine learning,” Geoscientific Model Development, vol. 11, no. 10, pp. 3999–4009, Oct. 2018.
[9] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, “Deep learning and process understanding for data-driven Earth system science,” Nature, vol. 566, no. 7743, p. 195, Feb. 2019.
[10] S. Scher and G. Messori, “Weather and climate forecasting with neural networks: using GCMs with different complexity as studyground,” Geoscientific Model Development Discussions, vol. 2019, pp. 1–15, 2019
[11] S. Rasp, M. S. Pritchard, and P. Gentine, “Deep learning to represent subgrid processes in climate models,” Proceedings of the National Academy of Sciences, p. 201810286, Sep. 2018.
[12] S. Rasp and S. Lerch, “Neural Networks for Postprocessing Ensemble Weather Forecasts,” Monthly Weather Review, vol. 146, no. 11, pp. 3885–3900, 2018.
[13] J. Molinder, H. Körnich, E. Olsson, and P. Hessling, “The use of uncertainty quantification for the modelling of wind turbine icing,” Submitted to Journal of Applied Meteorology and Climatology, 2019.
[14] Müller, M. Homleid, K.-I. Ivarsson, M. Køltzow, M. Lindskog, K. H. Midtbø, U. Andrae, T. Aspelien, L. Berggren, D. Bjørge, P. Dahlgren, J. Kristiansen, R. Randriamampianina, and M. Ridal, “AROME-MetCoOp: A Nordic Convective-Scale Operational Weather Prediction Model,” Weather and Forecasting, vol. 32, no. 2, pp. 609–627, 2017.
[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikitlearn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.