978-1-5386-5525- 2/18/$31.00 ©2018 IEEE

Improving wind power prediction with retraining machine learning algorithms

Mariam Barque, Simon Martin, Jérémie Etienne Norbert Vianin, Dominique Genoud, David Wannier
Institute of Information Systems
University of Applied Sciences, HES-SO Valais
Sierre, Switzerland
Email: mariam.barque@hevs.ch

***Abstract*—This paper presents a 48-h prediction methodology for wind power production using a machine learning algorithm and focuses on the optimization of the input dataset. While power-grid operators have to keep the production equal to demand, wind power depends on meteorological conditions. Therefore, the main issue of power-grid operators is to predict the wind production as precisely as possible. Our goal is to improve wind prediction accuracy based on the feedback of the relevant work on the subject. To this purpose, past power production and Numerical Weather Predictions (NWP) are used as input of a Gradient Boosting Tree algorithm. Our approach is to lay emphasis on the input data in order to extract as much knowledge as possible and remove as much error as possible from the source. Therefore, 20 days have been removed and additional values have been calculated to improve the algorithm accuracy. The novelty of our approach is in the constant retraining of the model to give the latest information available contrary to most studies, which use a fixed learning dataset. The idea is to combine the advantage of regression models and machine learning algorithms, which use large datasets, in order to learn the relationship between wind production, NWP, and date and time information. The whole methodology helps the model anticipate error of wind prediction and seasonal trends so that the overall prediction accuracy increases by 17 % compared to the persistance approach. Using 9 months of historical values from 2016.12 to 2017.08 and predicting from 2017.09 to 2018.01, a prediction accuracy of 83 % has been achieved; 17 % better than the persistence model. Result analysis show that more improvement can be achieved with a focus on low wind speed values, an improvement of weather prediction using real local weather measures and a reduction of the forecast horizon. The whole prediction process have been automatized and a visualization web page have been implemented.**

I. Introduction

Among new sources of renewable energy, wind energy is the fastest growing installed capacity over recent years and has become an alternative to fossil fuels in many countries. Its capacity has increased fivefold since 2007 and reaches approximately 539 GW, covering for more than 5 % of the world's energy consumption [1]. Contrary to fossil fuel power plants or pump-storage hydroelectricity, wind production is intermittent and only depends of weather conditions. As it is mandatory for grid operators to keep the production equal to demand, the high variability of wind speed and direction can lead to drastic evolution of the energy markets as seen in June 2013 with a negative price of -200 euros/MWh due to unexpected renewable production on the European area [2].

Therefore, improving the prediction of wind power plays a key role in anticipating the unit commitments and dispatching plans of grid operators. This paper focuses on day-ahead wind power prediction in order to support the local electricity distributors to plan the operation of other production plants and use energy markets to be able to meet the demand.

A wide range of studies focus on wind power prediction using either complex physical equations or data-driven statistical approaches and more recently, machine-learning algorithms that show interesting results. The physical approach uses the wind's power curve, speed and direction information to estimate the future production, whereas data-intensive models are based on historical production or forecasted values from a Numerical Weather Prediction model (NWP).The most common algorithms used in literature are regression methods like ARIMA time series, Support Vector Regression, K-NN and Neural networks [3] [4].

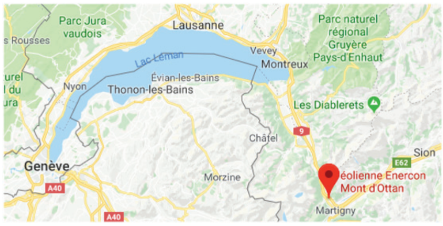
The European project on wind prediction ANEMOS, shows that NWP models outperform time series approaches for more that 3 to 6 hours prediction ahead [5]. J.P. Heinermann's research shows that SVR and neural networks outperform K-NN method [6]. J. Sousa and R. Bessa reach RMS errors for both forecasting systems ranged between 10 % and 25 % according to the forecasted horizon, with a mean value of 17 % over the three forecast days in terms of RMSE [7]. A few papers work on ensemble decision trees like L. Fugon [8] and X. Zhao [9] that obtained better results with a Random Forest algorithm compared to Neural Networks and SVR. They also point out ensemble decision trees are easier to use, as only a few parameters such as the number of trees in the forest have to be optimized. Gradient boosting ensemble trees are quite recent machine learning approaches that have proven to be highly effective with remarkable results for a vast array of problems so that they have gained popularity by winning numerous machine-learning competitions [11]. They are often more accurate than the Random forest algorithm but take more time to run. Moreover our recent project based on solar power prediction paper [12], has shown that Gradient Boosting outperforms Random Forests. To avoid reporting input dataset errors in the results, preparing the data is also an important step to gain accuracy from the data source and account for the first recommendations in the proceeding of 2012 Global Energy Forecasting Competition [13].

Our objective is to work on improving wind power prediction based on the different feedbacks of the latest relevant work on the subject. For this purpose, a GBT and a deep work on the input dataset have been explored and show interesting results. Our idea is to lay emphasis on the data processing step in order to find the appropriate column manipulation to help the machine-learning algorithm extract the most information possible. The novelty of our constantly retraining approach is the model always using the latest information available contrary to most studies, which use a fixed learning dataset. The idea is to combine an advantage of regression models which use the t-n latest value to predict the t value and the advantage of machine learning algorithms which use large datasets to learn the relationship between wind production, NWP and date and time information. Moreover, Physical knowledge from wind power is also used to tune the model.

The full methodology has been implemented for the day ahead prediction needs of the 7MW-wind farm of our local distributor. For this purpose, a daily updated web page visualizing the predictions has been implemented. This paper will describe the dataset set-up, the prediction methodology, and its results. The different steps of the prediction framework automation will also be presented.

II. Dataset settings

A. Wind production data



Construction Date	Capacity (MW)	Annual production (MWh, 2016)
2008	2	5044

Fig. 1. Wind turbine location and description

The total wind turbine installed capacity is 18 GW in 2016 in Valais (Wallis in German)[14]. Valais is a canton located in the south of Switzerland. It is in the Alps, home to world-renowned alpine resorts and vineyards in the Upper Rhone Valley. The topography of the area and the geographical effects make wind variables unstable and difficult to predict. Therefore, separate predictions for each wind turbine using the same methodology have been chosen. For the purpose of the paper, the prediction work presented is based on the wind turbine "Mont d'Ottan" owned by RhônEol SA and described in Figure 1.

Our model use both Numerical Weather Predictions (NWP) and historical production of wind turbines. The production is collected by smart meters and received from SEIC-Télédis Group, a power-grid operator in Valais. We receive the real 15-min measurement of the production though an FTP framework which is updated several times a day with the latest values. The data is available since 2013.

B. Weather forecast data

The NWP model is provided by MeteoSwiss, the national weather forecast company. We receive forecasted wind parameters, temperature and humidity with two different models: Cosmo-1, the more accurate, has a 33-h horizon forecast and is updated every tree hours with a 2km resolution. Cosmo-7 is updated every six hours and predicts the weather up to seven days ahead. The data received in a CSV file are stored and visualized in the Axibase framework. As a day-ahead prediction is performed, the latest forecasts of the two models are combined to cover the full horizon. Our use case aims to have a prediction around 10 AM using all information available until *h*-48, *h* being the hour to predict. Currently, grid operators have to send their production plan in the morning.

It is important to note that real weather measures in the area are not available. Therefore, considering weather information, only historical weather forecasts are used.

C. Construction of the dataset

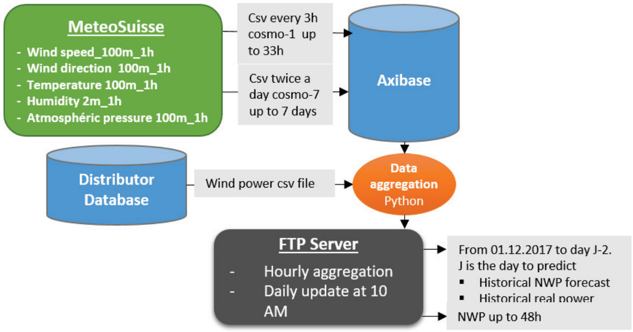
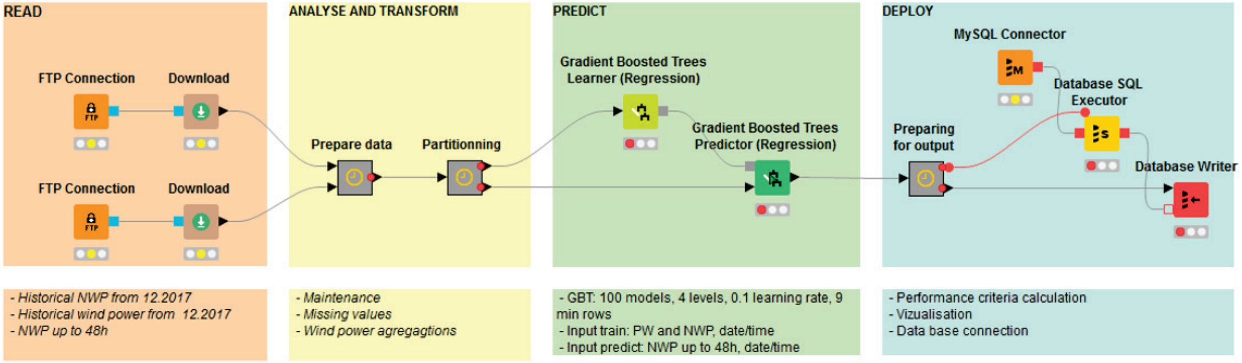


Fig. 2. Construction of the dataset

For the purpose of the paper, the data used ranges from 2016.12 to 2018.01 The dataset construction for the prediction is described in Figure 2.

A python script enables us to gather the NWP data from Axibase and the wind power data from the FTP client. For the wind power, a complete production file is collected per day and joined to the historical production received from the grid operator. For the NWP data, the first step was to build a historical dataset that matches our prediction horizon of 48h. Therefore, the appropriate prediction version of the cosmo-1 and cosmo-2 models have been selected and combined in one file to have a 48h prediction value for each location. As weather information is available for multiple sites around the wind farm, the formatted NWP data have been combined and sent to FTP using BizTalk framework.



- Historical NWP from 12.2017
- Historical wind power from 12.2017
- NWP up to 48h

- Maintenance
- Missing values
- Wind power agregagtions

- GBT: 100 models, 4 levels, 0.1 learning rate, 9 min rows
- Input train: PW and NWP, date/time
- Input predict: NWP up to 48h, date/time

- Performance criteria calculation
- Vizualisation
- Data base connection

Fig. 3. Prediction workflow using knime analytics platform

III. Prediction Methodology

Figure III presents the different stages of the prediction process. The first step is to collect the data from the FTP server. The data is collected since 2016.12. The data is then aggregated per hour, cleaned and analyzed in preparation of the prediction step using the open analytics KNIME 3.5.2 platform. Finally, the 48-h prediction values calculated from 2017.09 are pushed in a SQL database for visualization and storage. For the purpose of this paper, the predictions from 2017.09 to 2018.01 will be presented.

A. Preparing the dataset

$$P = c\rho v^3 \tag{1}$$

Equation 1 describes the wind power production formula: where P is the turbine's power output, c is a constant depending on the feature of the wind turbine, ρ is air density and v is the wind speed. At a first stage, this physical equation enables us to choose the key parameters that directly influence the production: wind speed, air temperature, humidity and density. The wind direction will influence the wind speed projection on the turbine's blades.

Cleaning the dataset means removing errors. Therefore, the first step is to deal with missing values, which occurs only for the real measurements of the power produced. Another source of misunderstanding for the model is the maintenance days that are not related to the weather conditions. As such, they have been manually removed. In total, 15 days' worth of invalid data have then been removed from the dataset due to missing values or maintenance (Dataset B).

In addition to Equation (1), the wind turbine will start and stop given a minimum and a maximum value of wind that is specific to each turbine. In our case, the production occurs between 4 to 20 m/s wind speed. However, due to the high variability of wind speed and direction in the area of the study, there is no particular constant time window where the wind is not strong enough to run the turbines. Moreover, due to errors in wind speed prediction, exploiting this information would lead to more errors as discussed in the results section. In order to help the model understand seasonal trends and anticipate false weather predictions, the average, maximum and minimum production per hour and per month are calculated based on the training dataset and used as new input columns for the training (Dataset C).

At the end of this step, four datasets are created, which are described in Table I. Dataset A is the raw data without cleaning, Dataset B includes the cleaning data step. Dataset C includes the average, maximum and minimum power per hour per month as new columns for training. The particularity of Dataset D is to add the power values of the *J*-2, *J*-3 and *J*-4 as new columns to the training (considering *J* is the day to predict). Therefore, Dataset D is only relevant with the retraining approach explained later in this section. A prediction is calculated based on each dataset and their results are discussed in the Results section.

TABLE I
TRAINING DATASETS USED FOR PREDICTION

Dataset A	**Dataset B**	**Dataset C**	**Dataset D**
Power NWP Date and time	Cleaned Power NWP Date and time	Dataset B Average power per hour per month Max power per hour per month Min power per hour per month	Dataset C Real production of the last three days

B. Gradient Boosting tree ensembles

For each dataset, a gradient boosting algorithm is used with the data described in Table 1. The principle of ensemble prediction is to generate multiple predictions with different extractions of the dataset. The predictors are then combined by voting for classification or averaging for regression [3]. The main advantage of averaging the predictions from several models is that it reduces the variance and prediction error. In our case, each model will be a decision tree.

The boosting method focuses on errors in order to improve the prediction. The weight of the different observations are updated based on the evaluation of the last prediction in order to give more weight to the less predictable items.

This enables us to focus on the most difficult observations to predict for each iteration in order to improve the global prediction accuracy [14]. The parameters to tune are the number of trees, maximum level, minimum number of rows and learning rate. An optimization loop on the Algorithm number of models and level was run to find the optimal parameters results. This resulted in optimal values of 100 Models, 4 levels, a learning rate of 0.1 and 9 minimum rows.

C. Retraining

The process described in Figure 3 is run daily around 10 AM so that each day, the training dataset is updated with the latest real values of wind production. This results in an increase of the training data with the latest values of wind power and weather predictions. It means that to predict day J, the training dataset ranges from 2016.12 to day J-2. Retraining the algorithm takes more time to run than a fixed training step but enables us to start the prediction with a relatively small dataset and improve the prediction as long as more data is collected. In order to keep a cleaned dataset, power value that remains at zero for more than five hours when the wind speed is above 4 m/s are deleted along with the missing values.

Finally, four calculations will be run with the Gradient Boosting Tree algorithm. One with each dataset A, B and C, which have the same granularity and ranges from 2016.12 to 2017.08 This enables us to show the impact of the input dataset in the result. The fourth prediction uses Dataset D and the retraining methodology. This enables us to show the impact of retraining on the prediction accuracy. All predictions are calculated from 2017.09 to 2018.01 with the same parameters of the Gradient Boosting Tree given in the subsection above.

IV. Results

A. Presentation of the results

The results are compared to the persistence approach, which is the most frequently used model to benchmark the performance of forecasting models. This model assumes the forecast *x* time ahead to be equal the real value at *t-x*. In our case, the prediction is equal to the real value 48 hours ahead. As a first stage, the global results are compared to the persistence model for each dataset presented in Table II. Errors distribution are presented in table III. Two graphs then compare the prediction with and without retraining for the best and worst prediction days. Finally, additional calculations are given to identify ways to improve the results. The chosen value criteria are the RMSE divided by 2MW, which is the capacity of the wind turbine (RMSEp), the standard deviation to estimate the error distribution and the hourly absolute maximum error (ErrPeak) as the grid operators sell or buy energy per MW each hour or quarter hour on energy markets.

The prediction slot is 5 months from September 2017 to January 2018. Without the retraining step explained in the section above, the training slot is 9 months from December 2016 to August 2017. With the retraining step, the training dataset will be from December 2016 to *h*-48 hours considering *h* is the time to predict.

B. Prediction performance description

TABLE II
Results comparison from 2017.09 to 2018.01. GBT (Gradient Boosting Trees)

Method	RMSEp	STDEV (MW)	ErrPeak (MW)
Persistance	34 %	0.46	2.5
GBT and Dataset A	25 %	0.31	2
GBT and Dataset B	24 %	0.29	2
GBT and Dataset C	23 %	0.28	1.6
GBT and Dataset D with retraining	17 %	**0.16**	**1.5**

The persistence model has the maximum error values. The Gradient Boosting prediction with the basic dataset beats the persistence by 9 %. Cleaning the dataset enables us to gain 1 % in accuracy by removing the small errors or misunderstandings from the input dataset. Dataset C includes hourly and monthly averages of production power that enable us to gain 1 % more accuracy. One percent of error saved for each wind turbine represents 70kWh saved for the wind park.

Results show that retraining with Dataset D helps us gain 6 % more accuracy on the 5 months prediction slot. The STDEV and ErrPeak also decrease significantly with the retraining step. With nine months' worth of training data, the retraining methodology runs in two minutes to predict the next 48 hours, contrary to one minute without retraining.

Finally, an improvement of 8 % is achieved with the retraining methodology compared to a fixed learning step with Dataset A. 17 % errors are saved with retraining comparing to the persistence model for a 48 hours ahead prediction.

TABLE III
Quartile errors distribution for dataset C with retrained prediction

Method	Q1 (MW)	Q2 (MW)	Q3 (MW)	Q4 (MW)
GBT and Dataset C	0.1	0.2	0.3	0.1
GBT and Dataset D with retraining	0.2	0.3	0.5	1.7

Analyzing the error distribution enables us to estimate how stable or unstable the results are. For the grid stability, two values are important: energy and power. The energy produced should be equal to the consumption, and the power demand lower than the installed capacity. As two separated markets on which the electricity prices change from an hour to another, avoiding high peak errors is important. Table III shows that retraining reduces the error deviation as well as the number of high errors with 75 % errors below 0.3 MW against 75 % below 0.5 MW for Dataset C without retrain.

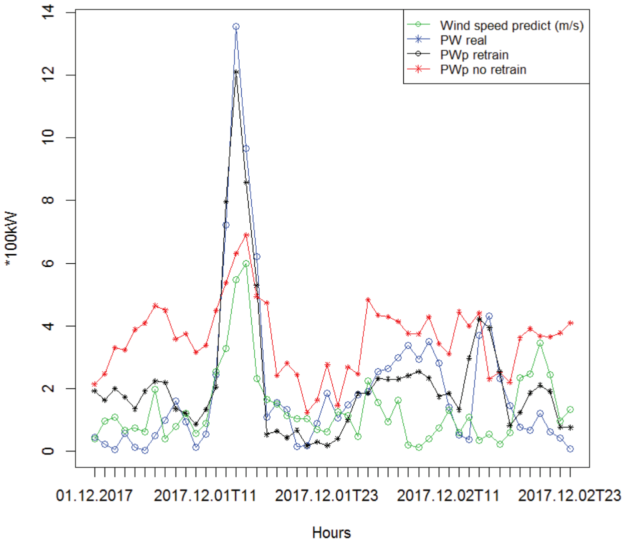


Fig. 4. Best prediction results on 2017.12.02 (day1) and 2017.12.03(day2)

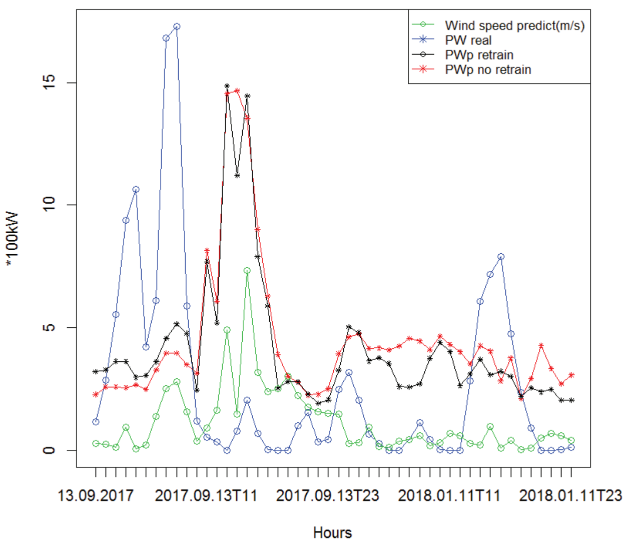


Fig. 5. Worst prediction days on 2017.09.01 and 2018.01.11

C. Results analysis

Figure 4 shows accurate prediction results for two days in december 2017. PWp retrain is the results of the retraining methodology, PWp no retrain is the prediction with the gradient boosting trees and dataset C. The prediction error for day one is 95 %. It was a high production day with a regular pattern of wind speed increasing around lunchtime. The retraining prediction outperforms the normal one in the production peak estimation but also in the low wind speed values. As the training data volume increases, the model extracts more information considering the correlation between wind speed and power production.

Moreover, the mean, minimum and maximum production values for each month and each hour have been given in order help the model anticipate the seasonal patterns and the production boundaries. Day two was a low production day with an accuracy of 96 %. The particularity of retraining prediction for day two is to anticipate best the variation of the production regardless to the wind speed prediction, which is completely underestimated. It shows that the closest information and the production patterns helps anticipating wind speed prediction errors.

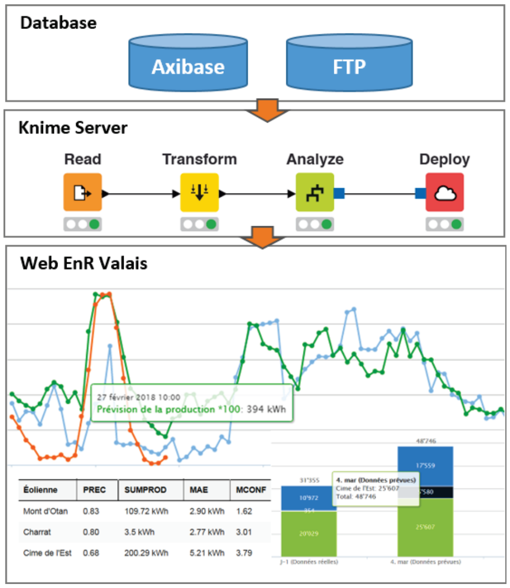
Figure 5 shows worst prediction days. For those days, the two predictions show similar results, which can be defined as a strong delay of the real production. The prediction follows the wind speed prediction, which is incorrect. Moreover, the more recent historical data given in addition to the production pattern is not enough to anticipate wind speed errors. Reducing the time horizon forecast could help correct the first day error as it could help the model to correct the wind speed prediction with the closest real conditions. The second day is a quite low production day which are the most difficult to predict. An incorrect wind prediction close to zero combined with a wind gust phenomenon can result in a drastic increase of the production during a few hours. Wind production turbine occurs with wind speed from 14 to 72 m/s. Day two's prediction shows why implementing this information given the wind speed prediction can result in huge errors.

D. Additional calculations

As the results show, reducing time horizon and focusing on small wind speed prediction can help improve power production. As a result, the accuracy increases by 7 % if the hours for a wind speed below 6 m/s are taken out from the entire dataset. A persistence model for a 1-hour-ahead prediction save 5 % in the prediction accuracy compared to to the retrained model, bringing the global RMSEp to 12 %. Therefore, Reducing the granularity of the prediction and uptdating predictions as often as possible is relevent to increase accuracy.

V. Automatizing

The methodology described above have been implemented for the three wind turbines of the electricity distributor. As wind variance is localized and the wind turbines are not close to each other (at least 10 km between each of them), adding separate predictions with local weather data gives more accuracy to the global prediction. Figure 6 shows the automation



Éolienne	PREC	SUMPROD	MAE	MCONF
Mont d'Otan	0.83	109.72 kWh	2.90 kWh	1.62
Charrat	0.80	3.5 kWh	2.77 kWh	3.01
Cime de l'Est	0.68	200.29 kWh	5.21 kWh	3.79

Fig. 6. Prediction automation process description

process from the collection of data to the visualization web page.

The NWP data from Axibase and the wind power data from the FTP server have been gathered and stored on our FTP virtual machine as explained in section 2. The prediction is calculated using the methodology described in section 3. The resulting data is stored in a MySQL database. This database is part of a three-tier architecture website based on the AMP (Apache-MySQL-PHP) environment. This website acts as a portal for the new renewable energy of the canton of Valais in Switzerland. The resulting data is integrated into this portal and presented in the form of graphics based on the JavaScript charting library 'HighCharts'. The whole process is run daily around 10 AM on KNIME Server, which provides deployment and management functionalities.

VI. Conclusion

The full methodology enables to reach an RMSEp of 17 % for a 48-hour ahead wind power prediction, which is 17 % better than the persistence model. The input dataset aggregations and the retraining steps in our methodology enables us to help the model anticipate wind error prediction and seasonal trends so that 6 % of errors are saved. Nevertheless, it is important to note that increasing the size of the dataset can potentially lead to overfitting. Therefore, the dataset should stop increasing once its optimal size has been reached. The dataset will then be moving to take the latest information and delete the oldest so as to keep the same number of rows. To continue improving the results, focusing on the small wind speed values can save 6 % errors as the results show. Moreover, as wind speed error forecasting is also responsible for huge errors, a great improvement of the weather forecast can be achieved if the real wind speed and direction measures are accessible for the location of the wind turbine. Currently, predictions are available within a 2 to 7km space resolution and having real local values would help anticipate wind gust phenomena and improve the weather forecast as shown in our last publication [15].

Our last recommendation would be to update the prediction as close as possible to realtime as the subsequent hour prediction can save 5 % more errors compared to a 48-hours ahead prediction.

VII. Aknowledgement

This research was undertaken as part of the Renewable Energy Data Integration (REDI) funded by the Institute of Information Systems of HES-SO Valais Wallis. The authors would also like to thank SEIC-Télédis Group and RhônEole SA, for providing the data case studies.

References

[1] GLOBAL WIND ENERGY COUNCIL, *Global wind statistics*, http://gwec.net/wp-content/uploads/vip/GWEC_PRstats2017_EN-003_FINAL.pdf, 2017.

[2] EXPEX SPOT, *Market Data, Day-ahead auction*, https://www.epexspot.com/en/market-data, 2018.

[3] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, *Current methods and advances in forecasting of wind power generation. Renewable Energy*, 37:18, 2012..

[4] THM. El-Fouly, EF. El-Saadany, MMA. Salama, *Grey predictor for wind energy conversion systems output power prediction*, IEEE Transactions on Power Systems 2006;21:1450–2.

[5] G. Giebel, R. Brownsword, RAL; G. Kariniotakis, ARMINES, *The State-Of-The-Art in Short-Term Prediction of Wind Power, A Literature Overview*, Project ANEMOS Contract No.: ENK5-CT-2002-00665, December 2003.

[6] J. P. Heinermann, *Wind Power prediction with Machine Learning Ensembles*, Universitat Olderburg, , p105-130, September 2016.

[7] J. Sousa and R. Bessa, *Comparison of two new short-term wind-power forecasting systems*, ELSEVIER Journal of Renewable Energy, December 2008.

[8] Lionel Fugon, Jérémie Juban, Georges Kariniotakis, *Data mining for wind power forecasting*, European Wind Energy Conference and Exhibition EWEC 2008, Mars 2008, Brussels, Belgium. EWEC, 6 p., 2008. <hal-00506101>

[9] Y. Mao, W. Shaoshuai, *A review of wind power forecasting and prediction*, IEEE, 10.1109/PMAPS.2016.7764085, October 2016

[10] M. De Giorgi, A. Ficarella, M. Tarantino, *Error Analysis of short term wind power prediction models*, Elsevier, Applied Energy 88 (2011) 1298–1311, November 2010

[11] M. He, j. Duan, S. Zheng, *Kaggle Competition: Product Classification*, Machine Learning CS933, 2015

[12] M. Barque, L. Dufour, D.Genoud, A. Zufferey, B. Ladevie and J-J. Bezian, *Solar production prediction based on nonlinear meteosource adaptation*, IEEE, IMIS Conference, Brazil, 2015.

[13] T. Hong, P. Pinson, S. Fan, *Global Energy Forecasting Competition 2012*, Elsevier, International Journal of Forecasting 30 (2014) 357–363, 2012.

[14] Canton du Valais, Production d'électricité,https://www.vs.ch/documents/87616/186657/Production+d'électricité+-+Elektrizitätserzeugung/044f5303-4e72-415f-bb4c-5e03f29bab99, 2016

[15] T. Xia, *Gradient Boosting Machine for High-dimensional Additive Models*, Stanford University, 2014

[16] Luc Dufour, *Contribution à la mise au point d'un pilotage énergétique décentralisé par prédiction*, CNRS, MEGEP UMR 5302 - RAPSODEE , mars 2017

[17] A. Kusiak, H. Zheng, and Z. Song, *Short-term prediction of wind farm power: A data mining approach*, IEEE Transactions on Energy Conversion, 24(1):125 136, 2009.