

Article

Probabilistic Forecasting of Wind Turbine Icing Related Production Losses Using Quantile Regression Forests

Jennie Molinder ^{1,*} , Sebastian Scher ² , Erik Nilsson ¹ , Heiner Körnich ³  and Hans Bergström ¹ 
and Anna Sjöblom ¹ 

¹ Department of Earth Sciences, Uppsala University, SE-75236 Uppsala, Sweden; erik.nilsson@met.uu.se (E.N.); hans.bergstrom@met.uu.se (H.B.); Anna.Sjoblom@geo.uu.se (A.S.)

² Bolin Centre for Climate Research and Department of Meteorology, Stockholm University, SE-106 91 Stockholm, Sweden; sebastian.scher@misu.su.se

³ Unit for Meteorology Research, SMHI, SE-60176 Norrköping, Sweden; heiner.kornich@smhi.se

* Correspondence: jennie.molinder@geo.uu.se or jennie.molinder@gmail.com

Abstract: A probabilistic machine learning method is applied to icing related production loss forecasts for wind energy in cold climates. The employed method, called quantile regression forests, is based on the random forest regression algorithm. Based on the performed tests on data from four Swedish wind parks available for two winter seasons, it has been shown to produce valuable probabilistic forecasts. Even with the limited amount of training and test data that were used in the study, the estimated forecast uncertainty adds more value to the forecast when compared to a deterministic forecast and a previously published probabilistic forecast method. It is also shown that the output from a physical icing model provides useful information to the machine learning method, as its usage results in an increased forecast skill when compared to only using Numerical Weather Prediction data. A potential additional benefit in machine learning for some stations was also found when using information in the training from other stations that are also affected by icing. This increases the amount of data, which is otherwise a challenge when developing forecasting methods for wind energy in cold climates.



Citation: Molinder, J.; Scher, S.; Nilsson, E.; Körnich, H.; Bergström, H.; Sjöblom, A. Probabilistic Forecasting of Wind Turbine Icing Related Production Losses Using Quantile Regression Forests. *Energies* **2021**, *14*, 158. <https://doi.org/10.3390/en14010158>

Received: 23 November 2020

Accepted: 24 December 2020

Published: 30 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: wind energy; icing on wind turbines; machine learning; probabilistic forecasting

1. Introduction

Wind energy production in cold climates has grown exponentially during the last decades and it is expected to continue to increase, owing to the good wind resources and low population density. The total amount of installed wind energy in cold climates at the end of 2015 was 127 GW and is expected to be 186 GW at the end of 2020, which is approximately 30% of the global share of forecasted wind energy capacity [1]. However, icing on the wind turbine blades can cause severe power production losses in these areas. Furthermore, ice increases the load, changes the aerodynamics of the blades, and generates vibrations, which, during severe icing events, can shorten the lifetime of the turbine if it is not shut down [2]. During these occasions, it is important for energy companies to forecast icing related power production losses in order to avoid increased trading costs in addition to production loss costs and keep the power grid stable. Specifically, IEA Wind Task 19 [3] points to the need for site specific forecasts of icing provided at least once a day. This is necessary to avoid sudden unplanned drops of power production leading to unstable grids and for the trading of energy on the spot market. It is also important for the planning of safe park maintenance.

Modelling icing and related power production losses are a challenge, both being due to the initial condition uncertainty as well as uncertainties in the modelling chain, such as model formulations, and due to the representation of the exact wind energy site conditions [4–7]. A common approach for this modelling is to employ an icing model that

forecasts ice load and icing intensity based on meteorological weather conditions that are forecasted by the Numerical Weather Prediction (NWP) model before estimating the power production loss (see modelling chain in Figure 1a) [5,8]. The effect of the different uncertainties in the modelling chain has previously been studied [4,6,9,10]. These studies show that the forecasted production loss can be greatly affected by, e.g., initial meteorological conditions, model formulations, and how the wind turbine sites are represented.

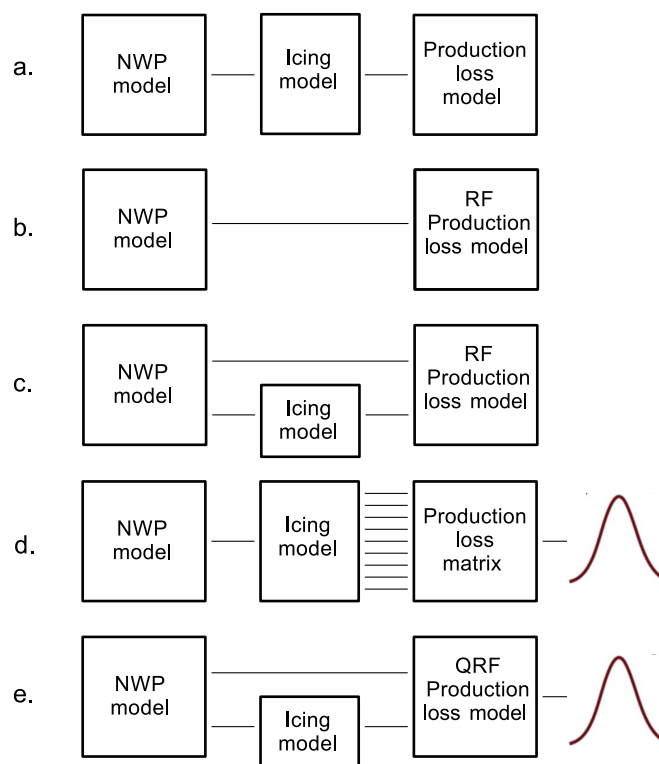


Figure 1. Modelling chain for different methods. (a). The baseline modelling chain for production loss forecasts. (b). The modelling chain used by Scher and Molinder [11]. (c). Adding the icing model to the modelling chain (d). Modelling chain according to Molinder et al. [8] and (e). The modelling chain for the QRF probabilistic forecast. Probability distribution in the end illustrate that both modelling chain d and e produce probabilistic output. RF = Random Forest, QRF = Quantile regression forest.

When dealing with uncertainties in weather predictions, a common approach is to use ensemble forecasts, which often are a form of probabilistic forecasting. In ensemble forecasting, not only one forecast is produced, but a range of possible forecasts that are based on the knowledge of the uncertainties in the modelling steps. Haupt and Monache [12] discuss the value of probabilistic approaches in wind energy in general and how machine learning approaches can be a beneficial tool in generating these forecasts. Probabilistic forecasting has several advantages: first, the ensemble mean forecast generally has a higher forecast skill than a deterministic forecast, because unpredictable parts of the forecast are filtered out; second, the range of the forecasts can be used probabilistically, which means that we obtain information regarding the forecast uncertainty in each forecast step [13]. This uncertainty forecast may be of great value for wind energy trading involving power production loss estimations, because it can provide cost-loss estimations for the decision maker.

Probabilistic forecasting has been used for wind energy purposes for the last two decades (e.g., [14–17]), focusing on the uncertainties in the wind speed forecast. Kim and Hur [16] showed the benefit of probabilistic wind speed forecasting, both in terms of better deterministic forecast results using the ensemble mean and the information of forecast uncertainty for the next-day power production. Wu et al. [17] proposed a probabilistic method that is based on observations and NWP forecasts of wind speed also improving the next-day wind power forecasts. Lately, probabilistic forecasting has also

been used for icing related production losses [4,8]. In Molinder et al. [8], an icing model ensemble was generated that was based on the knowledge of uncertain parts of the icing model. Using this ensemble resulted in a reduced forecast error and it was shown that the forecasted uncertainty in terms of ensemble spread, which was computed as the standard deviation, added valuable information. In Molinder et al. [4], an NWP ensemble and a neighbourhood method accounting for uncertainties in initial conditions and spatial representativeness of the model output was tested with success. Our primary aim in this study is to analyse the usability of machine learning for probabilistic forecasting of wind power production loss in cold climates. We selected the quantile regression forest (QRF) method as the specific method [18]. We examine whether this method can provide improved forecast skill and valuable forecast uncertainty estimations. Because this is done in the last modelling step, the probabilistic forecast, in this case, should account for uncertainties in the whole modelling chain.

Machine learning is a growing field and the use of machine learning models is increasingly frequent for geoscientific modelling in general [19] and it has successfully been applied, for example, in the post-processing step of weather forecasts [20,21] and for predicting the uncertainty in weather forecasts [22]. McGovern et al. [21] show the value of machine learning approaches for real-time decision making in a variety of meteorological applications, such as predictions of thunderstorms and for renewable energy. There are also ideas of making weather forecasts while using purely machine learning [23,24]. In Lee et al. [25], a machine learning based averaging method that was similar to a multi-model ensemble was used for wind energy purposes with good performance. Recently, machine learning has also been used in the modelling of icing and related production losses for wind energy [11,26]. Scher and Molinder [11] presented a machine learning based production loss model while using random forest regression. It was trained on a NWP model and observational data, and it resulted in a comparable forecast skill, as when using a physically based icing model. In Kreutz et al. [26], machine learning was used to forecast wind turbine icing. By using neural networks, they were able to capture most of the icing events. Machine learning has also previously been applied to construct the production loss model in the modelling chain [5]. If the amount of training data would be close to unlimited, it might be possible that a machine learning based model can perfectly forecast the icing itself without including information from an icing model. However, the general problem of the limited availability of both production observations and past NWP data greatly reduce the amount of training data for these purposes and, thus, it is not obvious whether the icing model could add value to the forecast or not. Therefore, a secondary aim in this study is to investigate the value of adding the icing model to the modelling chain (as done in Davis et al. [5]).

Section 2 describes the observational data and test periods, together with the NWP model and icing model. In the Method Section 3, the selected machine learning method is described, the input data specified, and the splitting of the training and test periods discussed. We examine the importance of taking auto-correlation into account when designing train-test splits and whether the data from different stations should be separated in the training or not. The Result Section 4 is divided into three sub-sections: the value of the icing model in the modelling chain, deterministic forecast skill, and probabilistic forecast skill while using the quantile regression forest method.

2. Data

2.1. Observations

The study is based on three periods during cold seasons between 2013 and 2015 with available NWP data, power production observations, and meteorological observations from four wind parks. The periods will further be referred to as period 1–3 and they are shown in Table 1. The studied time periods are somewhat limited by the need for both concurrent NWP model data and quality controlled wind turbine production data. NWP models can be computationally expensive to run for extended time periods, but the greatest limitations

are still in data-availability of concurrent trustworthy production observations. These observations are well-known to be difficult to obtain. The observed power production was related to an estimated potential production to calculate the observed production loss. Figure 2 presents the frequency of the relative production losses. The potential power production was estimated from the observed wind speed measured every full hour (10-min. averages) and power curves that are based on production data from two years, excluding all data with temperatures below 5 °C in order to avoid icing affecting the final curves. The frequency of the observed absolute production losses can be seen in Figure 3 (the zero-production loss data is in this case excluded). The higher production losses are clearly not as frequent as the lower. One problem that can occur with this method for the production loss estimation is that the wind speed measurements can be affected by the icing, and, thus, the estimated observed production loss can be over- or underestimated. The wind speed measurements were conducted while using a heated anemometer at the wind turbines nacelle, but some problems could, despite the heating, be seen with the observations occasionally. If the standard deviation of the 10 min. averaged wind speed was zero, the data were removed, and negative observed production losses were set to zero to account for this. Although our applied nominal quality control routine efficiently removes and adjusts for most problematic time periods for the anemometer data, it needs to be recognised that it is difficult to evaluate without additional complementary measurements that unfortunately were not available. Unfortunately, this is an unavoidable problem that might cause unreasonable forecast errors at times and also affect the training of the machine learning model. An additional potential problem that can cause production loss occurs when the power production is halted or reduced on purpose, because the power grid does not require that specific electricity.

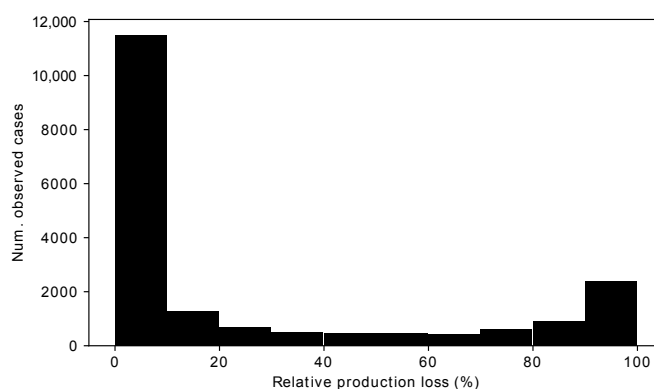


Figure 2. Frequency of observed relative production losses, including all stations and periods.

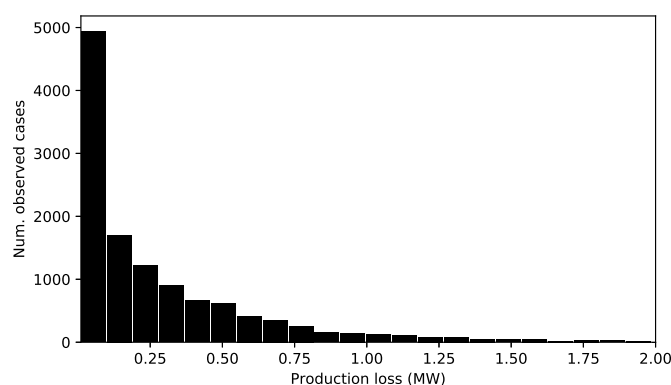


Figure 3. Frequency of observed production losses, including all stations and excluding zero production loss data.

Table 1. The three periods with available Numerical Weather Prediction (NWP) and observational data.

Period	1	2	3
Date	30 December 2013–28 February 2014	2 September 2014–12 January 2015	14 January 2015–18 February 2015

The exact locations of the four wind parks cannot be given due to contractual reasons, but they are located in Sweden and in hilly terrain. We will refer to the sites as sites A–D, where site A is located around 65° N, B in the middle of Sweden, and site C and D in northernmost Sweden. None of the sites have de-icing systems. For each site, the average power production observation from the currently operating turbines is used in the verification of the forecasts. The number of turbines in each park varies between 10 and 30. Turbines that were experiencing problems or were stopped, due to other reasons than icing, such as maintenance, were excluded while using error codes from the turbine data.

2.2. Numerical Weather Prediction Model

The NWP model is the HIRLAM-ALADIN Research on Mesoscale Operational NWP in Euromed Applications of Research to Operations at Mesoscale (HARMONIE AROME) model, cycle 40h11. The model is operationally used at several weather institutes in Europe, e.g., the Swedish Meteorological and Hydrological Institute (SMHI). Our setup contains a horizontal resolution of 2.5 km and 65 vertical levels and it is non-hydrostatic. Bengtsson et al. [27] further describes the model. Figure 4 shows the model domain. Here, the lateral boundary conditions come from the European Centre for Medium Range Weather Forecasts (ECMWF) and it has a horizontal resolution of 30 km.

For trading, it is especially important to forecast the power production for the next day as accurately as possible. Therefore, we use the 19- to 42-h forecasts that were initialised at 06 UTC with an hourly output. The 06 UTC forecast was chosen, because it provides a next-day forecast before 12 UTC, which is a common requirement in trading.

Even though the NWP model has a high resolution, the model terrain differs from the real wind turbine site terrain. Thus, the same height interpolations of the meteorological parameters were done, as in Molinder et al. [8].

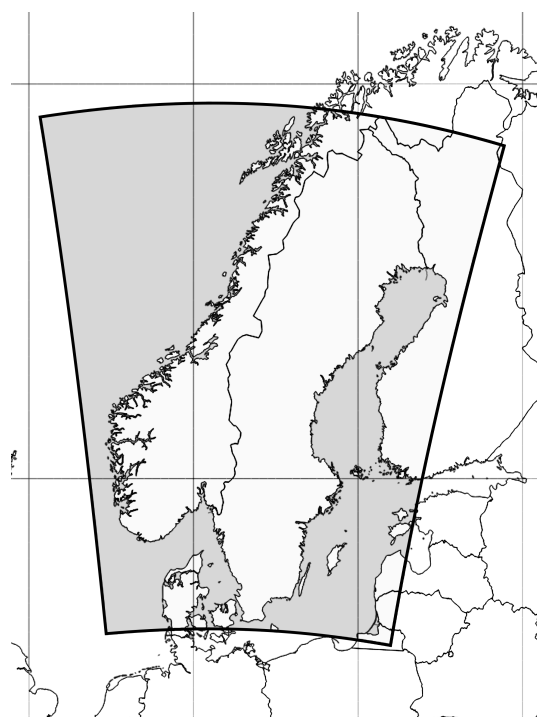


Figure 4. NWP model domain. The domain is centered over Sweden and cover a large part of the rest of Scandinavia.

2.3. Icing Model

The icing model is the one used in Molinder et al. [8]. It is based on the equation that was first described by Kuroiwa [28] and often referred to as the Makkonen equation [7], and it is built according to ISO standards [29] forecasting icing on a rotating rod. The equation used here has ice accretion due to snow, graupel (heavily rimed snow particles), cloud ice, and rain, in addition to cloud water. It also has an ice loss term that accounts for ice falling off the blade, sublimation, melting, and wind erosion.

$$\frac{dM}{dt} = \sum_{i=1}^5 \alpha_{1_i} \alpha_{2_i} \alpha_{3_i} W_i v A - L \quad (1)$$

where $\frac{dM}{dt}$ is the icing accumulation rate, further called icing intensity, the three α parameters are different efficiency coefficients [6–8,30], W is water content, A is the cross-sectional area of the object that increases with ice load, v is wind speed, and L is the ice loss term. The sum represents ice accumulation due to the different hydrometeors: cloud water, snow, graupel, cloud ice, and rain. The efficiency coefficients are: the collision efficiency, the sticking efficiency, and the accretion efficiency. The collision efficiency describes the ratio of e.g., cloud water on the wind ward side of the object that will collide with the object. The sticking efficiency determines the ratio of the colliding particles that will stick to the surface. For liquid water, the sticking efficiency is unity, while, for ice particles, it is lower. The accretion efficiency describes whether the ice or liquid particles freeze to the surface immediately, or whether some will fall off before freezing. A detailed description of the model can be found in Molinder et al. [8]. The model uses temperature, wind speed, water content of the different water components, the size of the hydrometeors according to median droplet volume, air pressure, and relative humidity as input.

3. Method

3.1. Machine Learning Method—Regression Forests

Decision tree based machine learning methods, such as the random forest regression method, is widely used machine learning methods [31–33]. In contrast to a simple regression tree, it fits several, often large number, of regression trees. Each tree is fit on a different random subset of the training data. Additionally, at each node in each tree, a random subset of input variables is used. The forecast is then based on the average forecast of the trees. One important property of random forest regression is that the range of the output variable is limited by the range of the target training variable. This means that the prediction cannot exceed the smallest and largest values in the training set. While, in some contexts, this is a limitation; in our case, it can be a desirable property. This is because, the output variable, for us, is production loss, which, per definition, is bounded. When using random forest regression, our prediction is thus automatically bounded, since, in the training set, the production loss is only varying in the range between zero and maximum power production.

In Scher and Molinder [11], four different methods were tested for the forecasting of icing related power production losses. Among the four methods, random forest regression [34] performed the best. Based on this result, we decided to use the probabilistic machine learning method quantile regression forest (QRF) [18]. This method is a probabilistic extension of random forests and it has previously been used in e.g., Rasp and Lerch [20]. Specifically, the method gives not only the conditional mean, as the general random forest method does, but also an estimation of the quantiles for the predictive values. Figure 5 shows the single tree in the forest. The model needs to be fitted only once. At prediction time, any desired quantile of the prediction can be specified. One advantage of this method is that it does not assume that the predicted value is normally distributed, or that it has a normally distributed uncertainty. We see this as a benefit, since normal distribution is not a good assumption for production loss. It should be noted that the QRF method does not provide a better mean forecast. Accordingly, with this method, the reason for probabilistic

forecast is really the forecast uncertainty estimation. This differs from a NWP ensemble, where the ensemble mean also often outperforms a deterministic forecast while using the same model.

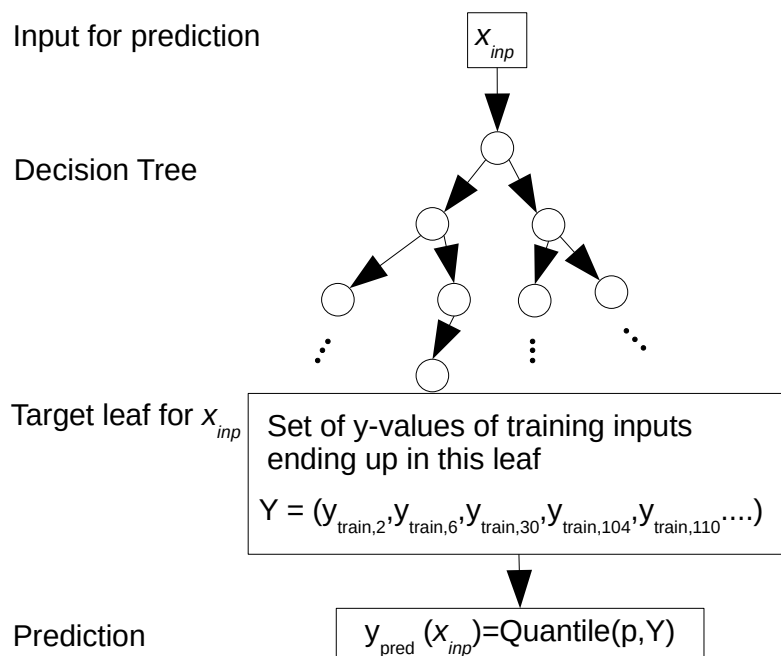


Figure 5. Illustration of how a single tree of a (trained) QRF is used to make a prediction. The decision tree is used to choose the leaf node that corresponds to the input via applying the learned decision rules. The leaf node contains the set of y (prediction) values of all training samples that would end up in this leaf node when predicting on the training data; here, an example of training sample 2, 6, 30, 104, and 110 ending up in the node is shown. The prediction value for the input is the desired quantile of this set. In a normal regression tree, the mean value of the set of y-values is used as prediction. In order to predict with the whole forest, each decision tree is parsed separately, and then a weighted quantile over all y-values in all the chosen leaf-nodes is computed.

Because of the random components in the training process, the fitted model, and, thus, also its prediction accuracy, will be slightly different each time that the training is repeated. Because of this, we train an ensemble of ten quantile regression forests each time we fit the model. The mean prediction is then used in the forecast verification and tests.

3.2. Training and Test Set Data

3.2.1. Training Features

Here, we tested adding the icing model from Molinder et al. [8] that is described above to the modelling chain before training the machine learning model. The output from the icing model, ice load, and icing intensity then form the input features (=input variables) of the machine learning model, together with the wind speed, wind direction, temperature, air pressure, and relative humidity from the NWP model. Absolute production loss is the target variable. Meteorological observations are not used as input to the machine learning model, only to estimate observed production loss. In Scher and Molinder [11], the random forest regression method was used in order to generate production loss forecasts trained only on NWP model data (Figure 1b). The normal random forest regression here uses the same training and test data, as in Scher and Molinder [11], to be able to compare the studies. The two modelling chains can be seen in Figure 1b,c. The icing model and training features above were further applied in the modelling chain for the QRF method (Figure 1e).

3.2.2. Splitting the Training and Test Data Using k-Fold cross Validation

Regarding the splitting of the data into training and test sets, there are some options and limitations. One question is whether it is better to train on the stations separately, while assuming that the location of the stations influences the resulting model to the extent that they should be treated separately, or whether the value of having four times as much data in the training when combining the stations compensates for this difference in location. Three options of separating the station data were tested. Firstly, we treated the station data as independent and trained on all data. Secondly, data from all the stations were still included in the training, but an additional feature, station name, specifying which site the data belong to was added in the training. This was realised with providing the QRF with an additional categorical input variable that has four unordered levels (representing the four sites). The third separating option was separating the training for each site, creating one model for each station. During validation, we also examine the performance for each station separately. The results from the three choices are shown and discussed in Sections 4.2 and 4.3.

Weather prediction data are autocorrelated, which lead to another splitting related issue, namely avoiding data leakage from the training into the test set. If we test on a certain day, and the training set contains the day before and after this day, then the skill on the test day might be over-optimistic, because the two days in the training set are so similar. Thus, choosing data randomly for training and test without any further separation would lead to an overestimation of model skill and an overfitted model. If this model was later used on completely new data, the skill would deteriorate. One way to avoid this is to divide the data into chunks of several consecutive days before the training. This was done in Scher and Molinder [11], who used 10-day chunks and showed similar skill as when performing an "on the fly" training on all the data from earlier dates and times. Here, we also examine how the model skill changes with different chunk lengths. Figure 6 shows the forecast error in terms of root mean square error (RMSE) for the test set for different choice of chunk lengths. It is clear that not adding any chunk lengths to the data splitting (which results in completely random splitting) gives the lowest RMSE, as expected, but the model is probably highly overfitted using this option. After choosing five or more days, the RMSE is not changing significantly anymore. We are aware that some data leakage occurs between the edges of the chunks, but, by using a chunk length longer than five days, the frequency for this is low and it should not significantly affect the average forecast skill. Even if the RMSE is higher for the longer chunks in the figure, this does not imply that the RMSE for new test periods would be better for shorter chunks. On the contrary, the model can, for shorter chunks, become overfitted to the training data and perform worse for new data.

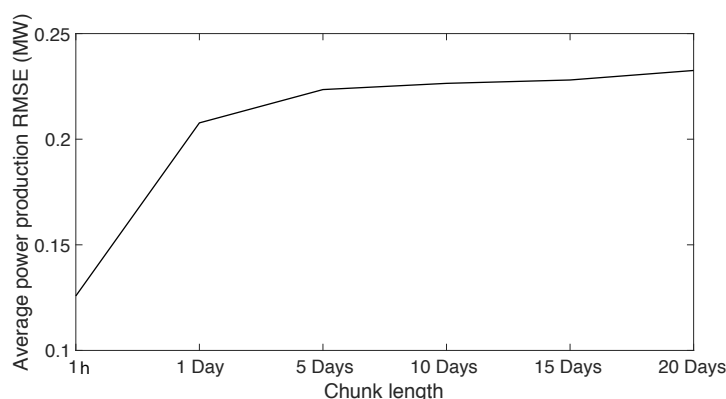


Figure 6. Root mean square error (RMSE) between forecasted and observed hourly production loss while using different lengths for the random chosen chunks of training and tests sets.

Another splitting option is to divide the data into two winter periods while using similar winter months to test and train on, but from different years. This approach was

conducted by splitting up the data into two periods: the first period, referred to as 2013–2014, is period 1 in Table 1, and the second period, referred to as 2014–2015, is a combination of periods 2 and 3 in Table 1, ranging from 1 December to 18 February. This means that only data from December, January and February are tested on, thereby excluding the autumn period in 2014, since this period, in general, has lower icing amounts than the winter months. Table 2 presents the results in terms of RMSE from these training and test sets, together with the RMSE calculated for the same periods but using a 10-day chunk method several times. Here, the 10-day chunk method has a higher forecast skill than the two-winter-period method. Given the differences between the two methods shown in Table 2, there is still some concern for data leakage when using 10-day chunks. However, the difference might result from the model having less data to train on with the two-winter-period method and that having data from only one winter period to train on is probably not enough to cover different weather scenarios. It should also be noted that, during the 2013–2014 period, there is a three-week period, where the forecasted weather by the NWP model has low temperatures and nearly no liquid water content. This led to almost no forecasted icing, while the observed production loss was still high. The too low forecasted liquid water content for low temperatures is a known problem with the employed version of the Harmonie NWP model [27]. This can also cause a higher forecast error, since this period is either only in the test data or only in the training data. With the 10-day chunk method we get some data from both winters in the training and also more data in general. Therefore, we choose to use the 10-day chunks in the study.

In Section 4, all of the results are based on the 10-day chunk method, training on 80% of the data and testing on 20%. This 10-day chunk split was done five times, choosing different 20% test sets and 80% training set, basically performing a k-fold cross validation with $k = 5$. For each different time, a different part of the data is in the test set, which means that, after five splits, all of the data have been in a test set once. We did this kind of cross validation both to be able to show the complete time series of the forecasts and also obtain a better estimate of the skill of the method.

Table 2. Splitting the training and test set into two winter periods compared to using the 10-day chunk method and only including the same periods when averaging the RMSE. RMSE between forecasted and observed hourly production loss.

Winter Period	2013–2014	2014–2015
Average RMSE two winter periods method	0.34	0.32
Average RMSE 10-day chunk method	0.29	0.27

3.2.3. Tuning of the Model

The machine-learning methods contain certain hyper-parameters that need tuning. These parameters determine, for example, the maximum depth of each tree. The number of estimators, maximum depth, maximum features, and minimum split of samples are the parameters that were tuned here. The other hyper-parameters were set to default. For the tuning, we used three-fold cross-validation. The data are split up into three equally large chunks. The model is then trained on 2/3 of the training data and evaluated on the remaining 1/3 for different hyper-parameter combinations. This process is repeated three times and the average skill on the evaluation sets is then computed. The hyper-parameter combinations that has the best average skill is then selected. This means that the test data are not used in the tuning at all, which is necessary for avoiding leakage into the test set. The tuning was done separately for all five different 80% and 20% splits of training and test sets, as explained above, to avoid the usage of test data in the tuning.

3.3. Reference Forecasts

A common reference forecast when analysing forecast accuracy of a new method is to compare the forecast accuracy with a persistence forecast accuracy. Here, we calculated

the persistence while assuming that the observed production loss in percentage at the initialisation time of the forecast is the same throughout the whole forecast length. Even if this is not the optimal reference forecast, it is not unreasonable. When the ice has accumulated on the turbine, it can stay for a few days before it melts or falls off. In this case, the persistence forecast would perform very well. Another reference forecast used here, mainly for the probabilistic forecast skill part, was the probabilistic forecasting method that was used in Molinder et al. [8] based on perturbations of the icing model parameters. In Molinder et al. [8], the icing model ensemble was used as the input to a statistically generated production loss matrix in order to obtain the production loss forecast according to the modelling chain in Figure 1d.

4. Results

The results are divided into three sections: Section 4.1 presents the possible benefit of adding the icing model to the modelling chain while using the basic random forest regression method, Section 4.2 presents the deterministic forecast performance while using the probabilistic QRF model, and Section 4.3 depicts the probabilistic forecast performance. The difference in forecast skill using the different station data separating options are presented and discussed in Section 4.2, and Section 4.3 presents the results from the QRF model. The training and test sets were randomly chosen from 10-day chunks of the data, training on 80% and testing on 20%, throughout Sections 4.2 and 4.3 according to the discussion above. All of the results are based on the 19- to 42-h NWP forecasts initialised at 06 UTC. The main evaluation metric used throughout the study is RMSE of the forecasted power production loss in Megawatt (MW).

4.1. Adding the Icing Model to the Modelling Chain and Relative Importance of Input Parameters

Here, we employed two approaches: one using only the NWP data and one using both NWP data and the output from the icing model. We calculated the RMSE for the same 20% test data as in Scher and Molinder [11], as the first approach is taken from this article. The results are presented in Table 3. When including the icing model, the RMSE for all stations is much lower than when only using NWP data. For stations B–D, the RMSE is even reduced by more than 50%. The results stress that the icing model provides valuable information to the modelling chain when using the random forest method to forecast icing related production losses. Our interpretation is that the icing model helps the machine learning algorithm to recognise icing conditions. With a larger amount of training and test data, this could change, since the machine learning algorithm could be able to find the correlations between the meteorological parameters from the NWP model and the icing conditions itself. Based on these results, the icing model was included in the modelling chain for the training of the probabilistic machine learning model QRF throughout the rest of the study.

Table 3. The forecast skill in RMSE tested on the same 20% set of the total amount of data, as in Scher and Molinder [11].

Station	Including Icing Model	No Icing Model (Scher and Molinder [11])
A	0.25	0.32
B	0.14	0.36
C	0.15	0.34
D	0.18	0.43

The importance of each input feature is an optional output when using the random forest regression algorithm [34]. According to this, the wind speed and the ice load (not shown) are the most important features when forecasting the production. After this follows the wind direction, temperature, air pressure, and icing intensity being of equal importance, while the relative humidity is of less value.

4.2. Deterministic Forecasts Skill

The results for the three station separating options from Section 3 are shown in Figure 7 for each of the wind park sites and their average. The figure presents the average production loss forecast RMSE in MW for the deterministic forecasts—i.e., only using the conditional mean—for the QRF model. The values in the parenthesis are the range (min-max) of the forecast error for the five test/training sets. It should be noted that this variation is highly dependent on the test data. Different meteorological conditions in the different test sets cause a considerable variation in the RMSE, while some variation can be attributed to the five QRF models that are tuned for each separate data set. For station C, there is, on average, no difference between the three separating options, while, for station B, the option of separating station data seems to be the best. This might occur, for example, due to biases in the forecasted data of wind speed or icing amount at station B that the machine learning algorithm can correct when separating the station data. Station D seems to benefit from using all station data in the training, while the results for station A are more similar to station B. Station D has, in general, slightly less icing related production losses compared to station A, which could explain this difference. However, station B has even lower production losses, thus this is not a conclusive explanation for the differences. In general, differences between the stations can be because of their geographical location that results in a more specific needed model for some stations.

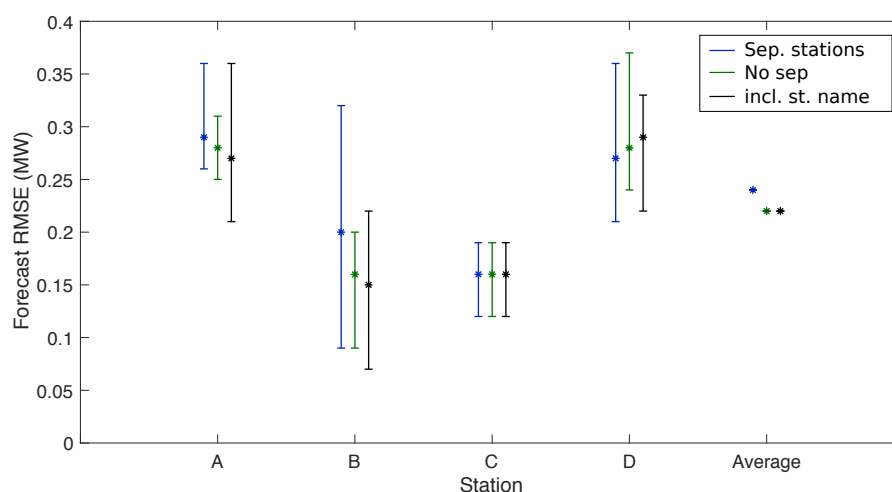


Figure 7. RMSE of the production loss forecast in MW for different sites and the three station data splitting options. The markers show the mean RMSE and the bars present the minimum-maximum of the five test/training sets. The “Average” markers is the average of the station data.

Based on the overall average RMSE presented in the figure at the right, there does not seem to be a difference in model performance between training on the station training data separately or while using all of the data, including the station name. The result however varies between the stations, for station B, for example, it might be good to exclude the other stations in the training, while station C shows equal results for the different options. The difference between the two best options according to the average RMSE and the RMSE for the option with no station separation is also relatively low, suggesting that any option could be chosen. Here, we are also interested in obtaining the forecasted uncertainty as accurately as possible. Therefore, when choosing a station separating option, the forecasted percentiles by the QRF model should first be validated. Section 4.3 discusses this further.

The mean QRF forecast error is compared to the persistence forecast error (Section 3.3) in Figure 8. The figure shows the average RMSE for each forecast length (19- to 42-h) while using the separating option including station name. The shaded area in the figure shows the uncertainty in terms of the 2.5–97.5th percentile of the RMSE estimate at each forecast length based on bootstrapping. Here, the QRF model outperforms the persistence forecast.

The average deterministic forecast skill of the QRF model also exceeds the skill of the ice model ensemble that was used in Molinder et al. [8] (not shown). A reason for this might be that the machine learning algorithm can correct for possible biases in the wind speed data. However, the correlation between the mean forecast and observations is also higher for the QRF model when compared to the method in Molinder et al. [8] (not shown), which implies that the QRF method provides a more reliable and valuable production loss forecast.

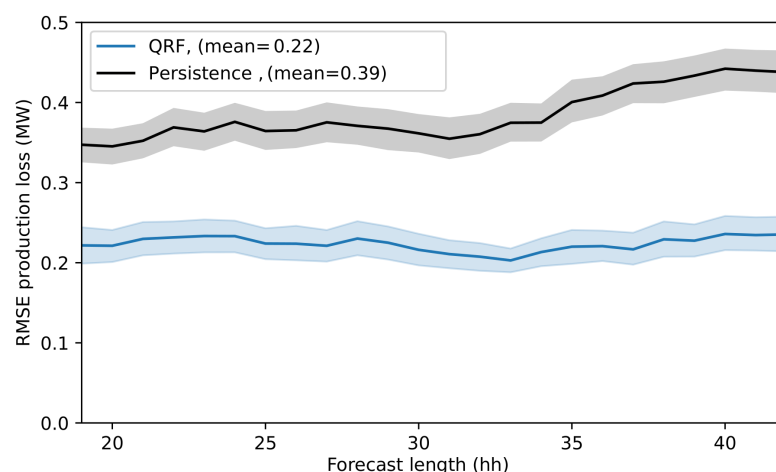


Figure 8. Persistence forecast RMSE vs QRF including station name option RMSE. The shading shows the 2.5–97.5 uncertainty range estimated with bootstrapping.

4.3. Probabilistic Forecast Skill

Figure 9 shows an example of the QRF forecasted time series for the separating option including station name in the training (site A). It also includes a light shaded area between the forecasted 10th and 90th percentile and a dark shaded area between the 30th and 70th percentile. A known limitation of random forest regression methods, which arises from the fact that the forecast is limited to the range of training values, is that the model will not be able to extrapolate to extremes that it has not seen before. It is clear that the model has a problem with the extremes for some dates, for example, around the 10 January and 2 February, while it captures other extremes, such as around 5 January. On average, for all sites and periods, the QRF mean forecast also displays lower variability of the production losses than the observations. When applying a random forest regression model with a limited quantity of data, this should be taken into account when employing the forecasts. More training data might amend this issue.

A commonly used method for verifying a probabilistic forecast in the meteorological community is to compare the forecast skill (RMSE) with the forecast spread; the spread is defined as the standard deviation of the ensemble members. Because we are forecasting percentiles from distributions that are not necessary gaussian, and not the actual standard deviation of these distributions, it is not possible to estimate an exact standard deviation and, thus, not possible to make this kind of direct comparison. However, we can still calculate a mean spread, defined here as the standard deviation between percentile 95, 85...5, for the different station separating options. This was used in order to determine whether the different station separating options result in different spreads between the percentiles in general. Figure 10 shows the average spread from the definition above and the RMSE of the QRF forecasts for each next-day forecast length. The two separating options including all station data, have, on average, a larger spread between the percentiles (black and green dashed lines) compared to the option of separating the station data (blue dashed line). This is to be expected, since the option separating the station data has less training data and, thus, less information regarding forecast variability than the other methods.

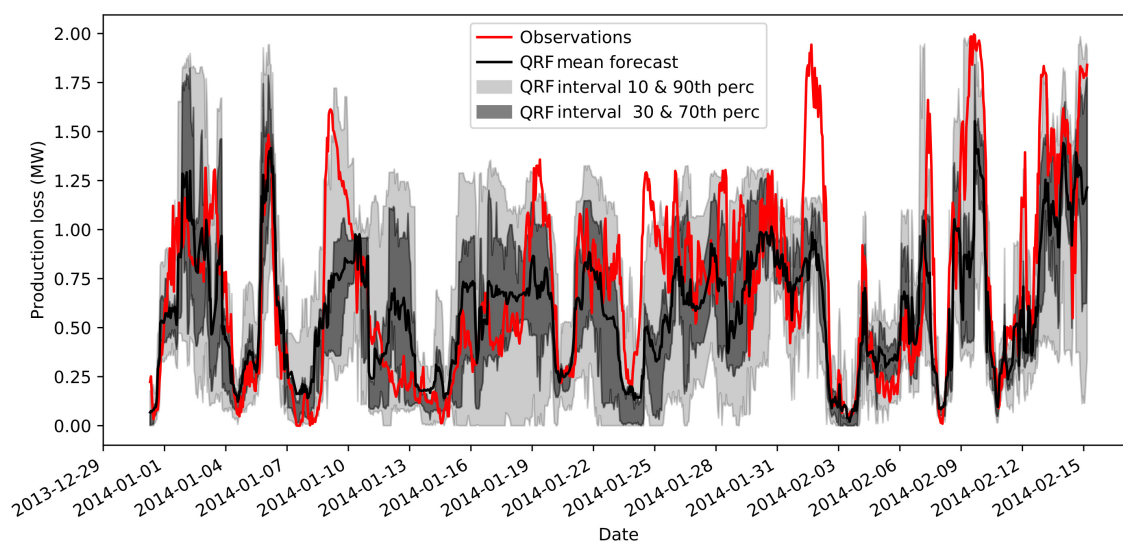


Figure 9. Probabilistic forecasts with the QRF (site A). All data and station name was used in the training. Light shaded area show 10–90th percentile range and dark shaded area 30–70th percentile of the forecast.

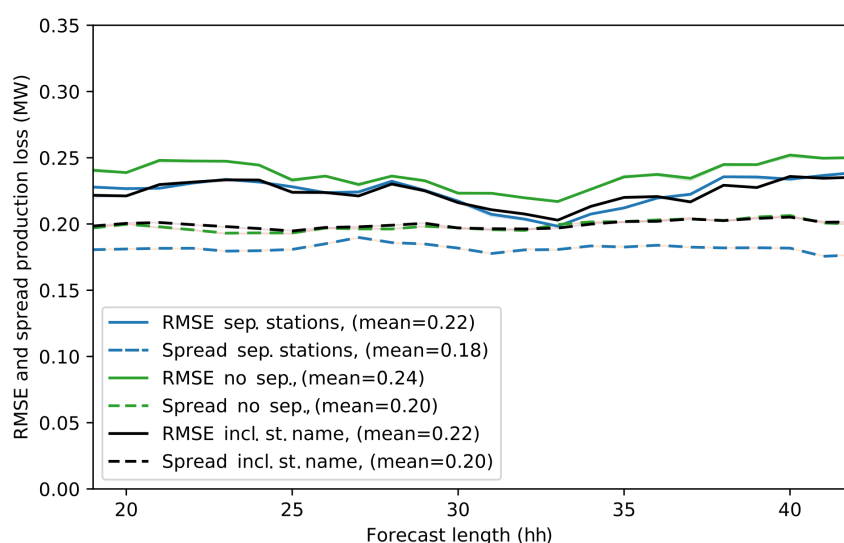


Figure 10. Spread in terms of the standard deviation of percentiles 95, 85 ... 5 and forecast error in RMSE of the production loss forecast relative to observations for the next-day forecast lengths.

While the separating option including the station name and the option of separating the station data, has the same mean RMSE (Figure 10, blue and black line), the option of not separating the station data (green line) has a higher RMSE, as discussed in Section 4.2. In order to determine which of the two separating options with lower RMSE is to be preferred as a probabilistic forecast, it first has to be determined if a higher spread is desirable or not. A probabilistic forecast should, if it is perfect, show the whole uncertainty distribution; otherwise, the forecast is under dispersive and the uncertainty is underestimated. However, it is also important that the uncertainty is not overestimated, since this would result in a useless uncertainty forecast.

A rank histogram can be used in order to determine whether a probabilistic forecast is under- or over dispersive. The rank histogram shows the relative frequency of observations occurring in equally large bins of the forecasted distribution. Each bin should optimally

have the same relative frequency of correct forecasts if the bins have equal size [13]. However, this is assuming that the observational error is zero, which we know, is not true. The estimation of the observational error is not trivial, since technical details of the wind parks have to be known. This is not possible here, and it has to be addressed elsewhere. Another assumption is that the forecast bias is small, since the forecasted uncertainty does not include a bias. Here, the forecast bias is low (not shown). The above limitations means that not all of the uncertainties are included in this verification. Thus, the rank histogram could show a slight convex form. Figure 11 shows the rank histogram for the two station separating options with the same average RMSE, i.e., the option including station name (black) and the option separating the stations in the training (blue). In this case, the bins are constructed from the percentiles 90, 80 ... 10, which result in equally sized bins. The relative frequency is clearly higher for the bins at the edges for both splitting options, which implies, while assuming that the observational error is low, that the probabilistic forecasts are slightly under dispersive. The separating option including all data but adding station name to the training is less under dispersive and is, thus, better calibrated and the best option according to the rank histogram, if a good probabilistic forecast is the aim.

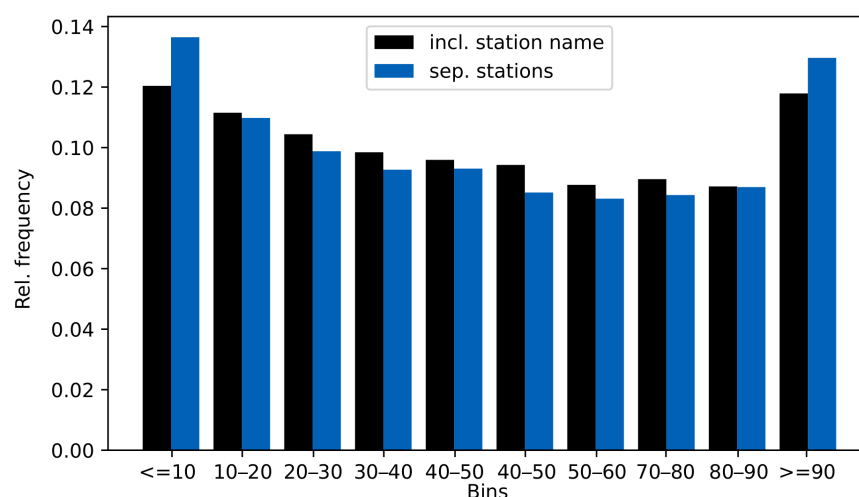


Figure 11. Rank histogram based on percentiles 90, 80, ... 10. Relative frequency of observations in each of these bins.

A rank histogram shows if the forecasted uncertainty is, on average, well calibrated. Based on the rank histogram, a higher spread between the percentiles are desirable, since the rank histogram shows that the forecasts are under dispersive. However, we also want the probabilistic forecast to have as high resolution as possible meaning that a high forecasted uncertainty occurs when the forecast is highly uncertain. If the forecast uncertainty is randomly forecasted, the probabilistic forecast is not useful and could, in the worst case, misguide the forecaster, for example, when applying the forecast that is uncertainty to make trading decisions. Thus, further verification is required.

The pinball loss function is a metric that can be used in order to assess the accuracy of a quantile forecast. The pinball loss (L) is defined as;

$$L = (y - q)\tau \quad \text{if } y \geq q \quad (2)$$

$$L = (q - y)(1 - \tau) \quad \text{if } q > y \quad (3)$$

where τ is the target quantile, q is the quantile forecast, and y is the observation. It is similar to the rank histogram, verifying the performance of each quantile, but it takes the size of the forecast error for the quantile into account. A low pinball loss is desirable. However, the pinball loss would never become zero. Figure 12 shows the pinball loss for the same splitting options as the rank histogram, separating the stations or including station name.

The pinball loss is lower for all quantiles for the splitting option separating the stations compared to the option including station name, which suggests that the splitting is the best option. Thus, the result is opposite to the rank histogram. This means that it is not possible to state which of them is preferable from this analysis. It should, however, be noted that the difference between the two splitting options is low for both the rank histogram and pinball loss.

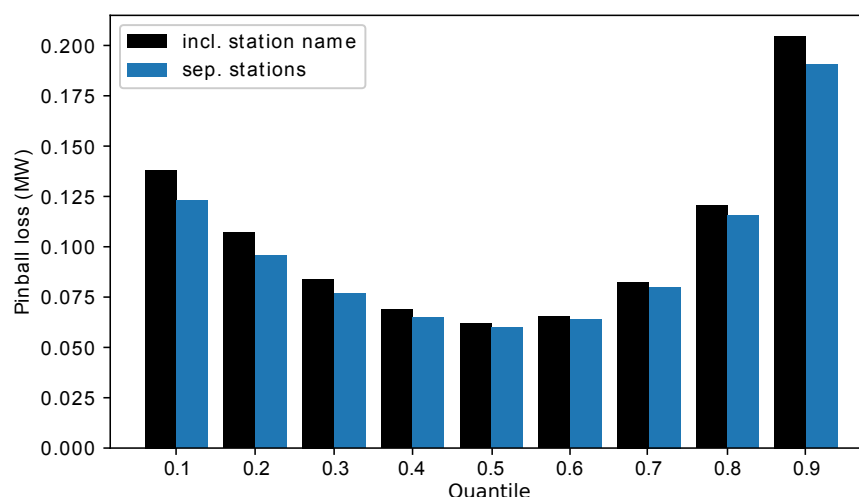


Figure 12. Pinball loss for quantile 0.1, 0.2 ... and 0.9 for two of the splitting options.

Table 4 shows the correlation between the forecast error and the forecasted uncertainty. The forecasted uncertainty for the QRF method is presented in terms of the difference between specific percentiles. For the icing model ensemble used in Molinder et al. [8], it is presented as the forecasted standard deviation. Here, we assume that the interval 84.15–15.85th is an approximation of the standard deviation of the QRF probabilistic forecast and, thus, comparable with the results for one standard deviation from Molinder et al. [8]. High positive correlation between the forecasted uncertainty and forecast error is desirable, since the forecasted uncertainty relates to the appropriate level of trust in the forecast quality. However, it should be noted that the correlation in absolute terms can be weak even for a perfect probabilistic forecast, since it depends on the weather situations [35]. According to Table 4, the forecasted uncertainty seems to be better for the QRF method than for the icing model ensemble presented in Molinder et al. [8], since it has a higher correlation between the forecast error and uncertainty. These results depend on the same periods and, thus, weather situations, and they are therefore comparable. The improved correlation for the QRF methods is significant (significance level 0.05, tested with bootstrapping and taking the non-independence of the methods into account). There is nearly no difference between the correlations for different station separating options, which is also the case if other percentile intervals are chosen as a measure of forecasted uncertainty (not shown).

Table 4. Average correlation (Pearson correlation) between forecast error and the 84.15–15.85th percentile interval for the three station data separating options and between forecast error and one standard deviation forecasted by the icing model ensemble utilised in Molinder et al. [8].

Intervall Compared to Forecast Error	84.15–15.85th perc.	1 std
No separation of station data	0.62	
Including station name as feature	0.62	
Separating station data	0.61	
Icing model ensemble		0.54

Receiver operating characteristics (ROC) can be examined to further verify the probabilistic forecasts. The ROC curve compares the false positive rate with the true positive rate. The probabilistic forecast performs well when the area under the ROC curve is high [13]. This means that, if the forecasted probability for a certain production loss is high/low, then the observation confirms that the loss occurred/did not occur. For a perfect deterministic forecast, the area under the ROC curve would be 1.0 and for a random forecast the area would be less than 0.5. A ROC curve can be constructed for a deterministic forecast, in that case the "probability" in the forecast is either 1 or 0. This can be used to compare it with the probabilistic forecast in order to verify that the probabilistic forecast is valuable. A ROC curve is based on a certain choice of threshold or classification. Here, we used the threshold of 0.5 MW production loss. This threshold was chosen, as its occurrence rate ensures a sufficient number of positive events to get a stable estimate of the ROC curve. Figure 13 shows the ROC curve for the probabilistic QRF forecasts for the splitting option including station name in the training (blue), its deterministic mean forecast (black), and the ROC curve for the probabilistic method used in Molinder et al. [8] (green). The area under the respective curves is given in the box within the figure.

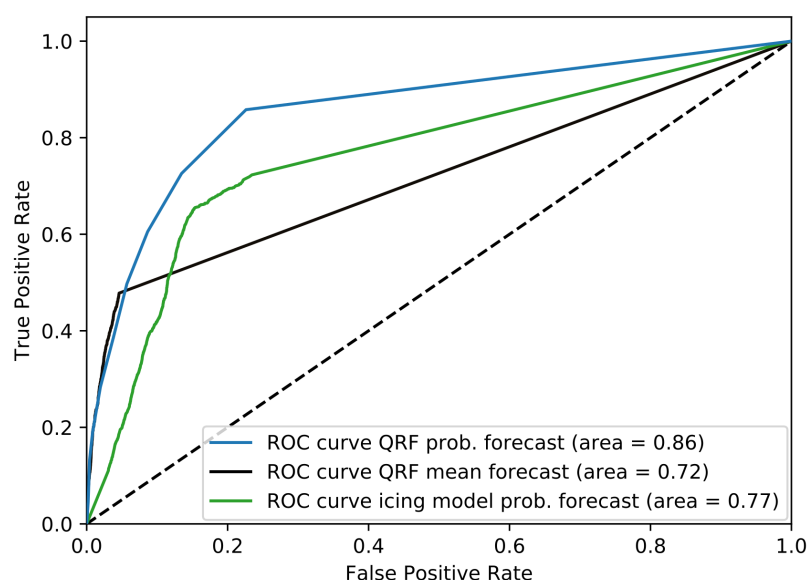


Figure 13. Receiver operating characteristics (ROC) curves for forecasting the probability of production loss > 0.5 MW for the splitting option, including station name, deterministically and probabilistic, and the icing model ensemble in Molinder et al. [8].

First of all, comparing the deterministic QRF forecast and the probabilistic QRF forecast gives confidence that the forecasted uncertainty adds value to the forecast, since the area under the QRF curve is larger. The area under the curve is even lower for the deterministic QRF forecast for higher thresholds, for example, 0.6 if 1 MW threshold is used, as compared to the probabilistic QRF forecast, which keeps a similar score for thresholds between 0.3 and 1 MW and highlighting the benefit of the probabilistic forecast. This means that the forecasted uncertainty is not randomly distributed, but that a high forecast uncertainty (or in this case probability of production losses higher than the chosen thresholds) is generally forecasted when it actually happens. The area under the curve is also higher for the QRF probabilistic forecast when compared to the method used in Molinder et al. [8], implying that the QRF method provides an improved probabilistic forecast.

The forecasted uncertainty for the separating option including station name, is on average more reliable according to the rank histogram than the option where the data are separated in the training, whereas the pinball loss is lower for the option separating the stations, which suggests that this is the best option, as discussed above. However, the area

under the ROC curve for the two splitting options is of similar size for thresholds between 0.3 and 1 MW (not shown). In order to evaluate the possible benefit of the higher spread regarding the splitting option including station name, it would have been valuable to be able to construct ROC curves for really high thresholds, which is not possible, since events exceeding such high thresholds do not occur often enough.

5. Discussion and Conclusions

A machine learning based probabilistic forecasting method, namely quantile regression forests, was used for the modelling of the next-day icing related power production losses for wind energy in cold climates. Additionally, the value of having an icing model in the modelling chain when compared to only using the NWP output in the machine learning was investigated. Different separating options for the four wind parks site data were tested and evaluated. The main conclusions are: 1. The forecasted uncertainty by the machine learning methods adds value to the forecast when compared to the deterministic forecast. 2. When compared to the icing model ensemble in Molinder et al. [8], the machine learning based method tested here shows a greater probabilistic forecast skill, but misses some extreme cases. 3. The icing model output adds value and it should be used in the machine learning method. 4. The model benefits from the information of the station name and from training models for the stations separately as compared to using all data, but not specifying the station.

The presented method increases the utility of predictions of icing related production losses and, therefore, are a valuable contribution to the field of wind energy in terms of improved decision-making in the trading process and can also be used when planning for site maintenance and for operations of de-icing systems. The method can be used for real-time trading, if the user has enough historical data from wind farms. The NWP data from the Harmonie-Arome 06 UTC forecast are available for users approximately three hours later and the icing model and machine learning algorithm are not computationally demanding and can be run on a laptop, if necessary. However, it is important to state that the results above are specific for the data used in this study and should only be seen as a proof-of-concept.

Training data are usually limited and we have seen that it can be beneficial to provide more information by an additional physical model, like the icing model, or to mix the data from several stations. The benefit of the icing model in the modelling chain might change if the amount of data is significantly higher. With more data, the machine learning model might be able to find the relationship between the NWP meteorological output data and icing itself. However, it is impossible to tell a-priori how much data one would need for this. Using data from all stations and adding the station name provides more possible outcomes to the machine learning method. This addresses a common problem in random forest methods, namely that the outcome range is limited to the range of the training data. However, the resulting probabilistic forecast skill only varies slightly between the different station separating options. The additional training data available when using the splitting option including station name, is an advantage for only two sites in terms of RMSE, and for the reliability of the probabilistic forecast according to the rank histogram, while other probabilistic verification tools show the opposite or no difference. Thus, the conclusion is that, when the data are limited—even if the results for only one site is the aim—it is worth testing, including data from other sites in the training. Therefore, we suggest deciding on a data separating option, depending on the application and aim.

The lower variability of the QRF forecasted production loss when compared to the observed production loss, which leads to some extremes not being forecasted, can potentially be a problem for users. Users of the forecast, such as energy companies, should be aware of this limitation in the forecasts. If more training data were available, then this problem would probably decrease—again, however, it is not possible to tell a-priori exactly how much data one would need to obtain an improvement. Finally, one could speculate based on meteorological reasoning on why the method works better for some extremes or not,

we have not attempted to do this here. The small sample size of the extremes would make a validation of these explanations impossible, and it would not be possible anyway to tell whether the difference in performance is pure chance or not.

In this study, we have only used the NWP data height-interpolated to the wind turbine hub height. Another approach could be to include several modelling heights or grid points data in the training, and provide the machine learning model with more information regarding the weather at the site. Other kinds of probabilistic approaches for icing related production loss modelling have also been previously tested, and it would be interesting to combine all methods to increase the probabilistic forecast skill or examine whether the machine learning method already optimises the probabilistic forecast.

Author Contributions: Conceptualisation, J.M. and S.S.; methodology, J.M. and S.S.; software, J.M.; validation, J.M.; formal analysis, J.M.; data curation, J.M.; writing—original draft preparation, J.M.; writing—review and editing, S.S., E.N., H.K. and A.S.; visualisation, J.M.; supervision, E.N., H.K., A.S. and H.B.; project administration, H.K. and A.S.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of the observational data of production losses. Data was obtained from wind park owners and can only be made available with the permission of data owners. Example code used in this study is available upon request from the corresponding author.

Acknowledgments: This work forms part of the Swedish strategic research program StandUp for Wind. We thank S.J. Coulson for his valuable comments on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
NWP	Numerical Weather Prediction
RF	Random Forest Regression
QRF	Quantile Regression Forest
PDF	probability density function
RMSE	root mean square error
ROC	Receiver operating characteristics
MW	Megawatt

References

1. Lehtomäki, V. IEA Wind Task 19, “Emerging from the Cold”. 2016. Available online: <https://www.windpowermonthly.com/article/1403504/emerging-cold> (accessed on 10 February 2020).
2. Kraj, A.G.; Bibeau, E.L. Phases of icing on wind turbine blades characterized by ice accumulation. *Renew. Energy* **2010**, *35*, 966–972. [CrossRef]
3. Lehtomäki, V.; Krenn, A.; Ordaens, P.J.; Godreau, C.; Davis, N.; Khadiri-Yazami, Z.; Bredesen, R.E.; Ronsten, G.; Wickman, H.; Bourgeois, S.; et al. IEA Wind TCP Task 19, Available Technologies for Wind Energy in Cold Climates Report. Technical Report; 2018. Available online: <https://community.ieawind.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=6697b7bd-b175-12b0-ecbf-2558c35d309b&forceDialog=0> (accessed on 28 December 2020)
4. Molinder, J.; Körnich, H.; Olsson, E.; Bergström, H.; Sjöblom, A. Probabilistic forecasting of wind power production losses in cold climates: A case study. *Wind Energy Sci.* **2018**, *3*, 667–690. [CrossRef]
5. Davis, N.; Pinson, P.; Hahmann, A.N.; Clausen, N.E.; Žagar, M. Identifying and characterizing the impact of turbine icing on wind farm power generation. *Wind Energy* **2016**, *16*, 1503–1518. [CrossRef]

6. Nygaard, B.E.K.; Àgústsson, H.; Somfalvi-Tóth, K. Modeling Wet Snow Accretion on Power Lines : Improvements to Previous Methods Using 50 Years of Observations. *J. Appl. Meteorol. Climatol.* **2013**, *52*, 2189–2203. [\[CrossRef\]](#)
7. Makkonen, L. Models for the growth of rime, glaze icicles and wet snow on structures. *Philos. Trans. R. Soc. Lond.* **2000**, *358*, 2913–2039. [\[CrossRef\]](#)
8. Molinder, J.; Körnich, H.; Olsson, E.; Hessling, J.P. The Use of Uncertainty Quantification for the Empirical Modeling of Wind Turbine Icing. *J. Appl. Meteorol. Climatol.* **2020**, *58*, 2019–2032. [\[CrossRef\]](#)
9. Davis, N. Icing Impacts on Wind Energy Production. Ph.D. Thesis, DTU Wind Energy, Lyngby, Denmark, 2014.
10. Thorsson, P.; Söderberg, S.; Bergström, H. Modelling atmospheric icing: A comparison between icing calculated with measured meteorological data and NWP data. *Cold Reg. Sci. Technol.* **2015**, *119*, 124–131. [\[CrossRef\]](#)
11. Scher, S.; Molinder, J. Machine Learning-Based Prediction of Icing-Related Wind Power Production Loss. *IEEE Access* **2019**, *7*, 129421–129429. [\[CrossRef\]](#)
12. Haupt, S.E.; Monache, D. Understanding Ensemble Prediction: How Probabilistic Wind Power Prediction can Help in Optimising Operations. *WindTech* **2014**, *10*, 27–29.
13. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 3rd ed.; Academic Press: Cambridge, MA, USA, 2011.
14. Pinson, P.; Madsen, H. Ensemble-based Probabilistic Forecasting at Horns Rev. *Wind Energy* **2009**, *12*, 137–155. [\[CrossRef\]](#)
15. Traiteur, J.J.; Callicutt, D.J.; Smith, M.; Baiday Roy, S. A Short-Term Ensemble Wind Speed Forecasting System for Wind Power Applications. *J. Appl. Meteorol. Climatol.* **2012**, *51*, 1763–1774. [\[CrossRef\]](#)
16. Kim, D.; Hur, J. Short-term probabilistic forecasting of wind energy resources using the enhanced ensemble method. *Energy* **2018**, *157*, 211–226. [\[CrossRef\]](#)
17. Wu, Y.; Su, P.; Wu, T.; Hong, J.; Hassan, M. Probabilistic Wind-Power Forecasting Using Weather Ensemble Models. *IEEE Trans. Ind. Appl.* **2018**, *54*, 5609–5620. [\[CrossRef\]](#)
18. Meinshausen, N. Quantile Regression Forests. *J. Mach. Learn. Res.* **2006**, *7*, 983–999.
19. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, M. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195. [\[CrossRef\]](#)
20. Rasp, S.; Lerch, S. Neural Networks for Postprocessing Ensemble Weather Forecasts. *Mon. Weather. Rev.* **2018**, *146*, 3885–3900. [\[CrossRef\]](#)
21. McGovern, A.; Elmore, K.L.; Gagne, D.J., II; Haupt, S.E.; Karstens, C.D.; Lagerquist, R.; Smith, T.; Williams, J.K. Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Am. Meteorol. Soc. Meteorol. Soc.* **2017**, *98*, 2073–2090. [\[CrossRef\]](#)
22. Scher, S.; Messori, G. Predicting weather forecast uncertainty with machine learning. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 2830–2841. [\[CrossRef\]](#)
23. Rasp, S.; Dueben, P.D.; Scher, S.; Weyn, J.A.; Mouatadid, S.; Thuerey, N. WeatherBench: A benchmark dataset for data-driven weather forecasting. *arXiv* **2020**, arXiv:2002.00469v1.
24. Weyn, J.A.; Durran, D.; Caruana, R. Can Machines Learn to Predict Weather Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data. *J. Adv. Model. Earth Syst.* **2019**, *11*, 2680–2693. [\[CrossRef\]](#)
25. Lee, J.; Wang, W.U.; Harrou, F.; Sun, Y. Wind Power Prediction Using Ensemble Learning-Based Models. *IEEE Access* **2020**. [\[CrossRef\]](#)
26. Kreutz, M.; Ait-Alla, A.; Varasteh, K.; Oelker, S.; Greulich, A.; Freitag, M.; Thoben, K.D. Machine learning-based icing prediction on wind turbines. *Procedia CIRP* **2019**, *81*, 423–428. [\[CrossRef\]](#)
27. Bengtsson, L.; Andrae, U.; Aspelien, T.; Batrak, Y.; Calvo, J.; De Rooy, W.; Gleeson, E.; Hansen-Sass, B.; Homleid, M.; Hartal, M.; et al. The HARMONIE—AROME Model Configuration in the ALADIN—HIRLAM NWP System. *Mon. Weather Rev.* **2017**, *145*, 1919–1935. [\[CrossRef\]](#)
28. Kuroiwa, D. *Icing and Snow Accretion on Electric Wires*; Research Report; Cold Regions Research and Engineering Laboratory (U.S.): Hanover, NH, USA, 1965; Volume 123.
29. ISO 12494. *Atmospheric Icing of Structures (ISO/TC 98/SC3)*; Technical Report; International Organization for Standardization: Geneva, Switzerland, 2001.
30. Dobesch, H.; Nikolov, D.; Makkonen, L. *Physical Processes, Modelling and Measuring of Icing Effects in Europe*; Zentralanstalt für Meteorologie und Geodynamik: Wien, Austria, 2005.
31. Gagne, D.J.; McGovern, A.; Brotzge, J. Classification of Convective Areas Using Decision Trees. *J. Atmos. Ocean. Technol.* **2009**, *26*, 1341–1353. [\[CrossRef\]](#)
32. McGovern, A.; Gagne, D.J.; Williams, J.K.; Brown, R.A.; Basara, J.B. Enhancing understanding and improving prediction of severe weather through spatiotemporal relational learning. *Mach. Learn.* **2014**, *95*, 27–50. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Williams, J.K. Using random forests to diagnose aviation turbulence. *Mach. Learn.* **2014**, *95*, 51–70. [\[CrossRef\]](#)
34. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
35. Hopson, T.M. Assessing the Ensemble Spread—Error Relationship. *Mon. Weather Rev.* **2014**, *142*, 1125–1142. [\[CrossRef\]](#)