Improving Wind Speed Uncertainty Forecasts Using Recurrent Neural Networks

Juri Backes and Wolfgang Renz

MMLab Research Group Distributed Adaptive Systems,
Hamburg University of Applied Sciences (HAW)
Berliner Tor 7, 20099 Hamburg, Germany

Abstract

For integration of growing amounts of volatile renewable energy in the European electricity system, reliable weather prognosis gains importance. But, depending on weather conditions, forecast reliability of wind speed for predicting wind power can vary drastically with time. Thus, relevance of risk-aware system operation strategies is increasing based on wind speed uncertainty a measure of which is provided by the standard deviations of ensemble forecasts of the German Weather Service. However, lacking validity of this measure is known as a long-standing problem.

Therefore, this work investigates how machine learning based on a suitably selected set of physical quantities of weather ensemble data as well as historic wind data allows for a more realistic uncertainty quantification. A recurrent neural network (RNN) based sequence-to-sequence architecture is implemented and probabilistic wind speed forecasts are generated for a region in northern Germany.

The results are evaluated and compared with the forecasts of the German Weather Service thereby revealing improved validity of such deep-learning based uncertainty measures.

1 Introduction

Weather services are typically based on Numerical Weather Prediction (NWP) [14]. For predicting wind power at certain locations, specialized services are provided by companies which use public NWP forecasts as input and provide wind power expectation values for future times. These forecasts are used for trading the expected generated energy and possibly for operation of energy storage, flexible producers or consumers. Uncertainties that come with any predicted values in general and with renewable energy generation forecasts in particular are nowadays tackled by intra-day trading activities until few hours and even 15-minute intervals before generation time. Since not all deviations between predicted and actual generated power can be eliminated by increasingly accurate forecasts, recent focus has been on predicting uncertainties based on confidence intervals that are not uniform over time and, therefore affect energy markets and finally energy costs. Once these uncertainties can be quantitatively predicted, risk-aware operation and trading strategies could be implemented and contribute to cost reduction [15].

Considering NWP, uncertainty has two origins. First, forecasts depend on initial values of numerical solutions of weather models, i.e. on measured values or results of previous forecasts that both are not exactly known. Also numerics cause errors in the resulting predictions, such *aleatoric* sources of uncertainty are typically described by random fluctuations of initial values. So-called *epistemic* uncertainties, on the other hand, arise from neglects or simplifications in the weather models used in NWP and can be hardly determined [4]. Aleatoric uncertainties are treated by ensembles predictions that combine the results of different but equally possible initial values of an NWP model.

Probabilistic forecasts that are derived from such ensembles form the basis for quantitative uncertainty predictions. However a known problem of ensemble predictions is that they systematically underestimate the uncertainty [8] and therefore are

1

not suitable for applying them directly to planning strategies that are constructed to benefit from uncertainty forecasts.

Consequently, the question of how uncertainty quantification of ensemble forecasts can be improved is explored in detail here. Machine learning (ML) methodology is a promising tool since there is a huge set of data available from ensemble predictions. These cover a multitude of, partially weakly correlated, meteorological variables in multiple altitude layers over the forecast horizon. The task then is to select a suitable set of meteorological quantities that are expected to influence wind speed forecasts in a relevant way, and to avoid inefficiency. This data is later used to approximate the forecast uncertainty using ML such that epistemic uncertainties are included, unlike the ensemble predictions. In the case study deep learning (DL) is applied to ensemble forecasts of the German Weather Service for several weather stations in northern Germany.

2 Related Work

Statistical post-processing of numerical weather predictions, such as (Ensemble) Model Output Statistics, which look at statistical relationships between forecast results and observed meteorological data, are common to improve the predictions of deterministic models [11, 2]. The goal is to reduce the deviation between the model prediction and observed data to the smallest possible level using linear, logistic and multiple regression. However, due to the linearity of the regression, the modelability of relationships is limited, and the informativeness of the uncertainty information hardly improves [10, 3].
Various Uncertainty Quantification (UQ) methods based on Neural Networks (NN) or Machine Learning (ML) techniques [12, 1] have been able to achieve good results in generating probabilistic predictions in various application domains. Uncertainty quantification is primarily used in classification problems, image processing, prediction of financial indicators, or renewable energy generation forecasts [13, 7], but rarely for meteorological predictions. Besides Bayesian NN's, common architectures are RNN and CNN (convolutional neural network), or mixtures of the two, allow both spatial and temporal relationships to be learned from the input data. Furthermore, various error functions can be found, which allow to generate probabilistic predictions, i.e. distributional forecasts.

A first feasibility study [9] shows how CNN-LSTM networks can be used to improve the uncertainty information of temperature ensemble forecasts, while focussing on the savings in computational time and power compared to regular forecasts.
In [17], a deep learning approach for Uncertainty Quantification of weather forecasts is presented, producing probabilistic forecasts of temperature, wind speed, and relative humidity for ten weather stations in Beijing. A combination of NWP forecasts and recent historic observations is used to feed a GRU-based (Gated Recurrent Unit, [6]) seq2seq [16], i.e. a RNN-based encoder-decoder architecture, which appeared best in comparison of different architectures studied. Their newly developed negative log-likelihood error loss function allows to learn the parameters of time-varying predictive normal distributions (μ, σ^2), and to predict the underlying forecast uncertainty this way. In comparison to the numerical predictions used as input, the accuracy was improved by 47.76 %, however, a small sample size was evaluated, which does not represent a general improvement of performance.
A modified configuration of [17] forms basis for this work, extending it, in order to support ensemble predictions as input data, using a different error function, and evaluate the model performance using more prediction points.

3 Conception

Aim of this work is to process and improve the uncertainty information of given ensemble wind speed forecasts (purely aleatoric), by learning about the model uncertainty (epistemic). For this purpose, existing numerical weather forecasts of several parameters for a forecast period and location as well as previous measurements are to be processed by a RNN architecture. In the following part, the general problem is presented and evaluation metrics are introduced.

3.1 Methodology

The following general problem formulation is inspired by [17]. Given historical ensemble forecasts from numerical weather models for predefined mete-
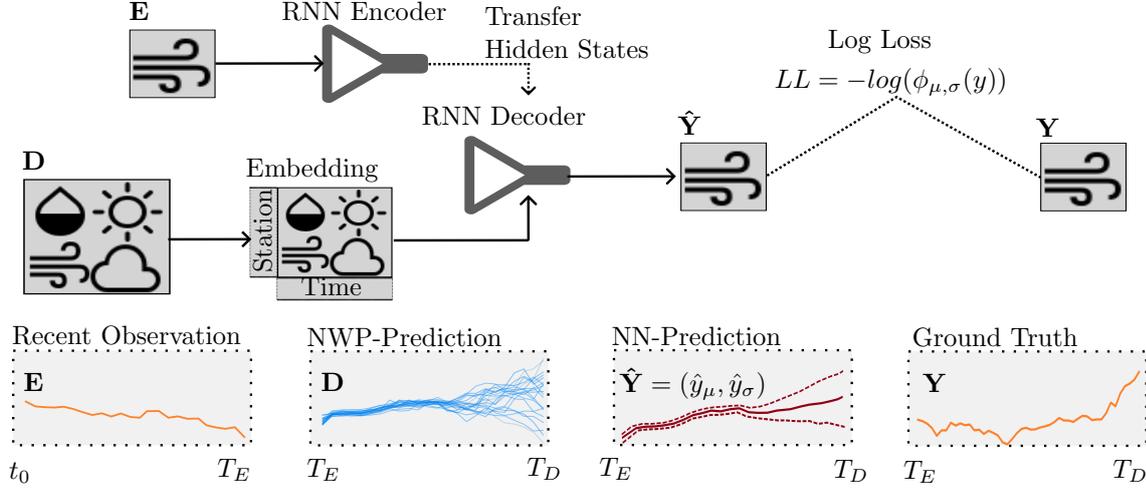
Figure 1: GRU-based Sequence-to-sequence architecture and data semantic for enhancing probabilistic wind speed forecasts, utilizing deep RNN's.

orological parameters and associated measurement data for a specific weather station s from S, the problem can be formulated as follows:

1. **Recent Observations**, representing a measurements time series, is denoted by the vector $\mathbf{e}(t) \in \mathbb{R}^{N_1}$ whereby the wind speed, is the only meteorological value of the set N_1 for the time $t = t_0 + 1, \ldots, T_E$.

2. Further the **NWP Forecasts** denoted by $\mathbf{d}(t) = [d_1(t), d_2(t), \ldots, d_{N_2 \cdot N_e}(t)] \in \mathbb{R}^{N_2 \cdot N_e}$ with $d_n(t)$ being the forecast values of one ensemble of the NWP parameter set N_2 over the forecast horizon $t = T_E+1, \ldots, T_E+T_D$, with each ensemble consisting of N_e members. Giving $T_D = T_{max}$. Further information on N_2 is given in section 4.1

3. The **Ground Truth** time series of the target variables denoted as $\mathbf{y}(t) \in \mathbb{R}^{N_3}$, where the set N_3 consists purely of the wind speed as target variable for $t = T_E + 1, T_E + 2, \ldots, T_E + T_D$.

4. The **Estimation** of $\mathbf{y}(t)$ is denoted as $\hat{\mathbf{y}}(t)$. Since the estimation errors of independent forecasts are expected to be distributed around the mean, the estimate is described by a normal distribution. Consequently, each estimation consists of $\hat{\mathbf{y}}_\mu(t)$, the prediction mean, and $\hat{\mathbf{y}}_\sigma(t)$ a corresponding standard deviation.

5. $\mathbf{E}_{T_E} = [\mathbf{e}(1), \mathbf{e}(2), \ldots, \mathbf{e}(T_E)] \in \mathbb{R}^{T_E \times N_1}$, the observation matrix, aggregating the recent historical observation vectors resulting in $T_E \times N_1$ dimensions.

6. $\mathbf{D}_{T_D} = [\mathbf{d}(T_E + 1), \mathbf{d}(T_E + 2), \ldots, \mathbf{d}(T_E + T_D)] \in \mathbb{R}^{T_D \times (N_2 \cdot N_e)}$, NWP forecast matrix, consisting of the aggregated NWP ensemble forecast vectors, corresponding to $T_D \times (N_2 \cdot N_e)$.

7. $\mathbf{X}_{T_E, T_D} = [\mathbf{E}_{T_E}; \mathbf{D}_{T_D}]$, concatenation of the observation matrix and NWP forecast matrix.

8. $\hat{\mathbf{Y}}_{T_D} = [\hat{\mathbf{y}}(T_E + 1), \hat{\mathbf{y}}(T_E + 2), \ldots, \hat{\mathbf{y}}(T_E + T_D)] \in \mathbb{R}^{T_D \times N_3}$ is the estimation matrix.

For each unique forecast horizon f from F all stations s from S are considered. With given $\mathbf{X}_{T_E, T_D}$, the primary goal is to estimate the ground truth $\mathbf{y}(t)$ by the mean value $\hat{\mathbf{y}}_\mu(t)$ as close as possible, further, the estimated standard deviation $\hat{\mathbf{y}}_\sigma(t)$ should give insights on the (un)certainty of the forecast, reflecting expected inevitable errors with a well parameterized time-varying probability density distribution.

Adopting the fusion methodology of [17], the proposed model includes recent measurements of relevant weather dynamics only $\mathbf{E}_{T_E}$ to account for the recent seasonal conditions. Furthermore, numerical ensemble wind speed forecast data $\mathbf{D}_{T_D}$ are used that contain trends and aleatoric uncertainties of the target variable. Together, $\mathbf{E}_{T_E}$ and $\mathbf{D}_{T_D}$ serve as an initial baseline for the predictions to be generated. Including forecasts of further meteorological parameters $\mathbf{D}_{T_D}$ that are physically related to the target variable will introduce useful information about the atmospheric and meteorological state.

3

3.2 Evaluation Metrics

Adopting the inflation factor r from [18], giving insights on the balance of standard deviation to actual error, with i being the index for each individual forecast point:

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}\sqrt{(\mu_i - \mathbf{y}_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma_i}} \tag{1}$$

For an optimal forecast model, an inflation value of $r = 1$ is expected. If $r < 1$, i.e. the mean standard deviation dominates over the mean error, meaning the standard deviations are systematically too large, otherwise, if $r > 1$ the standard deviation is too narrow.

To compensate for the uncertainty underestimation of the NWP model, the standard deviations are increased by the factor r for the entire evaluation. Further the mean absolute error (MAE) and the continuous ranked probability score (CRPS) [8] are considered:

$$MAE = \frac{\sum_{i=1}^{n}|\mu_i - \mathbf{y}_i|}{n} \tag{2}$$

$$CRPS(\Phi, \mathbf{y}) = \frac{\sum_{i=1}^{n}\int_{\mathbb{R}}(\Phi(\mu_i) - \mathbb{1}(\mu_i \geq \mathbf{y}_i))^2 d\mu_i}{n} \tag{3}$$

The CRPS can be seen as a generalization of the MAE for probabilistic predictions, with both MAE and CRPS being negatively oriented; accordingly, smaller values are desirable.

Probability distributions hold the exact information of the uncertainties, however, the margins of the distribution functions can extend to infinity or negative values, creating an unrealistic picture of the uncertainties. For this reason, forecast uncertainties are often described using prediction intervals, denoted γ, which can be determined directly from the parameters of the distribution density function:

$$\gamma = \frac{l - \mu_i}{\sigma_i} < \frac{\mathbf{y}_i - \mu_i}{\sigma_i} < \frac{u - \mu_i}{\sigma_i} \tag{4}$$

$$= \frac{l - \mu_i}{\sigma_i} < Z_i < \frac{u - \mu_i}{\sigma_i} \tag{5}$$

Where $\mathbf{y}$ corresponds to the measured value, l and u to predefined limits and Z being the standard score. Prediction intervals are given as percentages, e.g., $\gamma_{90\%}$, that it is expected, by given independent draws from the underlying distributions, 90% of the draws will be within the interval.

4 Case Study

In this work, a RNN is implemented and optimized to estimate the parameters of a probability density distribution $\hat{\mathbf{Y}}_{T_D}$, utilizing ability to learn temporal relationships as they are predominant in this application. The error function of the network has to depend on the two quantities $\hat{\mathbf{y}}_\mu(t)$ and $\hat{\mathbf{y}}_\sigma(t)$, as well as the true values $\mathbf{y}(t)$.

For this case study, wind forecasts of 11 selected weather stations in northern Germany are considered. In the following, the data source, data preprocessing, sample definition and the model architecture will be discussed.

4.1 Data

The wind speed measurements for $\mathbf{E}_{T_E}$ and $\mathbf{Y}_{T_D}$ are taken from the German climate-data-center [1]. The NWP data $\mathbf{D}_{T_D}$ is taken from daily 0, 6, 12, 18 o'clock forecasts of the ICON-EPS-D2 model [14] of the German Weather Service [2]. The NWP ensemble forecasts include our selected set of 18 parameters, hence $N_2=18$, (see table 1) related to the wind speed, partially considered for different altitudes indicated by their pressure level. Each ensemble forecast consists of 20 independent ensemble members, hence $N_e = 20$, with point predictions on every full hour over an forecast horizon of 24 hours, giving the number of the target times $T_D=24$. Accordingly, the lead times was chosen as $T_E=24$, so that they share the same length.

Considering the four daily forecasts over one year starting 07.2021, results in 1460 potential samples which include features of all seasons. Missing data on the edges or for more than four consecutive values, leads to the whole forecast horizon being discarded, otherwise being filled by linear interpolation, leaving $F=885$ forecasts. Overall following tensors are created:

$$\text{Input Tensors } \mathbf{E}_{T_E}, \mathbf{D}_{T_D}:$$
$$(F, T_E, S, N_1) = (885, 24, 11, 1)$$
$$(F, T_D, S, N_2 \cdot N_e) = (885, 24, 11, 360)$$
$$\text{Output Tensor } \mathbf{Y}_{T_D}:$$
$$(F, T_D, S, N_3) = (885, 24, 11, 1)$$

---

[1]Climate Data Center of the German Weather Service https://cdc.dwd.de/portal/
[2]OpenData Portal of the German Weather Service https://opendata.dwd.de/weather/nwp/

Observation and Ground Truth Description $\mathbf{E}_{T_E}, \mathbf{Y}_{T_D}$
Wind speed at $\left[\frac{m}{s}\right]$ at 2 m

NWP Parameter Description $\mathbf{D}_{T_D}$
Cloud coverage [%]
Relative humidity [%]   at 10 m and at pressures of 1000, 700, 500 hPa
Temperature [K]   at 2 m and at pressures of 1000, 700, 500 hPa
Wind speed in north-south direction $\left[\frac{m}{s}\right]$ at 10 m and at pressures of 1000, 700, 500 hPa
Wind speed in west-east direction $\left[\frac{m}{s}\right]$ at 10 m and at pressures of 1000, 700, 500 hPa
Maximum (directionless) wind speed $\left[\frac{m}{s}\right]$ at 10 m

Table 1: Listing of the considered meteorological quantities and corresponding altitudes.

Each of these sets is divided into a training set of 815, a validation set with 50 and a test set with 20 forecasts respectively. To speed up the learning process and minimize the possibility to get stuck in local minima, every parameter is normalized by a min-max rule to a value range of [0,1].

4.2 Model

The investigated architecture is based on a sequence-to-sequence model [16], as it can be seen in figure 1. The decoder and encoder network consists each of two deep layers, with 50 GRU-cells (Gated Recurrent Units, [6]) per layer. Utilizing the RNN-architecture, it is possible for the model to learn general interrelationships as well as temporal features from the data.

The encoder RNN-network extracts the current meteorological and atmospheric context of $\mathbf{E}_{T_E}$ and transfers its internal state to the decoder network. The decoder will then incorporate the NWP forecasts $\mathbf{D}_{T_D}$ embedded in station and time information via two embedding layers, making it easy for the model to learn the representation. Based on the given initial state, the decoder is able to generate sequential probabilistic estimations $\hat{\mathbf{Y}}_{T_D}$ consisting of the two parameters $\hat{\mathbf{y}}_\mu(t)$ and $\hat{\mathbf{y}}_\sigma(t)$, where the

standard deviation is guaranteed to be positive by utilizing the softplus activation function.

At train time, the model is learning to minimize the logarithmic loss (LL), making it possible to account for the standard score of the normal distribution $\phi_{\mu,\sigma}$ directly and optimizing probabilistic forecasts:

$$LL = -log\left(\phi_{\hat{\mathbf{y}}_\mu(t),\hat{\mathbf{y}}_\sigma(t)}(\mathbf{y}(t))\right) \qquad (6)$$

The principle of LL is based on the fact that correctly predicted events are neglected as small errors, while incorrect predictions are given more weight by applying the logarithm.

The training is done iteratively with a batch size of 512 and a maximum of 100000 iterations, with every 50 iterations testing the model against the validation set. Each sample corresponds to one forecast f of one station s, with 885 forecasts and 11 stations resulting in 9735 individual samples (training: 8965, validation: 550, test: 220). For every training step, the sample and corresponding station are chosen at random.

To prevent over-fitting, an early stop mechanism is implemented, a maximum of 10 consecutive tests on the validation set without improvement will stop the training. The model is realized using the tensorflow-interface keras [5].

5 Results
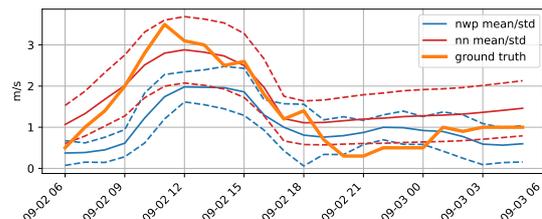
Figure 2: Comparison of NN and NWP probabilistic wind speed forecasts over a horizon of 24h, the ground truth is indicated by the orange line. Uncertainty underestimation of the NWP model can bee seen clearly.

The forecasts generated by the Neural Net (denoted NN) of the 220 test set samples, resulting in 5280 individual probabilistic predictions, will be compared with those of the NWP ensemble predictions and evaluated using established statistical

criteria. Corresponding designations used throughout this section were introduced in section 3.2. In order to make the ensemble forecasts comparable, velocities per wind direction are combined to one directionless velocity. Furthermore, a normal distribution is calculated from this directionless ensemble forecast. The generated NN forecasts are re-scaled to the original scale, a randomly selected example can be seen in figure 2.

NWP standard deviations are inflated by the factor r_{NWP}=1.579, for the NN model, a hypothetical inflation factor r_{NN}=0.775 (rather a deflation factor) can also be calculated, but the NN standard deviations remain unchanged for the evaluation.

In table 2 a comparison of evaluation metrics between the NWP model and the investigated NN approach is presented.

Metric	NWP	NN
Number of sample points	5280	5280
Inflation factor r	1.579	0.775
MAE	1.664 $\frac{m}{s}$	0.727 $\frac{m}{s}$
CRPS	0.896	0.523
$\gamma_{50\%}$	29.3%	48.9%
$\gamma_{75\%}$	47.4%	73.9%
$\gamma_{90\%}$	61.8%	90.03%
$\gamma_{99\%}$	79.1%	98.4%

Table 2: Comparison between NN generated predictions and NWP predictions using established statistical metrics shows the overall improvements in point predictions and uncertainty quantification using the deep learning method.

Considering the MAE, the prediction performance can be evaluated neglecting the uncertainty information. The MAE of the NWP model with 1.664 $\frac{m}{s}$ was improved by a factor of 56.4% using the deep learning application, resulting in an average error of 0.727 $\frac{m}{s}$.

When considering the CRPS that includes the uncertainty information, there still is a strong improvement of the NN model $CRPS_{NN}$=0.523 compared to the NWP prediction $CRPS_{NWP}$=0.896. The reason for this is twofold. In contrast to the NN model, the NWP model was not optimized for accurate uncertainty quantification since it only accounts for aleatoric uncertainties. On the other hand, predictions with deviating average and too

small standard deviation will receive strong penalties when evaluated with the CRPS, which appears often in NWP forecasts.

Resulting prediction intervals for the NWP model are well below the expected values, meaning that the uncertainties are systematically underestimated in agreement with the interpretation of the CRPS. For the prediction intervals of the NN model, on the other hand, the predictions appear to reflect the uncertainties better.

Besides the above quantitative comparisons it is common to evaluate probabilistic forecasts via visual methods. The histogram of the emerged standard score is shown in figure 3. The discussed prediction
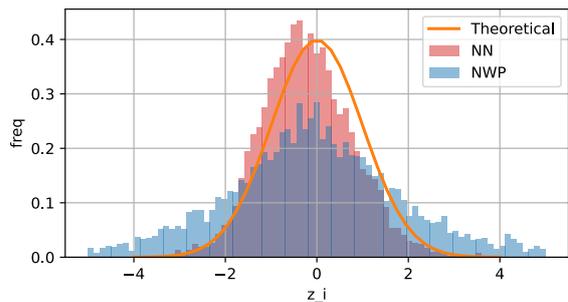
Figure 3: Histogram of the standard scores for the NN and NWP forecasts, also showing the theoretical optimum, a normal distribution with mean 0 and standard deviation 1.

behavior of the models, i.e., the strong underestimation of the uncertainties by the NWP model can be clearly seen in the graphical representation of the z-statistics (Figure 3). An ideal prediction corresponds to a standard normal distribution that is indicated by the orange colored line . A large part of the NWP distribution lies outside the ideal distribution indicating a lack of validity. In such cases, the uncertainty is underestimated, i.e., the standard deviation is too small, and the result is a to-wide z-statistic. The statistics for the NN model is closer to the ideal case but has some shift towards negative values, which indicates that predicted mean values larger than the measured values occur more frequently. Reliability plots are shown in figure 4.

The comparison of the observed frequency with the predicted proportion shows how well the predicted probabilities match the observations.
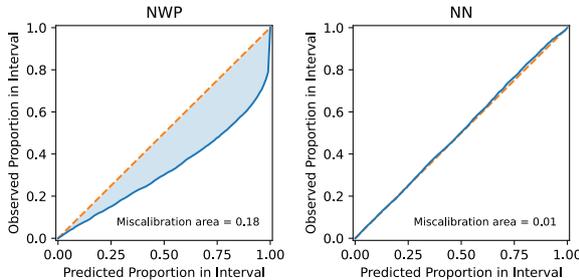
Figure 4: Reliability plot of the two forecast models. A perfectly calibrated forecast model is indicated by the orange dashed diagonal.

The reliability curve of the NWP, which lies below the diagonal line, again indicates too low predictive power (i.e., too small standard deviations); the NN model, on the other hand, corresponds almost perfectly to the ideal case and can thus be classified as reliable. Finally, it should be mentioned how the
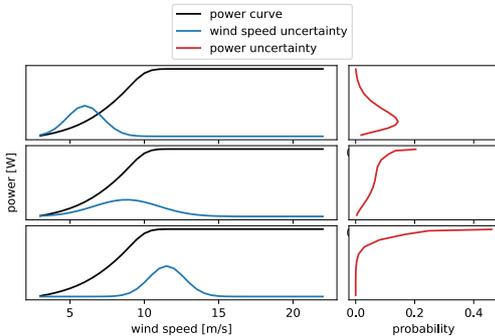
Figure 5: Illustration of how Gaussian wind speed forecast uncertainties can translate into different power output uncertainties. The transformation is nonlinear due to the saturation of the rated power curve of the wind energy plant.

wind uncertainty results can be applied to power generation. Power generation uncertainty can be derived from generated normal distributed forecasts, as exemplary shown in figure 5. That way it is obvious how the improvement of wind speed uncertainty forecasts translates straight into more accurate and reliable probabilistic power generation forecasts.

6   Conclusions

In this paper, we investigated how more realistic wind speed uncertainty quantification can be achieved by a deep learning approach based on a suitably selected set of meteorological quantities taken from the ensemble NWP and recent observations. The methodology can be used to significantly improve the actual point prediction and both the reliability and validity of the uncertainty forecasts. The results obtained exhibit validated uncertainty quantification that, in addition to the aleatoric causes included in the ensemble NWP, takes the epistemic uncertainty causes into account and overcomes the short-comings of the NWP uncertainty prediction.

Future work includes application of this approach to combine weather and energy market data for a more general risk quantification[15].

References

[1] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazade-gan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 2021. doi: 10.1016/j.inffus.2021.05.008.

[2] J. A. Baars and C. F. Mass. Performance of national weather service forecasts compared to operational, consensus, and weighted model output statistics. *Weather and Forecasting*, 2005. doi: 10.1175/WAF896.1.

[3] S. L. Benedikt Schulz, Mehrez El Ayari and S. Baran. Post-processing numerical weather prediction ensembles for probabilistic solar irradiance forecasting. *Solar Energy*, 2021. doi: 10.1016/j.solener.2021.03.023.

[4] R. J. Bessa, C. Möhrlen, V. Fundel, M. Siefert, J. Browell, S. Haglund El Gaidi, B.-M. Hodge, U. Cali, and G. Kariniotakis. Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry. *Energies*, 2017. doi: 10.3390/en10091402.

[5] F. Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

[6] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, 2014. doi: https://doi.org/10.48550/arXiv.1412.3555.

[7] O. Eikeland, F. Hovem, T. Olsen, M. Chiesa, and F. M. Bianchi. Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods: An arctic case. *Energy Conversion and Management: X*, 2022. doi: 10.1016/j.ecmx.2022.100239.

[8] T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 2007. doi: 10.1198/016214506000001437.

[9] P. Grönquist, T. Ben-Nun, N. Dryden, P. Dueben, L. Lavarini, S. Li, and T. Hoefler. Predicting weather uncertainty with deep convnets. *CoRR*, 2019. doi: https://doi.org/10.48550/arXiv.1911.00630.

[10] R. Hess. Statistical postprocessing of ensemble forecasts for severe weather at deutscher wetterdienst. *Nonlinear Processes in Geophysics*, 2020. doi: 10.5194/npg-27-473-2020.

[11] C. Junk, L. D. Monache, and S. Alessandrini. Analog-based ensemble model output statistics. *Monthly Weather Review*, 2015. doi: 10.1175/MWR-D-15-0095.1.

[12] H. M. D. Kabir, A. Khosravi, M. A. Hosen, and S. Nahavandi. Neural network-based uncertainty quantification: A survey of methodologies and applications. *IEEE Access*, 2018. doi: 10.1109/ACCESS.2018.2836917.

[13] A. Mashlakov, T. Kuronen, L. Lensu, A. Kaarna, and S. Honkapuro. Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. *Applied Energy*, 2021. doi: https://doi.org/10.1016/j.apenergy.2020.116405.

[14] D. Reinert, F. Prill, H. Frank, M. Denhard, M. Baldauf, C. Schraff, C. Gebhardt, C. Marsigli, and G. Zängl. Dwd database reference for the global and regional icon and icon-eps forecasting system, 2021.

[15] W. Renz, J. O. Sudeikat, J. Backes, and K. Eger. Assessment of functional risks for engineering adaptive smart grid applications. In *2022 IEEE 10th International Conference on Smart Energy Grid Engineering (SEGE)*, 2022. doi: 10.1109/SEGE55279.2022.9889752.

[16] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. NIPS'14. MIT Press, 2014. doi: 10.5555/2969033.2969173.

[17] B. Wang, J. Lu, Z. Yan, H. Luo, T. Li, Y. Zheng, and G. Zhang. Deep uncertainty quantification: A machine learning approach for weather forecasting. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19. Association for Computing Machinery, 2019. doi: 10.1145/3292500.3330704.

[18] D. S. Wilks. Smoothing forecast ensembles with fitted probability distributions. *Quarterly Journal of the Royal Meteorological Society*, 2002. doi: https://doi.org/10.1256/qj.01.215.