



HAL
open science

Statistical Post-Processing of Weather Forecasts for the Renewable Energy Production by Eolian Devices

Lionel Lenôtre

► **To cite this version:**

Lionel Lenôtre. Statistical Post-Processing of Weather Forecasts for the Renewable Energy Production by Eolian Devices. 2017. hal-01499295

HAL Id: hal-01499295

<https://hal.science/hal-01499295>

Preprint submitted on 31 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical Post-Processing of Weather Forecasts for the Renewable Energy Production by Eolian Devices

Lionel Lenôtre

CMAP UMR 7641 Ecole Polytechnique CNRS, Route de Saclay, 91128 Palaiseau Cedex France

lionel.lenotre@polytechnique.edu

This paper reviews different statistical methods dedicated to the post-processing of Numerical Weather Predictions and Ensemble Forecast. We focus on the application of the post-processing to problems linked to the production of electricity by eolian devices. The basic idea is to give a concise panorama of the methods commonly used nowadays. We pay a particular attention to the mathematics involved in the methods. We do not compare the methods and do not provide some preferences.

Classification. 62-02; 62P12.

Keyword. Statistical Post-Processing; Survey; Ensemble Forecast; Numerical Weather Predictions; Renewable Energy Production.

Acknowledgement. This work was supported and funded by the Agence National pour la Recherche through the research project CAESARS. We thank E. Gobet and C. Alasseur for their various comments and their help to improve this paper.

1 Introduction

Problematic. The renewable energy generation providers use various mathematical tools to manage operations on the spot (daily and intraday) electricity market[1, 9, 23]. Such tools must take into account the uncertainty of the future hourly production rates of renewable energy generation devices with an horizon of approximately 48 hours. For that purpose, the quantification of the uncertainty in weather predictions has to be solidly managed. The required forecast horizon for the spot market is called short-term in meteorology. For such an horizon, the future climatic conditions can be furnished by Numerical Weather Predictions and Ensemble Forecast [7] issued by meteorological centers. Both of these methods are mainly based on physic. Now other methods can be used to obtain

short-term weather forecasts such as time series trained on historical data [5] or neural networks [33]. These last two type of methods are statistical and rely on historical data. Hence, pure statistical methods has to be actualized regularly compared to Numerical Weather Predictions. Neural Networks had proved to be very satisfying since it perform online correction of itself. Inclusion of physical concerns can be performed in statistical methods. They often provide improvement and reduce the necessity of actualization. As explained in [7], it is very difficult to determine which methods are the most effective and must be preferred. Here we decide to only deal with numerical weather predictions and Ensemble Forecast.

A Numerical Weather Prediction is a scenario computed with a physical model. It contains forecast values of different weather variables at some time steps whose order is from minutes to few hours. This scenario is purely deterministic and it is not rare that it is completely wrong. A common manifestation of this fact is the lack of trust some people has on weather forecast given by news channel on television. The technique of Ensemble Forecast tries to render the variability of numerical weather predictions by producing a set of different possible scenarios using methods numerical weather predictions. Nevertheless, having different scenarios does not really measure the uncertainty and surely does not provides a clear quantification like a cumulative distribution function or a probability density function do it. Hence, numerical weather predictions and Ensemble Forecast cannot be incorporated directly into useful mathematical tools for operational management such as optimal stochastic control problems. This means that numerical weather predictions and Ensemble Forecast has to be post-processed.

Content. In this survey, we present a few suitable methods for the post-processing of Numerical Weather Predictions and Ensemble Forecast with the aim of solving problems linked to the generation of electricity using eolian devices. Some of them use simple Numerical Weather Predictions while others use Ensemble Forecast which are an improvement of Numerical Weather Predictions. However all the presented methods rely on the same fundamental idea which is to use the forecasts as hints. In other words, the methods infer statistical quantities or functionals of the conditional distribution of a weather variable of interest X knowing some forecasts $\mathbf{F} = (F_1, \dots, F_N)$ coming from an Ensemble Forecast containing N scenarios or from N weather variables correlated to the one of interest in the case of simple Numerical Weather Predictions. Among the methods, a few directly produces results from an Ensemble Forecast while most of them use historical data containing couples of forecasts and realizations. For pretty clear reasons, the methods for the post-processing of Numerical Weather Predictions are only of this last type.

Some the methods presented in this survey were not designed for weather variables directly linked to eolian devices. Since we decide to present the original article which introduce a method, it explain the presence of estimation of statistical quantities or functionals for weather variables such as sea level or temperature in spite of wind speed or wind direction. However, we provides references of the applications or modifications of these methods to the case of wind. At last but not least, we assume weather variables as being in one dimension all along this survey. The extension of some methods to weather variables in high dimension would misleadingly appeared to be very straightforward. Hence, we invite the reader to be very careful.

Two distinctive parts compose this survey. The first one exposed in §2 can be read independently. It gives a small account on models of numerical weather predictions and

point out some sources of uncertainty which are encountered in these model. It also present some basic facts about Ensemble Forecast. Precisely, it details the idea behind this method and explain the rendering of the uncertainty of numerical weather predictions by Ensemble Forecasts. In §3, we give defined some scoring rules for the assessment of statistical methods which are used in the sequel to measure the performances of the reviewed statistical methods. The second part is given in §4, §5 and §6. Precisely, we review methods for the computation of quantiles and confidence intervals and the estimation of cumulative distribution functions and probability density functions. We principally detail the mathematical part of these methods and the rationales. We make few comments on the performances of the methods, but we do not enter into details since a large number of articles such as [36, 38] provide benchmarks and comparisons.

2 Numerical Weather Predictions and Ensemble

In various fields of physics such as orbital mechanics, the future state of a system is completely determined by the present state of the system. The evolution of the system is completely governed by causal relationships, that is the laws of physics. This is the path of determinism. In meteorology, it is globally admitted that a fully deterministic method fails to predict future states of the atmosphere. For instance, it is often the case that weather predictions performed are not realized or even that some events cannot be predicted. In this section, we propose to provide some rationales behind the failure of determinism and to present a commonly used workaround solution.

Uncertainty in Numerical Weather Predictions. As we already said above, Numerical Weather Predictions are forecast values of different weather variables at some time steps whose order is from minutes to few hours and whose sum is usually around 48 hours. These forecasts are technically obtained using a numerical model which solves *in silico* a system of N differential equations coming from a spatial discretization of the laws governing the behavior of the atmosphere [20]. Such a numerical model is constructed as follows. The physical laws representing the dynamic of the atmosphere [18] which are

- the conservation of momentum,
- the conservation of mass,
- the equation of state for ideal gases,
- the conservation of energy,
- the conservation equation for water mass.

are combined into a physical model depicting a certain interpretation of the behavior of the atmosphere. A filtering approximations [18] is then performed to allow for the discretization process. The current approximation in use is mainly the hydrostatic one. It has progressively replaced the quasi-geostrophic approximation [18, 21]. Once these two steps are performed, a mathematical model of the dynamic of the atmosphere is obtained. In order to be solved *in silico*, the mathematical model is transformed into a numerical model. This step basically consist in choosing a set of numerical methods such as Finite Element or Finite Difference and to set consistent boundary conditions [18]. These choices are usually made depending on the available computing resources,

the desired computational time, etc... On balance, Numerical Weather Predictions comes from the derivation of three nested models, each of which resulting from a sequence of choices whose validity can always be impeached to some extent. In other words, the capability of the numerical model to truthfully render the dynamic of the atmosphere is actually completely uncertain. Moreover, truncation errors and numerical artefacts intrinsic to the algorithms within the numerical methods has to be added.

Once a numerical model has been derived, then it has to be initialized in order to obtain Numerical Weather Predictions. In few words, some initial conditions derived from an estimation of the present state of the atmosphere has to be plugged. The estimation procedure must be performed very carefully since the system of differential equations involved in the mathematical model is very chaotic and thus is the numerical model too. For instance, very different results can be obtained from a numerical model while it is simply initialized with multiple rounded estimates of the same initial conditions [22]. In practice, meteorological centers compute the initial conditions using statistical methods such as Kalman filtering on a combination of observations and very short-term forecast also called first guess [18]. Such a process is usually named *data assimilation* in the literature. It follows easily from the description above that the approximated initial conditions are subject to uncertainty coming from the errors in the measurement of observations and the error in the numerical model which is here used to provide the first guess. The iterative use of very short-term forecast can even be seen as the initiator of a particular propagation of the uncertainty through a sort of a loop [18, Fig. 5.1.2]. The uncertain initial conditions coupled with the amplification property of the numerical model are actually the most important source of uncertainty in the field of Numerical Weather Predictions [20].

Ensemble Forecast and Uncertainty Quantification. Monte Carlo methods are usually very practical to solve sets of differential equations with uncertain initial conditions. In meteorology, this approach is unfeasible. In fact, the number D of degrees of freedom of the numerical models used is rather too high ($D \approx 10^8 - 10^6$) [20]. Besides, the probability density function which is used to reflect the uncertainties in the initial conditions and to initialize the Monte Carlo methods is itself uncertain. It also has to be approximated as we are dealing with a problem possessing a high number of degrees of freedom.

Ensemble forecast was designed as a workaround solution in order to replace Monte Carlo methods. It recycles the idea of using random initial conditions. In a nutshell, a numerical model used to provide Numerical Weather Predictions is run several times using small variations of initial conditions obtained through the procedure of *data assimilation*. The computation of such small variations is performed using deterministic methods [6, 20]. Conceptually, the small variations represent the samples of the initial probability density which can not be clearly determined. Each run thus provides samples of each of the different weather variables at some time steps for a certain time lapse. Ensemble forecast can also deal with the uncertainty coming from choices performed in the physical, mathematical and numerical modeling. In fact, small variations can be plugged into different models [21]. On average, $N \approx 20$ to 50 ensemble members are produced due to the computational cost.

An Ensemble Forecast provides raw data which accounts for the uncertainty in the future weather condition. Hence, functionals representing the repartition of the uncertainty such as a cumulative or a probability density function has to be recovered. As multiple outputs from a numerical model, the data produced by an Ensemble Forecast are simply

values that can be possibly taken by a weather variable at regularly spaced consecutive times for the next 24 or 48 hours. These data are usually separated into Ensemble members, that is a sequence of values indexed by the time or a path coming from a particular run for each variable. Actually, the most common practice is to infer functionals of the marginal distribution of a weather variable (distribution at a given time step), that is to manipulate probability distributions on \mathbb{R}^d where $d \in \mathbb{N}$. Rarely, the law of the whole path are examined. In other word, statistical methods dealing with probability distribution on some space of functions or sequences are not really in the actual trend. Among the methods for the inference of marginal distributions, some use only the predicted paths of the variable of interest while others add the paths from other weather variables to account for potential correlation. However, almost all of them rely on the same idea of using the paths of Ensemble Forecast as predictors of weather variables. Basically, the common tactic is to infer the conditional marginal distribution of the weather variables knowing the samples coming from Ensemble Forecast. Every methods presented below used this tactic. As an illustration of this principle, the method detailed in §6 all aims at estimating the conditional density $f(x|\mathbf{F})$ of X knowing the N Ensemble members $(F_j)_{1 \leq j \leq N}$.

3 Measures of Performance.

In this section, we define the Rank-histogram which is tools designed to examine the qualities of an Ensemble Forecast. Indeed, even with the most advanced post-processing, the quantification of uncertainty can not be well done from an Ensemble Forecast if it is utterly biased or under-dispersive [14]. Then we provide the definition of the two most used scoring rules for the assessment of statistical post-processing. We do not go into detail. Some references are provided for more informations and interpretations of these two scoring rules.

Rank-Based Histogram. In [15], the author define Rank-Histograms and use it as a tool in order to assess the presence of some required probabilistic properties in an Ensemble Forecast. Rank-Histograms are very easy to derive from the data. Let N be the number of members of an Ensemble Forecast and $(\mathbf{F}^j, R^j)_{1 \leq j \leq M}$ be M historical pairs of forecast/realization where, for each $j = 1, \dots, M$, $\mathbf{F}_j = (F_1, \dots, F_n)$ is the vector formed with the Ensemble members. For each $j = 1, \dots, M$, incorporate R^j to the vector \mathbf{R}^j to form a new vector $\hat{\mathbf{F}}^j$ of dimension $N + 1$. Then sort the elements of $\hat{\mathbf{F}}^j$ in the ascending order and capture the rank of R^j . Repeat the procedure M times and form histogram using the relative frequency of apparition of each rank. According to [15], the closest the Histogram is to the one of a uniformly distributed random variable, the better the Ensemble Forecast is. Indeed, a rather high under-dispersitivity or a too large bias would directly overcharge some ranks/beans.

The Brier Score. For binary variables, the Brier Score (BS) is a practical equivalent of the squared errors [37]. It is defined as

$$BS = \sum_{j=1}^M (p_j - o_j)^2$$

where, for each $j = 1, \dots, M$, p_j is the forecast probability of a boolean event whose realizations are o_j . In the literature, this score is sometimes called the half-Brier score

[37] by some authors. The original Brier score together with an extension to categorical variables is given in [12].

The Brier Score is the most common score in use to measure the accuracy of the estimation of the quantiles of a predictive distribution. Basically, let q be the estimation of the θ -quantile q_θ of a variable X where $\theta \in [0, 1]$ using the method M , then the Brier score of M is

$$BS_M = \sum_{j=1}^M (p - \mathbb{1}(R_j \leq q_\theta))^2$$

where R_j are observations of X . The lesser the Brier score of M is, the better M is.

The Continuous Ranked Probability Score. The Continuous Ranked Probability Score (CRPS) is defined as

$$\text{CRPS}(F, R) = - \int_{\mathbb{R}} (F(y) - \mathbb{1}(y \geq R))^2 dy$$

where F is the cumulative probability function of the distribution used to forecast and R is a realisation [12]. By looking at the definition of BS_M above, one clearly see that the CRPS correspond to the integral of the Brier score extended to continuous random variable. Due to the difficulty to obtain a closed form for the CRPS, it is mostly numerically computed using the formula

$$\text{CRPS}(F, R) = \frac{1}{2} \mathbb{E}[|X - X'|] - \mathbb{E}[|X - R|]$$

where X and X' are independent random variables distributed according to the predictive distribution. Indeed, a numerical approximation using a Monte Carlo method can easily be obtain with such a formula.

The Continuous Ranked Probability Score offers the advantage of generalizing the absolute error [12]. Hence, the average of the CRPS of different pairs of forecasts/observations can be compared to the absolute error of a deterministic forecast. For more details on this last point, we refer to [16] which provides a clear presentation of the CRPS and its links with the Brier score and the absolute error.

4 Computation of Quantiles and Confident Intervals.

Let X be a random variable. In the context of the post-processing of Numerical Weather Predictions or Ensemble Forecast, X is either a weather variable subject to uncertainty or the error between a forecast and a realization considered as a noise. The θ -quantiles of the distribution of X where $\theta \in [0, 1]$ is defined as the value x such that

$$\mathbb{P}\{X \leq x\} = \theta.$$

Quantile forecasts of weather variables related to wind are used in various operational problems on the electricity market such as the quantification of the cost of a suboptimal decision [24]. In fact, optimal decision always relates on quantiles with a given coverage rate [11]. On the other side, quantiles of the error between forecasts and observations help to provide confidence intervals around some predicted values and thus assess the reliability of the forecast weather condition. Low and stable wind speed are very important for

maintenance of the production devices, while very high wind speed has to be taken into account since the devices stop in order not to break.

Numerous methods were developed to obtain quantile forecasts. Here we decide to focus on direct methods, that is those who do not involve cumulative or probability density functions. In fact, some weather variables cannot be handled using standard parametric distributions. For example, the distribution of the error in the case of the wind power is not Gaussian [25]. Two very different works are detailed here, one on classical quantiles, the other on confidence intervals. Each of these work presents minimal assumptions. Direct applications are also given in the paper where they were introduced. At last, both of these methods deal with wind power which can be naturally considered has a weather variable since the production rate of an eolian device mainly relates wind speed and direction, temperature and humidity.

Quantiles Estimation using Local Regression. In [2], the author proposes to estimate the θ -quantile $q_\theta(\mathbf{F})$ of the wind power E_p produces by eolian devices knowing the numerical weather predictions $\mathbf{F} = (F_1, \dots, F_K)$ of K weather variables which are relevant to predict wind power and thus serve as predictors. The author decides to use the local linear quantile regression [39] which approximate $q_\theta(\mathbf{F})$ by a linear functional of the predictors $\mathbf{F} \in \mathbb{R}^K$, that is

$$q_\theta(\mathbf{F}) = \alpha_0 + \alpha^T \mathbf{F},$$

where the regression parameters $\alpha_0 \in \mathbb{R}$ and $\alpha \in \mathcal{M}_K \mathbb{R}$, the space of square matrices of size K , are the solutions of the minimization problem

$$\arg \min_{(b_0, b)} \sum_{i=1}^M \rho(e_{p,i} - (b_0 + b^T (\mathbf{F}_i - \mathbf{F}))) w \left(\frac{\|\mathbf{F}_i - \mathbf{F}\|_2}{h_\lambda(\mathbf{F})} \right),$$

coming from an empirical version of

$$q_\theta(x) = \arg \min_{a \in \mathbb{R}} \mathbb{E}[\rho_\theta(E_p - a) | X = x]$$

using a local linear kernel weighted version of $\mathbb{E}[\rho_\theta(E_p - a) | X = x]$ [30, 35]. Here \mathbf{F}_i and $e_{p,i}$ $i = 1, \dots, M$ are respectively M historical values of the predictors \mathbf{F} and the wind power E_p , ρ_θ and w are respectively the "check" function and the tricube weight function given by

$$\rho_\theta(u) = \begin{cases} u\theta, & u \geq 0, \\ u(\theta - 1), & \text{otherwise,} \end{cases} \quad \text{and } w(u) = \begin{cases} (1 - u^3)^3, & u \in [0, 1], \\ 0, & \text{otherwise,} \end{cases}$$

$\|\cdot\|_2$ is the Euclidian norm and h_λ is the distance of \mathbf{F} to λM -nearest predictor value [30] where $\lambda \in (0, 1]$ and $\lambda M \in \mathbb{N}$.

The above estimation procedure of θ -quantiles is implemented by the author using numerical weather predictions given by Hirlam10. Numerical experiments are performed to show the importance of the selection of predictors. However, the author does not provide a comparison with an alternative method for the numerical computation of quantiles. Such a work is actually performed in an other article by the same author [3] where the local linear quantile regression is compared to a local Gaussian model and to the Nadaraya-Watson estimator for cumulative distribution function. Overall, the benchmark does not really discriminate any of the method. We can retain that the local linear quantile regression is a serious direct non-parametric method compared to other two which required the derivation of quantiles from the cumulative distribution function and has parametric assumptions.

Confident Intervals using Resampling. In [25], the authors propose a method to compute confidence intervals of wind power forecasts. One of the strength of this method is the possibility to perform On-line update of the confidence intervals. This property is of very high value for decision problems [24]. In this method, numerical weather predictions from an Ensemble Forecast are transformed into wind power forecasts which serve as predictors and can be considered as a set of scenario for wind power. This transformation is performed by plugging the relevant weather variables into a power curve of an eolian devices. To compute the confidence intervals, the authors estimate the quantiles of the distribution of the prediction errors

$$e_{t+k,t} = R_{t+k} - F_{t+k,t}$$

where $F_{t+k,t}$ is the wind power forecast for the look-ahead time issued at time t and R_{t+k} is the measured wind power at time $t + k$. The first step of the estimation is to collect the predictions errors made by the model in the past using a sliding windows defined on the index t . The resulting stack of errors is then sorted using fuzzy sets. In a nutshell, the range of values taken by the variable of interest is divided into sets and the errors coming from the previous forecast are kept only if the forecasts belong to the same set as the current one. Two types of sets are defined here. The first concerns the wind speed whose range is divided into two zones depending on the cut-off level, that is the threshold value of wind power where the eolian device stops in order no to break. The second one concerns the range of the wind power which is divided into three zones in order to capture the amplification of the error caused by the power curve. The author choose a particular slicing detailed in [25, pp. 122]. The second step of the method is a resampling procedure which must be performed multiple times. Basically, N values in the stack of errors are drawn randomly with replacement where N is the size of the stack. Then the values are sorted in the ascending order and the λ percent lowest and $(1 - \lambda)$ percent biggest values where λ represents the percentiles of interest are picked and recorded. Once the resampling step is performed a large number of times, the percentiles are simply given by the averages of the respectively lowest and biggest recorded values.

The method described above is tested by the authors on data coming from Irish wind farms. They provide measures of confidence on the computed intervals [25, Table II, pp. 130] which actually speak from themselves and provide a clear assessment of the method. In [26], they more deeply investigate the performance of the method. More precisely, they construct a measuring tool for the assessment of methods which aims at computing quantile forecasts and confidence intervals. They apply it to the present method. Now in [25], the authors also construct a method to narrow their confidence intervals. They introduce a so-called meteo-risk index which measures the spread of the forecasts. They use it to define a scale factor which allows to reduce confidence intervals in case of stable weather conditions. Overall, it is the great simplicity of this method which is very seducing compared the previous one we presented. We do no perform a benchmark to compared the methods, but it is clearly an interesting idea for a future work.

5 Cumulative Distribution Functions and Thresholds

The estimation of cumulative distribution functions is an essential step to numerically solve decision problem such as the Two-Price Market [24]. This functional is also a cornerstone to evaluate the probability of given threshold values, that is $\mathbb{P}\{X \leq x\}$, $\mathbb{P}\{X > x\}$

or $\mathbb{P}\{x_1 \leq X \leq x_2\}$ where X denotes a weather variable. Two very interesting threshold for the electricity production is the minimum required wind speed for an eolian device to work and the maximum wind speed before the cut-off of the device. Moreover, the cumulative distribution function can be used to compute the quantiles through inversion. This can be performed numerically with classical algorithm such as the Newton-Raphson methods. We would not present examples of such a process nor assessing any performance of this method versus the direct computation of quantiles.

Here we propose to review two different methods for the computation of cumulative distribution functions. The first one is deterministic while the second one is statistical. Both of them rely on very classical technology. Albeit, the choice of certain parameters or forms of the involved functional was performed carefully by the authors in order to deal with the characteristics of weather variables.

Direct Model Output. We present here one of the first method historically introduced to post-process Ensemble Forecast. It is thus worth to take a look at it. Moreover, it is a very simple method which can be implemented directly after the reading of this paragraph. The bad point is that this method is based on a plotting formula which is controversial since the validity and the reliability is not clearly assess mathematically.

Direct model output tries to estimate the quantities $\mathbb{P}\{X \leq x\}$ using a formula called the Tukey Plotting Position [37, (3.18) pp. 40 and Table 3.2 pp. 41]. The advantage of this formula which derive from the classical *democratic voting* method is that it correct deficiencies such as the assignment of zero probability to very small quantiles. It arises, for example, when the classical Gumbel formula is used [37, pp. 285].

Let us now explain the Direct model output. The N Ensemble members are regrouped with the value of the x -quantile which is here estimated empirically from the Ensemble members using any classical estimators. Then $\mathbb{P}\{X \leq x\}$ is simply estimated using the formula

$$\mathbb{P}\{X \leq x\} = \frac{Rank(x) - \frac{1}{3}}{(N + 1) + \frac{1}{3}}$$

where $Rank(x)$ is the position of x in the context of the Ensemble members.

The problem with such an estimator is that it does not correct any of the issues such as under dispersion. It basically leads to inaccurate and overconfident results without a very large ensemble [37].

Estimation using Logistic Regression. Among the first methods for the estimation of $\mathbb{P}\{X \leq x\}$ where X denotes as usual a weather variable, regression-based methods were very popular [37]. One of the most successful of such methods is certainly Logistic Regression [38]. Basically, when a weather variable X is measured, we either have $\{X \leq x\}$ or $\{X > x\}$. In other words, we have the realization of a Bernoulli law. The estimation of $\mathbb{P}\{X \leq x\}$ thus simply become an approximation of the parameter p of this Bernoulli law. The idea behind the Logistic Regression is to estimate the logit of p with a linear functional of the predictors $\mathbf{F} = (F_1, \dots, F_N)$ which are the values of the N Ensemble members. It follows

$$\text{Logit}(p) = \log\left(\frac{p}{1-p}\right) = b_0 + b_1 F_1 + \dots + b_N F_N$$

where the b_i 's are the regression's parameters to estimates and the F_i 's are the aforementioned predictors. In practice, the estimation of the b_i 's is performed using the maximum

likelihood estimator with

$$\Lambda(F_1, \dots, F_N) = \prod_{i=1}^M \frac{\mathbb{1}(R^i \leq q) \exp(b_0 + b_1 F_1^i + \dots + b_N F_N^i) + (1 - \mathbb{1}(R_i \leq q))}{1 + \exp(b_0 + b_1 F_1^i + \dots + b_N F_N^i)}$$

where the $(R_i)_{1 \leq i \leq M}$ are observed outcomes of X and $(F_j^i)_{1 \leq i \leq M, 1 \leq j \leq N}$ are historical Ensemble members. Here $\Lambda(b_1, \dots, b_N)$ is the joint likelihood function for the multiple occurrence of the Bernoulli experiments with p is replaced using the Logit formula by

$$\frac{\exp(b_0 + b_1 F_1 + \dots + b_N F_N)}{1 + \exp(b_0 + b_1 F_1 + \dots + b_N F_N)}$$

In [38], the authors propose a two-predictors logistic regression applied to daily temperature, medium-range temperature and precipitation forecasts. They choose to take the Ensemble's mean \bar{F} and the Ensemble standard deviation \bar{S} as predictors. With such a setting,

$$\mathbb{P}\{X \leq x\} = \frac{\exp(\alpha_0(x) + \alpha_1(x)\bar{F} + \alpha_2(x)\bar{F}\bar{S})}{1 + \exp(\alpha_0(x) + \alpha_1(x)\bar{F} + \alpha_2(x)\bar{F}\bar{S})}$$

where $\alpha_0(x)$, $\alpha_1(x)$ and $\alpha_2(x)$ are the parameters that must be estimated for each desired x . The authors called this method LR(2) and compared it to a logistic regression with a single predictor as well as other methods such as Nonhomogeneous Gaussian Regression [13]. However, we do not know the algorithm used to performed the Maximum Likelihood Estimation. The experiments can therefore not be reproduce exactly. The comparison shows that LR(2) performed quite well particularly in terms of rank probability score [38, Fig. 1 and 2] and Brier score [38, Fig. 3]. It also points out that LR(2) is a serious challenger to the Nonhomogeneous Gaussian Regression. Overall, the authors acknowledge that their method must be tested on the different weather variables before drawing completely reliable conclusions.

For the electricity market, the method we presented here is very relevant but actually not for direct weather variables. In fact, it was assessed to be a very powerful method for two categorical variables involved in the analysis of the impact of renewable energy generation on electricity markets and which depends on weather forecasts [24].

6 Estimations of Probability Density Functions

Probability density functions are certainly the most well-known probabilistic functionals. It provides the complete repartition of the mass of a random variable and provide thus more information than quantiles or cumulative distribution functions. Moreover, it has a strong physical meaning. Indeed, as we explain above when we describe the models of numerical weather prediction, a probability density function is the natural output of a Monte Carlo methods designed to numerically solve a model.

For the electricity marker, the probability density functions are of major interest since they are involved in the design of optimal offering strategies or in the optimal quantification of reserve requirements [24]. The problem of the two-price market can also be solved using probability density functions and the problem of the one-price market requires it.

Here we present a few methods of statistical post-processing of an Ensemble Forecast which provide a probability density function as an output.

Gaussian Kernel Dressing. In [38], the authors compare their method to compute cumulative distribution function to the integration of a density estimated using a kernel estimator combined with Gaussian ensemble dressing [29, 34]. The latter method is often called Gaussian kernel dressing [4]. In few words, this method estimate the conditional density $f_\sigma(x|\mathbf{F})$ of $\mathbb{E}[X|\mathbf{F}]$ where X is a weather variable and $\mathbf{F} = (F_1, \dots, F_N)$ is an Ensemble forecast of X with N members using [30, Chap. 2.4]

$$f_\sigma(x|\mathbf{F}) := \frac{1}{\sigma N} \sum_{i=1}^N G\left(\frac{x - F_i}{\sigma}\right)$$

where G denotes a Gaussian and σ is a bandwidth estimator determined by a method of Gaussian ensemble dressing. The density $f_\sigma(v|\mathbf{X})$ is basically estimated by a sequence of bumps, each of which is centered at X_i and has a length proportional to σ .

Originally, the methods of Gaussian ensemble dressing were designed to cheaply extend the number of members in an Ensemble Forecast and also remove the bias/errors in the members using historical pairs of forecasts-validations [29]. Here we examine the method of Gaussian ensemble dressing developed in [34] and combined with kernel dressing in [38]. In [34], the authors use the second moment to dress the Ensemble Forecast. This choice is made to priorly treat the problem of underdispersivity of Ensemble Forecast. It also overcomes some of the limitations of the dressing method developed in [29]. In a nutshell, the authors assume each ensemble members to be of the form

$$F_i = \bar{F}_i + F'_i$$

where \bar{F}_i denotes the mean of X and F'_i follows the recentered distribution of V . Hence, they consider that a dressed ensemble member has to be

$$\Psi_i = \bar{F}_i + F'_i + \varepsilon$$

where ε is Gaussian independent of $(F_i)_{1 \leq i \leq N}$. As a matter of fact, $\Psi \sim \mathcal{N}(\bar{F}, \sigma)$ with a σ which is computed using the second moment constraint. According to the computation performed in appendix of [34] and aggregated in the formula (9a),

$$\sigma = \sum_{k=1}^M (\bar{F}^k - R_k)^2 - \left(1 + \frac{1}{M}\right) \sum_{k=1}^M \sigma^k$$

where M denotes the number of pairs of available historical data belonging to the same season, $(R^k)_{1 \leq k \leq M}$ are the observed outcomes and $\bar{F}_{1 \leq k \leq M}^k$ and $\sigma_{1 \leq k \leq M}^k$ are respectively the empirical means and variance of the historical Ensemble Forecasts.

Part of the experiments on this methods were already commented right above. The weather variables use in those test are still daily temperature, medium-range temperature and precipitation. In term of results, the Brier scores of the Gaussian kernel dressing is given in [38, Fig. 6]. The most interesting part is certainly the good rank probability scores on a short training period obtained by Gaussian kernel dressing [38, Table 1]. Overall, it appears to be a very appealing methods in the case of a lack of historical data.

A method based on Gaussian kernel dressing were also developed in [27] to forecast the wind power using an Ensemble Forecast. The authors of this article model the variance of the gaussian with a Logistic function. The two parameters of the function are assumed to be the same for every kernels. For the estimation of the parameters, the authors build an

adaptative estimation using a recursive maximum likelihood estimation and the logit of the parameters. The authors compare forecasts produced with their method to raw Ensemble Forecast. Clearly, their Gaussian kernel dressing provide tighter and sharper forecast. However, a comparison of their method to [38] would have been very interesting. Now this work shows that Gaussian kernel dressing can be used on wind power with success. This is not the case for the different methods to compute densities presented above.

The Screening Regression Procedure. The procedure of multiple linear regression provides an estimate of a weather variable X by relating it to a set of N outputs F_1, \dots, F_n of a model of numerical weather predictions. As an equation,

$$X = a_0 + a_1 F_1 + \dots + a_N F_N,$$

where the parameters a_0, \dots, a_N minimize the sum of the squares of the estimation errors on the historical data, that is

$$(a_0, \dots, a_N) = \arg \min_{(b_0, \dots, b_N)} \sum_{j=1}^M (R^j - b_0 + b_1 F_1^j + \dots + b_N F_N^j)^2.$$

Here $R_{1 \leq j \leq M}^j$ and $(F_k^j)_{1 \leq j \leq M, k=1, \dots, N}$ represent the historical data of respectively the true realization of X and the outputs F_1, \dots, F_N .

The main problem with multiple linear regression is the selection of the predictors. Indeed, too few or too many of them can provide an estimation which performed poorly. In [10], the authors choose a method of selection called the screening procedure. First, the variable which is the most highly correlated to V is captured. Then, the others predictors are selected iteratively regarding their contribution to the variance reduction, that is their ability to minimize

$$\frac{\frac{1}{n} \sum_{j=1}^M (R^j - \bar{R})^2 - \frac{1}{n} \sum_{j=1}^M (R^j - \bar{F}^j)^2}{\frac{1}{n} \sum_{j=1}^M (R^j - \bar{R})^2} \quad (1)$$

where \bar{R} is the empirical mean of the true realizations $R_{1 \leq j \leq M}^j$ of X and \bar{F}^j are the estimate of X using the procedure of multiple linear regression with the predictors F_1, \dots, F_N and the historical predictions F_1^j, \dots, F_N^j , $j = 1, \dots, M$.

The multiple linear regression with the screening procedure was experimented in [10] on various weather variables such as surface wind, maximum temperature or cloud amount. The point of the tests is to prove the efficiency of the post-processing of numerical weather predictions. For the probability of precipitation, the authors provides a comparison of the Brier scores of their method versus basic forecasts at local station [10, Fig. 2]. On average, the multiple linear regression performed better. This fact is confirmed on maximum temperature when the method is compared to the so-called Klein-Lewis [10, Fig. 4]. Finally, we present this method in order to provide one which deals with Numerical Weather Predictions only, not Ensemble Forecast. In addition, the article which introduced this method provide good guidance on the selection of variables for weather predictions.

Nonhomogeneous Regression Techniques. Nonhomogeneous Regression Methods has received several attention as a method to post-process Ensemble Forecasts. Multiple variants of the basic concepts exposed in [13] and called EMOS were developed [19]. The classical regression method where a weather variable X is approximated by

$$X = a + b_1 F_1 + \dots + b_N F_N + \varepsilon$$

where ε is assumed to be Gaussian and F_1, \dots, F_N are the predictors, suggests to take has predictive probability density function a Gaussian whose mean and variance respectively equal the regression estimate and the mean of the square prediction errors. This is a basic estimation by the maximum likelihood estimator of the regression coefficients. In [13], the authors modify the above regression equation by assuming

$$\text{Var}(\varepsilon) = c + dS^2$$

where S is the empirical variance of the Ensemble Forecast and c and d has to be computed. They also decide to replace the classical estimation of the coefficients a, b_1, \dots, b_N, c and d by the maximum likelihood estimator. They choose to use take the parameters a, b_1, \dots, b_N, c and d which minimize the CRPS which can be here computed analytically, that is

$$(a, b_1, \dots, b_N, c, d) = \arg \min_{(A, B_1, \dots, B_N, C, D)} \frac{1}{M} \sum_{k=1}^M (C + DS_k^2)^{\frac{1}{2}} \left\{ Z_k [2\Phi(Z_k) - 1] + 2\phi(Z_k) - \frac{1}{\sqrt{\pi}} \right\}$$

where

$$Z_k = \frac{R^k - (A + B_1 F_1^k + \dots + B_N F_N^k)}{(C + DS_k^2)^{\frac{1}{2}}}.$$

Here Φ and ϕ respectively denote the cumulative distribution function and the probability density function of a standard Gaussian, $(F^k, F_1^k, \dots, F_N^k, S^k)$, $k = 1, \dots, M$ stand for the historical data of realizations, Ensemble members and empirical variance of the Ensemble members. . The estimation of the parameters is performed numerically by the authors using the Broyden–Fletcher–Goldfarb–Shanno algorithm which is implemented in the R language.

The method is tested using the University of Washington ensemble over the Pacific Northwest to forecast the sea level and the surface temperature. The basic scoring rules such as the Brier score and the CRPS are used to assess the performance of the method. The author also compare forecasts to raw ensemble, de-biased ensemble and climatological forecasts. The regression method provided by the authors clearly outperformed the climatological forecast. It more lightly beats the bias-corrected Ensemble Forecast [13, Table 4]. Some comparisons with other statistical methods are given in [38] using the rank probability score and the Brier score. In term of applications, this method was proven to be very relevant to forecast low and high wind regime when the kernel is respectively replaced by a truncated Gaussian kernel and a so-called GEV distribution [19]. Nonhomogeneous Regression can be easily combined with regime-switching. Hence, multiple wind regimes can be thus treated with the right kernel. This is clearly a great strength and thus a good encouragement to use this method on wind farm locate on places with irregular wind regimes.

Estimation using Bayesian Model Averaging. In [28], the authors propose to estimate the conditional density $\mathbb{E}[X|\mathbf{F}]$ using a simplification of Bayesian model averaging. As usual, X and $\mathbf{F} = (F_1, \dots, F_N)$ respectively denotes a weather variable and a vector of N Ensemble members [17]. The method can be seen as an improved kernel dressing. Boiled down to its essence, the idea is to estimate the conditional density $f(x|\mathbf{F})$ as a mixture, that is

$$f(v|\mathbf{F}) = \sum_{k=1}^N w_k g_k(x|F_k)$$

Here, for each $k = 1, \dots, N$, $g_k(x|F_k)$ is a conditional density which represents the information provided by F_k and w_k is a weight of the contribution of F_k . The weights w_1, \dots, w_n can be understood as the probability that F_k is the member of the ensemble which provides the most accurate forecast. In classical Bayesian model averaging, w_k is replaced by $\mathbb{P}(g_k(x|F_k)|X_D)$ where X_D represents the historical data, that is the posterior probability of the estimation of X using the kernel $g_k(x|F_k)$. For the estimation of temperature and sea level, the authors choose to take $\mathbb{E}[X|F_k] \sim \mathcal{N}(a_k + b_k F_k, \sigma^2)$. They directly estimates a_k and b_k where $k = 1, \dots, N$ using linear regression with paired historical data of true realizations and Ensemble Forecasts. It can be considered as a bias-correction process. The parameter σ and the weight are, on the other side, estimated using the maximum likelihood estimator and the same pairs of historical data. It has actually become usual to replace classical Bayesian estimation procedures by the maximum likelihood estimator in Bayesian model averaging to avoid expensive computations. In some instances, it provides a justified large-sample approximation [8].

The authors performed various benchmarks of their method where the EM algorithm is use to estimate the maximum likelihood. They compare their method to various raw Ensemble Forecasts using different measurement tools such as the CRPS or the RMSE. Overall, it is showed that the method performed very well [28, Table 4 to 6 and 8 to 12]. The authors also investigated the possibility to reduce the number of ensemble members by disqualifying those with a poor weight [28, Table 7]. Due to its performance, the present method was adapted to the prediction of wind speed [31]. In this article, several modification were performed since the kernels are chosen to be Gamma distributed. Different techniques of the estimation of the maximum likelihood were tried. However, it appears that the different methods perform similarly [31, Figure 4]. Some benchmarks which compared the method to basic Ensemble Forecast and climatology are also provided. The result are similar [31, Table 2] to those obtained in [28].

Bayesian Model Averaging is actually one of the leading method for the post-processing of Ensemble Forecast. It really takes into account the model uncertainty arising in Ensemble Forecast. However, several non-trivial improvements has to be performed in order to used this method to treat problems linked to the electricity market [32].

References

- [1] G. N. Bathurst, J. Weatherill, and G. Strbac. Trading wind generation in short term energy markets. *IEEE Trans. Power Syst.*, 17(3):782–789, 2002.
- [2] J. B. Bremnes. Probabilistic wind power forecasts using local quantile regression. *Wind Energy*, 7(1):47–54, 2004.
- [3] J. B. Bremnes. A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy*, 9(1-2):3–11, 2006.
- [4] J. Bröcker and L. A. Smith. From ensemble forecasts to predictive distribution functions. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.*, 60 A(4):663–678, 2008.
- [5] B. G. Brown, R. W. Katz, and A. H. Murphy. Time Series Models to Simulate and Forecast Wind Speeds and Power. In *J. Clim. Appl. Meteorol.*, volume 23, pages 1184–1195. 1984.

- [6] R. Buizza, P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei. A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Weather Rev.*, 133(5):1076–1097, 2005.
- [7] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. *Renew. Sustain. Energy Rev.*, 12(6):1725–1744, 2008.
- [8] D. Draper. Assessment and propagation of model uncertainty. *J. R. Stat. Soc. Ser. B*, 57(1):45–97, 1995.
- [9] A. Fabbri, T. G. San Román, J. R. Abbad, and V. H. M. Méndez Quezada. Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market. *IEEE Trans. Power Syst.*, 20(3):1440–1446, 2005.
- [10] H. R. Glahn and D. a. Lowry. The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. 11(8):1203–1211, 1972.
- [11] T. Gneiting. Quantiles as optimal point forecasts. *Int. J. Forecast.*, 27(2):197–207, 2011.
- [12] T. Gneiting and A. E. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.*, 102(477):359–378, 2007.
- [13] T. Gneiting, A. E. Raftery, Anton H. W., and T. Goldman. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Mon. Weather Rev.*, 133(5):1098–1118, 2005.
- [14] T. M. Hamill. Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Weather Rev.*, 129(3):550–560, 2001.
- [15] T. M. Hamill and S. J. Colucci. Evaluation of Eta–RSM Ensemble Probabilistic Precipitation Forecasts. *Mon. Weather Rev.*, 126(3):711–724, 1998.
- [16] Hans Hersbach. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather Forecast.*, 15(5):559–570, 2000.
- [17] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian Model Averaging: A Tutorial. *Stat. Sci.*, 14(4):382–417, 1999.
- [18] E. Kalnay. *Atmospheric modeling, data assimilation, and predictability*, volume 54. Cambridge University Press, 2003.
- [19] S. Lerch and T. L. Thorarinsdottir. Comparison of nonhomogeneous regression models for probabilistic wind speed forecasting. *Tellus A*, 1:1–17, 2013.
- [20] M. Leutbecher and T. N. Palmer. Ensemble forecasting. *J. Comput. Phys.*, 227(7):3515–3539, 2008.
- [21] J. M. Lewis. Roots of Ensemble Forecasting. *Mon. Weather Rev.*, 133:1865–1885, 2005.
- [22] E. N. Lorenz. Atmospheric predictability experiments with a large numerical model. *Tellus A*, pages 505–513, 1982.

- [23] H. Lund. Large-scale integration of wind power into different energy systems. *Energy*, 30(13):2402–2412, 2005.
- [24] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno. *Integrating renewables in electricity markets - Operational problems*, volume 205 of *International Series in Operations Research and Management Science*. Springer, 2014.
- [25] P. Pinson and G. Kariniotakis. On-line assessment of prediction risk for wind power production forecasts. *Wind Energy*, 7(2):119–132, 2004.
- [26] P. Pinson, G. Kariniotakis, H. A. Nielsen, T. S. Nielsen, and H. Madsen. Properties of quantile and interval forecasts of wind generation and their evaluation. *Proc. Eur. Wind Energy Conf. Exhib.*, (October 2015):2–6, 2006.
- [27] P. Pinson and H. Madsen. Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy*, 12(2):137–155, 2009.
- [28] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Mon. Weather Rev.*, 133(5):1155–1174, 2005.
- [29] M. S. Roulston and L. A. Smith. Combining dynamical and statistical ensembles. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.*, 55(1):16–30, 2003.
- [30] B. W. Silverman. *Density estimation for statistics and data analysis*, volume 37 of *Monogr. Stat. Appl. Probab.* Chapman Hall/CRC.
- [31] J. M. Slughter, T. Gneiting, and A. E. Raftery. Probabilistic Wind Speed Forecasting Using Ensembles and Bayesian Model Averaging. *Source J. Am. Stat. Assoc.*, 105(489):25–35, 2010.
- [32] J. M. Slughter, T. Gneiting, and A. E. Raftery. Probabilistic Wind Vector Forecasting using Ensembles and Bayesian Model Averaging. *Mon. Weather Rev.*, 105(489):120718065932001, 2012.
- [33] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal. A review of wind power and wind speed forecasting methods with different time horizons. *North Am. Power Symp.*, pages 1–8, 2010.
- [34] X. Wang and C. H. Bishop. Improvement of ensemble reliability with a new dressing kernel. *Q. J. R. Meteorol. Soc.*, 131(607):965–986, 2005.
- [35] G. S. Watson. Smooth regression analysis. *Indian J. Stat.*, 26(4):359–372, 1964.
- [36] D. S. Wilks. Comparison of ensemble-MOS methods in the Lorenz ’96 setting. *Meteorol. Appl.*, 13(3):243–256, 2006.
- [37] D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*, volume 100 of *Int. Geophys. Ser.* Elsevier, third edition, 2011.
- [38] D.S. Wilks and T.M. Hamill. Comparison of Ensemble-MOS Methods Using GFS Reforecasts. *Mon. Weather Rev.*, 135(6):2379–2390, 2007.
- [39] K. Yu and M. C. Jones. Local Linear Quantile Regression. *J. Am. Stat. Assoc.*, 93(441):228–237, 1998.