MDPI

Article

Probabilistic solar forecasting using quantile regression models

Philippe Lauret [1,†,‡]*, Mathieu David [1,‡] and Hugo T.C. Pedro [2,‡]

[1] PIMENT laboratory, Université de La Réunion,15 avenue René Cassin, 97715 Saint-Denis, France; philippe.lauret@univ-reunion.fr
[2] Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, Center for Energy Research University of California, San Diego, La Jolla, CA 92093, USA; hpedro@ucsd.edu
* Correspondence: philippe.lauret@univ-reunion.fr; Tel.: +262-262938127
† Current address: Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, Center of Excellence in Renewable Energy Integration, Center for Energy Research, University of California San Diego, La Jolla, CA 92093, USA
‡ These authors contributed equally to this work.

Academic Editor: name
Version September 11, 2017 submitted to Energies

Abstract: In this work, we assess the performance of three probabilistic models for intra-day solar forecasting. More precisely, a linear quantile regression method is used to build three models for generating 1h to 6h ahead probabilistic forecasts. Our approach is applied to forecasting solar irradiance at a site experiencing high variable sky conditions using the historical ground observations of solar irradiance as endogenous inputs and day-ahead forecasts as exogenous inputs. Day-ahead solar forecasts are provided by the European Center for Medium-Range Weather Forecast (ECMWF). This organization maintains and runs the Numerical Weather Prediction (NWP) model named Integrated Forecast System (IFS). Several metrics mainly originated from the weather forecasting community are used to evaluate the performance of the probabilistic forecasts. It is demonstrated that the additional exogenous inputs provided by the ECMWF clearly improves the quality of the intra-day probabilistic forecasts. It is also shown that the sky conditions experienced by a site may have an impact on the statistical performance of the probabilistic models.

Keywords: Probabilistic solar forecasting; quantile regression; ECMWF; reliability; sharpness; CRPS

1. Introduction

Solar forecasts are required to increase the penetration of solar power into electricity grids and also to ensure the security of the supply-demand balance ([1]). Also, the recent development of grid-connected storages associated with intermittent solar renewables necessitates solar forecasts in order to optimize their operational management ([2]). A forecast is inherently uncertain and in a context of decision-making faced by the operator, a point forecast plus a prediction interval is a true added-value. In other words, if appropriately incorporated in decision-making processes, uncertainty estimates may permit to significantly increase the value of solar power generation. In a previous study, Lauret *et al.* [3] showed that the combination of day-ahead ECMWF forecasts with past ground measurements clearly improved the accuracy of the intra-day point deterministic forecasts. In this work, it is investigated if such a combination could also improve the quality of the probabilistic forecasts. For that purpose, we build three probabilistic models based on the linear quantile regression method. The first model uses the deterministic point prediction generated by a linear model as unique explanatory input variable. The second one only makes use of past solar ground measurements while the third one combines past ground data with exogenous inputs

provided by the European Center for Medium-Range Weather Forecast (ECMWF). This organization maintains and runs the global Numerical Weather Prediction (NWP) model Integrated Forecast System (IFS)([4].

In order to evaluate the quality of the probabilistic forecasts, we propose to use the reliability diagram ([5], the rank histogram (or Talagrand histogram) ([6]) and the continuous ranked probability score (CRPS) ([7]) . We will also assess the sharpness of the predictive distributions by calculating the prediction intervals normalized average width (PINAW) ([8],[9]).

The remainder of this paper is organized as follows: Section 2 is devoted to the sites' description and data preprocessing. Section 3 sets out an in-depth analysis of the sky conditions experienced by each site while Section 4 describes the day-ahead solar forecasts provided by the ECMWF center. Section 5 presents the different probabilistic models based on the quantile regression method. Section 6 details the different metrics used to evaluate the quality of the solar probabilistic forecasts and section 7 presents the results based on the evaluation framework. Section 8 discusses the impact of the sky conditions on the quality of the predictive distributions. Finally, section 9 gives some concluding remarks.

2. Data

In this paper, we used data recorded at two sites that exhibit completely different sky conditions. The first site (Le Tampon) is located in La Réunion Island and experiences variable sky conditions while the second one is situated in the continental US and experiences an arid climate (Desert Rock). Table 1 lists the characteristics of the two sites. For this study, we computed 1-hour averages of global horizontal solar irradiance (GHI) directly from the raw 1 min-data for 2 consecutive years. One complete year (2012) was used to build the models (training dataset) and another one (2013) was used to evaluate the probabilistic forecasts (testing or evaluation dataset).

In the realm of solar forecasting community, it is now common practice to work with the clear sky index instead of the original GHI time series. Indeed, solar irradiance is characterized by diurnal and seasonal variations, thus clear sky model is used to remove this deterministic component in the GHI time series, which results in the clear sky index kt^*:

$$kt^*(t) = \frac{I(t)}{I_{clr}(t)} \tag{1}$$

This quantity corresponds to the ratio of the measured GHI $I(t)$ to the theoretical GHI observed under clear sky $I_{clr}(t)$. With this methodology, the models designed in this work are dedicated to the forecasting of the stochastic component of the global irradiance due to the cloud cover, leaving the geometric and the deterministic part to be modeled by the clear sky model. In this study, we use the clear sky data provided by the McClear model ([10]) publicly available on the SoDa website ([11]). This model uses aerosol optical depth, water vapor and ozone data from the MACC (Monitoring Atmospheric Composition and Climate) project ([12]).

Finally, it must be noted that the data corresponding to a solar zenith angle θ_Z greater than 85° are removed. This filtering removes less than 1% of the total annual sum of solar energy. Put in other words, night times and low solar elevations were not taken into account for the building and the test of the models. In addition, this filtering process discards less precise data as measurement uncertainties associated to pyranometers are typically much higher than 3.0% for $\theta_Z > 85°$.

Table 1. Main characteristics of the two sites under study.

Site	Le Tampon	Desert Rock
Provider	PIMENT	SURFRAD
Position	21.3°S 55.5°E	36.6°N 119.0°W
Elevation	550m	1007m
Climate type	Tropical humid	Arid
Training year	2012	2012
Testing year	2013	2013
Mean GHI of the testing year (W/m^2)	455	548
Annual solar irradiance (MWh/m^2)	1.712	2.105
Site variability ($\sigma(\Delta kt^*_{\Delta T=1h})$)	0.24	0.14

3. Site Analysis

Figure 1 plots the clear sky index distribution of each location. As shown by Figure 1, the continental station of Desert Rock experiences weather dominated by clear skies, evidenced by the high frequency of kt^* near 1. For the insular location Le Tampon, the occurrence of clear skies is much lower. In figure 1, one can notice a significant number of occurrences of clear sky index above 1. These events result from a phenomenon known as cloud edge effect and can be observed anywhere in the world ([13],[14]) .

Table 1 (last row) lists the solar variability of each site. This metric was defined by [15] as the standard deviation of the changes in the clear sky index $\sigma(\Delta kt^*_{\Delta T})$. The values of solar variability in Table 1 have been calculated for a time scale ΔT=1h (as we deal with hourly GHI records) and for a time span that corresponds to two years of hourly data. A site with variability above 0.2 is considered as experiencing very unstable GHI conditions [15]. The variability of the site of Le Tampon is above this threshold.

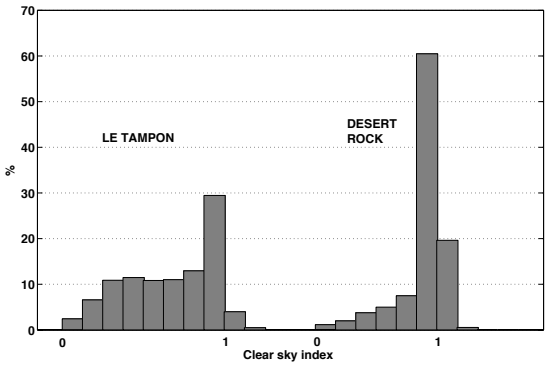

Figure 1. Statistical distribution of the 1-hour clear sky index for the insular and continental sites.

A previous paper related to point deterministic forecasts in a challenging insular context [3] showed that insular sites like Le Tampon, which exhibits higher solar variability due to clouds that are formed locally, are prone to have worse forecasting performance than less variable (continental or insular) sites. Section 8 below discusses the impact of the sky conditions on the forecasting skill of the different probabilistic methods.

4. ECMWF day-ahead forecasts

The European Center for Medium-Range Weather Forecast (ECMWF) is an intergovernmental organization that provides operational weather forecasts. It maintains and runs the numerical weather prediction (NWP) model Integrated Forecast System (IFS). NWP models outperform forecasts based on satellite data or ground measurements (only) for horizons longer than 5 h ([16]). In order to supply forecasts that can be used by the grid operator for day-ahead scheduling, we retrieved the data generated by the IFS at 12 h 00 UTC (16 h 00 in Réunion Island). The NWP forecasts provide hourly GHI data with a spatial resolution of $0.125° \times 0.125°$ (approximately 14 km $\times$ 14 km in Réunion — see Figure 2).

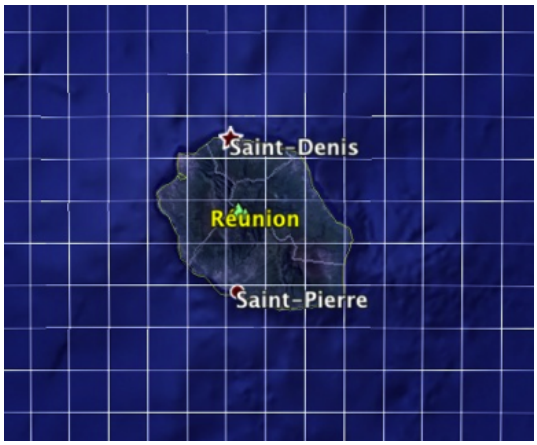

Figure 2. ECMWF grid for La Réunion Island.

5. Probabilistic models

In this study, we make no assumption about the shape of the predictive distributions. Consequently, the probabilistic forecasts are produced by non-parametric methods like quantile regression. In other words, the predictive distributions are defined by a number of quantile forecasts with nominal proportions spanning the unit interval ([17]).

In this work, we chose to use the simple linear quantile regression (QR) method proposed by [18] in order to build the probabilistic models. It must be stressed that possibly better results can be obtained with more sophisticated machine learning methods like quantile regression forest ([19]), QR Neural Networks ([20]) or gradient boosting techniques ([21]). But our goal here is to focus on the combination of ground telemetry and ECMWF forecasts as well as the impact on the sky conditions experienced by a site on the quality of the probabilistic forecasts.

5.1. The linear quantile regession method

This method estimates the quantiles of the cumulative distribution function of some variable y (also called predictand) by assuming a linear relationship between y and a vector of explanatory variables (also called predictors) $\mathbf{x}$:

$$y = \boldsymbol{\beta}\mathbf{x} + \epsilon, \tag{2}$$

where $\boldsymbol{\beta}$ is a vector of parameters to optimize and ϵ represents a random error term.

In quantile regression, quantiles are estimated by applying asymmetric weights to the mean absolute error. Following [18] the quantile loss function is:

$$\rho_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases} \tag{3}$$

with τ representing the quantile probability level.

The quantity $\hat{y}_\tau = \hat{\boldsymbol{\beta}}_\tau \mathbf{x}$ is the τ^{th} quantile estimated by the quantile regression method with the vector $\hat{\boldsymbol{\beta}}_\tau$ obtained as the solution of the following minimization problem:

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{N} \rho_\tau(y_i - \boldsymbol{\beta}\mathbf{x}_i), \tag{4}$$

where $(\mathbf{x}_i, y_i)$ denotes a pair of vector of predictors and corresponding observed predictand in the training set.

It must noted that the quantile regression method estimates each quantile separately (i.e. the minimization of the quantile loss function is made for each τ separately). As a consequence, one can obtain quantile regression curves that may intersect i.e $\hat{y}_{\tau 1} > \hat{y}_{\tau 2}$ when $\tau_1 < \tau_2$. To avoid this issue during the model fitting, we used the rearrangement method described by [22].

In the following, the predictand y is the clear sky index kt^*. Therefore, the output of the probabilistic model for each forecasting time horizon h is the ensemble of 9 quantiles defined by $\{\hat{kt^*}_\tau(t+h)\}_{\tau=0.1,0.2,\cdots,0.9}$ for each forecasting time horizon h=1,$\cdots$,6 hours. This set of $\{\hat{kt^*}_\tau(t+h)\}_{\tau=0.1,0.2,\cdots,0.9}$ quantiles can be transformed in GHI quantiles $\{\hat{I}_\tau(t+h)\}_{\tau=0.1,0.2,\cdots,0.9}$ by using Equation 1. As mentioned above, this set of quantiles represents the predictive distribution of the variable of interest (here GHI) at lead time $t+h$. This set of quantiles may form also what the verification weather community [23] calls an Ensemble Prediction System (EPS).

In addition, prediction intervals with different nominal coverage rates can be inferred from this set of quantiles. Prediction intervals give a range of possible values within which the true value of GHI is expected to lie with a certain probability, that is its nominal coverage rate. In this work, the $(1-\alpha)100\%$ central prediction interval is generated by taking the $\frac{\alpha}{2}$ quantile as the lower bound and the $1-\frac{\alpha}{2}$ quantile as the upper bound. More precisely, a prediction interval with $(1-\alpha)100\%$ nominal coverage rate produced at time t for lead time $t+h$ is defined by:

$$\hat{\mathrm{PI}}_{(1-\alpha)100\%}(t+h) = [\hat{I}(t+h)_{\tau=\frac{\alpha}{2}}, \hat{I}(t+h)_{\tau=1-\frac{\alpha}{2}}] \tag{5}$$

For instance, a $\hat{PI}_{80\%}$ is calculated from the two extremes quantiles of the forecasted irradiance distribution i.e. $\hat{I}_{0.1}(t+\Delta t)$ and $\hat{I}_{0.9}(t+\Delta t)$.

5.2. QR Model 1: Point forecast as unique explanotary variable

In this simple set-up, the deterministic point forecast produced by a simple linear auto-regressive moving average (ARMA) model constitutes the unique explanatory variable of the QR model. Therefore, we have:

$$\mathbf{x}(t) = [\hat{kt}^*_{ARMA}(t+h)] \tag{6}$$

Indeed, in a study related to solar power forecasting, Bacher *et al.*[24] showed that the level of uncertainty depends on the value of the point forecast. The reader is referred to [25] for details regarding the building of this ARMA model.

5.3. QR Model 2: Past ground measurements only

For this model, the vector of explanatory variables consists of the six past ground measurements i.e.:

$$\mathbf{x}(t) = [kt^*(t), \cdots, kt^*(t-6)] \tag{7}$$

5.4. QR Model 3: Combination of past ground measurements plus ECMWF forecast data

In addition to the six past ground measurements, the third QR model takes additional exogenous inputs provided by the day-head ECMWF forecasts i.e., the forecasted clear sky index $\hat{kt}^*_{ECMWF}(t+h)$ and the total cloud cover $\hat{TCC}_{ECMWF}(t+h)$ for the considered forecast horizon h:

$$\mathbf{x}(t) = [kt^*(t), \cdots, kt^*(t-6), \hat{kt}^*_{ECMWF}(t+h), \hat{TCC}_{ECMWF}(t+h)] \tag{8}$$

5.5. PersEn: Persistence ensemble model

Finally, we define the simple baseline persistence ensemble model (PersEn). This model ([25–27] is commonly used to provide reference probabilistic forecasts. In this work, the PersEn considers the GHI lagged measurements in the 10 hours that precede the forecasting issuing time. The selected measurements are ranked to define quantile values for the irradiance forecast.

6. Probabilistic error metrics

In [17] the authors emphasize several aspects related to the quality of a probabilistic forecasting system namely reliability, sharpness and resolution. These required properties for skillful probabilistic forecasts possess different meaning according a meteorologist's point of view or a statistician's point of view. The interested reader is referred to [28] or [29] for a definition of these properties in the realm of meteorology. As an illustration of these different definitions, the meteorological literature ([28]) defines the sharpness property as the ability of the forecasting system to produce forecasts that are able to deviate from the climatological value of the predictand while from a statistical point of view (as it will be the case here), the sharpness property refers to the concentration of the predictive distributions ([17],[8]).

As noted by [17], reliability is seen as a primary requirement when verifying probabilistic forecasts, since a lack of reliability would introduce a systematic bias in subsequent decision-making. A reliable EPS is said to be calibrated. More precisely, a probabilistic forecasting system is reliable if, statistically, the nominal proportions of the quantile forecasts are equal to the proportions of the observed value. In other words, over a testing set of significant size, the difference between observed and nominal probabilities should be as small as possible. This first requirement of reliability will be assessed with the help of reliability diagrams (see sub-section 6.1.1) or by calculating the prediction interval coverage probability PICP ([9]) (see sub-section 6.1.2) which permits to assess the reliability of the forecasts in terms of coverage rate. As a means to corroborate the analysis made with the reliability diagram, we will also provide rank histograms (see sub-section 6.1.3) that allow judging the statistical consistency of the ensemble constituted by the 9 quantiles members.

In this work, and similarly to ([17],[8]), the assessment of sharpness derives from a more statistical point of view with focus on the shape of the predictive distributions. For that purpose, the prediction interval average width (PINAW) metric (see sub-section 6.2) is used to evaluate the sharpness of the predictive distributions ([9]) .

Regarding the third property, namely resolution, it consists in evaluating the ability of the forecast system to issue different probabilistic forecasts (i.e. predictive distributions with prediction intervals that vary in size) depending on the forecast conditions ([17]). For instance, in the case of GHI, the level of uncertainty may vary according the sun's position in the sky (see for the instance the work of [30]). In this paper, however, we will not provide such a conditional assessment and will only focus on the most important properties of a skillful prediction system namely reliability and sharpness.

Finally, it must be noted that reliability can be corrected by statistical techniques also called calibration techniques ([31]), whereas the same is not possible for sharpness. Finally, the Continuous Rank Probability Score (CRPS) ([7]) provides an evaluation of the global skill of our probabilistic models.

6.1. Reliability property

6.1.1. Reliability diagram

The reliability diagram is a graphical verification display used to verify the reliability component of a probabilistic forecast system. In this work, we use the methodology defined by [5] that is especially designed for density forecasts of continuous variables.

This type of reliability diagram plots the observed probabilities against the nominal ones (i.e. the probability levels of the different quantiles). This representation is appealing because the deviations from perfect reliability (the diagonal) can be visually assessed [5]. However, due to the finite number of pairs of observation/forecast and also due to possible serial correlation in the sequence of forecast-verification pairs, it is not expected that observed proportions lie exactly along the diagonal, even if the density forecasts are perfectly reliable [5]. In this work, similarly to [32], consistency bars are computed in order to take into account the limited number of observation/forecast pairs of the evaluation (testing) set.

6.1.2. PICP

The prediction interval coverage probability (PICP) ([9]) is a metric that permits to assess the reliability of the EPS in terms of coverage rate. In addition, and contrary to what we did for reliability diagram, here we assess the PICP as a function of the forecast horizon. In this work, and as mentioned above, we propose to assess the empirical coverage probability of the central prediction intervals for different nominal coverage rate $(1-\alpha)100\%$. Equation 9 gives the definition of the PICP for a specific forecast horizon h and a particular nominal coverage rate :

$$\mathrm{PICP}(h,\alpha) = \frac{1}{N}\sum_{i=1}^{N} 1_{\{I(t+h)\in \hat{\mathrm{PI}}_{(1-\alpha)100\%}\}} \tag{9}$$

The indicator function $1_{\{u\}}$ has the value of 1 if its argument u is true and 0 otherwise.

6.1.3. Rank histogram

The rank histogram ([28]) is another graphical tool for evaluating ensemble forecasts. Rank histograms are useful for determining the statistical consistency of the ensemble, that is, if the observation being predicted looks statistically just like another member of the forecast ensemble ([28]). A necessary condition for ensemble consistency is an appropriate degree of ensemble dispersion leading to a flat rank histogram ([28]). In other words, a flat rank histogram shows that the members of an ensemble system are statistically indistinguishable from the observations. If the ensemble dispersion is consistently too small (underdispersed ensemble), then the observation (also called the verification sample) will often be an outlier in the distribution of ensemble members. This will result in a rank histogram with a U-shape. Conversely, if the ensemble dispersion is consistently too large (overdispersed ensemble) then the observation may too often be in the middle of the ensemble distribution. This will give a rank histogram with a hump shape. In addition, asymmetric rank histograms may suggest that ensemble may possess some biases. Ensemble bias can be detected from overpopulation of either the smallest ranks, or the largest ranks, in the rank histogram. An overforecasting bias will correspond to an overpopulation of the smallest ranks while an underforecasting bias will overpopulate the highest ranks. As a consequence, rank histograms can also reveal deficiencies in ensemble calibration or reliability ([28]). Again, care must be taken when analyzing rank histograms when the number of verification samples is limited. In addition, as demonstrated by [6], a perfect rank histogram does not mean that the corresponding EPS is reliable.

To obtain a verification rank histogram, one needs to find the rank of the observation when pooled within the ordered ensemble of quantile values and then plot the histogram of the ranks. For

a number of members M, the number of ranks of the histogram of an ensemble is $M+1$. If the consistency condition is met, this histogram of verification ranks will be uniform with theoretical relative frequency of $\frac{1}{M+1}$.

6.2. Sharpness

An EPS is sharp if prediction intervals are shorter on average than prediction intervals derived from naïve methods, such as climatology or persistence. The Prediction Interval Normalized Averaged Width (PINAW) is related to the informativeness of PIs or equivalently to the sharpness of the predictions. As for the PICP, we assess the sharpness of the forecasts by calculating the prediction intervals normalized width (PINAW) for different nominal coverage rates $(1-\alpha)100\%$. More precisely, this metric is the average width of the $(1-\alpha)100\%$ prediction interval normalized by the mean of GHI for the testing set. Equation 10 gives the definition of the PINAW metric for lead time h and for a specific nominal coverage rate.

$$\text{PINAW}(h,\alpha) = \frac{\sum_{i=1}^{N}\left(\hat{I}_{1-\frac{\alpha}{2}}(t+\Delta t) - \hat{I}_{\frac{\alpha}{2}}(t+\Delta t)\right)}{\sum_{i=1}^{N} I(t+\Delta t)} \tag{10}$$

As mentioned in [8], the goal of probabilistic forecasting is to maximize the sharpness of the predictive distributions subject to calibration. Put differently, ideally, predictive distributions should have PICPs close to the expected coverage rate and low PINAWs.

6.3. CRPS

The CRPS measures the difference between the predicted and observed cumulative distributions functions (CDF) ([7]. The formulation of the CRPS is:

$$\text{CRPS} = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}[\hat{P}^{i}_{fcst}(x) - P^{i}_{x_0}(x)]^2 dx \tag{11}$$

where $\hat{P}^{i}_{fcst}(x)$ is the predictive CDF of the variable of interest x (here GHI) and $P^{i}_{x_0}(x)$ is a cumulative-probability step function that jumps from 0 to 1 at the point where the forecast variable x equals the observation x_0 (i.e. $P_{x_0}(x) = 1_{x \geq x_0}$). The squared difference between the two CDFs is averaged over the N ensemble forecast/observation pairs. The CRPS has the same dimension as the forecasted variable. The CRPS is negatively oriented (smaller values are better), and it rewards concentration of probability around the step function located at the observed value ([28]). Thus, the CRPS penalizes lack of sharpness of the predictive distributions as well as biased forecasts.

Similarly to the forecast skill score commonly used to assess the quality of point forecasts, we also include the CRPSS (for continuous rank probability skill score). The latter (in %) is given by:

$$\text{CRPSS} = \left(1 - \frac{\text{CRPS}_m}{\text{CRPS}_0}\right) \times 100 \tag{12}$$

where CRPS_0 is the CRPS for the persistence ensemble model and CRPS_m is the CRPS for the model m (here the QR models). Negative values of CRPSS indicate that the probabilistic method fails to outperform the persistence ensemble model while positive values of CRPSS mean that the forecasting method improves on persistence ensemble. Further, the higher the CRPSS score, the better the improvement.

7. Results

In this section, we propose a detailed assessment of the 4 probabilistic models for the site of Le Tampon. In section 8 , we will give some results regarding the impact of the sky conditions on the quality of the probabilistic forecasts.

7.1. Reliability assessment

7.1.1. Analysis of the reliability diagrams

As suggested by [17], the first step of an evaluation framework is to analyze the reliability (calibration) of the probabilistic forecasts. Figure 3 shows the reliability diagrams of the four models for Le Tampon site. Notice that these reliability diagrams are averaged over all forecast horizons. The presence of consistency bars may help us to add more credibility in our judgment regarding reliability of the different models. First, we can state without any doubt that the persistence ensemble model is not reliable particularly for quantile forecasts ranging from 0.1 and 0.3 and for quantiles between 0.7 and 0.9. This lack of calibration will be further confirmed by the analysis of the rank histogram and values of the PICP metric for the PersEn Model. Reliability diagram for the QR Model 1 tends to suggest that this model is not reliable for quantile forecast between 0.4 and 0.8 as the corresponding observed proportions lie slightly outside the consistency bars. Conversely, all observed proportions of QR model 2 lie within the consistency bars hence suggesting QR model 2 is reliable. The same conclusion can be drawn for QR model 3 albeit slightly more significant deviations can be detected from the ideal case.

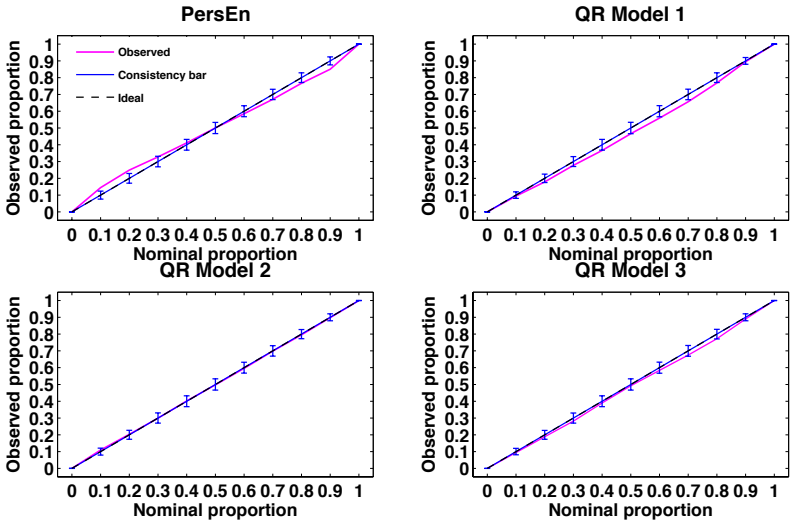

Figure 3. Reliability diagrams - Site of Le Tampon.

7.1.2. Analysis of the rank histograms

Figure 4 plots the rank histograms of the four different models for the site of Le Tampon. Again, it must be noted that, in order to provide condensed results, the rank histograms are averaged across all the forecasting time horizons. As mentioned above, rank histograms are designed to assess the consistency property of the different EPS. The rank histogram of the persistence ensemble model exhibits a U-shape, hence indicating a lack of spread of the predictive distribution or put differently, the ensemble persistence model is underdispersed. This result matches previous researches [25,27] that showed that the persistence ensemble model tends to underestimate the uncertainty. As a consequence, and as shown by Figures 5 and 6, the ensemble persistence prediction intervals have both low PICP and low PINAW because the ensemble members of the persistence ensemble tend to be too much like each other, and different from the observations. In agreement with the previous

evaluation of the reliability property made with the reliability diagrams, QR model 2 and (to a less extent) QR model 3 exhibit rather flat rank histograms. Conversely, a possibly (very) small bias is detected for QR model 1 indicated by a slight overpopulation of the highest ranks.

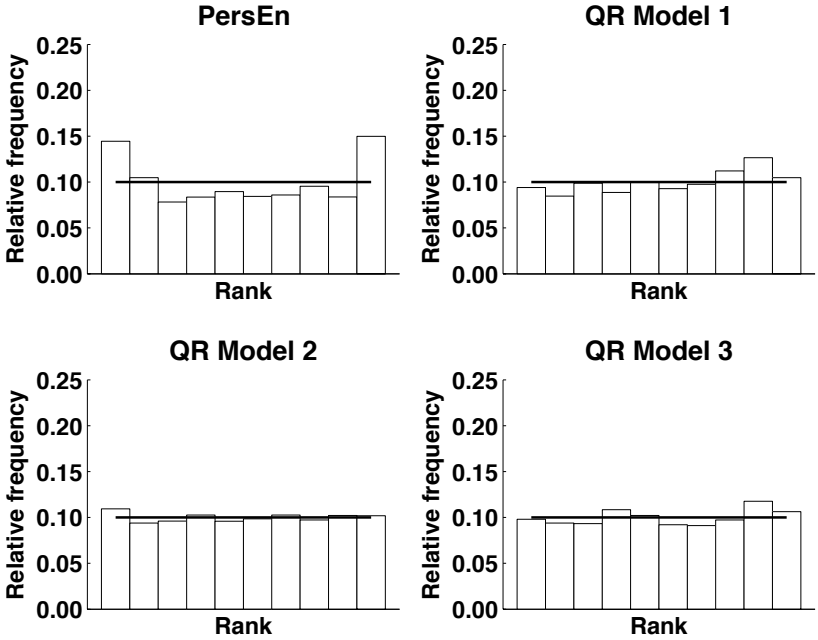

Figure 4. Rank histograms. The black horizontal line represents the theoretical relative frequency (here $\frac{1}{11}$). - Site of Le Tampon.

7.1.3. Analysis of PICP

As mentioned above, rank histograms and reliability diagrams have been averaged over all the forecast horizons. It is therefore also interesting to evaluate reliability as a function of the forecast horizon. Let us recall that calibrated predictive distributions should exhibit PICP values close to the nominal coverage rate. Figure 5 plots the PICP values of the four models for different nominal coverage rates and for all forecast horizons. The dotted black line in the graphs denotes the ideal case where observed coverage rates are equal to the nominal ones. Figure 5 shows that, except for QR model 2 for nominal coverage of 40% and forecast horizon of 1h, all PICP values for QR models 2 and 3 are rather close to the corresponding nominal coverage rates irrespective of the forecast horizon. Conversely, slight departures from the ideal line are noted for QR model 1. As expected following the previous analysis made with the reliability diagrams, the PICP values for the PersEn model do not match the nominal coverage rates. Further, the discrepancy increases with increasing nominal coverage rate.

Finally, this detailed analysis concerning reliability based on reliability diagrams, rank histograms and PICP values favors the selection of QR model 2 and QR model 3. The next step consists in extending the analysis to assess the sharpness property of the predictive distributions.

7.2. Sharpness assessment

Figure 6 plots the PINAW diagrams of the four models for different coverage rates and also for the six forecast horizons. As shown by Figure 6, the QR model 3 clearly leads to the narrowest predictions intervals irrespective of the nominal coverage rate and the forecast horizon. The PersEn model leads also to low PINAW values (except for lead time of 1h) but at the expense of not being reliable, as demonstrated above. We should normally discard this model at this point of the evaluation

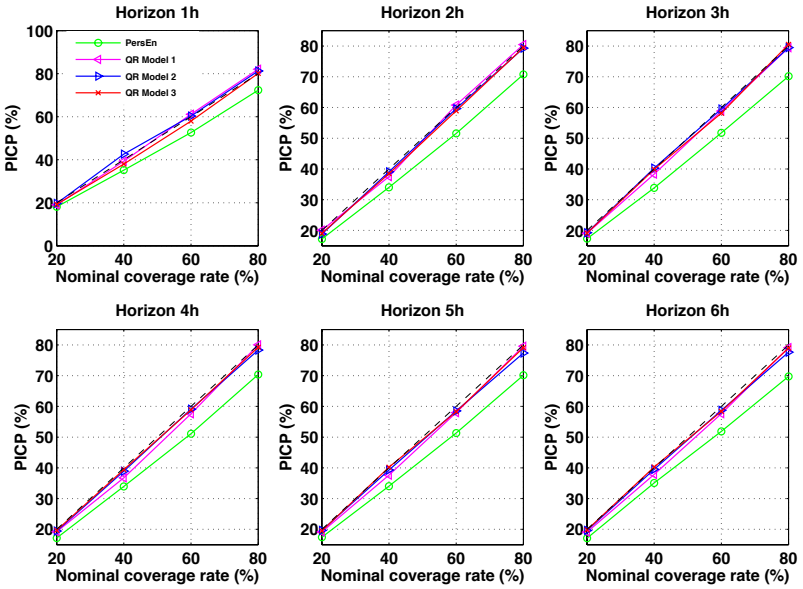

Figure 5. PICP for different nominal coverage rates and forecast horizons . The dotted black line represents the ideal case - Site of Le Tampon.

framework. As expected, PINAWs increase with increasing forecast horizon although they tend to stabilize for horizons greater than 3h (see also Figure 7(a)).

Figure 7(a) that details the evolution of the PINAW along the forecast horizon for a nominal coverage rate of 80% further reinforces the fact that QR model 3 yields the sharpest predictive distributions among the different models. One may however notice the rather high values of the PINAWs. Indeed, PINAWs are between 75% and 100% of the mean GHI for the testing set, corresponding to prediction intervals' widths between 341 Wm^{-2} and 455 Wm^{-2} (see Table 1 for the values of the mean of GHI for the testing period). One may conjecture that these large uncertainty intervals may come from the high GHI variability experienced at Le Tampon. Section 8 tries to shed some light on this conjecture by studying the performance of QR models 1 and 2 on a site (Desert Rock) that experiences high occurrences of clear and stable skies.

7.3. Overall probabilistic forecasting skill

In order to exhibit the best probabilistic model for this particular experimental set-up, we plot in Figure 8(a) the relative CRPS (i.e. CRPS divided by the mean of GHI for the testing period) as a function of the forecast horizon. As shown by 8(a), QR model 3 yields the best overall performance as it leads to the lowest CRPS for all forecast horizons. As mentioned above, the CRPS is an attractive metric due to its capacity to address reliability and sharpness simultaneously. In addition it establishes a clear-cut ranking of the different models that ranks QR model 3 first followed by QR models 1 and 2. Also, in order to assess the true skill improvement brought by the quantile regression models over the persistence ensemble model, Table 2 lists the CRPSS for the site of Le Tampon. Significant skill scores ranging from 12% to 36% are obtained with the QR models. One interesting point to note is that, contrary to skill scores related to point forecasts, here the CRPSS decreases with increasing lead time. This is mainly due to the fact that the CRPS of the models (including the PersEns model) level off after a lead time of 3h. One may notice also that QR model 3 yields greater skill scores than the two other QR models and particularly for lead times greater than 3h.

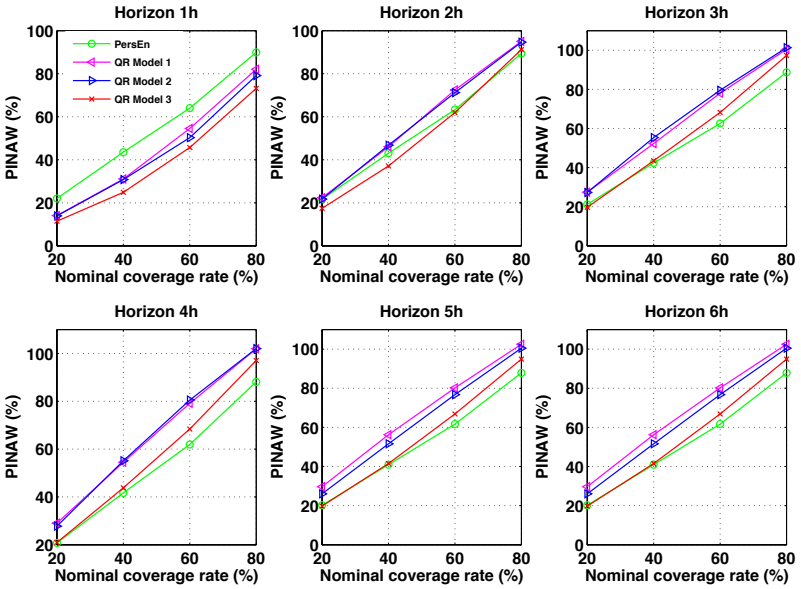

Figure 6. PINAW diagrams for different values of forecast horizon and nominal coverage rate - Site of Le Tampon.

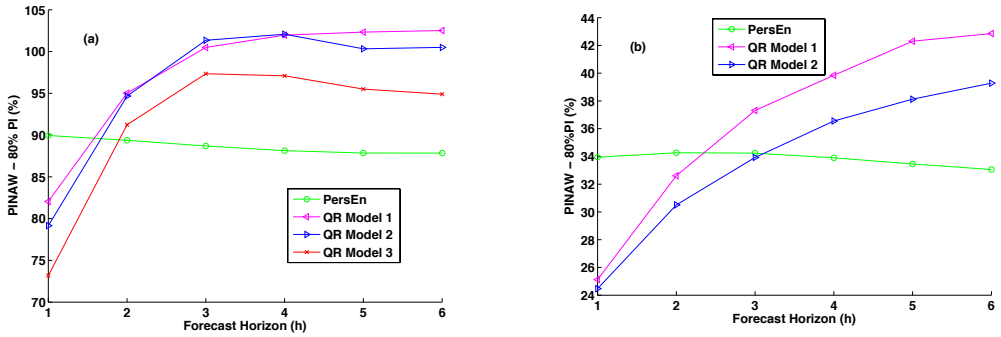

Figure 7. (a) PINAW for the 80% nominal coverage rate - Site of Le Tampon **(b)** PINAW for the 80% nominal coverage rate - Site of Desert Rock

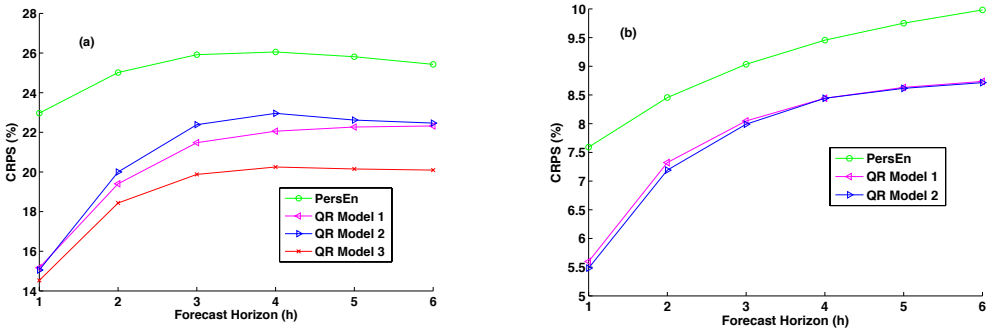

Figure 8. (a) Relative CRPS - Site of Le Tampon **(b)** Relative CRPS- Site of Desert Rock

Table 2. CRPS and CRPS skill score (CRPSS) for the site of Le Tampon. CRPS values are in Wm^{-2} and the CRPSS is in percentage.

Lead Time (h)	Persistence	QR model 1		QR model 2		QR model 3	
	CRPS	CRPS	CRPSS	CRPS	CRPSS	CRPS	CRPSS
1	104.7	69.1	34.0	68.6	34.5	66.2	36.7
2	114.0	88.4	22.4	91.1	20.1	84.0	26.3
3	118.1	97.9	17.1	102.0	13.6	90.6	23.3
4	118.7	100.5	15.3	104.6	11.9	92.3	22.3
5	117.6	101.5	13.7	103.1	12.4	91.8	21.9
6	115.9	101.7	12.2	102.4	11.7	91.6	21.0

Finally, in an attempt (the exercise here is obviously not exhaustive) to visually check the better skill of QR model 3, Figure 9 plots a few episodes (six days) of prediction intervals calculated by the four models for a lead time of 6h. The examination of these days is very instructive as it can be seen that, for day 6, several measurements lie outside of the uncertainty interval for QR models 1 and 2 but not for QR model 3.

In conclusion, and based on this detailed evaluation framework, we can argue that the inclusion of day-ahead ECMWF forecasts increases the forecasting quality of the probabilistic forecasts.

8. Impact of the sky conditions on the quality of the probabilistic forecasts

This section aims at assessing the impact of site variability on the probabilistic forecasting performance of the different methods. Table 3 gives the CRPS together with the CRPSS of the different models (except the QR model 3) for the site of Desert Rock. Let us recall that the site of Desert Rock experiences high occurrences of clear skies and exhibits a low annual variability.

As shown by Table 3, the performance of the models in terms of CRPS is obviously better for the site of Desert Rock than for Le Tampon site. Indeed, for this site, the CRPS values for the QR models range from 30 Wm^{-2} to 48 Wm^{-2} which translates into 5.5% to 8.7% for the relative counterparts (see also Figure 8(b)). Again, the skill scores of the QR models methods demonstrate that these techniques outperform the reference PersEns model whatever the site under study. In other words, we can argue that the QR models offer a significant improvement over the PersEn model by an amount that is fairly independent of the sky conditions experienced by a site.

Figure 7(b) shows the PINAW values for the 80% nominal coverage rate obtained for the site of Desert Rock. Here, PINAWs for the two QR models lie between 24% and 42% which correspond to prediction intervals' width between 131 Wm^{-2} and 230 Wm^{-2}. Let us recall that for Le Tampon

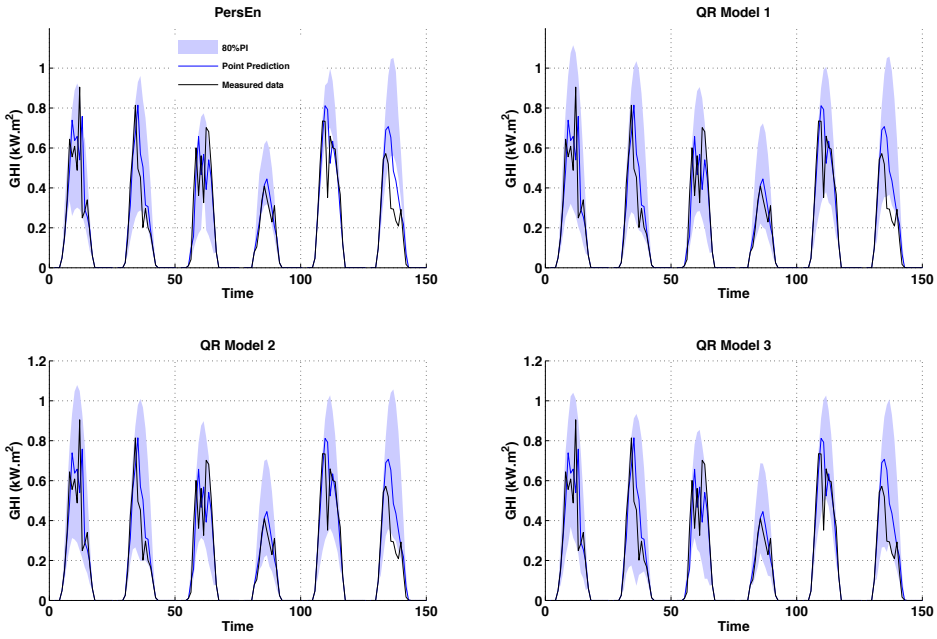

Figure 9. Measured and forecasted GHI time series and the corresponding 80%PI calculated by the four models for a lead time of 6h. The graphs shows 6 days that cover different weather conditions in the testing dataset.

site, the corresponding values were 341 Wm^{-2} and 455 Wm^{-2}. This comparison confirms that the sky conditions have a clear impact on the forecasting quality of the probabilistic models. Similarly to point forecasts, we can conjecture a link between the variability of the site and the performance of the probabilistic models.

Table 3. CRPS and CRPS skill score (CRPSS) for the site of Desert Rock. CRPS values are in Wm^{-2} and the CRPSS is in percentage.

Lead Time (h)	Persistence	QR model 1		QR model 2	
	CRPS	CRPS	CRPSS	CRPS	CRPSS
1	41.6	30.7	26.2	30.1	27.7
2	46.3	40.1	13.4	39.4	15.0
3	49.5	44.1	10.9	43.8	11.6
4	51.8	46.3	10.7	46.3	10.7
5	53.4	47.3	11.5	47.2	11.6
6	54.7	47.9	12.5	47.8	12.7

9. Main conclusions

This work proposed a comparison of three probabilistic solar forecasting models. The simple linear quantile regression method was used to build the models. It was demonstrated that the incorporation of additional exogenous inputs provided by ECMWF forecasts into the quantile regression model clearly improved the quality of the intra-day probabilistic forecasts. Also, preliminary results showed an impact of the site variability on the forecasting performance of the different methods. Insular sites like Le Tampon, which exhibits higher solar variability due to clouds that are formed locally, are prone to have the worse forecasting performance than less variable (continental or insular) sites. Future work will be devoted to the evaluation of other

more sophisticated quantile regression methods such as quantile regression Forest ([19]), or quantile regression based on Neural Networks ([20].

Acknowledgments: Dr Philippe Lauret has received support for his research work from Région Réunion under the European Regional Development Fund (ERDF) - Programmes Opérationnels Européens FEDER 2014-2020 - Fiche action 1.07. The authors would like also to thank the European Center for Medium-Range Weather Forecasts (ECMWF) for providing forecast data and the meteorological network SURFRAD for providing their ground measurements.

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

Abbreviations

The following abbreviations are used in this manuscript:

ECMWF	European Center for Medium-Range Weather Forecasts
GHI	Global horizontal solar irradiance
QR	Quantile regression
EPS	Ensemble prediction system
CRPS	Continuous ranked probabilty score
PI	Prediction interval
PICP	Prediction interval coverage probability
PINAW	Prediction interval normalized averaged width

References

1. Lorenz, E.; Heinemann, D. Prediction of solar irradiance and photovoltaic power. In *Comprehensive Renewable Energy*; Elsevier: Oxford, UK,, 2012; p. 239292.
2. Bridier, L.; David, M.; Lauret, P. Optimal design of a storage system coupled with intermittent renewables. *Renewable Energy* **2014**, *67*, 2–9.
3. Lauret, P.; Lorenz, E.; David, M. Solar Forecasting in a Challenging Insular Context. *Atmosphere* **2016**, *7*, 18.
4. ECMWF. Available online:https://www.ecmwf.int/en/research/modelling-and-prediction (accessed on 6 September 2017).
5. Pinson, P.; McSharry, P.; Madsen, H. Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation. *Quarterly Journal of the Royal Meteorological Society* **2010**, *136*, 77–90.
6. Hamill, T.M. Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly Weather Review* **2001**, *129*, 550–560.
7. Hersbach, H. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting* **2000**, *15*, 559–570.
8. Gneiting, T.; Balabdaoui, F.; Raftery, A.E. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2007**, *69*, 243–268.
9. Khosravi, A.; Nahavandi, S.; Creighton, D. Prediction Intervals for Short-Term Wind Farm Power Generation Forecasts. *IEEE Transactions on Sustainable Energy* **2013**, *4*, 602–610.
10. Lefèvre, M.; Oumbe, A.; Blanc, P.; Espinar, B.; Gschwind, B.; Qu, Z.; Wald, L.; Schroedter-Homscheidt, M.; Hoyer-Klick, C.; Arola, A.; Benedetti, A.; Kaiser, J.W.; Morcrette, J.J. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques* **2013**, *6*, 2403–2418.
11. SoDa (Solar radiation Data), McClear service for estimating irradiation under clear sky. Available online:http://www.soda-pro.com/web-services/radiation/mcclear. (accessed on 6 September 2017).
12. MACC project. Available online: http://www.gmes-atmosphere.eu/news/ (accessed on 6 September 2017).
13. Almeida, M.P.; Zilles, R.; Lorenzo, E. Extreme overirradiance events in São Paulo, Brazil. *Solar Energy* **2014**, *110*, 168–173.
14. Inman, R.H.; Chu, Y.; Coimbra, C.F. Cloud enhancement of global horizontal irradiance in California and Hawaii. *Solar Energy* **2016**, *130*, 128–138.
15. Hoff, T.E.; Perez, R. Modeling PV fleet output variability. *Solar Energy* **2012**, *86*, 2177–2189.

16. Perez, R.; Kivalov, S.; Schlemmer, J.; Hemker, K.; Renné, D.; Hoff, T.E. Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy* **2010**, *84*, 2161–2172.
17. Pinson, P.; Nielsen, H.A.; Møller, J.K.; Madsen, H.; Kariniotakis, G.N. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy* **2007**, *10*, 497–516.
18. Koenker, R.; Bassett, G. Regression Quantiles. *Econometrica* **1978**, *46*, 33.
19. Meinshausen, N. Quantile Regression Forests. *Journal of Machine Learning Research* **2006**, pp. 983–999.
20. Cannon, A.J. Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & Geosciences* **2011**, *37*, 1277–1284.
21. Friedman, J. Stochastic gradient boosting. *Computational Statistics & Data Analysis* **2002**, *38*, 367–378.
22. Chernozhukov, V.; Fernandez-Val, I.; Galichon, A. Quantile and Probability Curves Without Crossing. *Econometrica* **2010**, *78*, 1093–1125.
23. Candille, G.; Côte, C.; Houtekamer, P.; Pellerin, G. Verification of an Ensemble Prediction System against Observations. *Monthly Weather Review* **2007**, *135*, 2688–2699.
24. Bacher, P.; Madsen, H.; Nielsen, H.A. Online short-term solar power forecasting. *Solar Energy* **2009**, *83*, 1772–1783.
25. David, M.; Ramahatana, F.; Trombe, P.; Lauret, P. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Solar Energy* **2016**, *133*, 55–72.
26. Alessandrini, S.; Delle Monache, L.; Sperati, S.; Cervone, G. An analog ensemble for short-term probabilistic solar power forecast. *Applied Energy* **2015**, *157*, 95–110.
27. Chu, Y.; Coimbra, C.F. Short-term probabilistic forecasts for Direct Normal Irradiance. *Renewable Energy* **2017**, *101*, 526–536.
28. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences An Introduction.*; Elsevier Science: Burlington, 2014.
29. Jolliffe, I.; Stephenson, D. *Forecast Verification. A practitioner's guide in atmospheric science*; Wiley, 2003.
30. Grantham, A.; Gel, Y.R.; Boland, J. Nonparametric short-term probabilistic forecasting for solar radiation. *Solar Energy* **2016**, *133*, 465–475.
31. Gneiting, T.; Raftery, A.E.; Westveld, A.H.; Goldman, T. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review* **2005**, *133*, 1098–1118.
32. Bröcker, J.; Smith, L.A. Increasing the Reliability of Reliability Diagrams. *Weather and Forecasting* **2007**, *22*, 651–661.