



Probabilistic Solar Forecasting Using Quantile Regression Models

Philippe Lauret, Mathieu David, Hugo Pedro

► To cite this version:

Philippe Lauret, Mathieu David, Hugo Pedro. Probabilistic Solar Forecasting Using Quantile Regression Models. *Energies*, 2017, 10 (10), 10.3390/en10101591 . hal-01617867

HAL Id: hal-01617867

<https://hal.univ-reunion.fr/hal-01617867>

Submitted on 3 Jul 2018

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.



Distributed under a Creative Commons Attribution 4.0 International License

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; mathieu.david@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

Received: 11 September 2017; Accepted: 6 October 2017; Published: 13 October 2017

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 1 h–6 h-ahead probabilistic forecasts. Our approach is applied to forecasting solar irradiance at a site experiencing highly variable sky conditions using the historical ground observations of solar irradiance as endogenous inputs and day-ahead forecasts as exogenous inputs. Day-ahead irradiance forecasts are obtained from the Integrated Forecast System (IFS), a Numerical Weather Prediction (NWP) model maintained by the European Center for Medium-Range Weather Forecast (ECMWF). Several metrics, mainly originated from the weather forecasting community, are used to evaluate the performance of the probabilistic forecasts. The results demonstrated that the NWP exogenous inputs improve the quality of the intra-day probabilistic forecasts. The analysis considered two locations with very dissimilar solar variability. Comparison between the two locations highlighted that the statistical performance of the probabilistic models depends on the local sky conditions.

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. Accurate forecasts are important for several tasks. They can be used to ensure the balance between supply and demand necessary for grid stability [1]. At more localized scales, solar forecasts are used to optimize the operational management of grid-connected storage systems associated with intermittent solar resource [2].

A proper assessment of uncertainty estimates enables a more informed decision-making processes, which in turn, increases the value of solar power generation. Thus, a point forecast plus a prediction interval provide more useful information than simply the point forecast and are essential to increase the penetration of solar energy into the electrical grids. Given the recognized value of accurate probabilistic solar forecast, this work proposes techniques to improve their quality and performance.

Some recent works dealt with intra-day or intra-hour solar probabilistic forecasting. Grantham et al. [3] proposed a non-parametric approach based on a specific bootstrap method to build predictive global horizontal solar irradiance (GHI) distributions for a forecast horizon of 1 h. Chu and Coimbra [4] developed models based on the k-nearest-neighbors (kNN) algorithm for generating very short-term (from 5 min–20 min) direct normal solar irradiance (DNI) probabilistic forecasts. Golestaneh et al. [5] used an extreme machine learning method to produce very short-term PV power predictive densities for lead times between a few minutes to one hour ahead. David et al. [6] used a

parametric approach based on a combination of two linear time series model (ARMA-GARCH) to generate 10 min–6 h GHI probabilistic forecasts using only past ground data.

In a previous study, Lauret et al. [7] demonstrated that intra-day point deterministic forecasts can be improved by considering both past ground measurements and day-ahead ECMWF forecasts in the set of predictors.

In this work, it is investigated if such a combination could also improve the quality of the probabilistic forecasts. In order to test this conjecture, 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. These consist, as mentioned, of forecasted data obtained from the Integrated Forecast System (IFS) Numerical Weather Prediction (NWP) model [8].

Once the models are trained and applied to the testing set, we analyze the quality of the probabilistic forecasts using several qualitative and quantitative tools: reliability diagrams [9], the rank histograms (or Talagrand histograms) [10] and the continuous ranked probability score (CRPS) [11]. We will also assess the sharpness of the predictive distributions by calculating the prediction intervals normalized average width (PINAW) [12,13].

The remainder of this paper is organized as follows: Section 2 provides information about the locations studied, the data collection and data preprocessing. Section 3 analyzes the solar irradiance variability observed at the two locations under study. Section 4 describes the exogenous data collected from the ECMWF forecasts. 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 two consecutive years of 1-min data collected at two sites that exhibit completely different sky conditions. The first site (Le Tampon) is located on La Réunion Island and experiences variable sky conditions, while the second one is situated in the continental U.S. and experiences an arid climate (Desert Rock). Table 1 lists the characteristics of the two sites. For this study, we computed 1-h averages of global horizontal solar irradiance (GHI) directly from the raw 1 min-data. The first year (2012) constitutes the training dataset used to build the forecasting models. The second year (2013) was used to evaluate the probabilistic forecasts (testing or evaluation dataset).

Solar irradiance is characterized by deterministic diurnal and seasonal variations. Such a component can be removed from the analysis by working with the clear sky index (kt^*) instead of the original GHI time series, where kt^* is defined as:

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

This quantity corresponds to the ratio of the measured GHI $I(t)$ to the theoretical GHI observed under clear sky $I_{clr}(t)$. This preprocessing technique separates the geometric and the deterministic component of GHI from the stochastic component that is related to the cloud cover. The target variable for the forecasting model is then the stochastic component. Here, we used the clear sky data provided by the McClellan model [14] publicly available on the Solar radiation Data (SoDa) website [15]. This model uses aerosol optical depth, water vapor and ozone data from the MACC (Monitoring Atmospheric Composition and Climate) project [16].

Finally, data points corresponding to nighttime and low solar elevations (solar zenith angle θ_Z greater than 85°) were discarded from the dataset. This filtering removes less than 1% of the total

annual sum of solar energy and discards measurements with large uncertainty (uncertainties associated with pyranometers are typically much higher than 3.0% for $\theta_z > 85^\circ$).

Notice that, in addition to the filtering based on the zenith angle, a specific procedure was also used to preprocess the GHI data. This procedure is described at length in [6].

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	550 m	1007 m
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 one. 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 the clear sky index above one. These events result from a phenomenon known as the cloud edge effect and can be observed anywhere in the world [17,18].

Table 1 (last row) lists the solar variability of each site. This quantity is calculated as the standard deviation of step changes in the clear sky index $\sigma(\Delta kt_{\Delta T}^*)$ [19]. This metric was computed for hourly kt^* values ($\Delta T = 1 \text{ h}$) for the whole dataset (two years). The application of this metric to other locations [19] led to the conclusion that variability above 0.2 indicates very unstable GHI conditions. As the table indicates, variability for Le Tampon is above this threshold.

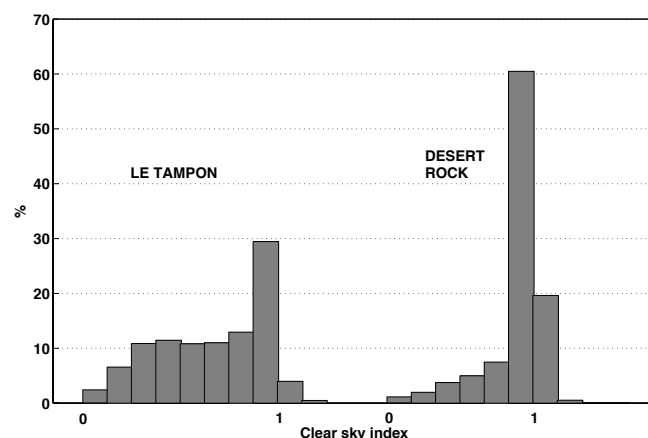


Figure 1. Statistical distribution of the 1-h clear sky index for the two locations under study, Le Tampon (left) and Desert Rock (right).

In [7], it was demonstrated that point deterministic forecasts for insular sites like Le Tampon are prone to have worse forecasting performance than less variable (continental or insular) sites. This fact results from the higher solar variability resulting from local cloud formation and dissipation. Section 8

below discusses the impact of the site variability on the forecasting performance of the different probabilistic models.

At this point, it must be stressed that the site of Desert Rock is only used to address the impact of the sky conditions on the quality of the probabilistic forecasts.

4. ECMWF Day-Ahead Forecasts

Besides ground data, the last model developed below also requires NWP forecasted data. As mentioned above, we used in this work data retrieved from the IFS model maintained and ran by ECMWF. Although NWP models outperform forecasts based on satellite data or ground measurements (only) for horizons longer than 5 h [20], they provide useful information as exogenous inputs. IFS provides hourly GHI data with a spatial resolution of $0.125^\circ \times 0.125^\circ$ (approximately $14 \text{ km} \times 14 \text{ km}$; see Figure 2). For the purpose of this work, we retrieve IFS data (GHI and total cloud cover (TCC)) generated at 12 h 00 UTC (16 h 00 for Réunion Island).



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 [21].

In this work, we chose to use the simple linear quantile regression (QR) method proposed by [22] 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 [23], QR neural networks [24] or gradient boosting techniques [25]. However, 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 Regression Method

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

$$y = \beta x + \epsilon, \quad (2)$$

where β 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 [22], the quantile loss function is:

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

with τ representing the quantile probability level.

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

$$\hat{\beta}_{\tau} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^N \rho_{\tau}(y_i - \beta x_i), \quad (4)$$

where (x_i, y_i) denotes a pair of vector of predictors and the corresponding observed predictand in the training set.

It must be 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 [26].

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 nine quantiles defined by $\{kt^*_{\tau}(t+h)\}_{\tau=0.1,0.2,\dots,0.9}$ for each forecasting time horizon $h = 1, \dots, 6$ h. This set of $\{kt^*_{\tau}(t+h)\}_{\tau=0.1,0.2,\dots,0.9}$ quantiles can be transformed into GHI quantiles $\{\hat{I}_{\tau}(t+h)\}_{\tau=0.1,0.2,\dots,0.9}$ by using Equation (1). As mentioned above, this set of quantiles represents the predictive distribution of the target variable (here GHI) at lead time $t+h$. This set of quantiles may form also what the verification weather community [27] calls an ensemble prediction system (EPS).

In addition, prediction intervals with different nominal coverage rates can be inferred from this set of quantiles. These intervals provide lower and upper bounds within which the true value of GHI is expected to lie with a probability equal to the prediction interval's 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 a $(1-\alpha)100\%$ nominal coverage rate produced at time t for lead time $t+h$ is defined by:

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

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

5.2. QR Model 1: Point Forecast as Unique Explanatory Variable

In this simple setup, 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:

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

Indeed, in a study related to solar power forecasting, Bacher et al. [28] showed that the level of uncertainty depends on the value of the point forecast. The detailed description of this ARMA model can be found at [6].

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.,

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

5.4. QR Model 3: Combination of Past Ground Measurements Plus ECMWF Forecast Data

The third QR model developed in this work considers as predictors the six past ground measurements plus exogenous data provided by the day-head ECMWF forecasts. As explained above in Section 4, the exogenous data consists of the forecasted clear sky index $\hat{kt}_{ECMWF}^*(t+h)$ and the total cloud cover $\hat{TCC}_{ECMWF}(t+h)$ predicted for time $t+h$. With these data, the vector of predictors is given by:

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

The selection of exogenous data follows a previous work [7] regarding point deterministic forecasts. There, it was shown that forecasts with improved accuracy can be obtained if one adds total cloud cover (TCC) as an explanatory variable.

5.5. Persistence Ensemble Model

Finally, we define the simple baseline persistence ensemble model (PersEn). This model [4,6,29] is commonly used to provide reference probabilistic forecasts. In this work, the PersEn considers the GHI lagged measurements in the 10 h that precede the forecasting issuing time. The selected measurements are ranked to define quantile values for the irradiance forecast.

6. Probabilistic Error Metrics

In [21], 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 meanings according a meteorologist's point of view or a statistician's point of view. The interested reader is referred to [30,31] for a definition of these properties in the realm of meteorology. As an illustration of these different definitions, the meteorological literature [30] 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 will be the case here), the sharpness property refers to the concentration of the predictive distributions [12,21].

The main requirement when studying the performance of probabilistic forecasts is reliability. As noted in [21], the lack of reliability introduces systematic bias that affects negatively the subsequent decision-making process. 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 Section 6.1.1) or by calculating the prediction interval coverage probability (PICP) [13] (see Section 6.1.2), which permits one 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 Section 6.1.3) that allow judging the statistical consistency of the ensemble constituted by the nine quantiles' members.

In this work, and similarly to [12,21], 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 Section 6.2) is used to evaluate the sharpness of the predictive distributions [13].

Regarding the third property, namely resolution, it consists of 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 [21]. 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 [3]).

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 [32], whereas the same is not possible for sharpness. Finally, the continuous rank probability score (CRPS) [11] 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 [9] 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). By doing so, deviations from perfect reliability (the diagonal) are immediately revealed in a visual manner [9]. 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 [9]. In this work, similarly to [33], 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) [13] is a metric that permits one 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 a 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:

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

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

6.1.3. Rank Histogram

The rank histogram [30] 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 [30]. A necessary condition for ensemble consistency is an appropriate degree of ensemble dispersion leading to a flat rank histogram [30]. 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 the 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 over-forecasting bias will correspond to an overpopulation of the smallest ranks, while an under-forecasting bias will overpopulate the highest ranks. As a consequence, rank histograms can also reveal deficiencies in ensemble calibration or reliability [30]. Again, care must be taken when analyzing rank histograms when the number of

verification samples is limited. In addition, as demonstrated by [10], 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 a theoretical relative frequency of $\frac{1}{M+1}$.

6.2. Sharpness

Sharp probabilistic forecasts must present prediction intervals that are shorter on average than the ones derived from naive 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 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.

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

As mentioned in [12], the objective of any probabilistic forecasting model is to maximize the sharpness of the predictive distributions (minimize PINAW) while maintaining a coverage rate (PICP) close to its nominal value.

6.3. CRPS

The CRPS quantifies deviations between the cumulative distributions functions (CDF) for the predicted and observed data. The formulation of the CRPS is:

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

where $\hat{P}_{fct}^i(x)$ is the predictive CDF of the variable of interest x (here GHI) and $P_{x_0}^i(x)$ is a cumulative-probability step function that jumps from zero to one 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 the concentration of probability around the step function located at the observed value [30]. Thus, the CRPS penalizes the 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:

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

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 four 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 [21], 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 the 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 the 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 to 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. The reliability diagram for 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 that QR Model 2 is reliable. The same conclusion can be drawn for QR Model 3, albeit that slightly more significant deviations can be detected from the ideal case.

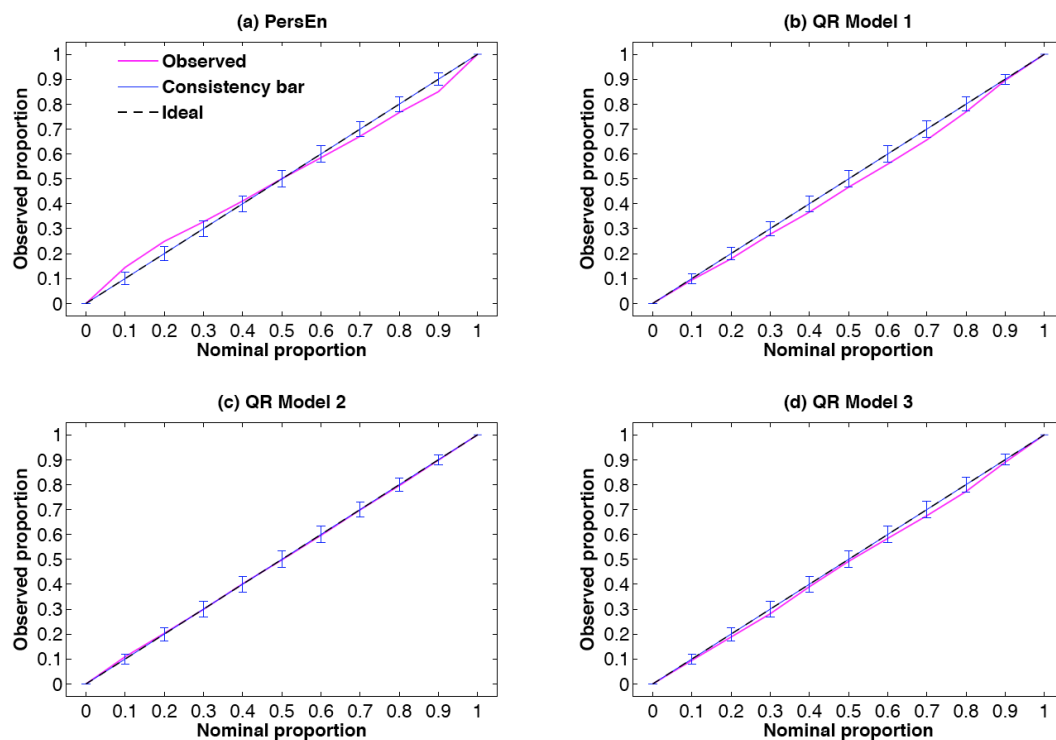


Figure 3. Reliability diagrams for the site of Le Tampon. (a) PersEn Model; (b) QR Model 1; (c) QR Model 2; (d) QR Model 3.

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 research [4,6] 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 lesser 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.

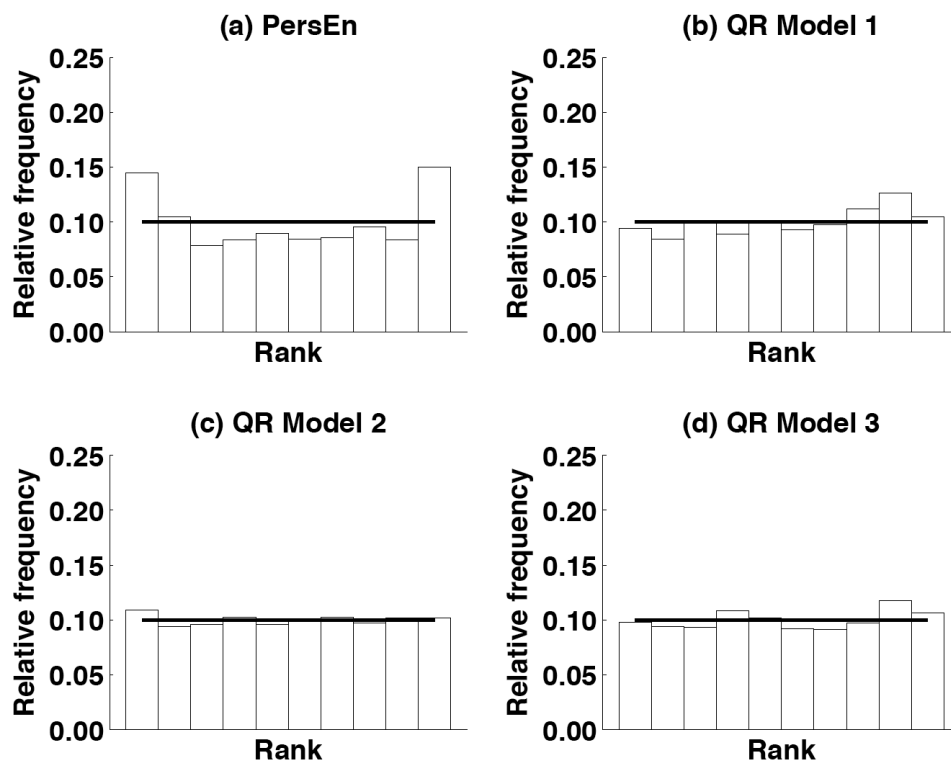


Figure 4. Rank histograms for the site of Le Tampon. (a) PersEns Model; (b) QR Model 1; (c) QR Model 2; (d) QR Model 3. The black horizontal line represents the theoretical relative frequency (here $\frac{1}{11}$).

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 a forecast horizon of 1 h, 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 the increasing nominal coverage rate.

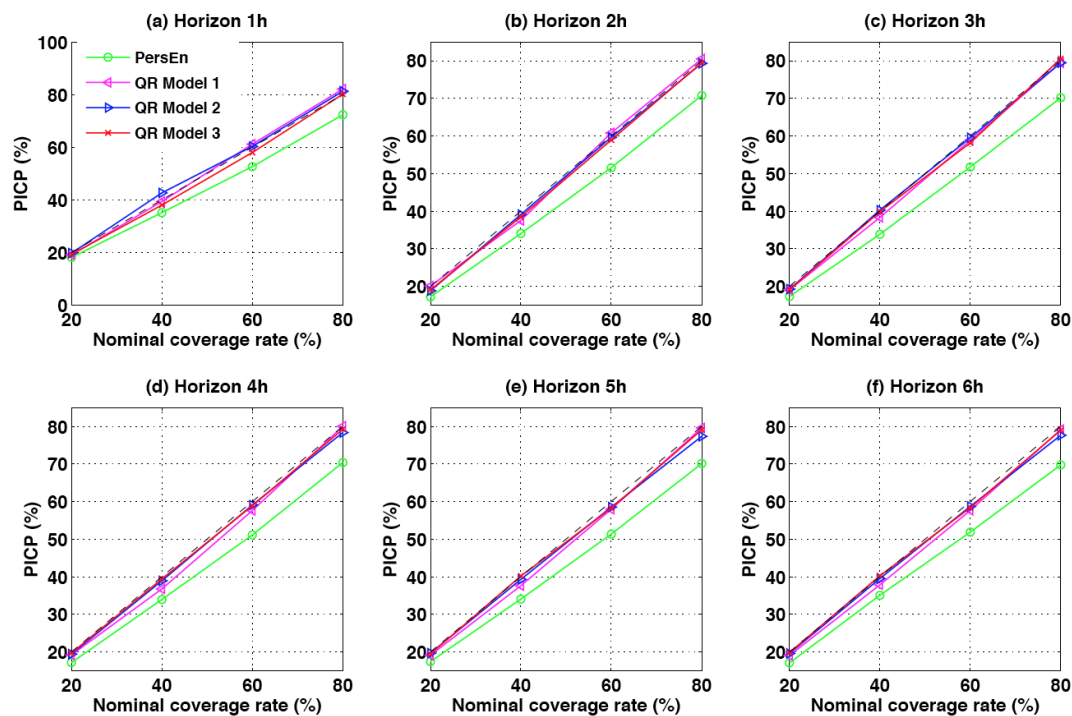


Figure 5. PICP for different nominal coverage rates and forecast horizons (site of Le Tampon). The dotted black line represents the ideal case. (a) Horizon 1 h; (b) Horizon 2 h; (c) Horizon 3 h; (d) Horizon 4 h; (e) Horizon 5 h; (f) Horizon 6 h.

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 of 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, 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 1 h), but at the expense of not being reliable, as demonstrated above. We should normally discard this model at this point of the evaluation framework. As expected, PINAWs increase with increasing forecast horizon, although they tend to stabilize for horizons greater than 3 h (see also Figure 7a).

Figure 7a, which 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 $341 \text{ W}\cdot\text{m}^{-2}$ and $455 \text{ W}\cdot\text{m}^{-2}$ (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.

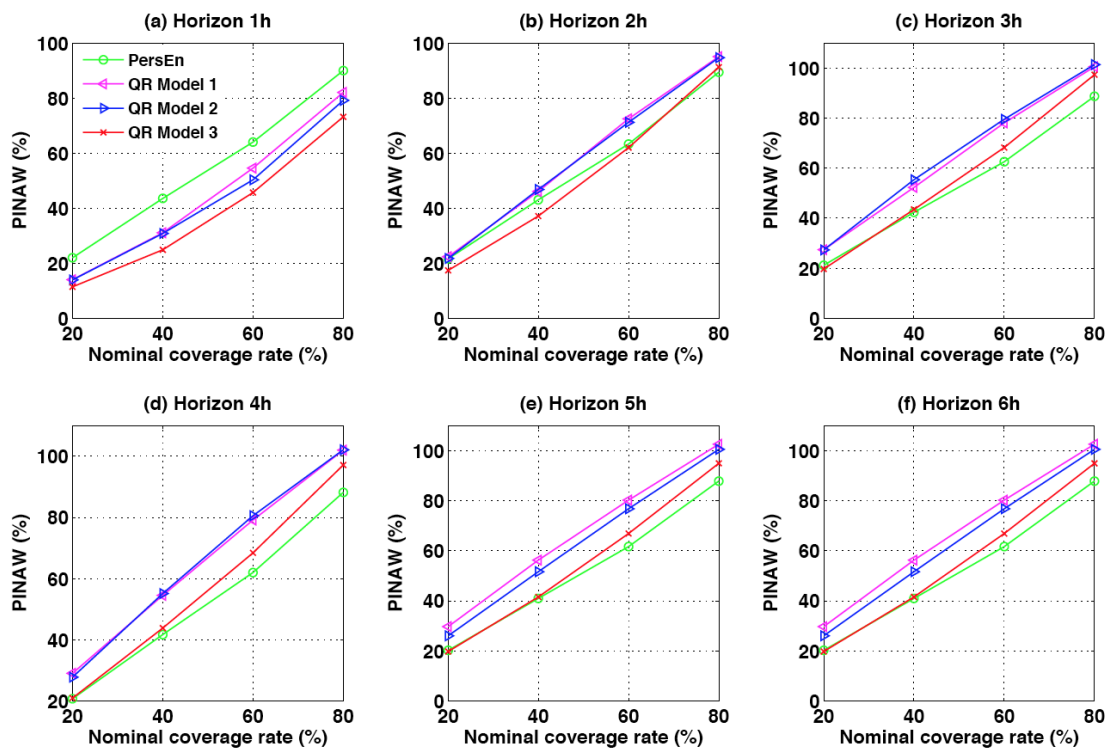


Figure 6. PINAW diagrams for different values of forecast horizon and nominal coverage rate (site of Le Tampon). (a) Horizon 1 h; (b) Horizon 2 h; (c) Horizon 3 h; (d) Horizon 4 h; (e) Horizon 5 h; (f) Horizon 6 h.

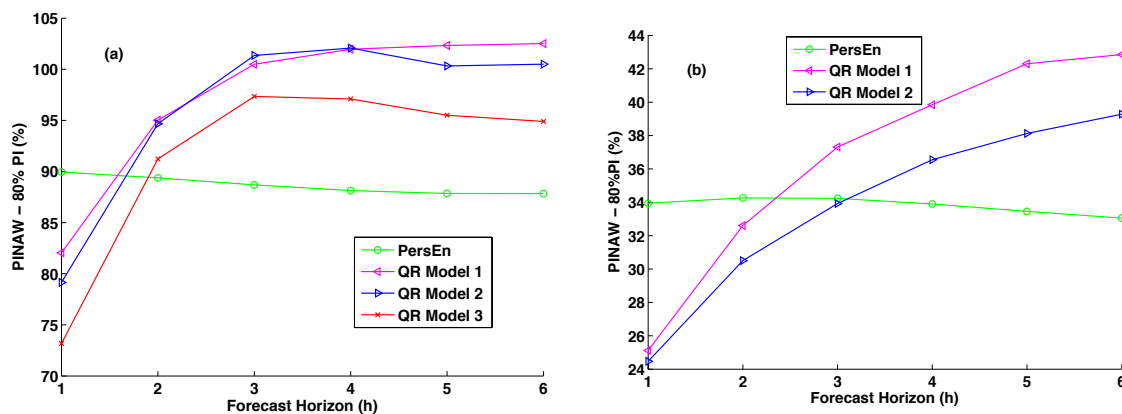


Figure 7. (a) PINAW for the 80% nominal coverage rate: site of Le Tampon; (b) PINAW for the 80% nominal coverage rate: site of Desert Rock.

7.3. Overall Probabilistic Forecasting Skill

In order to exhibit the best probabilistic model for this particular experimental setup, we plot in Figure 8a 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 Figure 8a, 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. Furthermore, 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 PersEn model) level off after a lead time of 3 h. 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 3 h.

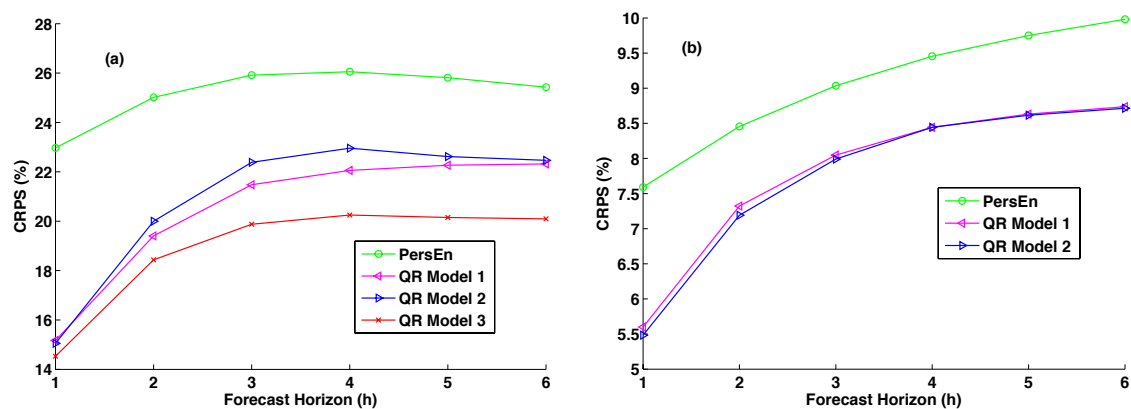


Figure 8. (a) Relative CRPS: site of Le Tampon; (b) Relative CRPS: site of Desert Rock.

Table 2. CRPS and CRPS skill score (CRPSS) for the site of Le Tampon. CRPS values are in $W \cdot m^{-2}$, 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 6 h. 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.

This result is in accordance with a previous work [7] related to point deterministic GHI forecasts where it was demonstrated that the incorporation of additional exogenous inputs provided by ECMWF improved the accuracy of the intra-day deterministic forecasts.

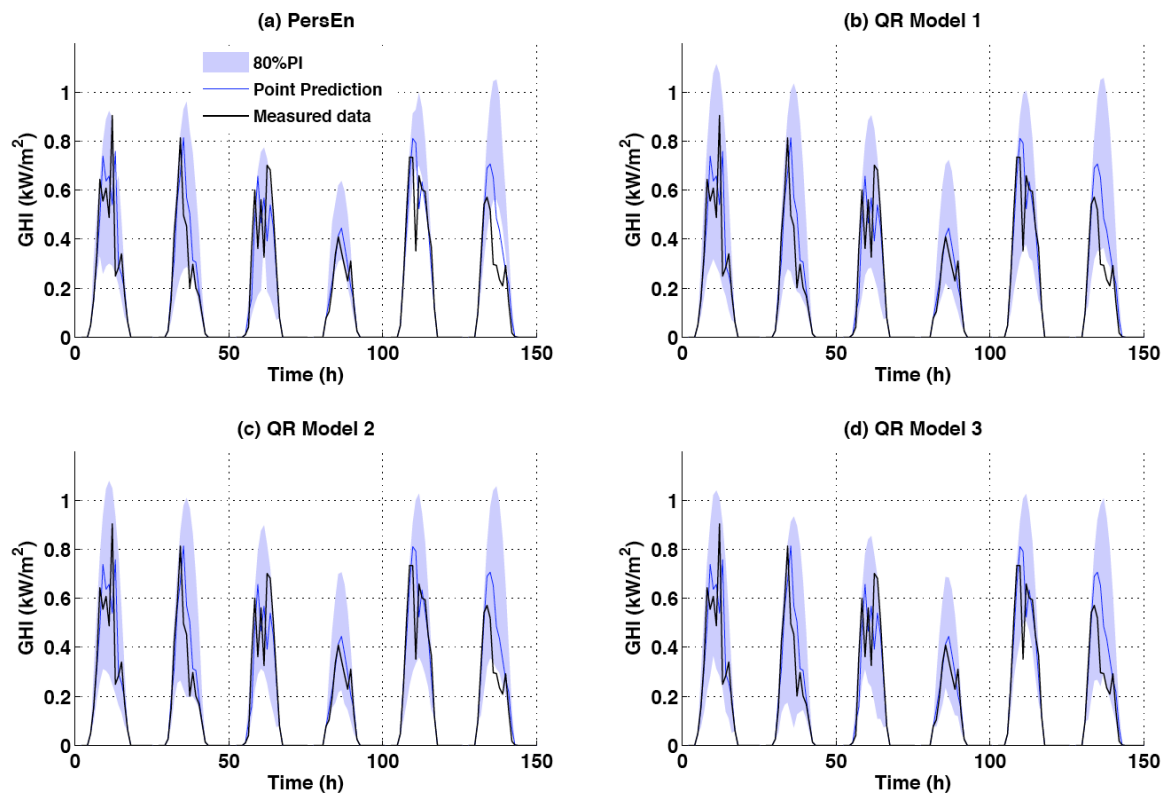


Figure 9. Measured and forecasted GHI time series and the corresponding 80%PI calculated by the four models for a lead time of 6 h. The graphs shows six days that cover different weather conditions in the testing dataset. (a) PersEn Model; (b) QR Model 1; (c) QR Model 2; (d) QR Model 3.

8. Impact of Data Variability on the Quality of the Probabilistic Forecasts

In this section, we study the impact of local sky conditions (as quantified by the site variability listed in Table 1) on the forecast performance.

Table 3 gives the CRPS together with the CRPSS of the different models (except QR Model 3) for the site of Desert Rock. As we have seen, this location displays low annual variability and a high frequency of clear skies ($kt^* \approx 1$).

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 the Le Tampon site. Indeed, for this site, the CRPS values for the QR models range from $30 \text{ W}\cdot\text{m}^{-2}$ to $48 \text{ W}\cdot\text{m}^{-2}$, which translates into 5.5–8.7% for the relative counterparts (see also Figure 8b). Again, the skill scores of the QR models demonstrate that these techniques outperform the reference PersEn model regardless of the site under study and the corresponding irradiance variability.

Figure 7b 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 $131 \text{ W}\cdot\text{m}^{-2}$ and $230 \text{ W}\cdot\text{m}^{-2}$. Let us recall that for Le Tampon site, the corresponding values were $341 \text{ W}\cdot\text{m}^{-2}$ and $455 \text{ W}\cdot\text{m}^{-2}$. This comparison confirms that the sky conditions have a clear impact on the forecasting quality of the probabilistic models.

Based on a previous study [6], we can conjecture a link between the typology of the solar irradiance experienced by a site (i.e., distribution of the clear sky index and variability) and the performance of the probabilistic models. An in-depth study with more sites (>2) is needed to establish definitively a link between solar variability 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 $W \cdot m^{-2}$, 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. The analysis of the forecasting performance demonstrated that the quality of the intra-day probabilistic forecasts benefits from including as supplementary explanatory variables the NWP data provided by ECMWF. The inclusion of these exogenous predictors led to the sharpest predictive distributions (i.e., narrowest predictions intervals) without degrading the reliability of the intra-day probabilistic forecasts and yielded larger skill scores than the other models that do not use exogenous inputs. By studying the forecast performance for two locations with very dissimilar solar variability, this work showed that this metric determines, to a large extent, the quality of the forecasts. Insular sites like Le Tampon, which exhibits high solar variability due to a very dynamic cloud formation and dissipation, show, in general, 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 [23] or quantile regression based on neural networks [24].

Acknowledgments: 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 ground measurements.

Author Contributions: The authors contributed equally to this work.

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

Abbreviations

The following abbreviations are used in this manuscript:

CRPS	Continuous ranked probability score
ECMWF	European Center for Medium-Range Weather Forecasts
EPS	Ensemble prediction system
GHI	Global horizontal solar irradiance
PI	Prediction interval
PICP	Prediction interval coverage probability
PINAW	Prediction interval normalized averaged width
PIMENT	Physics and applied mathematics for energy, environment and buildings laboratory
QR	Quantile regression
SoDa	Solar radiation Data
SURFRAD	Surface Radiation Budget Network
TCC	Total cloud cover

References

1. Lorenz, E.; Heinemann, D. Prediction of solar irradiance and photovoltaic power. In *Comprehensive Renewable Energy*; Elsevier: Oxford, UK, 2012; pp. 239–292.
2. Bridier, L.; David, M.; Lauret, P. Optimal design of a storage system coupled with intermittent renewables. *Renew. Energy* **2014**, *67*, 2–9.
3. Grantham, A.; Gel, Y.R.; Boland, J. Nonparametric short-term probabilistic forecasting for solar radiation. *Solar Energy* **2016**, *133*, 465–475.
4. Chu, Y.; Coimbra, C.F. Short-term probabilistic forecasts for Direct Normal Irradiance. *Renew. Energy* **2017**, *101*, 526–536.
5. Golestaneh, F.; Pinson, P.; Gooi, H.B. Very Short-Term Nonparametric Probabilistic Forecasting of Renewable Energy Generation—With Application to Solar Energy. *IEEE Trans. Power Syst.* **2016**, *31*, 3850–3863.
6. 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.
7. Lauret, P.; Lorenz, E.; David, M. Solar Forecasting in a Challenging Insular Context. *Atmosphere* **2016**, *7*, 18.
8. ECMWF. Available online: <https://www.ecmwf.int/en/research/modelling-and-prediction> (accessed on 6 September 2017).
9. Pinson, P.; McSharry, P.; Madsen, H. Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation. *Q. J. R. Meteorol. Soc.* **2010**, *136*, 77–90.
10. Hamill, T.M. Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Weather Rev.* **2001**, *129*, 550–560.
11. Hersbach, H. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather Forecast.* **2000**, *15*, 559–570.
12. Gneiting, T.; Balabdaoui, F.; Raftery, A.E. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2007**, *69*, 243–268.
13. Khosravi, A.; Nahavandi, S.; Creighton, D. Prediction Intervals for Short-Term Wind Farm Power Generation Forecasts. *IEEE Trans. Sustain. Energy* **2013**, *4*, 602–610.
14. Lefèvre, M.; Oumbe, A.; Blanc, P.; Espinar, B.; Gschwind, B.; Qu, Z.; Wald, L.; Schroedter-Homscheidt, M.; Hoyer-Klick, C.; Arola, A.; et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmos. Meas. Tech.* **2013**, *6*, 2403–2418.
15. 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).
16. MACC Project. Available online: <http://www.gmes-atmosphere.eu/news/> (accessed on 6 September 2017).
17. Almeida, M.P.; Zilles, R.; Lorenzo, E. Extreme overirradiance events in São Paulo, Brazil. *Solar Energy* **2014**, *110*, 168–173.
18. Inman, R.H.; Chu, Y.; Coimbra, C.F. Cloud enhancement of global horizontal irradiance in California and Hawaii. *Sol. Energy* **2016**, *130*, 128–138.
19. Hoff, T.E.; Perez, R. Modeling PV fleet output variability. *Sol. Energy* **2012**, *86*, 2177–2189.
20. 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. *Sol. Energy* **2010**, *84*, 2161–2172.
21. 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.
22. Koenker, R.; Bassett, G. Regression Quantiles. *Econometrica* **1978**, *46*, 33.
23. Meinshausen, N. Quantile Regression Forests. *J. Mach. Learn. Res.* **2006**, *7*, 983–999.
24. Cannon, A.J. Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Comput. Geosci.* **2011**, *37*, 1277–1284.
25. Friedman, J. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378.
26. Chernozhukov, V.; Fernandez-Val, I.; Galichon, A. Quantile and Probability Curves Without Crossing. *Econometrica* **2010**, *78*, 1093–1125.
27. Candille, G.; Côte, C.; Houtekamer, P.; Pellerin, G. Verification of an Ensemble Prediction System against Observations. *Mon. Weather Rev.* **2007**, *135*, 2688–2699.
28. Bacher, P.; Madsen, H.; Nielsen, H.A. Online short-term solar power forecasting. *Sol. Energy* **2009**, *83*, 1772–1783.

29. Alessandrini, S.; Delle Monache, L.; Sperati, S.; Cervone, G. An analog ensemble for short-term probabilistic solar power forecast. *Appl. Energy* **2015**, *157*, 95–110.
30. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences an Introduction*; Elsevier Science: Amsterdam, The Netherlands, 2006.
31. Jolliffe, I.; Stephenson, D. *Forecast Verification. A Practitioner's Guide in Atmospheric Science*; Wiley: Chichester, UK, 2003.
32. Gneiting, T.; Raftery, A.E.; Westveld, A.H.; Goldman, T. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Mon. Weather Rev.* **2005**, *133*, 1098–1118.
33. Bröcker, J.; Smith, L.A. Increasing the Reliability of Reliability Diagrams. *Weather Forecast.* **2007**, *22*, 651–661.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).