

PROBABILISTIC SOLAR FORECASTS EVALUATION PART 2: QUANTILE FORECASTS

Philippe Laurent¹, Faly Ramahatana-Andriamasomanana¹, Josselin Le Gal La Salle¹ and Mathieu David¹

¹ PIMENT Laboratory - University of La Reunion, Saint-Denis (Réunion Island)

Abstract

Two types of probabilistic forecasts are used in the field of solar energy forecasting. The first one takes the form of Ensemble forecasts commonly provided by meteorological utilities such as ECMWF. The second one is based on statistical methods that generate quantiles forecasts. The first type of forecasts is treated in the first part of this article. In this second part, we focus on the assessment of the quality of quantile forecasts characterized by two main attributes namely reliability and resolution. More precisely, we propose a verification framework based on diagnostic tools and quantitative scoring rules that are specifically designed for the evaluation of quantile forecasts. The probabilistic models are evaluated on two selected sites that experience very different climatic conditions. It is shown that the decomposition into reliability and resolution of a scoring rule, the Continuous Ranked Probability Score (CRPS), may help to obtain a detailed picture of the performance of the models and consequently may help in selecting the best probabilistic method.

Keywords: probabilistic solar forecasting, quantile forecasts, scoring rules, CRPS, reliability, resolution

1. Introduction

Integration of solar power generation capacities may lead to difficulties in the management of a power electricity grid. Indeed, the fluctuating character of solar energy may cause imbalances between electricity supply and demand. This requires the power system to either procure additional reserves or adjust the output of conventional generators so as to ensure the security of the supply-demand balance (Inman et al., 2013). One key element in facilitating the penetration of solar renewables consists in predicting at different time horizons the global horizontal solar irradiance (GHI) and the corresponding solar PV power output so that the grid operator is able to take appropriate actions to mitigate solar intermittency.

In order to improve the decision-making process of the grid operator, a point forecast plus an estimation of the associated uncertainty may contribute to a more efficient integration of intermittent sources in the electrical network (Morales et al., 2014).

Assessment of the performance of the point or deterministic forecasts is now quite standard in the solar forecasting community. Metrics like RMSE, MBE and MAE (Coimbra et al., 2013, Vallance et al., 2017) are routinely used to evaluate these point forecasts. Conversely, evaluation of probabilistic forecasts is quite more complex and currently not mature in the field of solar forecasting.

Probabilistic forecasts can take either the form of Ensemble Forecasts provided by meteorological centers or predictive distributions summarized by discrete quantiles (also called quantile forecasts). The assessment of the first type of forecasts is treated in part 1 of this article. The objective of this second part is to propose an approach to evaluating the quality of probabilistic solar forecasts that take form of quantile forecasts. Two main attributes namely reliability and resolution characterize the quality of the probabilistic forecasts (Wilks, 2006). The proposed evaluation framework is based on visual diagnostic tools as well as quantitative scoring rules like the Continuous Ranked Probability Score (CRPS) (Hersbach, 2000). The latter permits to have an idea of the overall skill of a forecasting method. In addition, a particular emphasis is put on the decomposition of the CRPS into reliability and resolution. Assessment of quantile regression models at two sites that exhibit different sky conditions will illustrate the application of the proposed framework.

2. Quantiles forecasts

Probabilistic forecasts correspond to the estimation of the statistical distribution of a future event. Thus, a probabilistic forecast can be represented by a cumulative distribution function (CDF). This CDF can be summarized by a set of discrete quantiles spanning the unit interval. In this work, these quantiles will be produced by non-parametric methods that make no assumption about the distribution of the future event.

Prediction intervals (PIs) with different nominal coverage rates can be inferred from this set of quantiles forecasts. PIs give a range of possible values within which the true value of the variable to forecast is expected to lie with a certain probability, that is its nominal coverage. Figure 1 shows an example of GHI probabilistic forecasts where point forecasts are enriched with PIs with different nominal coverage rates.

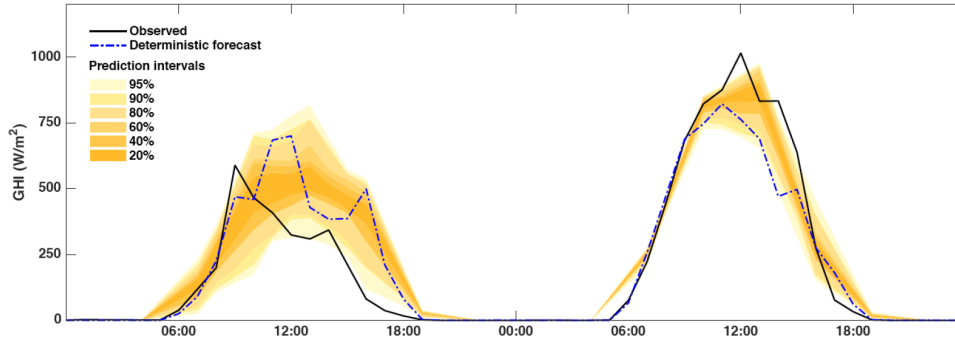


Fig. 1: Solar probabilistic forecasts. PIs with nominal coverage rates ranging from 20% to 95% are provided.

3. Verification tools

3.1. Diagnostic tools

For evaluating quantile forecasts, we propose first the use of reliability diagrams. The reliability diagram is a graphical verification display used to evaluate the reliability component of a probabilistic forecast system. As noted by (Pinson et al., 2007), reliability is a primary requirement when verifying probabilistic forecasts, since a lack of reliability would introduce a systematic bias in subsequent decision-making. In this work, a probabilistic forecasting system based on quantile forecasts is reliable if, statistically, the nominal proportions of the quantile forecasts are equal to the proportions of the observed value.

3.2. Scoring rules

We focus in this work on proper numerical scores (Bröcker and Smith, 2007) that provide summary measures for the evaluation of the quality of probabilistic forecasts. Among others, one can cite particularly the following scoring rules: CRPS, Ignorance Score, Interval score (Gneiting and Raftery, 2007). In this work, we focus on the CRPS, which measures the difference between the predicted and observed cumulative distributions functions (CDF). The CRPS score rewards concentration of probability around the step function located at the observed value (Wilks, 2006). In other words, the CRPS penalizes lack of resolution of the predictive distributions as well as biased forecasts. Notice that the CRPS is negatively oriented (smaller values are better) and has the same dimension as the forecasted variable. Hersbach (2000) proposed a specific method to calculate the CRPS for ensemble forecasts. Here, and specifically for methods that produce discrete quantile forecasts, we integrate (see eq. 1) the Brier Score (BS) over all possible values of the predictand x (here from 0 to the maximum GHI climatology). The Brier Score (BS) is a score used to evaluate probabilistic forecasts of binary predictand (Wilks, 2006):

$$CRPS = \int BS(x)dx = \int REL(x)dx - \int RES(x)dx + \int UNC(x)dx \quad (\text{eq. 1})$$

In addition, we provide a decomposition of the CRPS through the integration of the different terms of the BS. The reliability term (REL) provides an estimation of the forecast biases while the resolution term (RES) quantifies the improvement that results from issuing probability forecasts that are case dependent. The uncertainty term (UNC) cannot be modified by the forecast system and depends only on the observations variability (Wilks, 2006). Using this decomposition of the CRPS may lead to a detailed picture of the forecast performance of the different methods.

Figure 2 plots the Brier score and its related decomposition into reliability, resolution and uncertainty. The area under each curve corresponds to the related CRPS component. As the CRPS is negatively oriented, the goal of a forecast system is to minimize (resp. maximize) as much as possible the reliability term (resp. the resolution term).

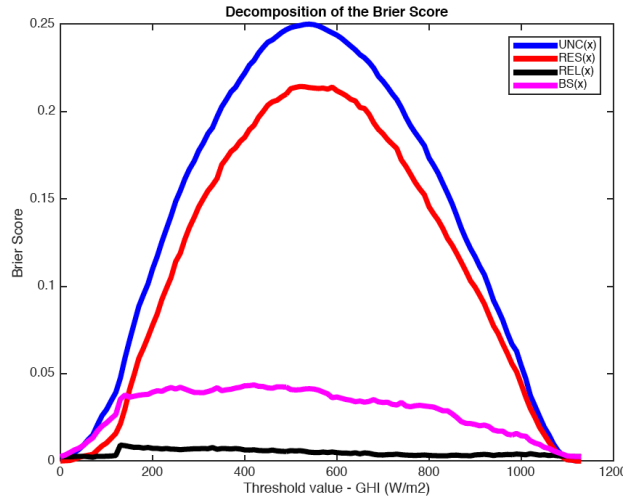


Fig. 2: Decomposition of the BS score - Integration of BS(x) for all threshold values x gives the CRPS.

4. Application of the proposed framework

As mentioned in Part 1 of this article, application of this verification framework is done on two sites that experience different sky conditions. The first site, Desert Rock (USA), has an arid climate with a very sunny and stable sky. The second site, Tampon (Réunion Island), is located in a tropical island and experiences a very variable sky. A data quality check based on the BSRN recommended procedures (Long and Dutton, 2019) was undertaken for the observation data of each of the two studied sites.

Intra-day forecasts with lead times ranging from 1 to 6 hours are provided by state of the art forecasting models that generate predictive distributions from a set of quantiles spanning the unit interval. The selected models are based on quantile regression techniques namely the quantile regression forest (QRF) and the Gradient Boosting (GB) methods that estimate directly the set of quantiles from a regression model $Y = f(X)$ that relates the response variable Y (here GHI) to a set of predictor variables X . Two variants of regression models with different sets of predictor variables are built. For the first variant described in (Lauret et al., 2017), the vector of explanatory variables X consists of the actual measurement plus five past ground measurements while the second one takes as additional inputs two geometrical solar features related to the course of the sun in the sky namely the cosine of the zenith angle ($\cos(SZA)$) and the cosine of the hour angle ($\cos(HA)$). The adding of the two variables originates from the following reasons. First, some authors (Grantham et al., 2016; Lorenz and Heinemann, 2012) showed a clear dependency of the forecasting error in relation to SZA. Second, we expect that the solar hour angle (which is the angular displacement of the sun at 15° per hour; 0° at solar noon, morning negative, afternoon positive) will bring some information regarding the asymmetry of the sky conditions between mornings and afternoons. This may be hold particularly for site like Le Tampon that experiences such a dichotomy between mornings and afternoons. Table 1 lists the acronyms of the resulting four quantile regression models.

Tab. 1: Acronyms related to the four quantile regression models

Quantile regression techniques	Variant 1	Variant 2
Quantile Regression Forest	QRF1	QRF2

Gradient Boosting	GB1	GB2
-------------------	-----	-----

4.1. Reliability assessment

As mentioned above, reliability diagrams allow to graphically assessing the reliability of a set of quantile forecasts. Figures 3 and 4 plot the reliability diagrams (averaged over all the forecasting horizons) for the two selected sites. Consistency bars for a 90% confidence level are individually computed for each nominal proportion. From the visual inspection of the reliability diagrams of Desert Rock, one can possibly state that the GB2 model is reliable as the observed proportions of all quantiles lie within the consistency bars. Conversely, for the others models, observed proportions of some quantiles lie outside the consistency bars. In particular, quantile forecasts generated by the QRF2 model should not be considered reliable. In addition, notice the particular signature of the QRF2 model that corresponds to an over dispersed predictive distribution (i.e. an underconfident model). For the site of Le Tampon, it seems that, except the GB2 model, all the other models lead to possible reliable quantile forecasts since all of their observed proportions lie within the consistency bars. Moreover, notice that the QRF1 model exhibits a reliability curve very close to the ideal diagonal case.

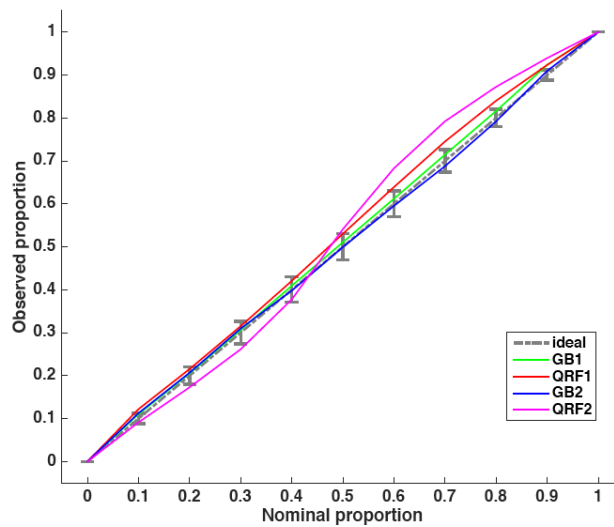


Fig. 3: Reliability diagrams - Site of Desert Rock.

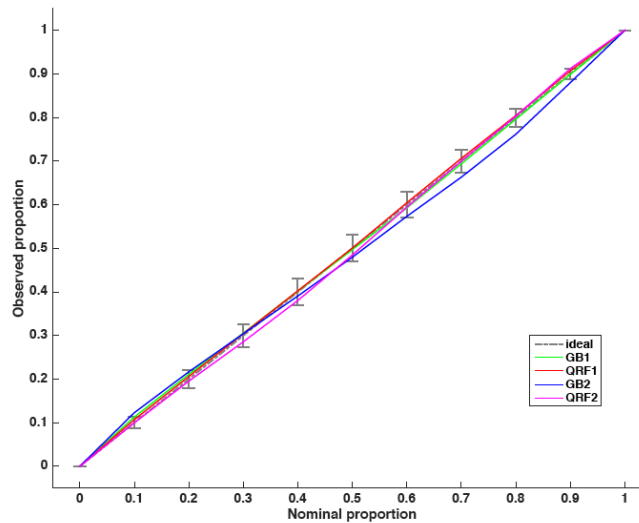


Fig. 4: Reliability diagrams - Site of Le Tampon

4.2. CRPS and its decomposition into resolution and reliability

This section proposes a detailed picture of the performance of the probabilistic models by plotting the CRPS and its associated decomposition into reliability and resolution. Figure 5 plots the relative CRPS in relation with the forecast horizon for the two considered sites. As expected, the performance of the models decreases as the lead-time increases (i.e. the lower the CRPS, the better the model). One also may note that the site of Le Tampon, which experiences variable sky conditions compared to Desert Rock, yields higher CRPS values. As shown by Figure 5, the two non-linear models that include the two solar geometric predictors namely zenith angle and hour angle (i.e. GB2 and QRF2 models) perform clearly better than the variant 1 models regardless the site. Thus, it appears that adding the two solar geometric variables brings a clear improvement and especially for a site like Le Tampon which is known to experience a morning/afternoon sky asymmetry.

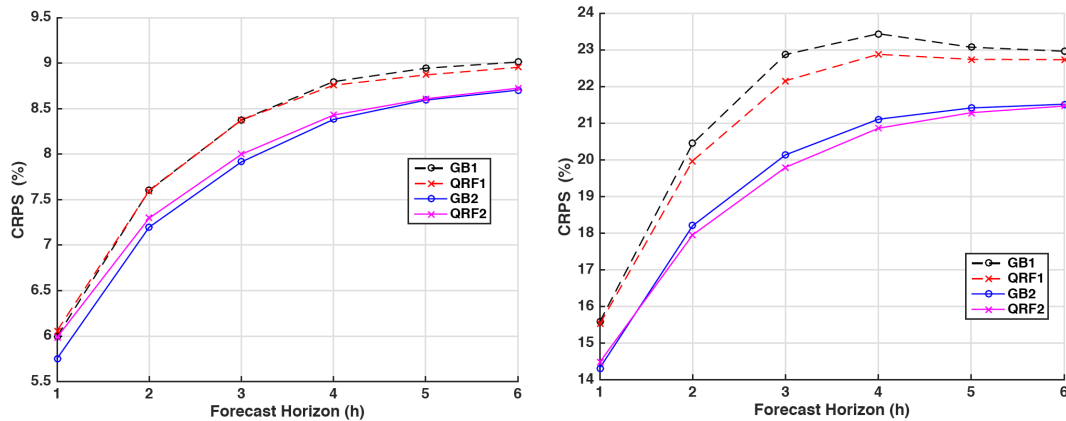


Fig. 5: CRPS in relation with the forecast horizon. Desert Rock (left)-Tampon (Right)

Unlike the previous separate analysis of reliability, CRPS establishes a clear-cut ranking of the models. However, some inconsistencies appear with the reliability analysis which showed that the QRF2 model (resp. the GB2 model) was non reliable for Desert Rock (resp. for Le Tampon). Therefore, in order to gain a better understanding of the CRPS results, we use the decomposition of the CRPS into reliability, resolution and uncertainty. This decomposition shows that the reliability component makes a small contribution to the CRPS and that the higher quality of the variant 2 models comes from the resolution attribute. Figure 6 plots the reliability component of the CRPS. Surprisingly, the reliability does not show a tendency to increase with the lead time. Indeed, we expect the reliability term to increase with increasing forecast horizon (we recall that the reliability term is negatively oriented i.e. a lower reliability value corresponds to a more reliable forecasts). However, in agreement with the reliability assessment, the GB2 model exhibits the lowest reliability for the site of Desert Rock while for Le Tampon, low reliability values are obtained with the QRF1 model. Nonetheless, it must be noted that the reliability component weakly contributes to the CRPS and that the higher quality of the probabilistic forecasts generated by the variant 2 models originates from the resolution attribute. Figure 7 shows the resolution part of the CRPS, which confirms the lack of resolution of the different models as the forecast horizon increases. Regarding resolution, the statements made regarding the CRPS still hold i.e. the two non-linear models (GB2 and QRF2) that include the solar geometric predictors lead to better resolution.

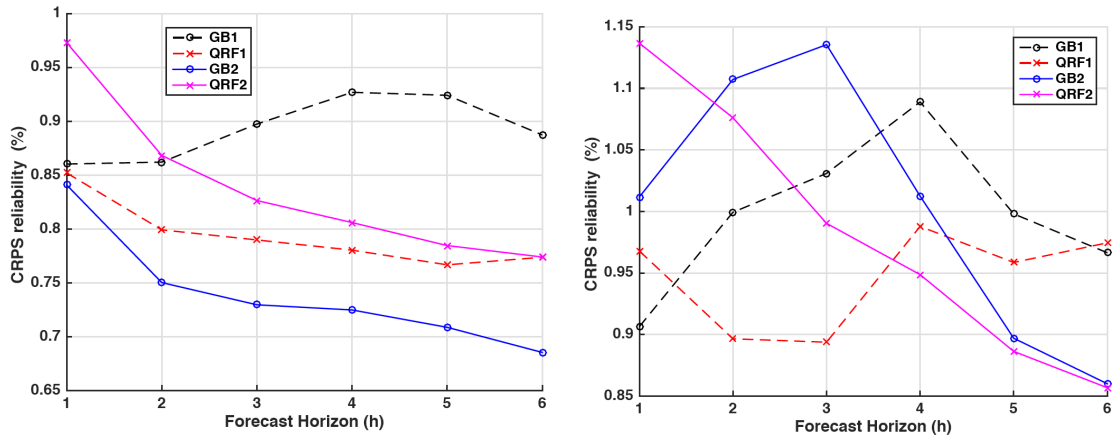


Fig. 6: CRPS reliability in relation with the forecast horizon. Desert Rock (left)-Tampon (Right)

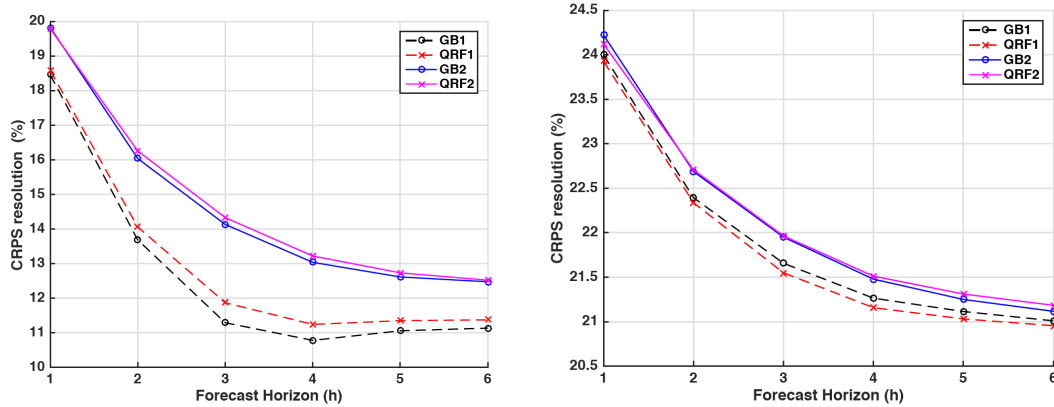


Fig. 7: CRPS resolution in relation with the forecast horizon. Desert Rock (left)-Tampon (Right)

5. Conclusion

This work proposed a set of metrics to evaluate predictive distributions summarized by quantiles forecasts. While visual diagnostic tools may help the user to qualitatively assess the quality of the probabilistic forecasts, use of quantitative score like CRPS and its associated decomposition into two important attributes namely reliability and resolution may help the user to select the best probabilistic model.

Therefore, we recommend to systematically compute an overall score i.e. the CRPS which, in our opinion, might be a standard in assessing probabilistic forecasts of continuous variable. This proper score allows ranking models and its relative counterpart (i.e. CRPS normalized by the mean irradiance) permit to carry out sites' comparisons. Furthermore, the decomposition of the CRPS into reliability and resolution may provide additional insight into the performance of a forecasting system.

6. References

- Bröcker, J., Smith, L.A., 2007. Scoring Probabilistic Forecasts: The Importance of Being Proper. *Weather Forecast.* 22, 382–388.
- Coimbra, C.F.M., Kleissl, J., Marquez, R., 2013. Overview of Solar-Forecasting Methods and a Metric for Accuracy Evaluation, in: *Solar Energy Forecasting and Resource Assessment*. Elsevier, pp. 171–194.
- Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.* 102, 359–378.

- Grantham, A., Gel, Y.R., Boland, J., 2016. Nonparametric short-term probabilistic forecasting for solar radiation. *Sol. Energy* 133, 465–475.
- Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather Forecast.* 15, 559–570.
- Inman, R.H., Pedro, H.T.C., Coimbra, C.F.M., 2013. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* 39, 535–576.
- Lauret, P., David, M., Pedro, H.T.C., 2017. Probabilistic Solar Forecasting Using Quantile Regression Models. *Energies* 10, 1591.
- Long, C.N., Dutton, E.G., 2019. BSRN Global Network recommended QC tests, V 2.0. https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf.
- Lorenz, E., Heinemann, D., 2012. Prediction of solar irradiance and photovoltaic power. In: *Comprehensive Renewable Energy*. Elsevier, Oxford, UK, pp. 239–292.
- Morales, J.M., Conejo, A.J., Madsen, H., Pinson, P., Zugno, M., 2014. *Integrating Renewables in Electricity Markets*, International Series in Operations Research & Management Science. Springer US, Boston, MA.
- Pinson, P., Nielsen, H.A., Møller, J.K., Madsen, H., Kariniotakis, G.N., 2007. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy* 10, 497–516.
- Vallance, L., Charbonnier, B., Paul, N., Dubost, S., & Blanc, P. 2017. Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric. *Solar Energy*, 150, 408-422.
- Wilks, D.S., 2006. *Statistical methods in the atmospheric sciences*. Academic Press, Amsterdam; Boston