EuroSun2026

PROBABILISTIC FORECASTING OF RENEWABLE ENERGY PRODUCTION AND ELECTRICITY LOAD USING HOURLY GAUSSIAN COPULAS ENSEMBLE FORECAST (GCH-En)

Vincent Barbera[1], Mathieu David[1], Philippe Lauret[1]

[1] *University of La Réunion - PIMENT laboratory, 15, avenue René Cassin, 97715 Saint-Denis, France*

vincent.barbera@univ-reunion.fr

Summary

Insular power systems must balance supply and demand in real time with a high share of intermittent generation and no interconnection. Operational decisions are driven by forecast uncertainty, which motivates probabilistic forecasting of PV power generation, wind power and load. Standard practice forecasts each variable independently (univariate case), ignoring the dependence structure between electrical load, photovoltaic (PV) generation and wind power (multivariate case). This work attempts to answer the following question: at which evaluation level and for which lead time does the multivariate approach deliver a measurable gain over the univariate one ? To this end, we benchmark a family of multivariate conditional Gaussian copulas (GCH-En) against strong univariate baselines methods (Quantile Random Forest, XGBoost), the study uses two years of 10-minute data collected at the Terre-Sainte campus on Reunion Island. The evaluation covers three target variables across six lead times, ranging from 10 minutes to 12 hours. We verify the forecasts with three increasingly dependence-sensitive metrics: the Continuous Ranked Probabilistic Score ($CRPS$), the Energy-Score and the Variogram-Score. The joint calibration evaluation is also implemented through the Band-Depth rank histograms. The XGBoost baselines lead on CRPS for each variable at every lead except wind power beyond 4h, hold a narrow Energy-Score advantage up to 4h and are overtaken on the Variogram-Score from 2h onwards. Joint calibration is the one level at which the separation is clear: the copulas are close to uniform at all lead times, whereas the univariate ensembles place the observation in the lowest 5% of band-depth ranks in 42% of cases, against 5% under calibration.

Keywords: probabilistic forecasting, Gaussian copula, renewable energy, ensemble forecast, multivariate dependence, CRPS, Energy-Score, Variogram-Score, multivariate rank histogram

Field of application for the finding: Power system management, renewable energy integration, smart grids

Addressed audience: Researchers and practitioners in solar energy forecasting, grid operators

1. Introduction

As the penetration of photovoltaic (PV) and wind generation increases, the residual load that conventional and storage assets must cover becomes increasingly variable and uncertain. In this situation, operational decisions depend not on a single best-guess trajectory but on the full predictive distribution: reserve sizing, storage dispatch and security margins depends on the full forecast distribution, of which the point forecast is only the centre.

Probabilistic forecasting addresses this need by delivering a predictive distribution per variable and horizon (Hong et al., 2016). Standard practice produces, however, produces these distributions one variable at a time: electrical load, PV generation and wind power are each forecast by an separate univariate model and the target forecasts are then used as if they were deemed jointly. The field's verification practice reflects this default: the reference framework for solar probabilistic forecasts is explicitly restricted to the univariate case, in which spatio-temporal dependencies are not accounted for (Lauret et al., 2019) and the standard benchmarks are likewise marginal (Yang, 2019). Multivariate formulations have emerged only recently, a dedicated multivariate benchmark having been proposed as late as 2021 (van der Meer, 2021b).

Multivariate probabilistic forecasting preserves the dependence structure by construction. Through Sklar's theorem, the joint distribution F of d-variate vectors decomposes into its univariate distributions F_j, also called marginals, and a copula C that carries all the dependence (Nelsen, 2006):

$$F(x_1, x_2, ..., x_d) = C(F_1(x_1), F_2(x_2), ..., F_d(x_d)), \tag{1}$$

allowing the dependence to be modelled explicitly and separately from the univariate marginal distributions. The trade-off is additional methodological and computational complexity, which raises a practical question for system operators: *is the added complexity justified and under what circumstances?*

Copula-based approaches to renewable energy forecasting fall into two classes. The first couples target predictive distributions a posteriori. The application of copula methods to renewable energy forecasting has evolved progressively over the past decade. Initially, (Pinson et al., 2009) introduced the Gaussian copula to model temporal dependencies in wind power trajectories. This concept was generalised by (Schefzik et al., 2013) into the Ensemble Copula Coupling (ECC) framework. Later, the ECC approach was extended to wide-area solar generation (Golestaneh et al., 2016; Woodruff et al., 2018) and to joint wind-solar time series (Pan et al., 2019). The second approach embeds the dependency structure (the copula) directly inside the forecasting model rather than applying it in a separate post-processing step. This internal is achieved either through Bayesian model averaging (Möller et al., 2013) or via recursive spatio-temporal conditioning (Munkhammar, 2024). Based on the theoretical foundation established by (Nelsen, 2006), this mathematical framework is used for multivariate time-series forecasting (Patton, 2013) and extends to post-processing in hydrological studies (Hemri et al., 2015).

Closest to this work, Van Der Meer et al. (2020) compares candidate copulas for clear-sky index trajectories and van der Meer (2021a) proposes a non-parametric multivariate benchmark, referenced here as MuPEn. Both address a single variable, the horizontal irradiance, across space and time. We depart from them in three aspects: the joint vector is a set of heterogeneous variables(load, PV and wind power), the dependence is intrinsic to the predictive mechanism rather than applied afterwards (Section 3.1) and the univariate baselines are competitive on univariate metrics. Reviews of solar forecasting (Inman et al., 2013) and of wind integration constraints (Pinson, 2013) provide the wider context. This paper gives a deliberately scoped answer. Rather than claiming a superiority of multivariate models, we show that the added value of modelling dependence is conditional on the metric used to evaluate the quality of the forecasts. We benchmark a family of conditional Gaussian copulas against state-of-art univariate models on a real insular dataset and we evaluate at three increasingly dependence-sensitive levels. The contributions of this work are (i) a like-for-like multivariate-versus-univariate comparison on insular data at six horizons and (ii) a clear empirical characterisation of where dependence modelling does and does not matter.

2. Data and Forecasting Targets

The dataset comprises 10-minute measurements collected on the Terre-Sainte Campus, La Réunion, over two full years, 2021 and 2022. The year 2021 is used for training ($52,559$ records) and 2022 for testing ($52,561$ records). Three targets are forecast jointly: total electricity load, total PV generation and wind power.

The wind target is not raw wind speed but the electrical power produced by a reference turbine. Using the *windpowerlib* model chain (Haas et al., 2019), the measured 10-meter wind speed is extrapolated to hub height under a logarithmic wind profile, the air density is corrected from temperature and pressure and the turbine power curve is evaluated. The resulting signal is non-linear in wind speed, which makes it a more realistic and more demanding forecasting target than wind speed itself.

Tab.1 summarises the training-set statistics of the three targets. The targets differ in distribution shape: load is positive and moderately dispersed, PV is zero at night with a strong diurnal cycle and wind power is highly intermittent with frequent zeros.

Tab. 1: training-set statistics of the three targets (2022)

Statistic	Load (kW)	PV (kW)	Wind power (kW)
Mean	93.76	30.98	69.95
Std. dev.	36.64	42.88	135.42
Median	83.00	0.16	9.69
Max	264.00	155.00	810.00
Uncertainty component (% observed mean)	18.90	21.53	54.48

3. Forecasting models

3.1 Multivariate models: conditional Gaussian copulas

The conditional Gaussian copulas model used in this work is inspired from the approach proposed by Pinson et al. (2009). The multivariate family (GCH-En) builds, for a given context, the joint distribution of the present state of the three targets together with their future states at all forecast horizons. The marginal distribution of each target variable is mapped to a standard Gaussian scale by the probability integral transform composed with the inverse standard normal CDF, using the empirical marginal F_j^{-1}:

$$z_j = \Phi^{-1}(F_j(x_j)), \tag{2}$$

where Φ is the standard normal CDF and x_j the targets. Stacking the Gaussian-scale present and future variables into a single vector $z = (z_P, z_F)$ (present P, future F), the dependence is summarised by a zero-mean Gaussian distribution with a correlation matrix Σ partitioned into present X and future Y blocks:

$$z \sim \mathcal{N}(0,\, \Sigma), \Sigma = \begin{bmatrix} \Sigma_{PP} & \Sigma_{PF} \\ \Sigma_{FP} & \Sigma_{FF} \end{bmatrix}. \tag{3}$$

The conditional distribution of the future given the observed present follows a Gaussian distribution, with conditional mean $\mu_{F|P}$ and covariance $\Sigma_{F|P}$ as:

$$\mu_{F|P} = \Sigma_{FP}\Sigma_{PP}^{-1}z_P, \tag{4}$$

$$\Sigma_{F|P} = \Sigma_{FF} - \Sigma_{FP}\Sigma_{PP}^{-1}\Sigma_{PF}. \tag{5}$$

Forecasts are produced by drawing $m = 200$ joint scenarios $\hat{z}_F^{(i)} \sim \mathcal{N}(\mu_{F|P}, \Sigma_{F|P})$ from this conditional Gaussian distribution and mapping each scenario back to physical units through the inverse applications done before, $\hat{x}_j^{(i)} = F_j^{-1}(\Phi(\hat{z}_j^{(i)}))$. By construction, every sampled scenario is a coherent joint trajectory across variables and horizons.

It is worth noting how this construction differs from copula-based scenario generation introduced by Pinson et al. (2009). In their approach, a predictive distribution is first produced independently for each variable and horizon, then a Gaussian copula is fitted afterwards to recombine these univariate forecasts into coherent joint trajectories. The dependence is applied a posteriori as a coupling step while each target variable already carries the full predictive content. The same a posterior-coupling logic underlies subsequent work, notably Van Der Meer et al. (2020), who compare several copula family for recombining univariate solar forecasts into clear-sky index space-time trajectories. Here the roles are reversed. The target values are static: within each conditioning layer they are the unconditional empirical climatology of the component and carry no predictive information of their own. All of the predictive content therefore comes from the present-to-future dependence held in the off-diagonal block Σ_{FP} of eq. (3), the forecast is Gaussian conditioning of the joint distribution pn the observed present, not the coupling pre-existing univariate forecasts. The dependence is therefore intrinsic to the mechanism rather than added after it. A direct consequence is that GCH-En carries no conditional univariate model, so its univariate resolution is bounded by what Gaussian conditioning can extract through correlation. This is the price paid for making the joint dependence the primary object and it is consistent with the univariate-versus-joint behaviour reported in Section 5. A second limitation follows from the Gaussian choice itself, made here for the closed-form conditioning it affords: the Gaussian copula carries no tail dependence (Genest and Favre, 2007).

Three stratifications of the copula are compared, differing only in how the conditioning context is defined: **GCH-En** fits one copula per time of day (HH:MM), **GCH-En-Sais** refines this to one copula per time of day (HH:MM) and season, and **GCH-En-Ex** additionally appends two exogenous variables, the current dry-bulb temperature (DBT) and relative humidity (RH), to the conditioning vector. Each layer is assumed stationary over the training year, so that a single correlation structure is estimated per conditioning key.

3.2 Univariate baselines

Two univariate baseline forecasting models are used. **Quantile Random Forest (QRF)** (Meinshausen, 2006) fits one forecast per (target, horizon) pair. **Quantile XGBoost** (Chen and Guestrin, 2016) fits one gradient-boosted model per (target, horizon, quantile) using the pinball score:

$$\rho_\tau(u) = \begin{cases} \tau u, \ u \geq 0 \\ (\tau - 1)u, \ u < 0 \end{cases}, u = y - q_\tau \tag{6}$$

for predicted quantile q at level τ. Both use the same feature set: the current observed values of the three targets plus a cyclic encoding (sine and cosine) of the time of day and of the day of year. A third baseline model, **XGBoost-Ex**, adds the current DBT and RH to the feature vector, providing the univariate analogue of GCH-En-Ex and enabling a fairer assessment of the impact of adding exogenous inputs. Predicted quantiles are treated as ensemble members so that all models are scored within an identical ensemble framework.

A non-parametric multivariate baseline, **MuPEn**, introduced by van der Meer (2021a), provides an unconditional reference. It is a climatological ensemble, so its forecasting performances are constant over the lead times and any conditional model is expected to improve on it.

4. Evaluation Framework

Forecast quality is assessed at three increasingly dependence-sensitive levels, plus joint calibration. The evaluation of scenario quality follows the principles set by Pinson and Girard (2012).

The Continuous Ranked Probability Score (CRPS) (Gneiting and Raftery, 2007) measures separately the quality of each target variable predictive distribution. For a predictive CDF $\hat{F}$ and observation y it is:

$$CRPS(\hat{F}, y) = \int_{-\infty}^{+\infty} (\hat{F}(t) - \mathbb{1}(t \geq y))^2 dt, \tag{7}$$

where $\mathbb{1}(.)$ is the *Heaviside* function. We report it normalised by the mean observation $nCRPS = CRPS/\overline{y}$. It is essentially blind to cross-variable dependence and only reflects reliability, resolution and uncertainty of a marginal distribution.

The Energy-Score is a strictly proper multivariate score that generalises the $CRPS$ to multivariate forecast (Gneiting and Raftery, 2007). For an ensemble prediction system $\{X^i\}_{i=1}^m$ and observation y it writes as follow:

$$ES(X, y) = \frac{1}{m}\sum_{i=1}^{m} ||X^{(i)} - y|| - \frac{1}{2m^2}\sum_{i,j=1}^{m} ||X^{(i)} - X^{(j)}||, \tag{8}$$

with $||.||$ the euclidean norm. In practice, it is dominated by the quality of the target values and is only weakly sensitive to the dependence structure.

The Variogram-Score is also a proper score (here with order $p = 0.5$) built on pairwise differences between components, designed to be sensitive to the dependence structure rather than to the quality of the predictive marginals (Scheuerer and Hamill, 2015):

$$VS^p = \sum_{k,l=1}^{d} w_{kl}(|y_k - y_l|^p - E|X_k - X_l|^p)^2, \tag{9}$$

where $E|X_k - X_l|^p$ is approximated by the ensemble mean and w_{kl} are non-negative weights. It is the most discriminating of the three for the question at hand.

The band-depth rank histogram (DB2) assesses ensemble calibration in the joint space rather than variable by variable. Since $\mathbb{R}^d$ admits no total order, each of the $m+1$ pooled vectors is first mapped to a scalar pre-rank and the observation is ranked among those (Gneiting et al., 2008). We use the band-depth pre-rank (Thorarinsdottir and Schuhen, 2018), which measures how often a vector is enclosed by pairs of the others. The pre-rank of member $X^{(i)}$ among the $m+1$ pooled vectors is:

$$db(X^{(i)}) = \frac{1}{d}\sum_{k=1}^{d}(r_k - 1)(m + 1 - r_k) \tag{10}$$

where r_k is the univariate rank of component k. The observation's pre-rank is then ranked among those of the members and accumulated into a histogram. A flat histogram indicates a calibrated joint forecast, and a "wall-shape" indicates joint under/over-dispersion. This diagnoses calibration in the full joint space rather than per variable.

5. Results

All results are reported on the 2022 test year, for the three targets and six forecast horizons (i.e. $10min$, $1h$,$2h$, $4h$, $6h$,and $12h$).

5.1 Univariate quality: the tie

Fig.1 reports the normalised CRPS. The XGBoost models give the best scores at short lead times across the three targets and they keep that advantage on load and PV up to 12h. Wind power behaves differently. There the conditional Gaussian Copulas overtake XGBoost-Ex beyond 4h, reaching 59.2% against 64, 6% nCRPS at 12h. QRF is the worst. Its nCRPS exceeds the climatological benchmark MuPEn from 6h onwards for PV, and by 12h it is well above it for wind power (81.7% against 67.3%). As shown in Fig.2, we attribute this high value of nCRPS to a lack of reliability rather than a poor resolution. The nCRPS decomposition shows QRF's reliability term growing by a factor of five for load and eighteen for wind power between 10min and 12h, while the copula's term stays almost flat. Averaged over lead times the two families are close: XGBoost-Ex gives

8.1 / 14.1 / 44.0% for load, PV and wind power, GCH-En 8.5 / 15.6 / 43.4%, on the nCRPS. For a decision involving one variable at one lead time, a well-tuned quantile univariate model is enough.

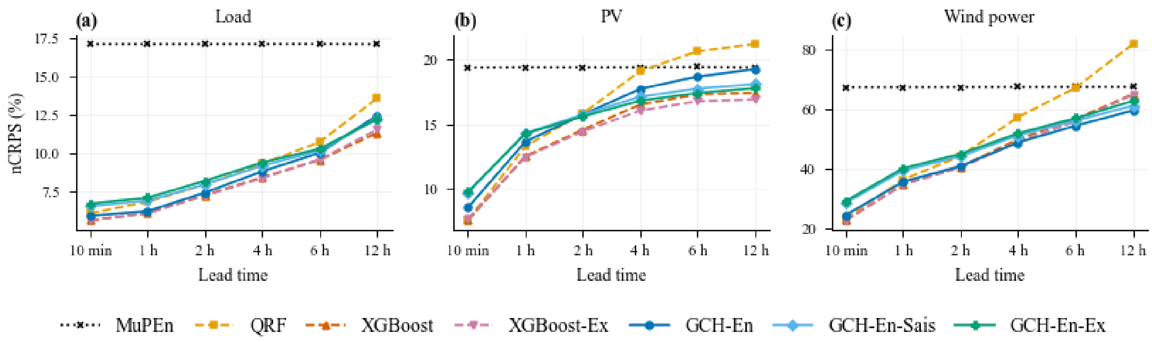

Fig. 1: Normalised $CRPS$ (% of mean observation) versus horizon for total load, PV generation and wind power.

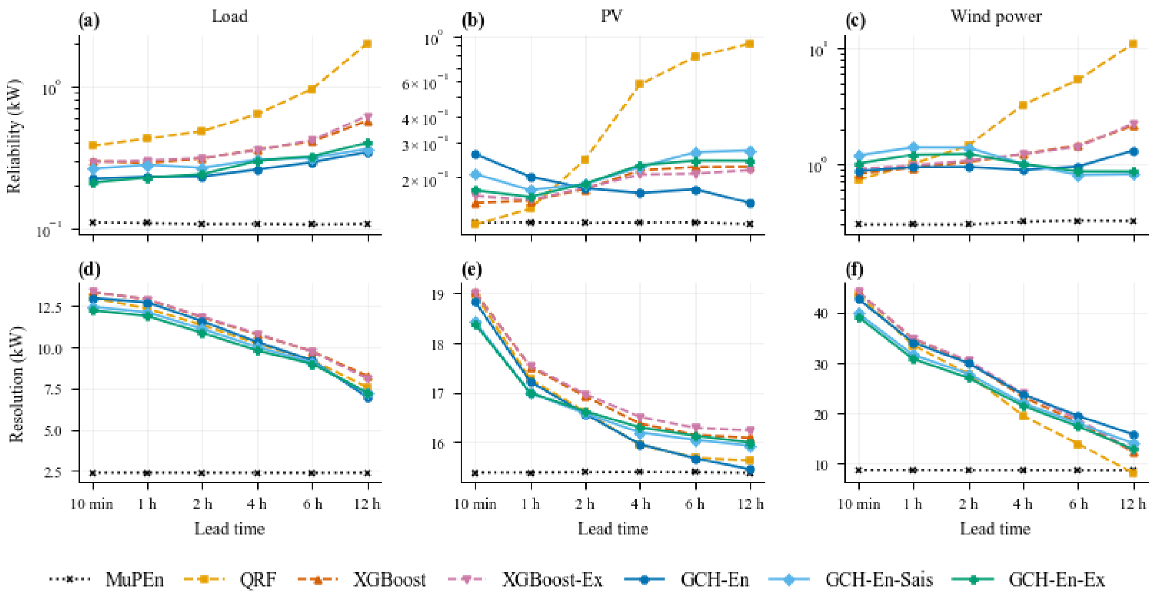

Fig. 2: Normalised $CRPS$ (% of mean observation) decomposition versus horizon for total load, PV and wind power

5.2 *Verification of the multivariate forecast*

Fig.3 reports the two multivariate scores. On the Energy-Score (Fig.3a), XGBoost and XGBoost-Ex lead up to 4h and the copulas take over from 6h. Averaged over lead times the two families are separated by less than one per cent (GCH-En 0.0830, XGBoost-Ex 0.0834, XGBoost 0.0841) and GCH-En beats the boosting baseline on that average, GCH-En-Sais (0.0860) and GCH-En-Ex (0.0869) falling behind. This is the expected behaviour of a score that generalises the CRPS but remains dominated by the univariate quality. It inherits equal outcome of section 5.1 and adds little discrimination of its own. The Variogram-Score (Fig.3b) moves the crossover forward. XGBoost still leads at 10min and XGBoost-Ex at 1h, but from 2h onwards the copulas hold the advantage at every lead time and the gap widens with the horizon. At 12h GCH-En-Sais scores 0.1158 against 0.1279 for XGBoost-Ex, a 9.5% gap and 0.1540 for QRF, a 25% gap. Against an Energy-Score of 4.1% at the same lead time, this is the signature of a score built to see the dependence structure. The same models, the same forecasts and the same test year therefore give a crossover at 4h on one score and at 2h on the other. What changes is not the forecast but the sensitivity of the metric. Neither score, however, delivers a uniform copula advantage. On the average over lead times the two families remain mixed on the Variogram-Score as well (GCH-En 0.0837, XGBoost-Ex 0.0868, GCH-En-Sais 0.0872, XGBoost 0.0883). So a conclusion drawn from a single metric at one lead time depends entirely on that specific setup.

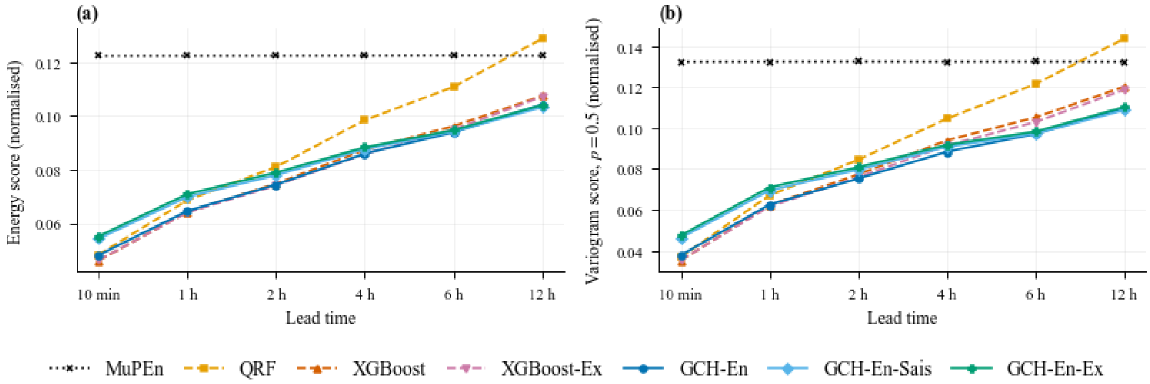

Fig. 3: Multivariate scores against lead time, both min-max normalised. (a) Energy-Score, (b) Variogram-Score ($p = 0.5$)

5.3 Joint calibration: the unconditional separation

The two scores of Section 5.2 disagree about when the copulas become preferable and neither separates the families on the average. Joint calibration shows an advantage for multivariate models at every lead time. The Band-Depth rank histograms of Fig.4 are close to uniform for the copulas and for MuPEn, but collapse for the univariate models throughout. The shape of that collapse is diagnostic. It is not the U-Shape of univariate calibration evaluation but a one-sided "wall". A calibrated ensemble would place the observation in the lowest 5% of band-depth in 5% of cases. GCH-En reaches 10%, whereas XGBoost reaches 42% and QRF 53%, and the upper ranks are left empty. Summarised by the reliability index (sum of the absolute deviation of the histogram from uniformity called DB2 index) of Tab.2, the univariate models score 1.24 to 1.48 against 0.34 to 0.39 for the copulas, a gap of a factor three to four. MuPEn scores lower still (0.279), as an unconditional climatology necessarily does. Calibration is only informative under a sharpness constraint (Gneiting et al., 2007). So the relevant comparison is among conditional models. The mechanism is built into the univariate construction. Sorting independent per-variable quantiles and reading them as ensemble members makes member k the k-th quantile level of every variable and every horizon at once. Therefore, the members are comonotone: they lie on a one-dimensional curve in the joint space, give each other high band depth and leave the observation at the minimum rank. A univariate quantile forecast is therefore not free of dependence assumptions. It commits to the strongest positive dependence that exists, the upper bound, which is also the least favourable coupling for any quantity formed as a difference between variables, such as net load.

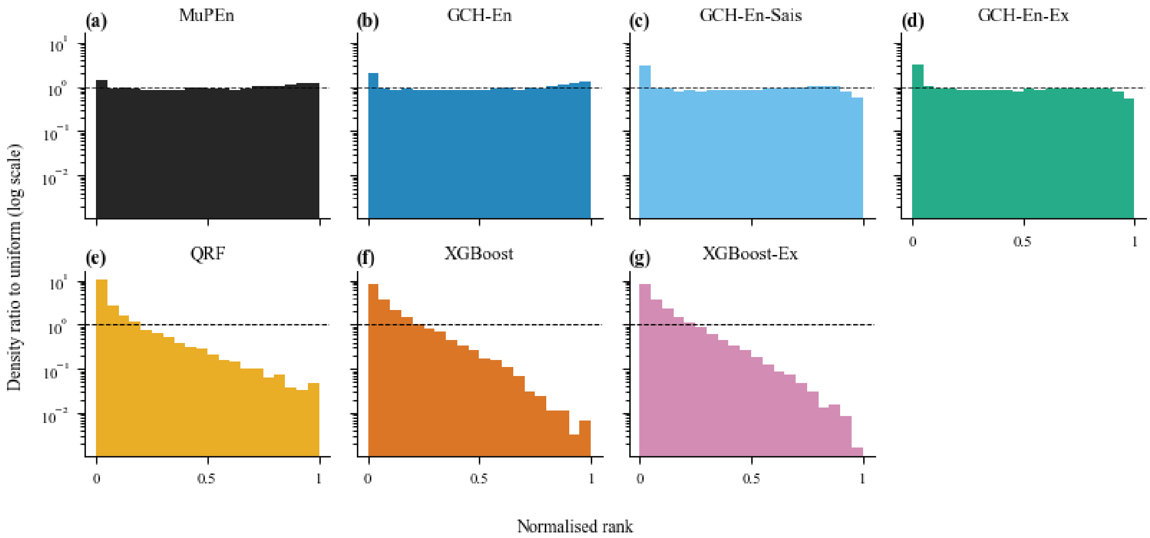

Fig. 4: Band-Depth (DB2) rank histograms, averaged over the six lead times and shown as the density ratio to uniformity on a log scale

5.4 Summary

Tab.2 collects the metrics per model. Read across a row, the pattern is a crossover rather than a ranking: the XGBoost baselines win the averages on the nCRPS and stay within one per cent of the best copula on both multivariate scores, while losing every lead time beyond 2h on the Variogram-Score. The one column in which copulas never lose is joint calibration.

Tab. 2: Headline metrics (test year 2022, averaged over the six lead times)

Model	nCRPS (% mean obs.)	Energy-Score	Variogram-Score	DB2 index
GCH-En	22.50	**0.0830**	**0.0837**	**0.336**
GCH-En-Sais	23.48	0.0860	0.0872	0.393
GCH-En-Ex	23.76	0.0869	0.0886	0.392
QRF	25.52	0.0948	0.0996	1.239
XGBoost	22.26	0.0841	0.0883	1.475
XGBoost-Ex	**22.05**	0.0834	0.0868	1.475
MuPEn	34.61	0.1295	0.1393	0.279

6. Conclusion

The added-value of the multivariate probabilistic forecasting methods tested in this work is neither absolute nor uniform. It depends on the level at which performance is measured and on the lead time. The more dependence-sensitive the score, the earlier the copula family of models take the lead. On nCRPS, XGBoost baselines win at every lead time except wind power beyond 4h. On the Energy-Score, the crossover occurs at 6h. On the Variogram-Score it moves to 2h and the gap then widens with the horizon, reaching 9.5% at 12h. Averaged over lead times the two families (i.e. XGBoost and copulas) remain within one per cent of each other on both multivariate scores, so neither can be declared the better forecaster on scores alone.

Joint calibration is the exception and it is where the practical argument lies. The univariate ensembles fail it at every lead time. Dependence can safely be ignored for decisions that concern a single variable at a short lead time, where a well-tuned univariate model is both simpler and slightly better. Multivariate approach becomes necessary beyond about 2h and it becomes unavoidable whenever a decision combines load, PV and wind independently.

7. Acknowledgment

This work received support from the Fine4Cast project funded by the French state and managed by the Agence Nationale de la Recherche under France 2030 (grant number ANR-22-PETA-0008).

References

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016. doi: 10.1145/2939672.2939785.

Christian Genest and Anne-Catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007. doi: 10.1061/(ASCE)1084-0699(2007)12:4(347).

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007. doi: 10.1111/j.1467-9868.2007.00587.x.

Tilmann Gneiting, Larissa I. Stanberry, Eric P. Grimit, Leonhard Held, and Nicholas A. Johnson. Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *TEST*, 17(2):211–235, 2008. doi: 10.1007/s11749-008-0114-x.

Faranak Golestaneh, Hoay Beng Gooi, and Pierre Pinson. Generation and evaluation of space–time trajectories of photovoltaic power. *Applied Energy*, 176:80–91, 2016. doi: 10.1016/j.apenergy.2016.05.025.

Sabine Haas, Birgit Schachler, and Uwe Krien. windpowerlib — a python library to model wind power plants, 2019. URL `https://doi.org/10.5281/zenodo.3254810`.

Stephan Hemri, Denisa Lisniak, and Bastian Klein. Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resources Research*, 51(9):7436–7451, 2015. doi: 10.1002/2014WR016473.

Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob J. Hyndman. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, 32(3):896–913, 2016. doi: 10.1016/j.ijforecast.2016.02.001.

Rich H. Inman, Hugo T. C. Pedro, and Carlos F. M. Coimbra. Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6):535–576, 2013. doi: 10.1016/j.pecs.2013.06.002.

Philippe Lauret, Mathieu David, and Pierre Pinson. Verification of solar irradiance probabilistic forecasts. *Solar Energy*, 194:254–271, 2019. doi: 10.1016/j.solener.2019.10.041.

Nicolai Meinshausen. Quantile regression forests. *Journal of Machine Learning Research*, 7:983–999, 2006.

Annette Möller, Alex Lenkoski, and Thordis L. Thorarinsdottir. Multivariate probabilistic forecasting using bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, 139(673):982–991, 2013. doi: 10.1002/qj.2009.

Joakim Munkhammar. Recursive copula-augmented space-correlated temporal very-short term spatiotemporal probabilistic forecasting of the clear-sky index. *Solar Energy Advances*, 4:100063, 2024. doi: 10.1016/j.seja.2024.100063.

Roger B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer Science+Business Media, Inc., New York, NY, 2nd edition, 2006. doi: 10.1007/0-387-28678-0.

Chaoqiong Pan, Can Wang, Ziyan Zhao, Jinhao Wang, and Zhaohong Bie. A copula function based monte carlo simulation method of multivariate wind speed and pv power spatio-temporal series. *Energy Procedia*, 159:213–218, 2019. doi: 10.1016/j.egypro.2018.12.053.

Andrew Patton. Copula methods for forecasting multivariate time series. In *Handbook of Economic Forecasting*, volume 2, pages 899–960. Elsevier, 2013. doi: 10.1016/B978-0-444-62731-5.00016-6.

Pierre Pinson. Wind energy: Forecasting challenges for its operational management. *Statistical Science*, 28(4):564–585, 2013. doi: 10.1214/13-STS445.

Pierre Pinson and Robin Girard. Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy*, 96:12–20, 2012. doi: 10.1016/j.apenergy.2011.11.004.

Pierre Pinson, Henrik Madsen, Henrik Aa. Nielsen, George Papaefthymiou, and Bernd Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009. doi: 10.1002/we.284.

Roman Schefzik, Thordis L. Thorarinsdottir, and Tilmann Gneiting. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, 28(4):616–640, 2013. doi: 10.1214/13-STS443.

Michael Scheuerer and Thomas M. Hamill. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321–1334, 2015. doi: 10.1175/MWR-D-14-00269.1.

Thordis L. Thorarinsdottir and Nina Schuhen. Verification: Assessment of calibration and accuracy. In *Statistical Postprocessing of Ensemble Forecasts*, pages 155–186. Elsevier, 2018. doi: 10.1016/B978-0-12-812372-0.00006-6.

Dennis van der Meer. A benchmark for multivariate probabilistic solar irradiance forecasts. *Solar Energy*, 225:286–296, 2021a. doi: 10.1016/j.solener.2021.07.032.

Dennis van der Meer. A benchmark for multivariate probabilistic solar irradiance forecasts. *Solar Energy*, 225:286–296, 2021b. doi: 10.1016/j.solener.2021.07.010.

Dennis Van Der Meer, Dazhi Yang, Joakim Widén, and Joakim Munkhammar. Clear-sky index space-time trajectories from probabilistic solar forecasts: Comparing promising copulas. *Journal of Renewable and Sustainable Energy*, 12(2):026102, 2020. doi: 10.1063/1.5140604.

David L. Woodruff, Julio Deride, Andrea Staid, Jean-Paul Watson, Gerrit Slevogt, and César Silva-Monroy. Constructing probabilistic scenarios for wide-area solar power generation. *Solar Energy*, 160:153–167, 2018. doi: 10.1016/j.solener.2017.11.067.

Dazhi Yang. A universal benchmarking method for probabilistic solar irradiance forecasting. *Solar Energy*, 184:410–416, 2019. doi: 10.1016/j.solener.2019.04.018.