Scenario generation and probabilistic forecasting analysis of spatio-temporal wind speed series with multivariate autoregressive volatility models

Scenario generation and probabilistic forecasting analysis of spatio-temporal wind speed series with multivariate autoregressive volatility models

Carlo Lucheroni [*] John Boland [†] Costantino Ragno [‡]

February 3, 2019

[*]School of Science and Technologies, University of Camerino, Via M. delle Carceri 9, 62032, Camerino (MC), Italy. Email: carlo.lucheroni@unicam.it. Corresponding author.

[†]School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes Boulevard, Mawson Lakes, SA 5095, Australia. Email: john.boland@unisa.edu.

[‡]School of Science and Technologies, University of Camerino, Via M. delle Carceri 9, 62032, Camerino (MC), Italy. Email: costantino.ragno@unicam.it.

Abstract

This paper evaluates the application of a family of VAR-mGARCH (Vector Auto-Regressive with multivariate AutoRegressive Conditional Heteroskedasticity) volatility models to the problem of modeling (i.e. generating correct in-sample scenarios) and forecasting in a probabilistic way three univariate but mutually dependent wind speed hourly series. The evaluation starts from the consideration that optimal modeling and optimal forecasting are not necessarily attained by the same models, and that they are assessed in different ways. The proposed VAR-mGARCH family, originated in Finance, consists of the VAR-BEKK (VAR - Baba, Engle, Kraft and Kroner) model and the VAR-DCC (VAR - Dynamic Conditional Correlation) model, the latter seen as the simplified and more scalable version of the former. These models were never used in wind speed studies before. Both models will be used with Gaussian innovations, and the VAR-DCC model will be also used with Student's t innovations. These models, which are able to dynamically couple volatilities, are compared to two benchmark model sets, the first set consisting of three independent univariate AR-GARCH (i.e. non-vector) models, the second set consisting of an individual VAR model without the GARCH sector. These benchmark models cannot dynamically couple volatilities. Both benchmark model sets will be used with Gaussian innovations. The time series used for the evaluation are taken from three North-American wind metering sites located at few tens of kilometers from each other. In order to highlight the usefulness of choosing autoregressive coupled volatility schemes for modeling and forecasting spatially and temporally varying wind speed data, a controlled variance electricity generation portfolio example of Markowitz type, typical of Energy Finance, is also discussed at some length. In all, it is concluded that the VAR-DCC model with Student's t innovations is the most balanced choice for this data set when the three goals of modeling, forecasting and scalability in the number of considered sites are taken into account at once.

JEL classification: C33; C53; C58

Keywords: wind dynamics, stochastic dynamical systems, forecasting models; renewable energy finance, risk assessment and management

1 Introduction

In electricity markets, integration into generation of wind as a renewable source is as much supported by policy makers as it is a challenge for power systems management and price systems stability. This is due to both random intermittency of this kind of production and associated limited modeling and forecasting ability of current approaches. These two integration issues, the first mainly linked to engineering, the second to finance, are basically due to the fact that at a given production spot the weather variability can induce strongly nonlinear dynamic variations of wind intensity, autocorrelated in time in a very complicated way with many time scales involved. In order to alleviate these problems, a lot of effort is put not only into the development of hardware and infrastructure for buffering and smoothing out production, but also in accurately modeling and forecasting wind speed dynamics in the short term from empirical data [1]. It is not easy to accurately model or forecast wind random intermittency at just one site, but the modeling issue becomes even more difficult when space and time dimensions get coupled. This can be the case of an array of wind metering sites scattered across a territory in such a way that wind cannot blow on all sites at once with same intensity and direction. In this case using a representative site for the whole array could be not enough. One of the consequences would be that if the time series from a site is modeled independently of all other sites' time series, a lot of useful information already present at the other sites at preceding times would be lost. This happens for example when a site is hit by a gust of wind which starts far away from it, and that gust has hit already other sites. In this case the individual site model misses available and useful forecasting information present at the other sites. Besides Ref. [1], also Ref. [2] can be useful as a literature review on short term wind forecasting at an individual site. In order to fully exploit wind speed time series information in an efficient way, spatio-temporal approaches are thus needed, in which the dynamics at all space locations is modeled at once and not independently. Such models can be fully probabilistic [3], or discriminative and quasi-linear [4], or nonlinear [5]. It should be noticed that, as a consequence of its spatio-temporal nature, each of these models has to find its own way to cope with upscaling, i.e. with what happens

to the model when the number of sites becomes large. A common strategy to take on this problem is that of extracting the most relevant degrees of freedom from the spatio-temporal wind series and work only with them, like in Ref. [6] where a least absolute shrinkage and selection operator (LASSO) technique typical of linear systems is used, or Ref. [7] in which a compressive sensing approach is applied. Once a robust, reliable and scalable model is obtained for the spatio-temporal wind series, this model can be in turn used by system controllers for safer dispatching and by market participants for more efficient production scheduling and market pricing. The aim of this paper is thus to present a modeling frame for spatio-temporal wind speed series which extends current literature in this respect. In the following, we will evaluate a selection of multivariate econometric models on wind speed data, and quantitatively assess their performance in-sample (i.e. for scenario generation, by sampling from estimated models) and out-of-sample (i.e. probabilistic forecasting) abilities. Before our work, as far as we know, these spatio-temporal volatility-coupled models were never used in wind speed modeling literature. In describing these models, our paper will take into account interlinked aspects from finance, engineering and environmental science in an interdisciplinary way. As we will try to show, this evaluation is indeed complex and multifaceted, especially if one wants to take into account the fact that data modeling is not the same as forecasting, and that models that can do good forecasting not necessarily do good data modeling. This fact is often overlooked in related wind literature.

Usually, modeling approaches to wind speed data build from the empirical evidence that, when a univariate wind speed data series taken at a given time interval and coming from an individual metering spot is unconditionally aggregated, the resulting static (i.e. unconditional) univariate distribution of wind speeds is well approximated by a Weibull or a Burr distribution. But when the same univariate series is studied dynamically, its behavior turns out to be complex, showing steady wind speed periods, recurring bursts of air with sharp speed peaks, and in some cases long periods of zero wind speed. In order to model this strongly nonlinear behavior, a range of dynamical models can be tried, from physical models [8] to purely statistical models, which include advanced machine learning models [2]. A drawback of machine learning models is that in most cases they depend on

meta-parameters, like the number of layers and neurons in the case of neural networks, or regularization coefficients in the case of support vector machines, to which it is difficult to give an interpretation. In contrast, simpler statistical approaches can be much more transparent. If hence a simple statistical dynamical approach is chosen for this reason, one could start by considering a single-site model, such as a simple univariate linear autoregression (AR) model. An AR model is good enough for replicating some basic features of the time series, like a certain random variability, and for doing basic forecasting. A little more sophisticated approach can be that of a univariate AR-GARCH (AR - Generalized AutoRegressive Conditionally Heteroschedastic) model [9], which allows one to indirectly include some mild nonlinearity in the otherwise linear AR model through a time-varying variance of its innovations. This improvement allows for example for volatility clustering [10], a nonlinear effect obtained in a very simple way. When a model of this type is estimated on data and fits them well, it can in turn be sampled and used to generate synthetic series with statistical features similar to the original data set, volatility clustering included. This in-sample operation can be called data modeling, or modeling for scenario generation, and the obtained model can be used to replace original data. Synthetic series from data models can be used for assessing wind energy production risk in some specific area, or for assessing trade strategies, or to price financial contracts on electricity production using Monte Carlo techniques. The other use of models, alternative or concurrent to scenario generation, is forecasting. Point or probabilistic forecasts useful for production or financial dynamic risk management can be made by exploiting the simple but well understood conditional point forecasting ability of linear autoregressions and the conditional variance forecasting ability of the GARCH sector. For these reasons, in the current literature this AR-GARCH setting is considered a benchmark model, and more nonlinear (but still univariate) extensions of it are being explored, like for example in Ref.[11] where a switching GARCH model is studied. It is important here to make explicit the fact that although AR-GARCH modeling improves modeling and forecasting of volatility over simple AR modeling, it does not improve AR point forecasting itself. Different research communities like engineering and finance can be interested in different properties of these models.

Notice that not necessarily models that are good at forecasting are also good at data modeling and scenario generation, nor the converse is necessarily true. This can be for example gathered from the machine learning generative vs. discriminative modeling debate, in which it is discussed how generative (i.e. fully probabilistic) models, often very good at detecting and reproducing data patterns, are not always good enough for point forecasting, a mainly discriminative task [12]. This aspect is rarely made explicit in most of the wind literature, where often a model is assessed either for its forecasting ability or its data generation ability only. Overlooking this aspect can have implications when the model is put at work. For example, risk assessment and energy derivative pricing depend often on scenario generation properties, whereas dynamic risk control depends much more on forecasting properties. It would thus be better not to use a model good at forecasting but weak at data modeling for risk assessment. Thus, when models are evaluated on data, it is also important to tell in-sample (i.e. data modeling and scenario generation) from out-of-sample (i.e. forecasting) properties, in order to be aware of consequences on possible applications of a studied modeling choice, especially in a field of interest for different disciplines like wind modeling. For sure, good general purpose wind models, and some special purpose models, need to have both good data modeling and good forecasting properties.

When wind speed measurements come from spatial networks of meters but are studied statically and in an unconditional way, from a scenario generation point of view they can be thought as vector samples sampled from the multivariate probabilistic static joint distribution of a wind speed vector, where the marginals individually model site speed distributions. The marginals of this joint distribution will have to match Weibull or Burr type distributions. These marginals should also be in general different from each other. Notably, in this static case the distribution with all its features is itself the spatial wind model, like in the case of static copula models. In contrast, dynamic probabilistic models for spatial networks require in general far more sophisticated modeling machinery than static models (see for example Ref. [4] and references therein). For example, at a given time an important piece of information which has to be in some way included in the machinery is the instant cross-correlation among sites. In addition, for each individual site the on-spot autocorrela-

tion along time is as important as well. Finally, a good multivariate model must in principle be able to encapsulate dependency between a site in place A at time t_A and a site in place B at (possibly different) time t_B, but still generate meaningful static marginals.

An interesting approach which is able to do all that in a transparent way is multivariate AR-GARCH modeling. This approach comes mainly from financial econometrics literature. Such a multivariate dynamic modeling frame, preliminarily discussed in [13] in relation to wind speed time series modeling, seems not having ever been considered in the wind literature, nor having ever been investigated in that context in any theoretical and numerical depth. This quasi-linear approach is partly discriminative and partly probabilistic, because it dynamically models some parameters of its conditional distributions, i.e. the covariance matrix, using lagged information. In this approach, a Vector AutoRegression (VAR) sector models the lagged interactions among sites (like in Ref. [4]), with the trade-off of making the structure of innovations, i.e. the VAR residuals, cross-dependent. In a standard VAR approach (i.e. without the multivariate GARCH sector) a static covariance matrix of the residuals is obtained. This matrix can be used in turn as an error forecast in point forecasting. This approach can hence be improved by modeling the residuals by a suitably chosen multivariate GARCH (in short, mGARCH) sector, or layer. A further modeling freedom of this approach is the possibility of deciding the type of the innovation distributions, choosing usually between the two well known elliptic cases of Gaussian and Student's t innovations. Student's t innovations should be indeed explored in the wind context because they have thicker tails than Gaussian distributions, so that in principle they can better deal with extreme speed events, individual or coincident. mGARCH models come in many different flavors and with different coupling properties for the residuals, described in detail in standard textbooks like Refs. [14], [15] and [16]. Among them the BEKK model [17] by Engle and Kroner[1] and the Dynamic Conditional Correlation model (DCC) by Engle [18, 19] and Tse et al. [20] are particularily well suited to the wind spatio-temporal modeling

[1]This econometric model is called BEKK because the paper in which it was introduced, written by Engle and Kroner only, was based on unpublished joint work by Y. Baba, R. Engle, D. Kraft and K. Kroner. The two authors acknowledged that in a note, and have been calling the model BEKK instead of EK since then.

problem. Both BEKK and DCC schemes model in a dynamic way the covariance matrix of the residuals, which includes volatilities (variances) on the diagonal and co-volatilities (covariances) off the diagonal. BEKK and DCC differ in the way they take into account the interactions among sites, in the number of parameters, and in the way they scale against the number of sites in terms of speed of computation and stability of the numerical solution. Because the BEKK model allows for direct interactions among site volatilities, it can model so-called volatility spillovers, i.e. situations in which volatility at one lag is transferred from one site to another at the next lag. This property is very attractive, but it implies bad scaling properties for the model, so that the BEKK model is not suitable for datasets with more than few sites. The DCC model keeps individual site volatilities separated, i.e. acts as a set of individual independent GARCH models, which makes it much better in terms of scaling (in Ref. [18] it was applied to 50 financial assets and still easily estimated on year 2001 hardware), but it is not able of modeling volatility spillovers. This notwithstanding, the DCC model can extract from the univariate GARCH components some extra information about the whole correlation matrix, i.e. about the off-diagonal components of the covariance matrix. The DCC model can thus be seen as a sort of well scaling limit of the BEKK model. All in all, these mGARCH models have the same limitations and advantages of univariate models, i.e. they add nonlinearity and, even though they don't improve point forecasting over VAR modeling, they can be used for dynamic volatility modeling and forecasting. But in addition, more importantly, they return dynamic co-volatilites and their forecasting, a kind of extra information which can be very interesting for renewable energy portfolio problems [21]. Models like GARCH-copula models, which can be considered even closer to a fully probabilistic approach, could also be tried, but from the preliminary analysis carried on in Ref. [13] it seems that they don't add very much to what more simple and standard mGARCH models like a DCC can already do. Models like switching mGARCH models [22] could on the contrary be very interesting in this context, but they are much more difficult to estimate.

The bi-sector structure of the models chosen (i.e. the VAR and mGARCH sectors) and the fact that the models could be seen in some way as nested (in the bottom-to-

top four levels hierarchy AR-GARCH, VAR, VAR-DCC, VAR-BEKK), are convenient for understanding which part of modeled information goes in where, that is if it goes into either the scenario generation or the forecasting model ability, or into both at once. In the paper, special care will be taken to make clear this aspect. The four AR-GARCH, VAR, VAR-DCC, VAR-BEKK models with Gaussian innovations plus a VAR-DCC model with Student's t innovations, called VAR-t-DCC, will be estimated on hourly wind speed vectors from three suitably chosen metering sites in the US, and compared among each other in their modeling and probabilistic forecasting features. The models will be estimated with the help of the `rmgarch` package [23] by A. Ghalanos and the `MFE` Matlab Toolbox [24] by K. Sheppard. The evaluation of probabilistic forecasting ability will be made with the help of the `scoringRules` package [25].

The plan of the paper is the following. After this Introduction, Sec. 2 illustrates the wind speed dataset used and the reasons for the choice of this specific dataset. Sec. 3 discusses at some length the two benchmark models. These are the AR-GARCH model, also called univariate model in the following, which actually consists of three uncoupled AR-GARCH models, and the VAR model. The discussion will pay special attention to their behavior in relation to scatterplot and autocorrelation modeling, i.e. in-sample properties, and the point and probabilistic forecasting features of the two models, i.e. out-of-sample proprieties. Sec. 4 reviews the mathematical features of the BEKK and the DCC models, and presents the reasons why these two models can be in some sense considered as part of a hierarchy. Sec. 3 and Sec. 4 are somewhat preparatory for Sec. 5, which is dedicated to understanding which part of the hierarchy is useful to which specific aspect of spatio-temporal wind speed series modeling, either scenario generation or forecasting, and in which sense one model of the hierarchy can be considered better than another on this specific dataset and in more general cases. Sec. 6 discusses by means of an example the possibility of using this kind of financial model for controlling the variability of the power output of one fictional wind plant formed by wind generators placed in the three studied sites. This Section, which makes use of Markowitz porfolio theory, tries also to put together all understanding collected in the previous Sections. Sec. 7 concludes. Auxiliary information

about the Student's t distributions used in the models, and about the battery of metrics used in the paper to assess probabilistic forecasting properties, are collected in Appendix A and Appendix B, respectively.

2 Wind data

The wind scalar-speed data used in the paper are taken from the US National Renewable Energy Laboratory (NREL) web site [26]. They are hourly data for the full years 2012 and 2013 expressed in meters per second (m/s) for three metering sites placed at a 10 meters height in the Denver (Colorado, USA) area. The sites are called respectively Solar TAC (site A), Lowry Range Solar Station (site B), and M2 Tower (site C). The distances among the sites are: 15 km between A and B, 55 km between A and C, 62 km between B and C. These three sites are shown on the Denver area map of Fig. 1, where a scale is marked in the bottom left corner. The reason for this choice of sites is the following. A gentle breeze of the Beaufort wind speed scale corresponds to $3.4 - 5.5$ m/s. This makes $12 - 19$ km per hour, and it means that, during a gentle breeze, the distance between sites A and B can be covered in one hour by an air disturbance. Hence, these two sites can be coupled at one lag of the modeling series already whereas site C remains uncoupled. On the contrary, fast winds can 'instantly' (i.e. with lag zero) couple all three sites. This spatial geometry, even though small, is thus suitable to challenge the full cross-sectional behavior of mGARCH models already at few lags. Yet, it should be noted that too few lags cannot fully capture on-site autocorrelations, especially effects implicitly due to the periodic 24-hours temperature and insolation variation. Thus, in the next Sections, all models will be studied using 24 AR or VAR lags. From a purely econometric point of view, the number of lags to be included in each model studied could also be estimated by the Bayes Information Criterion (BIC). This would return different values for each of the models studied. In order to focus the comparison on the structural properties of the models only, no such optimization will be discussed, and the more geophysically and meteorologically motivated choice of using 24 lags will be considered satisfactory. For paper length reasons the very interesting case of a

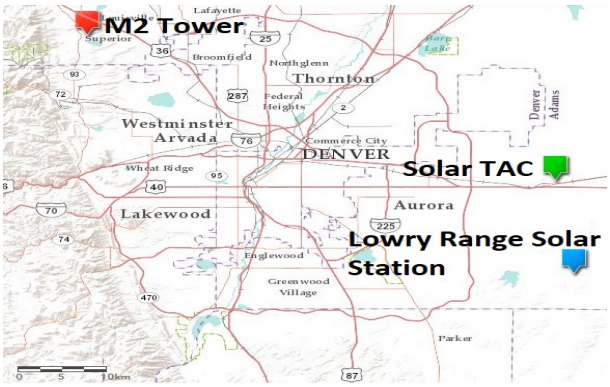

Figure 1: Wind site locations in the Denver (CO, USA) area. Site A (Solar TAC) in green, site B (Lowry Range Solar Station) in blue, site C (M2 Tower) in purple. On the bottom left corner, the 10 km scale mark.

number of sites larger than three won't be openly discussed either, even though this issue will be always maintained in the discussion background.

As an example of the time series studied, in Fig. 2 wind speeds in the time period between Feb. 11, 2013, and Feb. 21, 2013 are shown. Dynamic cross-dependency and its breakdown is visible in this Figure, where in some periods the speeds u_t^A, u_t^B and u_t^C seem to move in the same way (for example before Feb. 13), in some others they clearly break free (for example on Feb. 16).

Before being estimated with the chosen models, the three wind speed series, which have positive support, must be transformed to a full line support, since standard AR, VAR and mGARCH models use errors with full line support. In this paper the invertible Box-Cox mapping

$$x^i = \frac{(u^i)^\gamma - 1}{\gamma} \tag{1}$$

will be used. Notice that if in Eq. (1) γ is set to 0, one gets a logarithm. Thus x_t^i, the value of x^i in time, will be called 'bcspeed', a name chosen in analogy to the term logspeed. More importantly, if $\gamma \neq 0$, this transformation has no problems with zero wind speeds. The common value $\gamma = 0.1$, not far from zero, will be thus used used for the three series. Data for estimation and in-sample analysis will span one year from January 2012 to December 2012, and data for forecast analysis will span one year from January 2013 to December

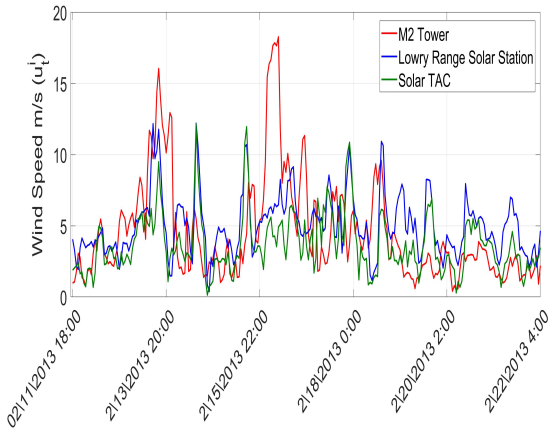

Figure 2: Hourly wind speed u_t^i in m/s for the three sites in the period between Feb. 11, 2013, and Feb. 21, 2013. On the abscissa, time in hours and information about the time of the year.

		M2 Tower	Lowry Range	Solar TAC
2012	mean	3.85	5.16	3.74
	variance	8.61	5.47	4.72
2013	mean	3.71	4.29	3.47
	variance	7.23	4.78	3.9

Table 1: Unconditional yearly mean (in m/s) and variance for the three series, for year 2012 (training and in-sample data) and 2013 (testing and out-of-sample data). NREL data.

2013. This choice allows preserving the integrity of the geophysical yearly cycle in both estimation and testing phases. Unconditional yearly mean and variance for the three series are reported in Tab.1, for the two years separately.

3 AR-GARCH and VAR modeling: the benchmark models

In Subs. 3.1 of this Section some basic properties of univariate AR-GARCH modeling will be recalled, in order to introduce the first benchmark model consisting of a set of three independent univariate AR-GARCH models with Gaussian innovations, which will be col-

lectively called either AR-GARCH model or univariate model. Next, in Subs. 3.2, the second benchmark model, a Gaussian VAR model without a GARCH sector will be introduced, which will be called VAR model. These two benchmark models will be estimated on the first year (year 2012) of the dataset, sampled, and discussed in their in-sample data modeling ability. Scatterplots (i.e. bivariate empirical marginals) will be used to study their cross-sectional behavior, and autocorrelation will be used to study their longitudinal behavior. Once studied in-sample, the forecasting ability of these models will be assessed [27], [28] on the out-of-sample dataset (year 2013). Plots showing how the realized values of bcspeeds compare with error intervals centered on the forecast values, and a self-consistent collection of mutually related metrics for point and probabilistic forecasting will be used. For reference, Appendix A contains formulas related to the Student's distributions used, and Appendix B contains the definition and a discussion of the used metrics.

3.1 Univariate AR-GARCH and AR-t-GARCH models

In the case of a single site, the sample bcspeeds $\hat{x}_t$ $(t = 1, \dots T)$ can be modeled with a univariate Gaussian AR(N)-GARCH(P,Q) model, assuming for the bcspeeds the N-lags AR form

$$x_{t+1} = \mu_t + \epsilon_{t+1} \tag{2}$$

where

$$\mu_t = a_0 + \sum_{j=1}^{N} a_j x_{t-j+1}, \tag{3}$$

ϵ_t are independent $\forall t$ and $E_t[\epsilon_{t+1}] = 0$. The bcspeed $l = 1$ steps ahead point forecast at t is obtained by taking the expectation of Eq. (2) at t, which gives

$$x_t^{f,1} = E_t[x_{t+1}] = \mu_t. \tag{4}$$

Once the AR sector is estimated on the bcspeed series $\{\hat{x}_t\}$, an estimated innovations (i.e. residuals) series $\{\hat{\epsilon}_t\}$ becomes available. Innovations $\{\epsilon_t\}$ are in turn modeled in the GARCH sector by the factorized form

$$\epsilon_{t+1} = \sqrt{h_{t+1}}\, z_{t+1} \tag{5}$$

	M2 Tower	Lowry Range	Solar TAC
ω	0.052215	0.036656	0.075434
α	0.087699	0.272771	0.172117
β	0.634015	0.341171	0.455921

Table 2: Parameters of the GARCH sector of the AR-GARCH model according to Eq. (6) with $P = Q = 1$, estimated on the 2012 NREL data.

with $h_{t+1} > 0 \ \forall t$. In Gaussian AR-GARCH models the z_t of Eq. (5) are assumed to be i.i.d. standardized (i.e. Normal) Gaussian stochastic variables, whereas in AR-t-GARCH models they are standardized Student's t stochastic variables, each distributed as p_{ust} of Eq. (A.2) with parameter ν. Being standardized, in both cases these variables have conditional mean $E_t[z_{t+1}] = 0$ and conditional variance $\mathrm{var}_t[z_{t+1}] = 1$. This implies that h_{t+1} in Eq. (5) represents the conditional (at t) variance of x_{t+1} and of the innovations, because $E_t[(\epsilon_{t+1})^2] = h_{t+1}$. In GARCH(P,Q) models the time $t+1$ variances h_{t+1} are made dependent on previous variances and innovations as

$$h_{t+1} = \omega + \sum_{j=1}^{P} \alpha_j \epsilon_{t-j+1}^2 + \sum_{i=1}^{Q} \beta_i h_{t-i+1}, \tag{6}$$

which gives them dynamics, and where stationarity requires that $\sum_{j=1}^{P} \alpha_j + \sum_{i=1}^{Q} \beta_i < 1$. In this paper it will be assumed $P = Q = 1$ for all GARCH and mGARCH models, hence, from now on, the GARCH sector of each of the component univariate GARCH models will be characterized by the parameters set Θ, in which besides $\{\omega, \alpha, \beta\}$ there can be also ν. This piece of information about t-GARCH models will be used later when discussing the VAR-t-DCC model. When α or β are large the variance chain has memory, and periods can form in which it keeps a large value before reverting to a smaller asymptotic constant ω. This is the mechanism behind volatility clustering. Tab. 2 shows the GARCH parameters estimated on the in-sample dataset. The β coefficient is rather high for all three sites, indicating the necessity of memory and a consistent volatility clustering effect in the data.

The main feature of GARCH forecasting analysis depends on the fact that even though the h_t chain in Eq. (6) is stochastic (it depends on past innovations), at time t the value

h_{t+1} is already available once h_t is known. Thus, the $l = 1$ steps ahead bcspeed variance forecast at t is

$$h_t^{f,1} = h_{t+1} = \omega + \alpha\epsilon_t^2 + \beta h_t = \omega + \alpha(x_t - \mu_{t-1})^2 + \beta h_t, \tag{7}$$

where h_t, even though unobservable, can be computed recursively.

From a phenomenological point of view, an AR-GARCH model allows one to add non-linearity to an otherwise linear AR model in a conveniently simple way, just working on the residuals sector. This means that, when synthetically generated series are obtained from the estimated model by sampling, they display volatility clustering and other interesting features in their unconditional distribution. In the wind speed case, after a sampled series (back-transformed with the Box-Cox inverse) is obtained from an estimated univariate AR(24)-GARCH(1,1) model, the shape of the histogram that unconditionally (i.e. statically) collects these sampled values (after Box-Cox inversion), looks rather satisfactorily like a Weibull distribution

$$p_W(u; a_1, a_2) = \frac{a_2}{a_1}\left(\frac{u}{a_1}\right)^{a_2-1} e^{-\left(\frac{u}{a_1}\right)^{a_2}} \tag{8}$$

or much better like a Burr type XII distribution

$$p_B(u; c_1, c_2, c_3) = \frac{\frac{c_1 c_2}{c_3}\left(\frac{u}{c_3}\right)^{c_2-1}}{\left(1 + \left(\frac{u}{c_3}\right)^{c_2}\right)^{c_1+1}} \tag{9}$$

(where $u, c_1, c_2, c_3 > 0$). This is shown in Fig. 3 for one of the three sites, Solar TAC (the histograms for the three sites all display the same features). Realized data for year 2012 are shown in the left panel, and synthetic data obtained from an AR(24)-GARCH(1,1) univariate model estimated on year 2012 are shown in the center panel. The model can thus capture a basic stylized fact of this kind of data. Dynamic information on the sequence of bcspeeds captured by the model is shown in Fig. 4 where the autocorrelation functions (ACFs) for the three sites, on measured data and on sampled data obtained from individually estimated models are displayed. The shapes are very well reproduced, and the bump at $N = 24$ lag captures the day/night cycle. The two sites closest to each other share the same shape, as

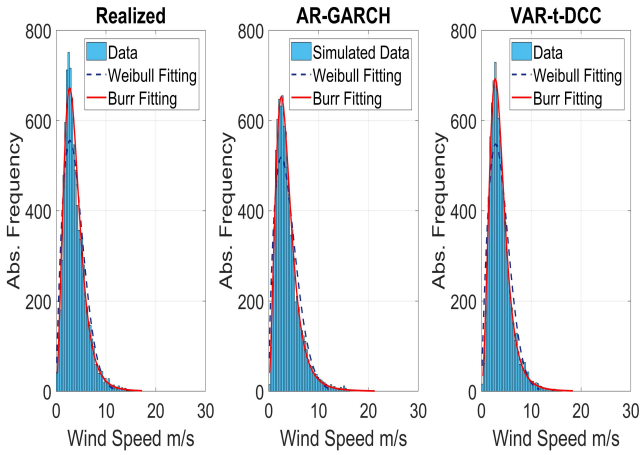

Figure 3: Histograms (unconditional empirical distributions) of one year of the wind speed realized series of the Solar TAC site (l.h.s. panel), and sampled from the univariate AR(24)-GARCH(1,1) (center) and the VAR(24)-t-DCC (r.h.s.) models both estimated on year 2012 data. Superimposed, fitted Burr and Weibull distributions. Fitted Burr values: realized, $c_1 = 2.12$, $c_2 = 2.44$, $c_3 = 4.90$; AR-GARCH, $c_1 = 1.70$, $c_2 = 2.42$, $c_3 = 4.30$; VAR-t-DCC, $c_1 = 1.68$, $c_2 = 2.70$, $c_3 = 4.32$.

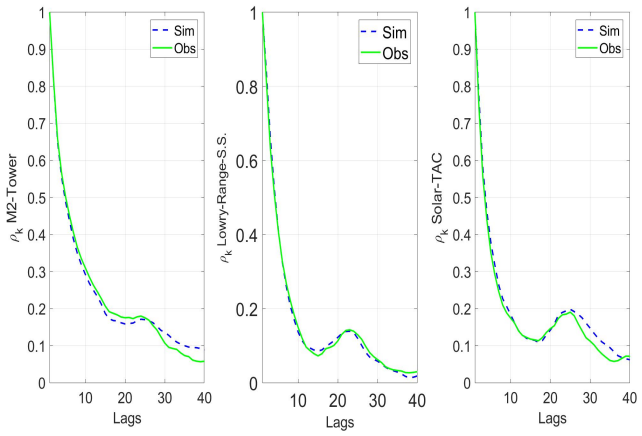

Figure 4: ACFs from original data and synthetic data obtained from univariate AR(24)-GARCH(1,1) models for the three sites. Lags in hours.

it is expected from sites subject to very similar weather conditions. Estimated univariate models can also be used to generate bidimensional scatterplots that gather outcomes of pairs of wind speeds for a pair of given sites, i.e. bivariate marginals. Univariate models can be very good at reproducing ACFs, but grossly miss site-to-site dependency features of observed data. This can be seen in Fig. 5, where both bcspeed data values and synthetic values sampled from the two independent univariate models for the Solar TAC - Lowry Range pair, the two closest sites of the set, is shown. Since they are close they should correlate, but the sampled series don't. This scenario generation aspect will be discussed again at the end of the the next Subsection on multivariate models, in which case dynamic information will include cross-site correlation.

Using the out-of-sample data, point forecasts $x_t^{f,1}$ can be computed by means of Eq. (4). A quantitative assessment of point forecasting ability of the univariate model comes from Tab. 3, where the Mean Square Error (MSE), the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) of $x_t^{f,1}$ against realized values $\hat{x}_{t+1}$ is computed for the full year 2013 for each of the three sites and for all models studied. As it should, RMSE $\geq$ MAE always. The first row of Tab. 3, which refers to the univariate model, shows that these errors are low, even though not the same for all the sites. MAE values should

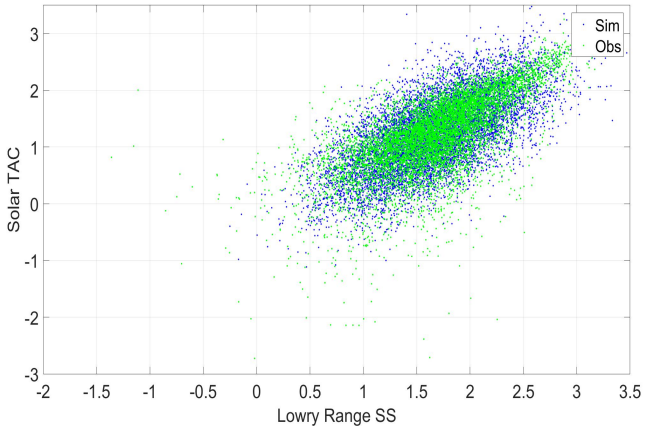

Figure 5: Univariate models: Solar TAC - Lowry Range wind speeds scatterplot. Axes in bcspeed units. Observed (lighter, green dots) and synthetic (darker, blue dots) data.

be compared to the corresponding Continuous Ranked Probablity Loss (CRPL) [33] values of the AR-GARCH row of Tab. 4. As discussed in Appendix B, the CRPL metric is a loss measure, like MSE and MAE, which means that the lower it is the better it is. The CRPL metric of Eq. (B.2) can be used in two ways. It can be used to compare different models among each other, ranking them from the best (which corresponds to the minimum loss) to the worst. This ranking function will be used in the following sections. CRPL can also be used to assess how probabilistic forecasting improves on point forecasting, because when applied to point forecasts CRPL reduces to MAE. Moreover, being a loss, CPRL $\leq$ MAE. In the case of Tab. 4, this comparison shows that in all three cases the addition of a dynamically changing sector improves on point forecasting seen as a simplified form of probabilistic forecasting. By means of Eq. (7), the forecast in t of the dynamically changing one-standard-deviation corridor around $x_t^{f,1}$ can be computed as $x_t^{f,1} \pm \sqrt{h_t^{f,1}}$, to be compared with $\hat{x}_{t+1}$. Such a relationship is shown in Fig. 6, where the result for about two weeks for the M2 Tower site is displayed. From this figure the good forecasting ability of the univariate model is made pictorially evident, because the realized value apparently almost never exits its forecast corridor. The same feature appears in longer time series and for all three sites. A simple way to assess this feature in a more quantitative way is the Ranked

	M2 Tower			Lowry Range			Solar TAC		
	MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE	MAE
AR-	0.1793	0.4234	0.3246	0.0973	0.3119	0.2205	0.2493	0.4993	0.3358
VAR-	0.1806	0.4250	0.3219	0.0939	0.3064	0.2215	0.2326	0.4823	0.3458

Table 3: Comparison of the point forecasting ability of the models by MSE, RMSE and MAE. AR- stands for AR-GARCH, VAR- stands for VAR, VAR-DCC, VAR-BEKK, VAR-t-DCC.

	CRPL			EL
	M2 Tower	Lowry Range	Solar Tac	
AR-GARCH	0,2333	0,1627	0,2547	0,4441
VAR	0,2350	0,1631	0,2493	0,4386
VAR-BEKK	0,2345	0,1617	0,2474	0,4367
VAR-DCC	0,2344	0,1613	0,2471	0,4361
VAR-t-DCC	0,2352	0,1626	0,2489	0,4381

Table 4: Comparison of the probabilistic forecasting ability of the models by the univariate Continuous Ranked Probablity Loss (CRPL) and the multivariate Energy Loss (EL) metrics.

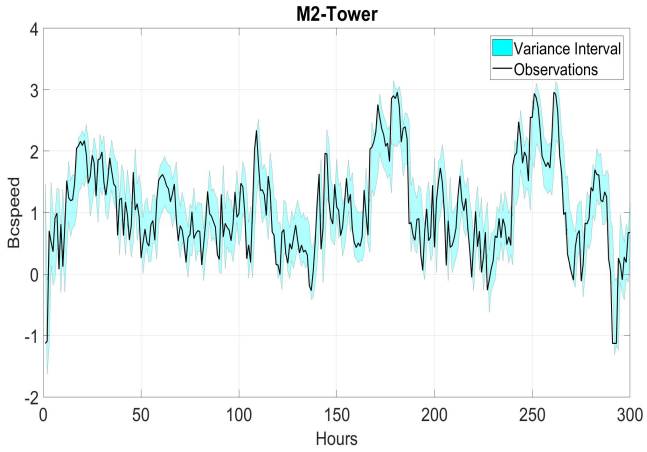

Figure 6: Univariate model: forecasting of M2 Tower 2014 series. Dynamically changing forecast one-standard-deviation corridor and realized data. Ordinate axis: bcspeed units.

Probability Plot (RPP) discussed at the end of Appendix B. In this plot the effect of a quantile-wise perfect forecast is compared to the empirically computed frequency RP_F^α that the point forecast hits within the forecast quantile, as indicated in Eq. (B.7). RP_F^α depends on the forecast parametric cumulative distribution F and on the quantile-defining probability mass α, and is a type of coverage measure. For a perfect forecast, $\mathrm{RP}_F^\alpha = \alpha$ should hold. In Fig. 7 RP_F^α is plot against α. In this Figure the bisector is superimposed to the plot, and corresponds to the perfect forecast. Deviation from the bisector are thus due to forecast errors. For the AR-GARCH model Fig. 7 shows that the AR-GARCH model is pretty good at forecasting in probability. Its S shape shows also that one side of the forecast distribution falls short of the true distribution, the missing probabilistic weight being recovered on the other side. This is to be expected, because wind speeds are not Gaussian, and their distribution remains somewhat skewed even after the Box-Cox mapping of Eq. (1). Elliptic (i.e. symmetric) distributions like the Gaussian and the Student's t distributions will not be ever able to exactly match bcspeeds. Thus Fig. 7 shows that the univariate model does its best, within its modeling limits. It can be anticipated here, as it can be seen directly from the Figure, that all studied models display the same behavior. In addition, it is clear that all of them perform more or less the same under the RPP, except for the VAR-t-DCC model,

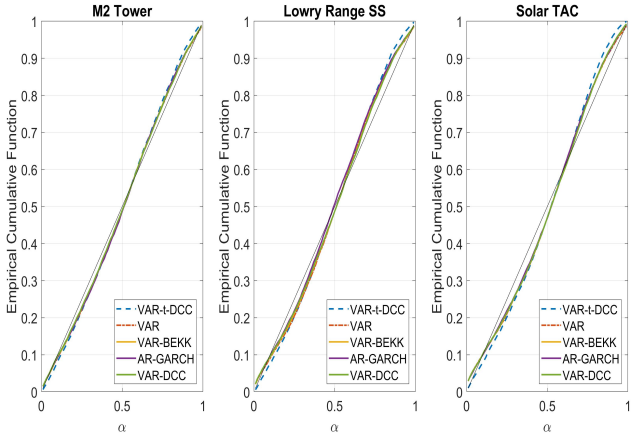

Figure 7: Comparison of the probabilistic forecasting ability of all discussed models by the Ranked Probability Plot. On the y-axis the empirical cumulative function RP_F^α corresponding to Eq. (B.7), as discussed in Appendix B. On the x-axis the probability mass α. The diagonal line (i.e. the bisector) indicates the locus of a perfect forecast.

which has a slightly worst performance. As discussed in Appendix B, the CRPL metric measures the deviation from the bisector. This is consistent with the results of Tab. 4, in which values of CRPL are very similar for all models, within the statistical variability.

Summing up, the AR-GARCH model spends all its parameters to fit and forecast very well each series individually, but grossly misses - as expected - any relationship among the three series. Some form of coupling is needed.

3.2 The VAR model

Intersite interaction is better expressed in a vector notation. For K sites and network data x_t^k $(k = 1, \ldots, K)$, at each time $t = 1, \ldots, T$ the data can be grouped in the T vectors $\vec{x}_t = \{x_t^1, \ldots, x_t^K\}$. The general N-lags VAR is

$$\vec{x}_{t+1} = \vec{\mu}_t + \vec{\epsilon}_{t+1} \tag{10}$$

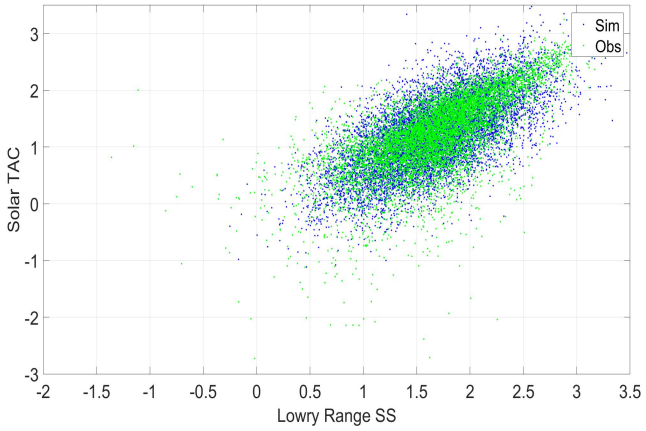

Figure 8: VAR model: Solar TAC - Lowry Range wind speeds scatterplot. Axes in bcspeed units. Observed (lighter, green dots) and synthetic (darker, blue dots) data.

where $\vec{\epsilon}_{t+1}$ are i.i.d. vector innovations, in this case with a joint Normal distribution. In Eq. (10) the vectors

$$\vec{\mu}_t = \vec{\Phi}_0 + \sum_{j=1}^{N} \bar{\Phi}_j \vec{x}_{t-j+1}, \tag{11}$$

contain $\bar{\Phi}_j$ unconstrained square matrices that couple bcspeeds at the same lagged instant j, and $\vec{\Phi}_0$ is a constant vector. For a large number of sites the matrices in the set $\{\bar{\Phi}_j\}$ can become very large and sparse, and special techniques like the LASSO or the technique developed in Ref. [5] have to be devised to take care of that. Yet, for the three sites of the NREL dataset this Gaussian VAR model can be still easily estimated with $N = 24$ as before, sampled and evaluated for in-sample modeling ability with scatterplots and autocorrelations. Its forecasting ability can be tested on the out-of-sample dataset using MSE, RMSE, MAE and RPP as in the univariate case. In addition, here a further metric called Pinball Loss will be introduced.

Data modeling properties are better than the univariate model. This is seen especially in scatterplots, as it is shown in Fig. 8, to be compared to Fig. 5. Only in the case of the VAR model the scatterplot of synthetically generated data is similar to the original data, which the univariate model grossly misses. Autocorrelations are slightly worse than in the univariate case, as it is shown in Fig. 9. This could be expected, because now the model

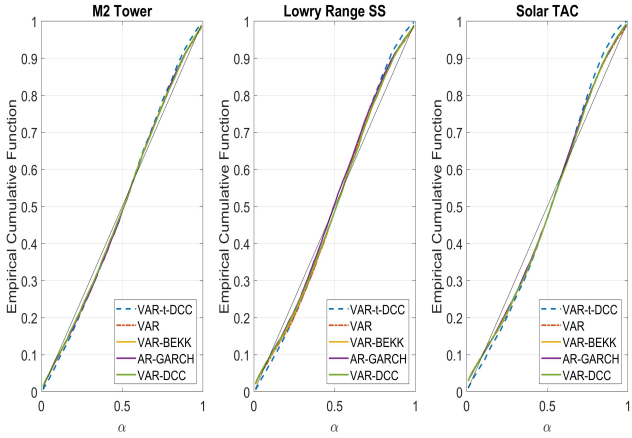

Figure 9: ACFs from original data and synthetic data obtained from a VAR(24) model for the three sites. Lags in hours.

must share its fitting ability between longitudinal (i.e. autocorrelation) and panel (i.e. cross-correlation) dimensions, with a trade-off. Moreover, notice that in VAR models nonlinearity is not included as it is in AR-GARCH models.

Point and probabilistic forecasting can be assessed by MSE, RMSE, MAE and RPP like in the case of the univariate model. MSE, RMSE and MAE results are shown in the second row of Tab. 3. These values are very similar to those of the univariate model. This means that the multivariate VAR structure does point forecast neither better nor worst than the univariate model. As to probabilistic forecasting, the RPP shows a behavior almost identical to that of the univariate model, with the same S shape. In principle, an interesting difference with the AR-GARCH case could be that in the VAR case the width of the corridor is constant, allowing for a lesser dynamic precision in forecast with respect to the AR-GARCH case, but this feature cannot be detected by CRPL or in the RPP. A slight superiority of the VAR model (and more in general of all models with a VAR sector) over the AR-GARCH model can be on the contrary detected using the Pinball Loss metric PL_F^α, defined in Eq. (B.5) and discussed in Appendix B. Even though a plot of PL_F^α vs. α cannot show how far from the perfect forecast a model is, PL_F^α is more accurate for ranking models in terms of quantile forecasting precision. Indeed, Fig. 10 shows that in the specific

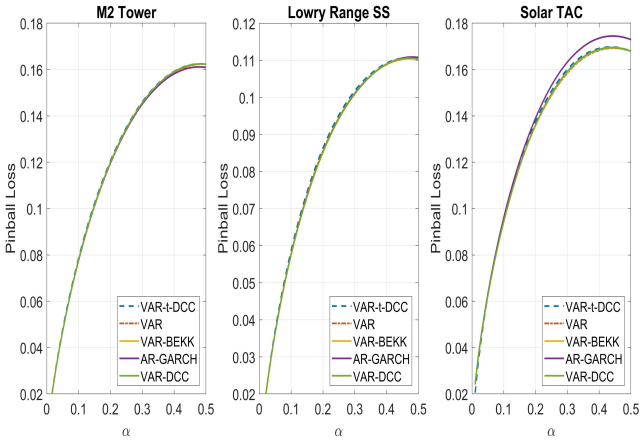

Figure 10: Comparison of the probabilistic forecasting ability of the models by the averaged Pinball Loss metric PL_F^α. On the $y-$axis PL_F^α corresponding to Eq. (B.5). On the $x-$axis the probabilistic mass α. The small values of α on the l.h.s. of each graph correspond to where it is easier to make few errors, which correspond to low losses.

case of the Solar TAC site (third panel from the left) the univariate model scores slightly worst than all other models which include a VAR structure.

Summing up, the multivariate VAR model is essentially neither better nor worst than the univariate model from a forecasting point of view. This weakness in forecast properties is yet compensated for by a better data modeling ability. This is a clear example of the difference that can be found between data modelling and forecasting ability of models. Choosing an AR-GARCH model over a VAR model only because the AR-GARCH scores the same (according to Tab. 3 and RPP) in forecasting but it is simpler, directly leads to weak data modeling. Moreover, all of this might suggest that the addition of a multivariate GARCH sector to the VAR sector could only improve the features of the model, which thus could include at once both longitudinal and panel required features. A VAR-mGARCH model, with more parameters, would be then naturally better than a multivariate AR-GARCH or a VAR model for both generation and forecasting purposes. Interestingly, neither this turns out to be necessarily true.

4 VAR-mGARCH models

In order to discuss the mGARCH sector, it is necessary to introduce some extra matrices and vectors. Let $\vec{z}_{t+1}$ be the standardized stochastic vectors of length N such that $E_t[\vec{z}_{t+1}] = 0$ and $\mathrm{cov}_t(\vec{z}_{t+1}) = \bar{1}$, where $\bar{1}$ is the unit matrix. The bar above a capital letter will henceforth indicate a matrix. Usually the distribution of $\vec{z}_t$ is multivariate Normal. Let

$$\bar{H}_{t+1} = \mathrm{cov}_t(\vec{\epsilon}_{t+1}), \tag{12}$$

be the conditional covariance matrix of the residuals, required to be positive definite, so that $\vec{w}'\bar{H}_t\vec{w} \geq 0\ \forall\vec{w}$. From $\vec{z}_t$, set the standardizing matrix $\bar{S}_t$ (sometimes denoted by $\bar{H}_t^{1/2}$) such that

$$\vec{\epsilon}_{t+1} = \bar{S}_{t+1}\,\vec{z}_{t+1}, \tag{13}$$

so that the conditional covariance here assumed symmetric (then Hermitian)

$$\bar{H}_{t+1} = \bar{S}_{t+1}\,\mathrm{cov}_t(\vec{z}_{t+1})\bar{S}'_{t+1} = \bar{S}_{t+1}\bar{S}'_{t+1} \tag{14}$$

becomes Cholesky factorized. Set the devolatilizing matrix $\bar{D}_t$ as a diagonal matrix by collecting the conditional standard deviations $\sqrt{h_t^{i,i}} \in \bar{H}_t$ in its diagonal (thus $\bar{D}_t = \bar{D}'_t$). From $\bar{D}_t$, write $\bar{S}_t$ in terms of the matrix $\bar{G}_t$ as

$$\bar{S}_t = \bar{D}_t\,\bar{G}_t. \tag{15}$$

If one defines

$$\bar{R}_t = \bar{G}_t\,\bar{G}'_t, \tag{16}$$

it follows that

$$\bar{H}_t = \bar{D}_t\,\bar{G}_t\,\bar{G}'_t\,\bar{D}_t = \bar{D}_t\,\bar{R}_t\,\bar{D}_t. \tag{17}$$

Since in components Eq. (17) becomes

$$h_t^{i,j} = \sqrt{h_t^{i,i}}\,r_t^{i,j}\,\sqrt{h_t^{j,j}}, \tag{18}$$

i.e. the typical covariance/correlation relationship, $\bar{R}_t$ of Eq. (16) is the correlation matrix of the residuals. Through $\bar{D}_t$, by

$$\vec{\epsilon}_{t+1} = \bar{D}_{t+1}\,\vec{\xi}_{t+1} \tag{19}$$

an additional devolatilized stochastic vector $\vec{\xi}_t$ is defined. By inversion of Eq. (19) it can be seen that $\vec{\xi}_{t+1}$ is linked to the conditional correlation matrix, because

$$\mathrm{cov}_t(\vec{\xi}_{t+1}) = \bar{D}_{t+1}^{-1} \bar{H}_{t+1} \bar{D}_{t+1}^{-1} = \bar{R}_{t+1}. \tag{20}$$

An mGARCH model for $P = Q = 1$ is obtained adding to the VAR Eq. (10) a dynamic equation for the components of $\bar{H}_t$. In the case of the BEKK model this equation is, as discussed in Refs. [16] and [17],

$$\bar{H}_{t+1} = \bar{O}\,\bar{O}' + \bar{U}\;\vec{\epsilon}_t \odot \vec{\epsilon}_t\,\bar{U}' + \bar{L}\bar{H}_t\bar{L}', \tag{21}$$

where $\vec{\epsilon}_t \odot \vec{\epsilon}_t$ stands for a matrix with $\epsilon_{i,t}\,\epsilon_{j,t}$ on its entries. The parameters of this equation are subject to static constraints in order to keep $\bar{H}_t$ positive definite and to make the system identifiable. Since the r.h.s. of Eq. (21) involves only quadratic terms, a sufficient condition for positive definiteness is that, given a suitable initial condition, $\bar{O}$ is lower triangular with positive elements on the diagonal [17]. Moreover, if $U_{1,1} > 0$, and $L_{1,1} > 0$, there is only one solution. Eq. (21) is a matrix equation, which looks similar to the univariate Eq. (6), but which actually couples dynamically all components of $\bar{H}_t$ among each other. This allows for volatility spillovers, i.e. transfers of volatility at one site at time t to volatility at another site at time $t + 1$, which can be a very appealing feature for spatio-temporal wind series modeling. Yet notice that it is not easy to read off causative volatility effects in the site-i-to-site-j channel directly from matrices $\bar{L}$ and $\bar{U}$, since all of their parameters contribute to each spillover channel. Moreover, this flexibility is traded off with the problem of a number of parameters K, those of the matrices $\bar{O}, \bar{U}$ and $\bar{L}$, which grows as K^2 with the number of sites K, making estimation at large K less reliable. Finally, if the matrices $\bar{O}, \bar{U}$ and $\bar{L}$ are diagonal, there is no coupling among the components of $\bar{H}_t$, thus no spillover. Another mGARCH model seems more suitable for scalable wind problems, the more parsimonious DCC model, which is defined by the vector equation

$$\vec{h}_{t+1} = \vec{\omega} + \bar{A}\;\vec{\epsilon}_t \otimes \vec{\epsilon}_t + \bar{B}\vec{h}_t \tag{22}$$

where $\otimes$ denotes a product between vectors which allows to form a vector with $\epsilon_{i,t}\,\epsilon_{i,t}$ on its entry i, $\vec{\omega}$ is a vector, and $\bar{A}$ and $\bar{B}$ are positive diagonal matrices. This equation doesn't mix

components of $\vec{h}_t$, hence it doesn't allow for volatility spillovers. What it anyway can do, like the diagonal form of the BEKK model, is transfering volatility at one site at time t to volatility at the same site at time $t+1$, thus it can at least sustain the same kind of volatility clustering as that of univariate models. Thus, the values on the diagonal of $\bar{A}$ and $\bar{B}$ will be collectively indicated with $\vec{\alpha}$ and $\vec{\beta}$ using the same convention of Eq. (6), because they control the univariate 'marginal' subsectors of the mGARCH sector. The vector Equation (22) takes care of the diagonal components of $\bar{H}_t$, but doesn't take care of its off-diagonal components. Thus, in order to complete the whole $\bar{H}_t$ and keep it positive definite an extra dynamics is added for the accessory matrix $\bar{Q}_t$ [14], defined by means of the devolatilized variables of Eq. (19) as the solution of

$$\bar{Q}_{t+1} = (1-a-b)\,\Xi + a\,\vec{\xi}_t \odot \vec{\xi}_t{}' + b\,\bar{Q}_t, \tag{23}$$

$$a, b \geq 0, \tag{24}$$

$$1-a-b > 0. \tag{25}$$

In Eq. (23) it is usually assumed that $\Xi = \mathrm{cov}_{\mathrm{sample}}(\vec{\xi}_t)$ [14, 23], the unconditional covariance of the vectors $\vec{\xi}_t$, which by Eq. (20) is an estimate of the average of $\bar{R}_t$. Constraints (24) force $\bar{Q}_t$ to stay positive definite (but not necessarily positive) and constraint (24) makes the $\bar{Q}_t$ dynamics stationary. After the diagonal positive matrices $\bar{Q}_t^*$ are defined as having the diagonal elements of $\bar{Q}_t$ on their diagonal and being 0 elsewhere, the dynamics of $\bar{Q}_t$ can be linked to the dynamics of $\bar{R}_t$ by

$$\bar{R}_t = (\bar{Q}_t^*)^{-1} \bar{Q}_t \, (\bar{Q}_t^*)^{-1}, \tag{26}$$

since in this way the dynamics of $\bar{R}_t$ are stationary, and all elements of $\bar{R}_t$ are in $[-1, 1]$ so that $\bar{H}_t$ is positive definite too. Taken together, Eq. (22) and Eq. (23) have $3K + 2$ parameters in all, i.e. the DCC model scales linearly in K, differently from the BEKK model which scales as K^2. Notice that setting $a = b = 0$ in Eq. (23) makes $\bar{Q}_t$ constant, and so $\bar{R}_t$ too. This special case with constant $\bar{R}_t$ is the so-called Constant Conditional Correlation (CCC) model by Bollerslev [35], which will not be discussed here.

The DCC model can also be estimated requiring that the distributions of its standardized variables are Student's t distributions (see Appendix A), in order to better care for

fat tails. In this case the full model can be called VAR-t-DCC. Notice that in this case the full model has the VAR sector estimated with a Gaussian distribution at first [23], then its residuals are passed to its mGARCH sector in which because of Eq. (22) K independent univariate Student's t Eq. (A.2) are used, and finally $\bar{R}_t$ is obtained from Eq. (23) where the multivariate Student's t p_{mst} from Eq. (A.11) is used, which contains one extra ν. Thus the model returns $K + 1$ parameters ν, which refer to different aspects of fat tails, and which could be called the parameters of the 'marginals' and of the 'joint'. Hence, the point forecasting ability of the VAR-t-DCC model is exactly the same as that of the Gaussian VAR and that of the (Gaussian) VAR-DCC model. On the contrary, volatility forecasting features of the VAR, VAR-BEKK, VAR-DCC and VAR-t-DCC models are different, as it will be shown in the next Section.

5 VAR-mGARCH models on bcspeed data

Being based on a VAR sector, all VAR-mGARCH models discussed above deal well with cross-correlations, because they process cross-correlation using two layers (VAR and mGARCH) instead of one. Can they be considered better than the univariate model, wich has no layer dedicated to cross-correlation, or the bare VAR models, which has one dedicated layer only? That depends.

Their scenario generation ability can be at first be evaluated from a static point of view by looking at Fig. 3, third panel on the right. That panel shows the histogram of a series of one year of wind speed values for the Solar TAC site sampled from a VAR-t-DCC model estimated on year 2012. That graph has to be compared with the outcome of the univariate model, center panel, and the original data, left panel. Like in the AR-GARCH synthetic data case, these generated data are well fitted. This multi-site VAR-mGARCH model, like all other models of this type, has hence a performance on static, unconditional data comparable to that of the single-site univariate model. Scatterplots from sampled data are not shown here for paper length reasons, but they are very similar to those of the VAR model. Autocorrelation plots are very similar to those of Fig. 9, too. Thus, the models of

$\bar{\Omega} = \bar{O}\bar{O}'$			$\bar{U}$			$\bar{L}$		
0.0671	0.0022	0.0048	0.2851	0.0029	0.0057	0.7512	0.0024	0.0027
0.0022	0.0290	0.0171	-0.0024	0.3060	0.0058	-0.0027	0.7508	0.0018
0.0048	0.0171	0.0645	0.0014	**-0.0127**	0.2998	0.0007	-0.0018	0.7515

Table 5: The three matrices $\bar{\Omega}$, $\bar{U}$ and $\bar{L}$, side by side, obtained from the estimate of the parameters of the mGARCH sector of the VAR-BEKK model. Rows/columns 1, 2, 3 of each matrix refer to sites M2 Tower, Lowry Range, Solar TAC. The only off-diagonal element in the $\bar{U}$ and $\bar{L}$ matrices which has a value larger than 0.009 (i.e. -0.0127, in bold) is the Solar TAC (row) to Lowry Range (column) entry, which can hint to a spillover effect.

this type are better than the univariate model in terms of generation ability.

In terms of point forecasting ability, these models are by design the same as the bare VAR model, which does the same as the univariate model, as already noticed when discussing Tab. 3. In terms of probabilistic forecasting, CRPL (Tab. 4), RPP (Fig. 7) and PL_F^α (Fig. 10) don't detect anything special. The Energy Loss (EL), a summary measure similar to CRPL but specialized for multivariate systems, introduced in Ref. [33], shown in Eq. (B.4), discussed in Appendix B, and reported in Tab. 4, doesn't detect anything either. Only the Interval Loss (IL), shown in Eq. (B.6) and discussed in Appendix B, does. Before discussing this result in detail it is better first discussing the estimation results of the mGARCH sector for the different models.

Tab. 5 shows the estimated parameters of the BEKK model, which clearly show that matrix $\bar{L}$ of Eq. (21) is almost diagonal whereas matrix $\bar{U}$ is not, hinting at the possibility of some weak spillover effect. It should be indeed recalled here that, by choice, the dataset includes the possibility of gentle breeze gusts passing from one site to another in one or more time lags. Tabs. 6 and 7, where mGARCH parameter estimates of the VAR-DCC and VAR-t-DCC models are reported, confirm that data do require at least a univariate GARCH sector (the β coefficient is high in both cases) and in a lesser way fat tails (in Tab. 7 all ν values are high, indicating that deviation from Gaussian behavior is not so large). Significance of all coefficients (not reported in the tables) is high.

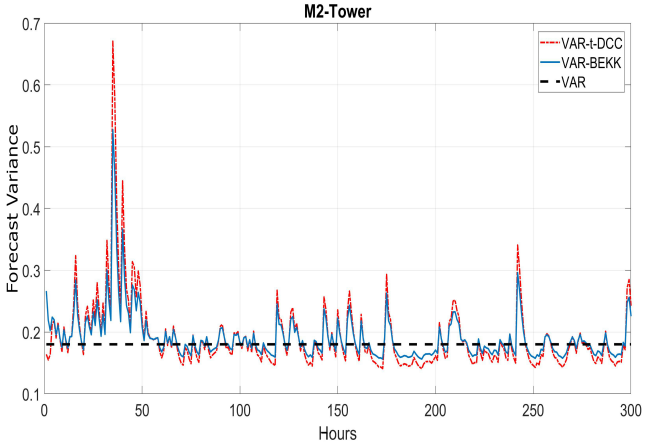

Figure 11: On site forecast variance obtained with VAR (static, flat broken line), VAR-t-DCC (dynamic, dotted red line, the highest one) and VAR-BEKK (dynamic, solid blue line) for the M2 Tower site.

What the mGARCH sector in facts adds to VAR modeling is dynamically adapting error forecast 'corridors', i.e. a possibly more precise assessment of forecasting power at individual sites. A direct visual comparison of what kind of forecast volatility paths one gets from VAR, VAR-BEKK and VAR-t-DCC models is shown in Fig. 11. In this Figure the flat line represents volatility from VAR, which is static and sets a fixed reference volatility corridor amplitude. VAR-t-DCC amplitude is often smaller in value, leading frequently to more precise forecasts, but can sometimes grow much larger when volatility bursts in clusters (as at hours $25 - 50$ in the Figure). These two effects balance each other, leading to CRPL values, RPP shapes and Pinball Loss values that are very similar to each other and to those of the bare VAR case. Rather unexpectedly, the VAR-BEKK model, with all its parameters and notwithstanding its ability to exploit intersite volatility couplings, doesn't do better than the VAR-DCC model, and doesn't seem to be able to exploit volatility information from neighbor sites at previous lags. P and Q BEKK values larger than 1 were also tried, but that didn't change the picture.

Yet, it should be noted that assessment by these probabilistic forecasting measures is blind to the advantage of having corridors which are as small as possible, but not so small that too many forecast values fall outside them. If this feature is indeed considered

	M2 Tower	Lowry Range	Solar TAC	joint
ω	0.054551	0.036334	0.064107	-
α	0.100312	0.239875	0.147401	0.02414
β	0.607098	0.355482	0.514638	0.74726

Table 6: VAR-DCC GARCH sector parameters. ω, α and β refer to the 'marginal' GARCH subsectors, i.e. they are found with Eq. (6) with $P = Q = 1$ independently for each site. In the case of the column 'joint', α refers to the parameter a and β refers to the parameter b of the accessory dynamics Eq. (23).

	M2 Tower	Lowry Range	Solar TAC	joint
ω	0.057926	0.031627	0.070324	-
α	0.115523	0.295808	0.181037	0.032202
β	0.578168	0.373677	0.455469	0.759689
ν	8.568022	5.391234	5.167480	6.871423

Table 7: VAR-t-DCC GARCH sector parameters. ω, α and β refer to the 'marginal' GARCH subsectors. In the case of the column 'joint', α refers to the parameter a and β refers to the parameter b of the accessory dynamics Eq. (23). In the fourth data row the first three Student's t parameters ν refer to the 'marginal' subsectors, and the fourth to the 'joint'.

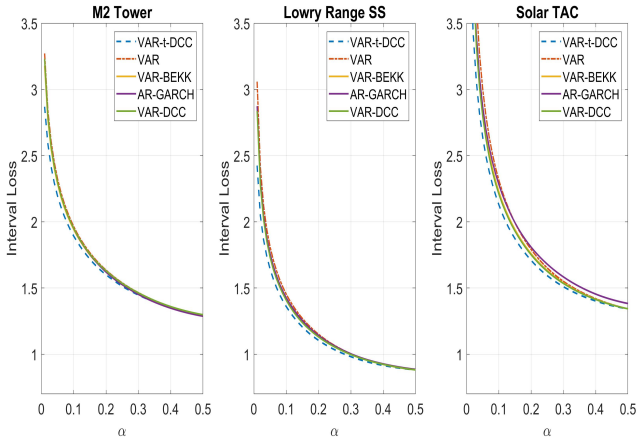

Figure 12: Comparison of the probabilistic forecasting ability of the models by the Interval Loss metric. On the $x-$axis the probabilistic weight α, on the $y-$axis IL_F^α at the corresponding $\alpha-$quantile, where α controls the width of the tested intervals. The small values of α on the l.h.s. of each graph correspond to very small forecasting intervals, which originate large errors and correspond to high losses.

an advantage, a more discriminating metric is needed, like the averaged Interval Loss IL_F^α. As discussed in Appendix B this loss metric penalizes both too large forecasting corridors and too many forecasts falling outside them. In the IL scheme (Eq. (B.6)) the forecasting window width is controlled by choosing an $\alpha-$quantile pair (upper and lower) that can be reconstructed from the forecast variance. For α close to zero large corridors are tested, whereas for α close to 0.5 small corridors are tested. Since the IL is a penalization score, lower values of IL_F^α are better. In Fig. 12 all VAR models are compared to each other and ranked quantile-wise by IL, thus assessesing their respective corridor adaptability, site by site. Fig. 12 shows that, on-site, differences among the VAR models in probabilistic forecasting performance are now detectable, even though rather small. The least penalized model is the VAR-t-DCC, which clearly does better than the VAR model and than the VAR-DCC itself, whereas the most penalized one is the AR-GARCH, at all quantiles. This is especially evident in the Solar TAC graph (third from the left). This analysis shows that even choosing models on the basis of their forecasting performance only (thus neglecting

their data modeling performance), requires to first decide which forecasting feature one is interested in. Here it was shown that for example a large number of metrics are blind to corridor adaptability, a feature that can be on the contrary important for engineering and financial applications.

A last feature that the mGARCH models have is the possibility of extracting dynamic information about conditional co-volatilities, i.e. off-diagonal elements of the conditional covariance matrix, forecasting them, and adding adaptive corridors to them. This is a feature that neither the univariate model nor the VAR model can extract (the VAR model can only extract static covariances). Fig. 13 shows for the DCC model a short stretch of the values of dynamic conditional correlations between pairs of sites against horizontal lines which represent static conditional correlations obtained from the VAR model. The Figure shows how the two closest sites cross-correlate in a positive way on average and more strongly in volatility bursts tens of hours long, whereas each pair of far sites can have almost zero correlation or even anticorrelate for a while. The BEKK model can give similar information (think about the values in matrix $\bar{U}$ of Tab. 5) but with worse scaling properties in K, that is, if this information is needed a DCC model is probably enough. Information on off-diagonal elements is generative dynamic information that can help better modeling and better risk assessment, and which still cannot be found in existing wind literature in the authors' knowledge. Notice that with relation to the wind dataset studied, a specific configuration in which cross-correlation among sites is highly asymmetric and almost zero for some pairs was purposely chosen, as remarked in Sec. 2. This implies that this dataset is very suitable to show limitations in data modeling of univariate models and start the discussion of this paper, but it is not very good to study the importance of off-diagonal volatility components. The univariate model can be in fact seen as the diagonal limit of the DCC model, which is turn is a sort of limit of the BEKK model. When most of the off-diagonal entries are not important, the univariate model can be indeed effective. Thus, before deciding to choose an mGARCH model over an univariate model it is better to use a bare VAR model to assess presence and intensity of covariance among residuals at the time scale chosen.

Figure 13: Cross-site correlation obtained with the VAR-t-DCC model for the the three pairs of sites.

Is it in general beneficial to gather all this accurate dynamic cross-correlation information, against the simplicity of use of the sole static VAR sector? It is worthwhile under certain conditions like the obvious ones mentioned above, and probably mainly if this information is needed for dynamic control problems when on-time cross-correlation and forecast co-volatility can be used for anticipative control. This is characteristic of risk control purposes, in which both system knowledge and forecasting ability are used. Otherwise, for standard risk assessment cases, a VAR model without a mGARCH sector can be more than enough.

6 An example of anticipative control

The pieces of information that can be obtained with wind univariate or VAR-mGARCH models can be used alternatively for setting up reliable simulations of real systems, or for forecasting their future behavior. But there are also uses in which both system knowledge and forecasting ability can be required together at once. An example of that is risk shaping obtained by anticipative control of volatility.

In anticipative control the information about the system is dynamically used to forecast scenarios using current data, decide which scenario is best among the forecast ones, and

prepare action consequently. In the more specific case of volatility models, these scenarios are based on the forward projection of a hidden variable, i.e. volatility, not even observable at current time, but anyway computed on current observed wind data. This kind of control is hence pretty sophisticated.

In the following demonstrative thought experiment of volatility control, a wind energy producer owns the three sites and wants to bid electricity production to the market. Typically, bids are made some time before delivery, and here it will be fictionally assumed that the bid in quantity is submitted one hour before delivery. Typical European markets penalize deviations from the submitted schedule [30]. As to this, it is fictionally assumed that this penalization is symmetric around the committed quantity. It is also assumed that the producer owns a wind turbine at each site, and that for each wind turbine a cubic relationship between local wind speed u_t^i and locally produced power P_t^i of the form

$$P_t^i = \lambda \, (u_t^i)^3 \tag{27}$$

holds. In Eq. (27) high speed cutoffs are deliberately not taken into account and λ is a constant with the same value at all sites. The producer knows that he or she can model and forecast the farm production and its variance by VAR-mGARCH models. The models can be estimated on the three in-sample NREL series transformed to power P_t^i by Eq. (27), and collected in the column vector $\vec{P}_t = (P_t^1, P_t^2, P_t^3)'$. One-step-ahead forecast power components will be denoted $P_t^{i,f,1}$, collected in the column vector $\vec{P}_t^{f,1}$, and the associated forecast conditional covariance matrix will be $\bar{\Pi}_t^{f,1} = \bar{\Pi}_{t+1}$.

The producer can bid the total aggregated forecast production $P_t^F = \sum_{i=1,2,3} P_t^{i,f,1}$, or just a part of it. Define $w_t^i \in [0,1]$ as weights, to be collected in the weight vector $\vec{w}_t = (w_t^1, w_t^2, w_t^3)'$. The generic forecast production, aggregated in an energy portfolio of weighted components $w_t^i \, P_t^i$, is thus $P_t^f = \vec{w}_t' \, \vec{P}_t^{f,1}$. If the producer chooses all of the weights as $w_t^i = 1$, all forecast power is bid and $P_t^f = P_t^F$. If not, a percent of power has to be shed off or stored.

Because of the financial penalization, the producer who bids P_t^F can be concerned about aggregate production forecasting errors. Thus he can try to use his knowledge of the

system to compute one hour in advance, i.e. at t, those weight values $\vec{w}_t^*$ that would keep the portfolio variance at $t+1$ fixed at (or below) a chosen target level σ_T^2 and maximize output power. He hence accepts to trade off some missed income (from missed production) for avoided penalization, at a break-even level which is set by the specific market rules. He thus sets up at each time t a Markowitz-like maximum problem [32] for the optimal weights

$$\vec{w}_t^* = \arg\min_{\vec{w}_t} \vec{w}_t' \, \vec{P}_t^{f,1} \quad \text{s.t.} \quad 0 \leq \vec{w}_t' \, \bar{\Pi}_t^{f,1} \, \vec{w}_t \leq \sigma_T^2 \quad \text{and} \quad w_t^i \in [0,1]. \tag{28}$$

In Eq. (28) the data are the power components $\vec{P}_t^{f,1}$ forecast by the VAR sector and the conditional covariance matrix $\bar{\Pi}_t^{f,1}$ forecast by the mGARCH sector. The optimal production to be bid is hence $P_t^{f*} = \vec{w}_t^{*\prime} \, \vec{P}_t^{f,1}$. At each time this problem is convex, since the objective to be maximized and all constraints are convex. Notice that the $0 \leq \vec{w}_t' \, \bar{\Pi}_t^{f,1} \, \vec{w}_t$ part of the variance constraints is automatically satisfied because BEKK and DCC generate positive definite matrices. Notice also that the typical Markowitz relationship $\sum_i w_t^i = 1$ is not imposed here, and that the value of λ doesn't affect $\vec{w}_t^*$, being just a scale in the objective. It is straightforward that this approach is not very much suitable to the univariate AR-GARCH model, which gives no information about off-diagonal covariance matrix components.

It is important to highlight the fact that differently from what was discussed in the preceding Sections, this approach actively uses volatility forecasting to exploit on-time global information on future covariances in order to compute the future expected portfolio variance, instead of using it to just estimate adaptive corridors (i.e. for mere on-site forecasting purposes). Notice yet that, specifically for this reason, this construct could be in principle rather unreliable. For example, in standard Markowitz problems the covariance matrix is an observed sample quantity, whereas here $\bar{\Pi}_t^{f,1}$ is a forecast unobservable entity (obtained from other unobservable entities like current volatilities, not even easily estimable ex-post), strongly linked to the underlying model, i.e. both on the VAR and mGARCH modeling choices.

Yet, surprisingly enough, using the NREL data and $\gamma = 0$ (i.e. logtransforms) on the cube of wind speeds, for the two target values of $\sigma_T^2 = 0.1$ and $\sigma_T^2 = 0.8$, the output variance is satisfactorily reduced. This is seen in in Fig. 14, looking at the distribution of

the output (i.e. the quantity risk profile), which is suitably squeezed. More specifically, in Fig. 14 the histograms of the realized (i.e. out-of-sample) aggregate output power $\vec{w}_t^{*\prime}\vec{P}_{t+1}$ (in log-power units, for the sequence of 8583 hours of year 2013) obtained with the VAR-DCC model are plotted both for the two values of the control variance and without control (for reference). In both controlled cases the variance of the controlled output is indeed reduced, even though it is about twice as large as the imposed target (see also data reported in the Figure caption). This variance measures the realized error around the bid quantity, and it is generated mainly because of the errors in point and variance forecasting. This mismatch happens because, even though the numerical routine that computes the optimal weights is able to keep the forecast aggregate variance exactly at the required level, the weights obtained at t are applied to the power realized at $t+1$. This consideration shows also that this exercise can also be used as a qualitatively indirect way to assess forecasting errors.

From Fig. 14 notice also that in exchange for controlling variance the producer must accept a reduction in the bid quantity, which in principle can have a large value. As in all Markowitz problems, here too a return is traded off for (reduced) risk, and in this specific case the break-even risk reduction point depends on market parameters. Notice also that if market penalities are not symmetric or the producer is concerned of only one side of the risk distribution, other risk measures (like CVaRD, i.e. CVaR Deviation, see [31]) can be used. The other models behave more or less in the same way.

Of course, different schemes and more realistic variations of this scheme could be analyzed in more depth, but again, this numerical experiment is intended mainly to show on real data that VAR-mGARCH modeling can also be used and evaluated in situations in which models that are good at both scenario generation and forecasting at once are required.

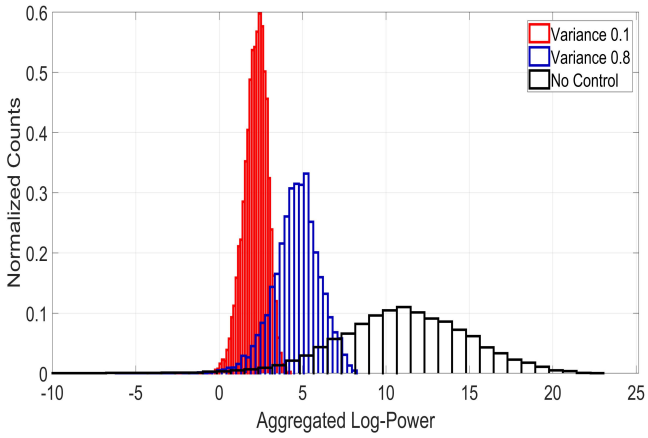

Figure 14: Histograms (normalized counts) of controlled and uncontrolled aggregated power output from the three sites, in log-transformed power units. Control obtained with the VAR-t-DCC model. Maximum variance allowed are $\sigma_T^2 = 0.1$ (leftmost, in red: mean 2.18, variance 0.48, standard deviation 0.69) and $\sigma_T^2 = 0.8$ (center, in blue: mean 4.69, variance 1.86, standard deviation 1.36). Uncontrolled output distribution too (rightmost, in black: mean 11.00, variance 15.81, standard deviation 3.98) is reported for comparison.

7 Conclusions

The paper tried to answer the question about which model in the univariate, VAR, VAR-DCC, VAR-BEKK, VAR-t-DCC hierarchy is better for estimating spatio-temporal wind speed time series. The answer is that the comparison should at least be grounded on choosing first what is the criterion for valuation in which one is interested. The choice of the criterion implies a choice of appropriate metrics, quantitative or qualitative. In turn, this implies that when the modeling goal of either synthetic generation or forecasting is chosen, or they are considered together, not necessarily a unique model can be identified as the best. The answer of course depends also on the dataset studied, and more subtly even on the model-dependent slice of information one is interested in. In facts, it depends on the way in which information flows through the model. Yet, according to the results discussed above, and limited to the studied dataset, one can conclude that a good multipurpose model is the VAR-t-DCC model.

An obvious extension of this study is hence the evaluation of all considered models on different wind speed datasets. Even better, these models should be evaluated on a dataset with a larger number of sites, in order to understand their scaling properties in the number K of the sites. All of this is left for further study.

References

[1] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, C. Draxl, *The state-of-the-art in short-term prediction of wind power: A literature overview*, ANEMOS.plus, 2011.

[2] J. Jung R. P. Broadwater, *Current status and future advances for wind speed and power forecasting*, Renewable and Sustainable Energy Reviews, v. 31, pp. 762–777, 2014.

[3] L. Yang, M. He, J. Zhang, V. Vittal, *A Spatio-Temporal Analysis Approach for Short-Term Forecast of Wind Farm Generation*, in: Spatio-Temporal Data Analytics for Wind Energy Integration, SpringerBriefs in Electrical and Computer Engineering, 2014.

[4] A. Lenzi, I. Steinsland, P. Pinson, *Benefits of spatio-temporal modelling for short term wind power forecasting at both individual and aggregated levels*, Environmetrics v. 29(3), e2493, 2018.

[5] J. Tastu, P. Pinson, E. Kotwa, H. Madsen, H. A. Nielsen, *Spatio-temporal analysis and modeling of short-term wind power forecast errors*, Wind Energy, v. 14, n. 1, pp. 43–60, 2011.

[6] L. Cavalcante, R. J. Bessa, M. Reis, J. Browell, *LASSO vector autoregression structures for very short-term wind power forecasting*, Wind Energy, v. 20, pp. 657-–675, 2017.

[7] B. M. Sanandaji, A. Tascikaraoglu, K. Poolla, P. Varaiya, *Low-dimensional Models in Spatio-Temporal Wind Speed Forecasting*, arXiv preprint arXiv:1503.01210v1, 2015.

[8] S. Emeis, Wind Energy Meteorology, Springer, 2013.

[9] T. Bollerslev, *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, v. 31, pp. 307–327, 1986.

[10] A. Lojowska, D. Kurowicka, G. Papaefthymiou, L. van der Sluis, *Advantages of ARMA-GARCH wind speed time series modeling*, in: IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 2010.

[11] Z. Shen, M. Ritter, *Forecasting volatility of wind power production*, Applied Energy, v. 176, pp. 295–308, 2016.

[12] T. Jebara, *Machine Learning: Discriminative vs. Generative*, Kluwer Academic Publishers, 2004.

[13] C. Lucheroni, C. Ragno, *Modeling of Wind Speed Spatio-Temporal Series by Multivariate-GARCH and Copula/GARCH models*, in: Proceedings of the 2017 14th International Conference on the European Energy Market (EEM), IEEE, 2017.

[14] *Handbook of volatility models and their applications*, ed. by L. Bauwens, C. Hafner, S. Laurent, Wiley, 2012.

[15] C. Alexander, *Market Risk Analysis Volume II*, Wiley, 2008.

[16] J. Daníelsson, *Financial Risk Forecasting*, Wiley, 2011.

[17] R. F. Engle, K. F. Kroner, *Multivariate Simultaneous Generalized ARCH*, Econometric Theory, v. 11(01), pp. 122–150, 1995.

[18] R. F. Engle, *Dynamic Conditional Correlation*, Journal of Business and Economic Statistic, vol. 20(3), pp. 339–350, 2002.

[19] R. F. Engle, K. Sheppard, *Theoretical and Empirical porperties od Dynamic Conditional Multivariate GARCH*, NBER working paper, 2001.

[20] Y. K. Tse, A. K. C. Tsui, *A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations*, Journal of Business and Economic Statistic, vol. 20(3), pp. 351–362, 2002.

[21] C. Lucheroni, C. Mari, *Optimal Integration of Intermittent Renewables: A System LCOE Stochastic Approach*, Energies, vol. 11(3), pp. 549–570, 2018.

[22] M. Haas, Markus, J.-C. Liu, *Theory for a Multivariate Markov–switching GARCH Model with an Application to Stock Markets*, 2015 Annual Conference (Münster) 'Economic Development - Theory and Policy', n. 112855, Verein für Socialpolitik / German Economic Association, 2015.

[23] A. Ghalanos, *The* `rmgarch` *models*, 2018. Software `rmgarch` available on CRAN.

[24] Software available at

`www.kevinsheppard.com/MFE_Toolbox`

[25] A. Jordan, F. Krüger, S. Lerch, *Evaluating Probabilistic Forecasts with scoringRules*, Journal of Statistical Software, forthcoming. Available also as arXiv preprint arXiv:1709.04743. Software `scoringRules` available on CRAN.

[26] National Renewable Energy Laboratory (NREL) web site `www.nrel.gov`

[27] H. Madsen, P. Pinson, G. Kariniotakis, H.A. Nielsen, T.S. Nielsen, *Standardizing the performance evaluation of short term wind power prediction models*. Wind Engineering, v. 29, pp. 475-–489, 2005.

[28] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, *Non-parametric Probabilistic Forecasts of Wind Power: Required Properties and Evaluation*, Wind Energy, v. 10, pp. 497-–516, 2007.

[29] R. L. Winkler, *A decision-theoretic approach to interval estimation*, Journal of the American Statistical Association, vol. 67(337), pp. 187—191, 1972.

[30] P. Pinson, J. Juban, G. N. Kariniotakis, *On the Quality and Value of Probabilistic Forecasts of Wind Generation*, 2006 International Conference on Probabilistic Methods Applied to Power Systems, pp. 1–7., 2006. doi: 10.1109/PMAPS.2006.360290.

[31] C. Lucheroni, C. Mari, *Risk shaping of optimal electricity portfolios in the stochastic LCOE theory*, Computers and Operations Research, vol.96, pp. 374–385, 2018.

[32] A. Pachamanova, F. J. Fabozzi, *Simulation and Optimization in Finance: Modeling with MATLAB, @Risk, or VBA*, Wiley, 2010.

[33] T. Gneiting, A. E. Raftery, *Strictly Proper Scoring Rules, Prediction, and Estimation*, Journal of the American Statistical Association, v. 102, n. 477, pp. 359–378, 2007.

[34] B. Liu, J. Nowotarski, T. Hong, R. Weron, *Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts*, IEEE Transactions on Smart Grid, v. 8, n. 2, pp. 730–737, 2017.

[35] T. Bollerslev, *Modelling the coherence in short-run nominal exchange rates: A multivariate generalizec arch model*, Review of Economics and Statistics, vol. 72(3), pp. 498–505, 1990.

A Student's t distributions

The general univariate Student's t (also called location-scale) distribution for the stochastic variable x has the form

$$p_{ugt}(x;\nu,m,\lambda) = \left(\frac{1}{\lambda\pi\nu}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left[1+\frac{(x-m)^2}{\lambda\nu}\right]^{-\left(\frac{\nu+1}{2}\right)} \tag{A.1}$$

For this variable $E[x] = m$ (the expectation integral is undefined for $\nu \leq 1$) and $\mathrm{var}(x) = \lambda\frac{\nu}{\nu-2}$ (undefined for $\nu \leq 2$). Requiring that $E[x] = 0$ and $\mathrm{var}(x) = 1$ implies that $m = 0$ and $\lambda = (\nu-2)/\nu$, so that the standardized univariate distribution of the standardized variable z is

$$p_{ust}(z;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi(\nu-2))^{1/2}} \left[1+\frac{z^2}{(\nu-2)}\right]^{-\left(\frac{\nu+1}{2}\right)}. \tag{A.2}$$

Stochastic variables x and z are thus linked as

$$x = \sqrt{\frac{\lambda\nu}{\nu-2}}\, z + m = s\, z + m, \tag{A.3}$$

where m is called location of x and $s = \sqrt{\frac{\lambda\nu}{\nu-2}}$ is called its scale. If x was Gaussian, s would have been its standard deviation, which is yet a quantity not always defined for Student's t variables.

The ν-dependent multivariate general Student's t distribution for the K components stochastic vector $\vec{x}$ is defined as

$$p_{mgt}(\vec{x};\nu,\vec{m},\bar{\Sigma}) = \left(\frac{1}{|\bar{\Sigma}|}\right)^{\frac{1}{2}} \left(\frac{1}{\pi\nu}\right)^{\frac{K}{2}} \frac{\Gamma\left(\frac{\nu+K}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left[1+\frac{(\vec{x}-\vec{m})\bar{\Sigma}^{-1}(\vec{x}-\vec{m})'}{\nu}\right]^{-\left(\frac{\nu+K}{2}\right)} \tag{A.4}$$

where $'$ means transpose, and $\bar{\Sigma}$ and $\vec{m}$ are respectively a positive invertible matrix with determinant $|\bar{\Sigma}|$ and a vector of parameters. The stochastic vector $\vec{x}$ has then the property that

$$\mathrm{cov}(\vec{x}) = \frac{\nu}{\nu-2}\bar{\Sigma} \tag{A.5}$$

If $\bar{D}$ is a matrix with the elements of the diagonal of $\bar{\Sigma}$ on its diagonal and zero elsewhere (with inverse $\bar{D}^{-1}$), defining

$$\bar{R} = \bar{D}^{-1}\bar{\Sigma}\bar{D}^{-1} \tag{A.6}$$

(i.e. $\bar{\Sigma} = \bar{D}\bar{R}\bar{D}$ so that $\bar{\Sigma}^{-1} = \bar{D}^{-1}\bar{R}^{-1}\bar{D}^{-1}$) the distribution of Eq. (A.4) becomes

$$p_{mgt}(\vec{x};\nu,\vec{m},\bar{\Sigma}) = \left(\frac{1}{\pi\nu}\right)^{\frac{K}{2}} \frac{\Gamma\left(\frac{\nu+K}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$$
$$\left(\frac{1}{|\bar{D}|^2\,|\bar{R}|}\right)^{\frac{1}{2}} \left[1 + \frac{(\vec{x}-\vec{m})\bar{D}^{-1}\bar{R}^{-1}\bar{D}^{-1}(\vec{x}-\vec{m})'}{\nu}\right]^{-\left(\frac{\nu+K}{2}\right)}. \tag{A.7}$$

Notice that Eq. (A.6) expresses the usual relationship between correlation and covariance. This suggests to define the devolatilized variable

$$\vec{\xi} = \bar{D}^{-1}(\vec{x} - \vec{\mu}) \tag{A.8}$$

distributed as

$$p_{cgt}(\vec{\xi};\nu,\bar{R}) =$$
$$\left(\frac{1}{|\bar{R}|}\right)^{\frac{1}{2}} \left(\frac{1}{\pi\nu}\right)^{\frac{K}{2}} \frac{\Gamma\left(\frac{\nu+K}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left[1 + \frac{\vec{\xi}\,\bar{R}^{-1}\vec{\xi}'}{\nu}\right]^{-\left(\frac{\nu+K}{2}\right)} \tag{A.9}$$

and such that

$$\mathrm{cov}(\vec{\xi}) = \frac{\nu}{\nu-2}\bar{R} \tag{A.10}$$

The standardized multivariate Student's t distribution for the stochastic vector $\vec{z}$ such that $E[\vec{z}] = 0$ and $\mathrm{var}(\vec{z}) = \bar{1}$ is obtained for $\vec{m} = 0$ from Eq. (A.4) with $\bar{\Sigma} = \frac{\nu-2}{\nu}\,\bar{1}$, or from Eq. (A.7) with $\bar{D} = \bar{1}$ and $\bar{R} = \frac{\nu-2}{\nu}\,\bar{1}$ (where $\bar{1}$ indicates the unit matrix), as

$$p_{mst}(\vec{z};\nu) = \frac{\Gamma\left(\frac{\nu+K}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi(\nu-2))^{K/2}} \left[1 + \frac{\vec{z}\vec{z}'}{(\nu-2)}\right]^{-\left(\frac{\nu+K}{2}\right)}. \tag{A.11}$$

Marginal densities of the multivariate standardized Student's t distribution are univariate standardized Student's t distributions of Eq. (A.2). All these distributions will be useful when discussing t-VAR-mGARCH modeling.

Notice that for $\nu \to \infty$ the univariate and multivariate Student's t distributions tend to a Gaussian, $\mathrm{cov}(\vec{x}) \to \bar{\Sigma}$ and $\mathrm{cov}(\vec{\xi}) \to \bar{R}$, whereas $\mathrm{cov}(\vec{z}) = \bar{1}$.

B Loss metrics used in the paper

The quality of probabilistic forecasting can be assessed by many different metrics. In Ref. [33] a unifying theory of an important and self-consistent group of them is discussed. The main idea of Ref. [33] is that all metrics of this group hit an upper bound (usually zero) when the distribution F_t, forecast at time t for time $t+1$, matches the unknown true distribution Q_{t+1}. If F_t is parametric in mean and variance, the analytic form of F_t can be obtained by the point forecast x_t^f and the variance forecast h_t^f, whereas Q_{t+1} can only be empirically inferred from the realized values $\hat{x}_{t+1}$. A metric is thus a function $s(F_t, \hat{x}_{t+1})$. The higher is the metric's value, the closer the forecast distribution is to the true one, and for this reason these metrics are called scores. They can be also considered in the reverse direction (i.e. the lower the better), like in the present paper, and in that case they are called losses or penalizing scores. Each metric is actually used in its averaged form

$$S_F = \frac{1}{T}\sum_{t=1}^{T} s(F_t, \hat{x}_{t+1}). \tag{B.1}$$

The averaged metrics S_F don't give direct information on the distance of F_t from Q_{t+1}, they hence cannot be used in an absolute way. Yet, they can be used to rank forecasting models among themselves. Some of them summarize forecasting quality in one number, like the Continuous Ranked Probability Score (CRPS) and the Energy Score (introduced in Ref. [33] itself), other detail quantile dependency like the Pinball and the Interval[2] Score. The CRPS, the Pinball Score and the Interval Score are designed for univariate distributions, whereas the Energy Score is designed for multivariate distributions. When these metrics are evaluated on parametric univariate distributions, a direct and efficient use of the analytic form of the forecast distribution can be made. As to the Energy Score, its evaluation can only be made numerically, on multivariate samples drawn from F_t [33]. This evaluation is computationally much heavier than the analytic one and has bad scaling properties.

[2] In electricity econometrics literature like Ref. [34] this score is often called Winkler Score and is related to Ref. [29], but in the Reference [33], which is the one followed here, the name Winkler Score is used for another measure. To avoid confusion, this score will be called Interval Score.

The Continuous Ranked Probability Loss (CPRL) is defined as the integral

$$\mathrm{CRPL}(F_t, \hat{x}_t) = \lim_{z\to\infty} \int_{-\infty}^{z} \left(F_t(z') - 1[\hat{x}_t \leq z']\right)^2 dz', \tag{B.2}$$

where $1[\hat{x}_t \leq z']$ is the indicator function, always equal to 0 except for being 1 when $\hat{x}_t \leq z'$. This integral is well defined since for any finite $\hat{x}_t$ the integrand is close to zero both at $z' \to \pm\infty$. Incidentally, notice that the CRPS is defined as $-\mathrm{CRPL}(F_t, \hat{x}_t)$. The CPRL can also be equivalently computed as the expectation

$$\mathrm{CRPL}(F_t, \hat{x}_t) = E_{F_t} |\Phi_1 - \hat{x}_t| - \frac{1}{2} E_{F_t,F_t} |\Phi_1 - \Phi_2| . \tag{B.3}$$

In Eq. (B.3) Φ_1 and Φ_2 are independent random variables both with distribution F_t, and $E_{F_t}|\bullet|$ indicates the expectation of the absolute value with respect to the distribution F_t. Eq. (B.3) shows that the $\mathrm{CRPL}(F_t, \hat{x}_t)$ can be numerically evaluated by sampling F_t and replacing the expectation with an average. The form of Eq. (B.3) also indicates that if the forecast is a point forecast, i.e. with zero width around the forecast, the average $\mathrm{CPRL}_F = \sum_{t=1}^{T} \mathrm{CRPL}(F_t, \hat{x}_t)/T$ coincides with the Mean Absolute Value MAE, because all forecasts are the same. Thus the CPRL_F can be directly compared to MAE. If CPRL_F $<$ MAE strictly, passing from point to probabilistic forecasting improves the forecast.

The Energy Loss (EL) for multivariate distributions is defined in a way similar to Eq. (B.3), but directly in a numerical form for D vector samples $\vec{\hat{\Phi}}_d$ $(d = 1, \ldots, D)$ of F_t, as

$$\mathrm{EL}(F_t, \vec{\hat{x}}_{t+1}) = \frac{1}{D} \sum_{d=1}^{D} \|\vec{\hat{\Phi}}_d - \vec{\hat{x}}_{t+1}\| - \frac{1}{2D^2} \sum_{i=1}^{D} \sum_{j=1}^{D} \|\vec{\hat{\Phi}}_i - \vec{\hat{\Phi}}_j\|, \tag{B.4}$$

where $\vec{\hat{x}}_{t+1}$ are the observed vectors and $\|\bullet\|$ indicates the Euler norm. Like before, $\mathrm{EL}_F = \sum_{t=1}^{T} \mathrm{EL}(F_t, \vec{\hat{x}}_{t+1})/T$.

The Pinball Loss (PL) makes use of the α-quantiles r^α of the forecast univariate distribution. It is defined as

$$\mathrm{PL}^\alpha(r_t^\alpha, \hat{x}_{t+1}) = \begin{cases} (1-\alpha)(r_t^\alpha - \hat{x}_{t+1}) & \hat{x}_{t+1} < r_t^\alpha \\ \alpha(\hat{x}_{t+1} - r_t^\alpha) & \hat{x}_{t+1} > r_t^\alpha, \end{cases} \tag{B.5}$$

where $\alpha \in [0, 1]$. Notice that Eq. (B.5) corresponds to the very definition of quantile, in the sense that $r^\alpha = \arg\min_x E[\mathrm{PL}^\alpha(x, X)]$ defines the α-quantile of the distribution of the

random variable X. In contrast to CPRL and EL, PL^α details forecast errors along all profile of the cumulative distribution. At small values of α, to which very negative quantiles r_t^α correspond, the condition of the second line of Eq. (B.5) is much more frequently satisfied than the condition of the first line, so that the corresponding penalization is low. Like before, $\mathrm{PL}_F^\alpha = \sum_{t=1}^T \mathrm{PL}^\alpha(r_t^\alpha, \hat{x}_{t+1})/T$.

The Interval Loss (IL) is based on quantiles as well, but penalizes for too large forecasting intervals, i.e. corridors. Be L_t^α the lower forecast $\alpha-$quantile and U_t^α the upper forecast $(1-\alpha)-$quantile, seen as the lower and upper values of the forecasting interval, and $\delta_t^\alpha = U_t^\alpha - L_t^\alpha$ the interval forecast width. The IL is defined as

$$\mathrm{IL}^\alpha(U_t^\alpha, L_t^\alpha, \hat{x}_{t+1}) = \begin{cases} \delta_t^\alpha & \hat{x}_{t+1} \in [L_t^\alpha, U_t^\alpha] \\ \delta_t^\alpha + \frac{2}{\alpha}(L_t^\alpha - \hat{x}_{t+1}) & \hat{x}_{t+1} < L_t^\alpha \\ \delta_t^\alpha + \frac{2}{\alpha}(\hat{x}_{t+1} - U_t^\alpha) & \hat{x}_{t+1} > U_t^\alpha. \end{cases} \tag{B.6}$$

Here the penalization has two terms: δ_t^α which penalizes large corridors, and a residual term e_t^α which penalizes the cases in which realized values fall outside the forecast corridor. For example, for $\alpha = 0$, since L_t is the quantile corresponding to 0 and U_t is the quantile corresponding to 1, one gets the largest windows possible centered on the forecast $x_t^{f,1}$. All realized values will fall into this interval, thus $e_t^\alpha = 0$ and $W_t^\alpha = \delta_t^\alpha$. For α slightly higher than zero the rare errors are heavily penalized by the denominator in the e_t^α term. For $\alpha = 0.5$, the window closes on the median (which for a Gaussian corresponds to the average), $\delta^\alpha = 0$, and $W_t^\alpha = e_t^\alpha$. For $\alpha > 0.5$, L_t^α passes U_t^α, and everything happens in reverse in a symmetric way. Like before, the metric is used in the averaged from $\mathrm{IL}_F^\alpha = \sum_{t=1}^T \mathrm{IL}^\alpha(U_t^\alpha, L_t^\alpha, \hat{x}_{t+1})/T$. A low IL_F^α thus signals a good average trade-off between a small corridor (with a high risk of outside values) and a large corridor, large $\hat{\delta}^\alpha$ in average.

Finally, Eq. (B.5) can be changed into the simpler quantile-based score defined as

$$\mathrm{RP}^\alpha(r_t^\alpha, \hat{x}_{t+1}) = \begin{cases} 1 & \hat{x}_{t+1} < r_t^\alpha \\ 0 & \hat{x}_{t+1} > r_t^\alpha, \end{cases} \tag{B.7}$$

not belonging to the group of measures discussed in Ref. [33]. The comparison of this form with Eq. (B.2) suggests a simple graphical assessment technique, actually used in the

literature under varied names. The average $\mathrm{RP}_F^\alpha = \sum_{t=1}^{T} \mathrm{RP}^\alpha(r_t^\alpha, \hat{x}_{t+1})/T$ can be seen as the empirical coverage corresponding to the theoretical coverage $F_t(r_t^\alpha) = \alpha$. A perfect forecast would thus give $\mathrm{RP}_F^\alpha = \alpha$ for each α. A plot of RP_F^α against α can thus assess the forecasting error quantile-wise. The amount of the squared difference $F_t(r_t^\alpha) - \mathrm{RP}^\alpha(r_t^\alpha, \hat{x}_{t+1})$ integrated on all values of r_t^α (i.e. on all values of α) and averaged is in turn assessed by CRPL_F. In the context of Ref. [33] this plot could be called Ranked Probability Plot (RPP). RP_F^α is clearly less accurate than the Pinball Loss in assessing quantile profile forecasting quality.