Downloaded from orbit.dtu.dk on: Dec 20, 2017

Technical University of Denmark

Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix

Tastu, Julija; Pinson, Pierre; Madsen, Henrik

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Tastu, J., Pinson, P., & Madsen, H. (2013). Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix. Kgs. Lyngby: Technical University of Denmark (DTU). (DTU Compute-Technical Report-2013; No. 14).

DTU Library
Technical Information Center of Denmark

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Technical Report–2013–

Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix.

Julija Tastu[a,*], **Pierre Pinson**[b], **Henrik Madsen**[a]

[a] Department of Applied Mathematics and Computer Science, Technical University of Denmark
[b] Department of Electrical Engineering, Technical University of Denmark

Abstract

The emphasis in this work is placed on generating space-time trajectories (also referred to as scenarios) of wind power generation. This calls for prediction of multivariate densities describing wind power generation at a number of distributed locations and for a number of successive lead times. A modelling approach taking advantage of sparsity of precision matrices is introduced for the description of the underlying space-time dependence structure. The proposed parametrization of the dependence structure accounts for such important process characteristics as non-constant conditional precisions and direction-dependent cross-correlations. Accounting for the space-time effects is shown to be crucial for generating high quality scenarios.

Keywords: wind energy, scenarios, trajectories, space-time, precision matrix, directional CAR

1 Introduction

Large scale integration of wind energy into power grids induces difficulties in operation and management of power systems due to the stochastic nature of wind, with its variability and limited predictability [1]. For optimal integration of wind energy into power systems high quality wind power forecasts are required [21]. A history of short-term wind power forecasting and an overview of the state-of-the-art methodology are given in [9] and [14], respectively.

Owing to the complexity of the related decision making tasks, it is preferable that the forecasts provide the user not only with the expected value of the future power generation, but also with the associated uncertainty estimates. This calls for probabilistic, rather than point forecasting [16]. Applications of probabilistic forecasts to power grid operations include trading wind energy [6], economic load dispatch and stochastic unit commitment [25, 7, 24], optimal operation of storage [12], reserve quantification [4] and assessment of operating costs [28].

* *Contact author:*
Technical University of Denmark
Department of Applied Mathematics and Computer Science
Matematiktorvet, Building 322, DK-2800 Kgs. Lyngby, Denmark

Usually probabilistic wind power forecasts are generated on a per-site and per- look-ahead time basis. As a result, they do not inform about the interdependence structure between forecast errors obtained at different times and/or at different sites.

Addressing each site of interest individually is motivated by the fact that power curves describing the conversion of meteorological variables to power are often given by complex non-linear functions of meteorological conditions, number and type of the considered wind turbines, their interposition within the wind farm, some topographical particularities of the area, etc. The fact that wind power dynamics is so site-specific makes it more complicated to issue high quality forecasts for a large number of sites simultaneously, because the local particularities (if to be respected) keep the dimension of the problem high.

Similarly, a common practice is to issue direct power forecasts for each of the time horizons of interest individually, rather than addressing the joint distribution. This can be explained by the fact that such direct forecasts are more robust to model misspecification. Iterated multistep-ahead predictions as a rule are more efficient if the model is correctly specified. Given the complexity of the underlying process, a correct specification is hard to achieve in practice, therefore direct forecasts are often preferred.

As a result, what is often available in practice for the decision maker is a set of marginal predictive distributions for N sites of interest and T lead times. For some decision tasks marginal densities are a suboptimal input, since the joint behaviour of power generation at all sites and the considered lead times might be of interest.

Having a set of marginal distributions, the joint density can be restored using a copula approach. One important feature of copulas is that they can be used to model dependency between stochastic variables independently of the marginal distribution functions. This is important because, as mentioned previously, modelling wind power generation at individual sites while targeting a specific lead time is already a complex task. Therefore, it is an advantage to decouple the problem of estimating marginal densities from the estimation of the space-time dependence structure.

Copulas have been widely used in many fields for modelling the dependence between stochastic variables, including a number of problems related to wind power. As an example in [5], predictive densities for wind power generation were built by modelling the relation between wind speed and wind power using copulas. In [30], copulas have been used to estimate system net load distribution when accounting for the dependence structure between wind activity at different locations and its relation to the system load. In [19], a copula has been used to model the dependence between wind speed at a number of sites.

In [31], the authors focused on a single wind farm. A Gaussian copula, fully characterized by an empirical covariance structure, has been used to derive joint predictive distributions (multivariate in time) from the set of marginal densities. Furthermore, in [26], the author placed emphasis on wind power generation at a pair of sites and, considered different types of copulas for modelling the dependence between wind power generation at these sites for a given lead time. The present study generalizes these works by looking at the interdependence of wind power generation in time and in space. It is aimed at issuing joint predictive density of wind power generation from a set of marginal predictive distribution. The problem then boils down to specifying and estimating a suitable dependence structure.

In this work a modelling approach taking an advantage of sparsity of precision matrix is introduced for the description of the underlying dependence structure. In order to make the methodology mathematically tractable in high dimensions, a parametrization of the precision matrix is proposed. This proposal goes beyond the conventional assumptions of homogeneous stationary Gaussian Random fields, since the presented parametrization accounts for the boundary points and considers non-constant conditional variances and direction-dependent conditional correlations.

The paper has the following outline. Section 2 introduces the data set used in the study. The methodology is described in Section 3. It consists of some preliminaries and definitions, introduction to copula modelling and explanation on how precision matrices relate to the Gaussian copula approach. Further, Section 4 presents the proposed parametrization of the dependence structure. The estimation process is discussed in Section 5, while the empirical results are given in Section 6. The paper finishes with the conclusions and perspectives presented in Section 7.

2 Data

The case study relates to western Denmark, including the Jutland peninsula and the island of Funen, which produces approximately 2.5 GW, or 70% of the entire wind power capacity installed in Denmark.

Besides the significant share of wind generation, a reason for placing the focus on Denmark is given by its climate and terrain characteristics. That is, the territory is small enough for the incoming weather fronts to affect all its parts. In addition, the terrain is smooth, therefore passing weather fronts do not meet any obstacles when propagating over the country. These aspects make the test case an ideal candidate for understanding space-time effects before moving to more complex cases.

The data selected for this work comes from 15 groups of wind farms spread throughout the territory of Western Denmark. The chosen grouping corresponds to the resolution map used by the Danish Transmission System Operator. For all 15 groups measurements of wind power production with an hourly resolution are available, along with the related marginal predictive densities, derived using the adaptive re-sampling method as described in [29]. This forecasting method is one of the state-of-the-art approaches which yields reliable probabilistic forecasts with high skill. [32].

Figure 1: *Geographical locations of the 15 zones of wind farms*

The available data covers a period from the 1st of January, 2006 to the 24th of October, 2007. The data set has been divided into two subsets. The first of them covering a period from the 1st of January, 2006 to the 30th of November, 2006 has been used for data analysis, the model building and the estimation. The second subset covering a period from the 30th of November, 2006 to the 24th of October, 2007 has been used for evaluating the predictive performance of the models.

3 Methodology

The objective of the methodology introduced here is to generate multivariate probabilistic forecasts describing wind power generation at a number of distributed locations and for a number of successive lead times.

The proposed approach follows two main steps. First, a state-of-the-art forecasting system is used to issue probabilistic forecasts for each location and each lead time individually. Subsequently, these are upgraded to full multivariate predictive densities using a copula function.

The focus in this work is on parametrization of the copula function.

3.1 Preliminaries and definitions

In general, the problem has the following setup. At every time step t the interest is in predicting wind power generation for times $t+1,\ t+2, \cdots, t+T$ at N distributed locations. That is, there are in total $n = NT$ quantities of interest which are denoted in the following by $Y_{t,1},\ Y_{t,2}, \cdots, Y_{t,n}$. The enumeration is done so that $Y_{t,1}, \cdots, Y_{t,T}$ represent wind power generation at the first location for the lead times $1, \cdots, T$, then $Y_{t,T+1}, \cdots, Y_{t,2T}$ represent wind power generation at the second location for the lead times $1, \cdots, T$, and so on.

Uppercase letters represent stochastic variables, while lowercase letters denote the corresponding observations. Bold font is used to emphasize vectors and matrices. For example, $\mathbf{y}_t = [y_{t,1},\ y_{t,2}, \cdots, y_{t,n}]^\top$ stands for the realization of $\mathbf{Y}_t$.

The aim of the forecaster is to issue a multivariate predictive distribution F_t, describing a random vector $\mathbf{Y}_t = [Y_{t,1},\ Y_{t,2}, \cdots, Y_{t,n}]^\top$

$$F_t(y_1, y_2, \cdots, y_n) = P(Y_{t,1} < y_1, Y_{t,2} < y_2, \cdots, Y_{t,n} < y_n) \tag{1}$$

There are two different families of approaches to probabilistic forecasting: parametric and non-parametric ones. The parametric approach refers to a distribution-based methodology, which requires an assumption on the shape of predictive densities. The non-parametric one refers to the distribution-free techniques, i.e. to the ones that are based on estimating the predictive densities directly from the data, without any constraints on the shape of the resulting distribution. An advantage of the non-parametric approach is given by the fact that it is fully data driven and, thus, can account for any level of asymmetry, any dependence structure, etc. The drawback, however, is that in high dimensions a fully non-parametric approach becomes intractable, even if only a climatological distribution is considered. If one wishes to issue conditional predictive densities, the curse of dimensionality becomes even more evident. Therefore, some parametrization ought to be proposed in order to make the estimation of predictive densities mathematically tractable.

Parametrizing F_t directly implies a simultaneous description of both marginal densities as well as the space-time interdependence structure. Considered distributions should account for the non-Gaussian, bounded nature of wind power generation as well for non-constant wind power variability. Unfortunately, there is no obvious distribution function which could address all the required aspects together. Copulas propose a solution by decomposing the problem of estimating F_t into two parts.

First, the focus is on marginal predictive densities, $F_{t,i} = P(Y_{t,i} < y_i),\ i = 1, 2, \cdots, n$, describing wind power generation at each location and for each lead time individually. As opposed to multivariate predictive densities, for which not many proposals exist in the literature, marginal predictive densities for wind power generation have been considered more. Thus, at this point the forecaster might take advantage of the state-of-the-art methods for probabilistic wind power forecasting. In this thesis an adaptive resampling has been used for obtaining marginal predictive densities $F_{t,i}$. The method was first described in [29]. The results documented both in [29] and in [32] confirm that it yields reliable wind power forecasts with high skill.

Subsequently, the marginal predictive densities are upgraded to F_t using a copula function. Mathematically the foundation of copulas is given by Sklar's theorem in [37]. The theorem states that: For any multivariate cumulative distribution function F_t with marginals $F_{t,1}$, $F_{t,2}$,...,$F_{t,n}$ there exists a copula C such that

$$F_t(y_1, y_2, ..., y_n) = C(F_{t,1}(y_1), F_{t,2}(y_2), ..., F_{t,n}(y_n)) \tag{2}$$

This means that, given a set of marginal distributions, the task of getting the joint distribution boils down to finding a suitable copula function.

3.2 Copulas for wind power data

In general, copulas can be classified into parametric and non-parametric. In this work focus is on the former ones, since the latter become intractable in very high dimensions.

Several parametric copula types have been considered for wind power data. Namely, in [31] the authors advocate that a Gaussian copula is an adequate choice when generating multivariate in time predictive densities when describing wind power generation at a single location.

In parallel, in [26] the author has considered different copula types for modelling the dependence between wind power generation at two sites when focusing on a single lead time. The results have shown that a Gumbel copula performs best, however Gaussian and Frank copulas also fit the data adequately.

When moving to higher dimensions, the construction of Archimedean copulas (e.g. Gumbel) becomes complex. For instance, a traditional approach for constructing the n-variate Gumbel copula requires the n^{th} order derivative of the inverse of the process generating function. Even considering explicit formulas for those derivatives given in [20], the complexity remains high compared to the Gaussian copula approach. Moreover, in Ref. [10] Guzman shows that in higher dimensions Gaussian copulas outperform their Gumbel's counterparts. However, the results should be interpreted with care as they depend on the site characteristics as well as on the type of the marginal predictive densities considered.

The works mentioned above indicate that the Gaussian copula is an adequate choice for describing spatial and temporal dependencies which are present in wind power data. However, these works have not considered spatio-temporal dependencies. Thus, the first step in this study involved a preliminary data examination to verify whether the Gaussian copula was consistent with the observed space-time dependence structure.

For example, consider $Y_{t,5*43+5}$ and $Y_{t,4*43+4}$ which represent wind power generation at zone 6 at time $t+5$ and wind power generation at zone 5 at $t+4$, respectively. The dependence between random variables $Y_{t,5*43+5}$ and $Y_{t,4*43+4}$ can be graphically represented looking at the ranks of the uniform variables $F_{t,5*43+5}(y_{t,5*43+5})$ and $F_{t,4*43+4}(y_{t,4*43+4})$.

The scatterplot of the corresponding ranks characterizes the dependence structure between $Y_{t,5*43+5}$ and $Y_{t,4*43+4}$, while the overlaying contour plot represents the so called empirical copula [13]. The empirical copula is then compared to the corresponding Gaussian copula and the results are illustrated in Fig. 2. Both patterns are very similar, and this is an indication that the Gaussian copula is appropriate for describing the spatio-temporal dependence structure. The results obtained while considering different pairs of variables have been qualitatively similar.

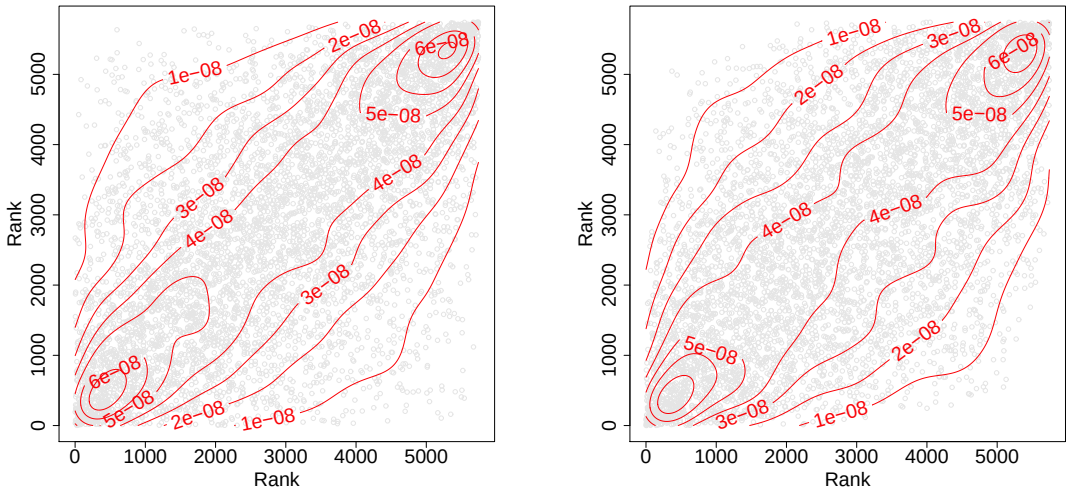
One should note, that the considered verification scheme does not guarantee that the Gaussian copula is the best choice for modelling the dependence structure. It should be only seen as an indication that there are no obvious inconsistencies between the Gaussian copula and the data. The reason not to consider other copula types has been given by a strong preference to use Gaussian copulas, since they have an advantage of being simple to use in high dimensions, widely used and having a strong theoretical linkage to a large class of mathematical theories.

3.3 Gaussian Copula

Gaussian copula is given by

$$C(F_{t,1}(y_1),...,F_{t,n}(y_n)) = \Phi_{\mathbf{\Sigma}}(\Phi^{-1}(F_{t,1}(y_1)),\cdots,\Phi^{-1}(F_{t,n}(y_n))) \tag{3}$$

where Φ^{-1} denotes the inverse of the univariate standard Gaussian distribution function and $\Phi_{\mathbf{\Sigma}}(.)$ is the n-variate Gaussian distribution function with zero mean, unit marginal variances and correlation matrix $\mathbf{\Sigma}$.



(a) Observed ranks

(b) Ranks of a simulated Gaussian process

Figure 2: *Left: Scatterplot with contour overlay given by the ranks of* $F_{t,5*43+5}(y_{t,5*43+5})$ *and* $F_{t,4*43+4}(y_{t,4*43+4})$*. Right: Scatterplot with contour overlay of the simulated bivariate Gaussian process having the same rank correlation as the observed data illustrated on the left.*

That is, the copula is built by transforming wind power generation $y_{t,i}$ to the latent standard Gaussian variable $x_{t,i}$ by applying the following:

$$x_{t,i} = \Phi^{-1}(F_{t,i}(y_{t,i})) \tag{4}$$

The resulting $\mathbf{x}_t = [x_{t,1}, ..., x_{t,n}]^\top$ are realization of the corresponding random process $\mathbf{X} = [X_1, ..., X_n]^\top$ which is distributed as multivariate Gaussian with zero mean, unit marginal variances and a correlation matrix $\mathbf{\Sigma}$, i.e.

$$\mathbf{X} \sim \mathcal{N}(0, \mathbf{\Sigma}) \tag{5}$$

In other words, it is assumed that a joint multivariate predictive density for $\mathbf{Y}_t$ can be represented by the multivariate Gaussian density in the transformed domain given by $\mathbf{X}$:

$$F_t(y_1, ..., y_n) = \Phi_{\mathbf{\Sigma}}(\Phi^{-1}(F_{t,1}(y_1)), \cdots, \Phi^{-1}(F_{t,n}(y_n))) \tag{6}$$

Note, that in this setup, even though the marginal distributions $F_{t,i}$ as well as the joint distributions F_t are time-dependent, the underlying dependence structure is fully represented by the time-invariant correlation matrix $\mathbf{\Sigma}$, thus there is no time index in the notation of the random variable $\mathbf{X}$.

The goal is to propose a sensible parametrization for $\mathbf{\Sigma}$. This is done by focusing on $\mathbf{X}$.

3.4 Modelling as a conditional autoregression

Consider, a set of available wind power observations corresponding to $\mathbf{y}_t$, $t = 1, \cdots, T$. The observations are transformed to the latent Gaussian variables $\mathbf{x}_t$ and the covariance structure of the latter ones is

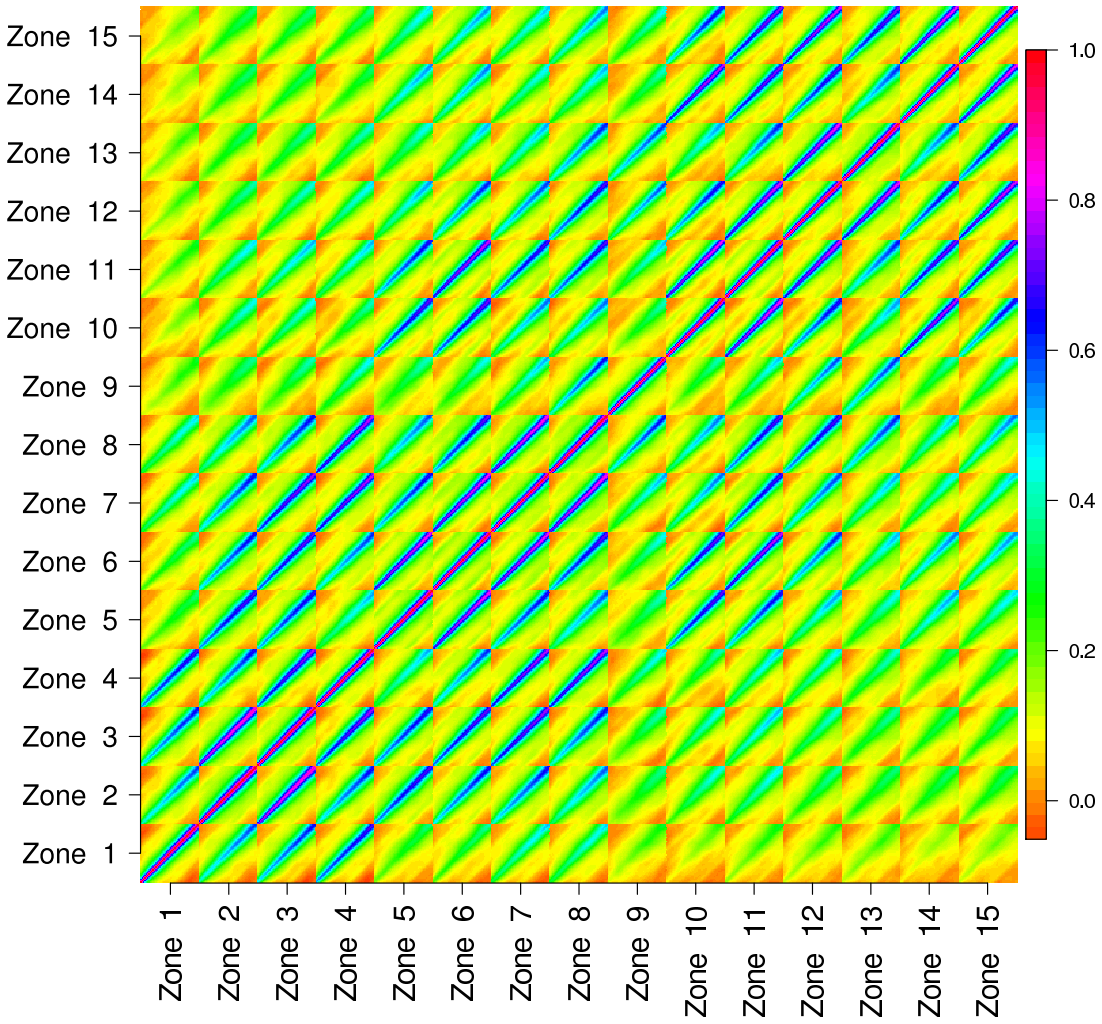


Figure 3: *Sample correlation matrix*

studied. As can be seen from Fig. 3, the sample covariance matrix, $\mathbf{\Sigma}$ is dense. This implies that inference with such a matrix has a computational complexity of $\mathcal{O}(n^3)$. In order to make the proposed methodology applicable for problems of high dimension, instead of modelling the covariance matrix directly, we focus on its inverse, denoted by $\mathbf{Q}$ [36]. The inverse of a covariance matrix is called a precision matrix.

The sample precision matrix (see Fig. 4) is very sparse. This opens the doors to the framework of Gaussian Markov Random Fields (GMRF), allowing us to benefit from computationally efficient algorithms derived for the inference with sparse matrices. More specifically, by switching from a dense covariance matrix to its sparse inverse, we reduce the computational complexity from $\mathcal{O}(n^3)$ to the range from $\mathcal{O}(n)$ to $\mathcal{O}(n^{3/2})$, depending on the process characteristics [36].

In contrast to covariance structure which informs of marginal dependence between variables, the precision matrix represents conditional interdependencies. The elements of the precision matrix have a useful conditional interpretation.

The diagonal elements of $\mathbf{Q}$ are the conditional precisions of X_i given $\mathbf{X}_{-i} = [X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n]^\top$ while the off-diagonal elements, with a proper scaling, provide information about the conditional correlations between variables. For a zero mean process, the following holds:

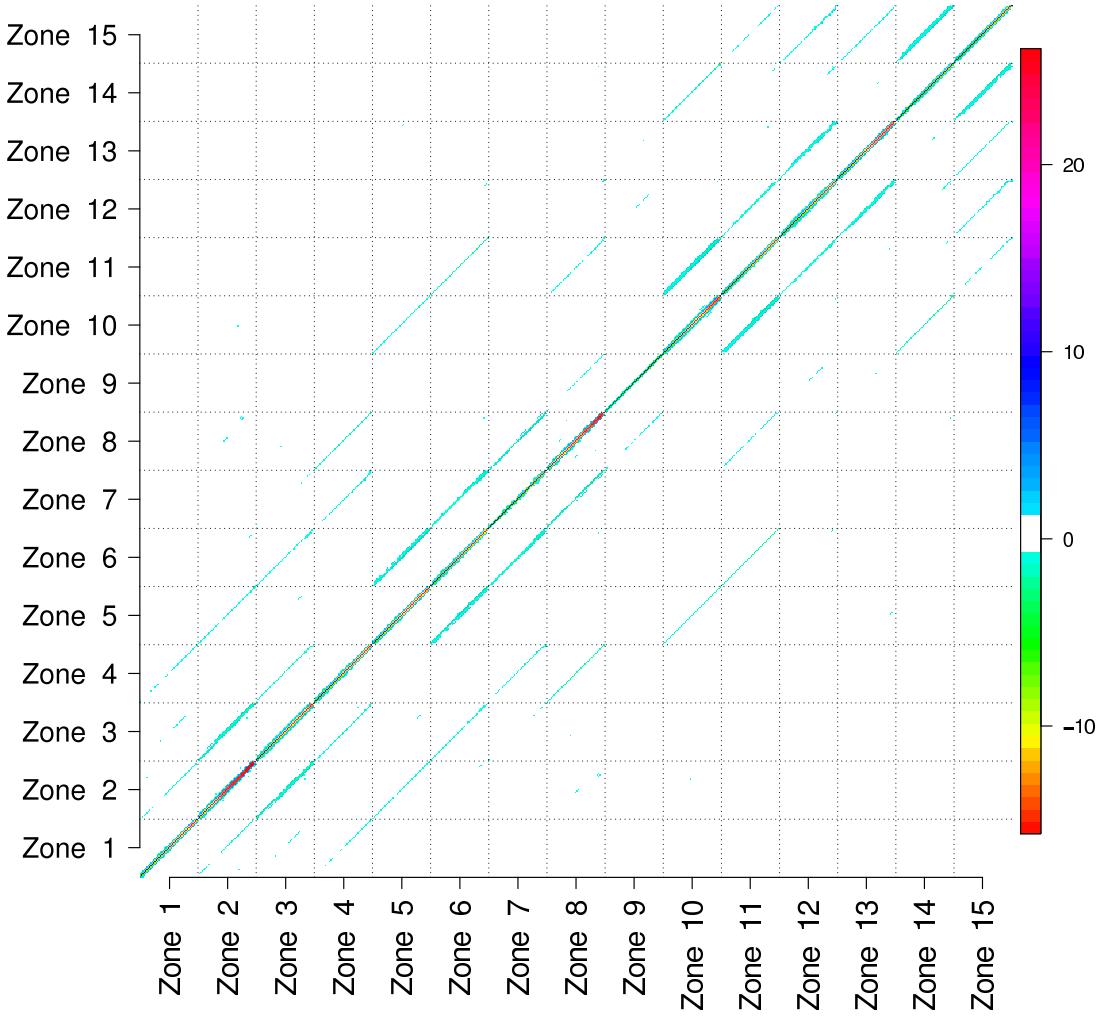


Figure 4: *Sample precision matrix*

$$E(X_i|\mathbf{X}_{-i}) = -\frac{1}{Q_{ii}} \sum_{j \neq i} Q_{ij} X_j \tag{7}$$

$$Var(X_i|\mathbf{X}_{-i}) = 1/Q_{ii} \tag{8}$$

A very important relation is that $Q_{ij} = 0$ if and only if elements X_i and X_j are independent given the rest, $\mathbf{X}_{-\{i,j\}}$. This means that the non-zero pattern of $\mathbf{Q}$ determines the neighbourhood of the conditional dependence between variables and can be used to provide parametrization of the precision matrix. Of course, one still has to keep in mind that $\mathbf{Q}$ is required to be symmetric positive-definite (SPD).

The relationship given by eq. (8) is sometimes used for an alternative specification of Gaussian Markov Random Field through full conditionals. This approach was pioneered by Besag in [3] and the resulting models are also known as conditional autoregressions, abbreviated as CAR. When specifying GMRF through CAR, instead of considering the entries of $\mathbf{Q}$, Q_{ij}, directly, focus is on modelling terms $\kappa_i = Q_{ii}$ and $\beta_{ij} = Q_{ij}/Q_{ii}$, $i, j = 1, \cdots, n$.

From eq. (8) it is seen that elements β_{ij} are given by the coefficients of the corresponding conditional autoregression models, while κ_i informs on the related variances.

This translates to the following equality:

$$\mathbf{Q} = \boldsymbol{\kappa} * \mathbf{B}. \tag{9}$$

where $\boldsymbol{\kappa}$ denotes a diagonal matrix of dimension $n \times n$, the diagonal elements of which are given by κ_i, $i = 1, \cdots, n$. $\mathbf{B}$ is a coefficient matrix consisting of coefficients β_{ij}. In other words, $\mathbf{B}$ equals the precision matrix standardized by its diagonal.

CAR specification is sometimes easier to interpret and we will use it to propose a parametrization for $\mathbf{Q}$ in this work.

4 Parametrization of the precision matrix

The CAR specification (see eq. 8) decouples the problem of describing $\mathbf{Q}$ into two parts. First, parametrization of conditional precisions, $\boldsymbol{\kappa}$ is discussed. Then, parametrization of the coefficient matrix $\mathbf{B}$ is presented.

4.1 Structure of the diagonal elements

Conventionally, CAR models are given by stationary GMRF. Stationarity implies rather strong assumptions on both the neighbourhood structure and the elements of $\mathbf{Q}$. Firstly, the structure of the neighbourhood allows for no special arrangement for the boundary points. Secondly, the full conditionals have constant parameters not depending on i. In other words, the conditional precisions given by κ_i, $i = 1, \cdots, n$ are assumed to be constant. However, the data analysis has shown this assumption would be very restrictive in this case.

The diagonal of the sample precision matrix $\mathbf{Q}$ is depicted in Fig. 5. One can note, that its elements are not constant. Their variation has some structure, which is captured in the following.

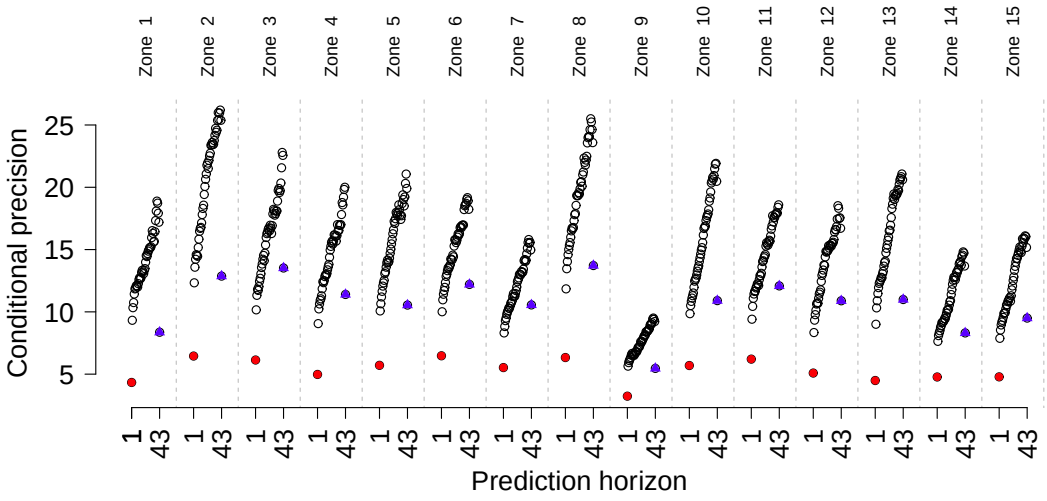


Figure 5: *Diagonal elements of the sample precision matrix,* $\mathbf{Q}$*. Boundary points given by the conditional precisions related to the horizons of 1 and 43 hours ahead are marked with red and blue circles, respectively*

4.1.1 Conditional precisions for different zones

First, it can be seen that the conditional precisions describing different zones are rather similar. The most significant deviation from the global picture is observed for zone 9. This is also in line with the results shown in [38] and could be explained by two main factors. On the one hand, it could be caused by the fact that group 9 covers a smaller territory compared to the other zones. This leads to more significant local variations, which results in the lower conditional precisions.

Another possible explanation is that zone 9, in contrast with the rest of the territories, is situated off the mainland. Therefore, it is very likely that the dynamics in zone 9 are different from the rest of the considered region.

If looking at the rest of the zones, then the observed pattern of conditional precisions is very similar. The differences are present, however, as of now, we have not been able to explain them by any of the available explanatory variables. An assumption that the precision pattern could depend on whether a zone is located in the centre of the considered territory or on the boundary has not been supported by the data. It has been also considered that patterns of conditional precisions could depend on the overall level of power variability. This, however, has not found a support in the data, either. Further investigation of this matter is left for future work. In this study it is considered that the pattern is the same for all zones. That is, any potential differences are disregarded and the following parametrization is proposed:

$$diag(\mathbf{Q}) = [\kappa_1, \kappa_2, \cdots, \kappa_{645}]^\top = \tag{10}$$

$$= [\kappa_1, \kappa_2, \cdots, \kappa_{43}, \kappa_1, \kappa_2, \cdots, \kappa_{43}, \cdots, \cdots, \kappa_1, \kappa_2, \cdots, \kappa_{43}]^\top \tag{11}$$

where $\mathbf{K} = [\kappa_1, \kappa_2, \cdots, \kappa_{43}]^\top$ is a vector of conditional precisions corresponding to a single zone.

4.1.2 Conditional precisions for different lead times

Since it has been assumed that, at zone level, the conditional precisions follow the same dynamics, focus is on a single zone. The corresponding conditional precisions are given by $\mathbf{K} = [\kappa_1, \kappa_2, \cdots, \kappa_{43}]^\top$. κ_1 and κ_{43} correspond to the temporal boundary and this explains why they stand out from the general pattern

as show in Fig. 5. The temporal boundary for κ_1 and κ_{43} is given by the fact that we do not consider lead times of less than 1 hour ahead and more than 43 hours ahead.

Further from the temporal boundaries, i.e. for the lead times from 2 to 42 hours ahead, Fig. 5 suggests that the conditional precisions increase with the lead time. For accounting for this effect, the following parametrization is proposed:

$$\kappa_i = \kappa_{i-1} * \rho \tag{12}$$

for $i = 2, \cdots, 42$. Here ρ is a ratio parameter.

4.1.3 Final parametrization of the conditional precisions

Summarizing the reasoning presented in the previous sections, the following parametrization for the diagonal elements of the precision matrix is proposed:

$$\boldsymbol{\kappa} = \begin{array}{c} \text{zone} \\ 1 \\ 2 \\ \vdots \\ 15 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & 15 \end{array} \\ \begin{pmatrix} \mathbf{K} & & & \\ & \mathbf{K} & & \\ & & \ddots & \\ & & & \mathbf{K} \end{pmatrix} \end{array} \tag{13}$$

where

$$\mathbf{K} = \begin{array}{c} \text{lead time} \\ 1 \\ 2 \\ 3 \\ \vdots \\ 42 \\ 43 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & \cdots & 42 & 43 \end{array} \\ \begin{pmatrix} q_1 & & & & & \\ & \rho & & & & \\ & & \rho^2 & & & \\ & & & \ddots & & \\ & & & & \rho^{41} & \\ & & & & & q_{43} \end{pmatrix} \end{array} \frac{1}{\sigma^2} \tag{14}$$

Thus, the diagonal of $\mathbf{Q}$ can be described with four parameters. q_1 and q_{43} describe the temporal boundary conditions, ρ describes a proportional increase in conditional precisions and σ^2 represents a base level of variation.

4.2 Structure of the standardized precision matrix

Next step is to propose a parametrization for $\mathbf{B}$. This requires understanding the neighbourhood structure of $\mathbf{Q}$, i.e. identifying which elements are non-zero.

4.2.1 Spatial neighbourhood

Consider a single zone, further denoted by A. A careful look at Fig. 4 reveals that information at zone A is only dependent on local information at A and on the four closest neighbouring zones: Northern (N), Eastern (E), Southern (S) and Western (W) neighbours of A (see Fig. 6).

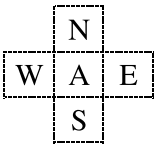


Figure 6: *Neighbourhood specification of a single zone. The focus zone is marked A, while W, N, E and S denote its Western, Northern, Eastern and Southern neighbours, respectively.*

4.2.2 Temporal neighbourhood

Fig. 4 shows that information observed at zone A at time t is only dependent on a very small amount of elements at zones A, N, E, S, and W.

Since precision matrices ought to be symmetric, it is sufficient to focus on the dependency between A and its Western and Southern neighbours, without direct consideration of the Eastern and Northern neighbours. Let us zoom-in to some relevant blocks of the sample coefficient matrix $\mathbf{B}$ when focusing on zone 6.

From the results depicted in Fig. 7 one can note that the corresponding conditional correlations of zone A with its North and the West side neighbours differ. Information at zone A observed at time t is conditionally dependent only on the simultaneous information at zone N. Meanwhile, the conditional correlation with zone W is significant at times $t-j$, $j=-2,\cdots,2$. This difference in the dependency pattern can be partly explained by the fact that in Denmark prevailing winds are westerly. Thus, forecast errors most often propagate from West to East, as discussed in [15]. This means that usually zones A and N are influenced by the upcoming weather front at similar time, while zone W is exposed to it earlier. Of course, one should also keep in mind, that in our test case distances between zones A and N are larger than those between A and W, and this can be another factor influencing different patterns of the related dependencies.

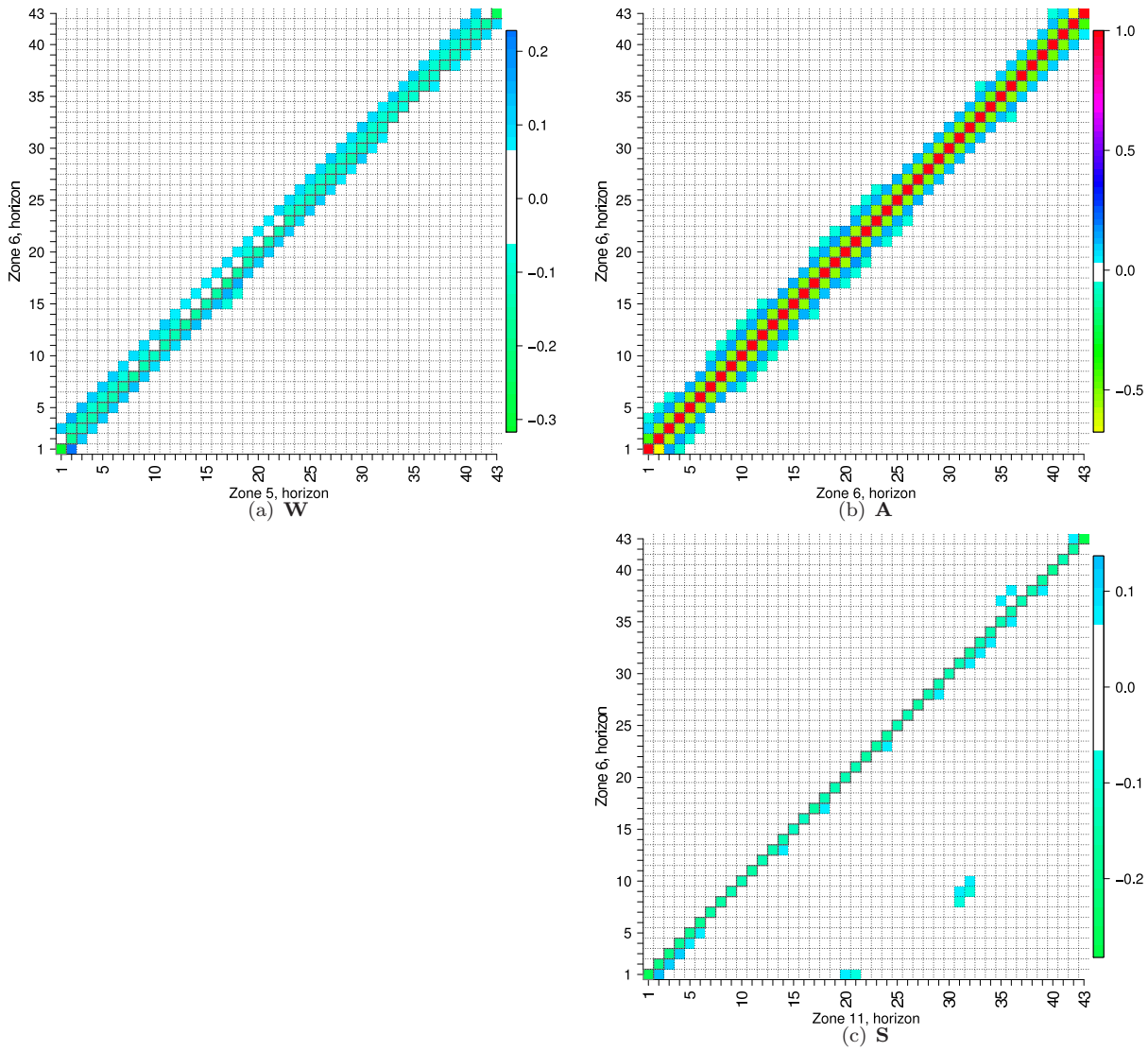
In general, the results depicted in Fig. 7 show that information corresponding to lead time h for zone A is dependent on the variables at the neighbouring zones corresponding to lead times $h-j$, where $j=-2,\cdots,2$. Thus, visually the data suggests a second order (temporal) process. In this work both the second ($j=-2,\cdots,2$) and the first order ($j=-1,0,1$) models have been considered. Since the corresponding difference in the performance of the resulting predictive densities was rather minor, in this study the focus is on the first order model ($j=1$). Extension to higher order models is rather straight-forward and all the discussed parametrization and estimation procedures apply.

In this work a directional non-stationary CAR model, abbreviated as DCAR, is considered. That is, the conditional correlations are made direction-dependent. In this respect the work is inspired by [22] where the authors consider a directional (in space) CAR model. We refer the reader to that work for a clear description of the modelling approach. The current proposal can be viewed as a generalization of the work presented in [22] since space-time neighbourhoods are considered along with the non-constant precisions.

When considering DCAR models, directional neighbourhoods should be chosen carefully so that each of them forms a (directional) clique. That is, consider two elements from the full random vector $\mathbf{X}$: X_i and X_j. Then, given that X_i is a "west-side" and "one-hour-ago" neighbour of X_j, X_j should be assigned as the "east-side" and "one-hour-ahead" neighbour of X_i. This is essential for ensuring the symmetry of the precision matrix.

4.2.3 Final parametrization of the standardized matrix

Summarizing, data analysis has suggested that information coming from zone A at lead time h conditionally depends only on information coming from zones N, E, S, and W with lead times $h-1$, h, $h+1$ and on the local situation at zone A for lead times $h-1$ and $h+1$. In terms of the CAR specification



(a) **W** (b) **A** (c) **S**

Figure 7: *Zoomed in blocks of the standardized (by its diagonal) sample precision matrix*

given in eq. (8) this translates to:

$$E(x_h^{(A)}) = -\sum_{j=\{-1,1\}} a_j x_{h+j}^{(A)} - \sum_{j=\{-1,0,1\}} (b_j x_{h+j}^{(W)} + b_j^* x_{h+j}^{(E)} + c_j x_{h+j}^{(N)} + c_j^* x_{h+j}^{(S)}) \tag{15}$$

Here $x_h^{(\cdot)}$ refers to a single element from the latent Gaussian vector $\mathbf{x}$ corresponding to the information obtained at zone "." when considering marginal forecasts for lead times h. a_j, b_j, b_j^*, c_j and c_j^* denote the corresponding coefficients which are the building blocks for $\mathbf{B}$.

Data analysis has shown that a_j, b_j, b_j^*, c_j and c_j^* do not depend on the considered lead time h. It can be also seen in Fig. 7 that there is no indication of any increase/decrease of coefficient values with the lead time. The only values which drop out from the constant picture are the ones corresponding to the temporal boundaries and this will be accounted for when scaling by the corresponding conditional precisions.

In this work it is assumed that the corresponding coefficients are constant for all zones. Further work could be done in order to explain spatial variation in the coefficient values.

Some restrictions have to be imposed on the parameters in $\mathbf{B}$ to ensure that the resulting $\mathbf{Q}$ is symmetric positive definite. Imposing symmetry reduces the parameter space significantly, since coefficients a_1 can be derived from a_{-1}, b_j^* from b_{-j} and c_j^* from c_{-j}, $j = -1, 0, 1$. This will be formulated below in eq. (23).

4.3 Final parametrization of the precision matrix

The precision matrix is given by:

$$\mathbf{Q} = \boldsymbol{\kappa}\mathbf{B} \tag{16}$$

where $\boldsymbol{\kappa}$ represents the diagonal elements which are assumed to be independent on the considered zone, but dependent on the lead time:

$$\boldsymbol{\kappa} = \begin{array}{c} \text{zone} \\ 1 \\ 2 \\ \vdots \\ 15 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & 15 \end{array} \\ \begin{pmatrix} \mathbf{K} & & & \\ & \mathbf{K} & & \\ & & \ddots & \\ & & & \mathbf{K} \end{pmatrix} \end{array} \tag{17}$$

where $\mathbf{K}$ describes how conditional precisions change with the lead time. It is given by:

$$\mathbf{K} = \begin{array}{c} \text{lead time} \\ 1 \\ 2 \\ 3 \\ \vdots \\ 42 \\ 43 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & \cdots & 42 & 43 \end{array} \\ \begin{pmatrix} q_1 & & & & & \\ & \rho & & & & \\ & & \rho^2 & & & \\ & & & \ddots & & \\ & & & & \rho^{41} & \\ & & & & & q_{43} \end{pmatrix} \end{array} \frac{1}{\sigma^2} \tag{18}$$

Standardized by the diagona) precision matrix is given by $\mathbf{B}$:

$$\mathbf{B} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \end{array} \begin{array}{c} \begin{array}{ccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \end{array} \\ \begin{pmatrix} \mathbf{A} & & & \mathbf{S} & & & & & & & & & & & \\ & \mathbf{A} & \mathbf{E} & & \mathbf{S} & & & & & & & & & & \\ & \mathbf{W} & \mathbf{A} & \mathbf{E} & & \mathbf{S} & & & & & & & & & \\ \mathbf{N} & & \mathbf{W} & \mathbf{A} & & & & \mathbf{S} & & & & & & & \\ & \mathbf{N} & & & \mathbf{A} & \mathbf{E} & & & & \mathbf{S} & & & & & \\ & & \mathbf{N} & & \mathbf{W} & \mathbf{A} & \mathbf{E} & & & & \mathbf{S} & & & & \\ & & & & & \mathbf{W} & \mathbf{A} & \mathbf{E} & & & & & & & \\ & & & \mathbf{N} & & & \mathbf{W} & \mathbf{A} & \mathbf{E} & & & \mathbf{S} & & & \\ & & & & & & & \mathbf{W} & \mathbf{A} & & & & \mathbf{S} & & \\ & & & & \mathbf{N} & & & & & \mathbf{A} & \mathbf{E} & & & \mathbf{S} & \\ & & & & & \mathbf{N} & & & & \mathbf{W} & \mathbf{A} & \mathbf{E} & & & \mathbf{S} \\ & & & & & & & \mathbf{N} & & & \mathbf{W} & \mathbf{A} & \mathbf{E} & & \\ & & & & & & & & \mathbf{N} & & & \mathbf{W} & \mathbf{A} & & \\ & & & & & & & & & \mathbf{N} & & & & \mathbf{A} & \mathbf{E} \\ & & & & & & & & & & \mathbf{N} & & & \mathbf{W} & \mathbf{A} \end{pmatrix} \end{array} \tag{19}$$

where $\mathbf{W}$, $\mathbf{N}$ represent the blocks of conditional dependencies between the focus zone and its Western and Northern neighbours, respectively, while $\mathbf{A}$ represent local dependencies at zone A itself. The blocks are parametrized in the following way:

$$\mathbf{A} = \begin{array}{c} \text{lead time} \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 41 \\ 42 \\ 43 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & \cdots & 41 & 42 & 43 \end{array} \\ \begin{pmatrix} 1 & \frac{\rho}{q_1}a_{-1} & & & & & & \\ a_{-1} & 1 & \rho a_{-1} & & & & & \\ & a_{-1} & 1 & \rho a_{-1} & & & & \\ & & a_{-1} & 1 & & & & \\ & & & & \ddots & & & \\ & & & & & 1 & \rho a_{-1} & \\ & & & & & a_{-1} & 1 & \rho a_{-1} \\ & & & & & & \frac{\rho^{42}}{q_{43}}a_{-1} & 1 \end{pmatrix} \end{array} \tag{20}$$

$$\mathbf{W} = \begin{array}{c} \text{lead time} \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 41 \\ 42 \\ 43 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & \cdots & 41 & 42 & 43 \end{array} \\ \begin{pmatrix} b_0 & \frac{b_1}{q_1} & & & & & & \\ b_{-1} & b_0 & b_1 & & & & & \\ & b_{-1} & b_0 & b_1 & & & & \\ & & b_{-1} & b_0 & & & & \\ & & & & \ddots & & & \\ & & & & & b_0 & b_1 & \\ & & & & & b_{-1} & b_0 & b_1 \\ & & & & & & \frac{\rho^{42}}{q_{43}}b_{-1} & b_0 \end{pmatrix} \end{array} \tag{21}$$

$$\mathbf{N} = \begin{array}{c} \text{lead time} \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 41 \\ 42 \\ 43 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & \cdots & 41 & 42 & 43 \end{array} \\ \begin{pmatrix} c_0 & \frac{c_1}{q_1} & & & & & & \\ c_{-1} & c_0 & c_1 & & & & & \\ & c_{-1} & c_0 & c_1 & & & & \\ & & c_{-1} & c_0 & & & & \\ & & & & \ddots & & & \\ & & & & & c_0 & c_1 & \\ & & & & & c_{-1} & c_0 & c_1 \\ & & & & & & \frac{\rho^{42}}{q_{43}}c_{-1} & c_0 \end{pmatrix} \end{array} \tag{22}$$

with

$$\begin{aligned}\mathbf{E} &= \mathbf{K}^{-1}\mathbf{W}^{\top}\mathbf{K} \\ \mathbf{S} &= \mathbf{K}^{-1}\mathbf{N}^{\top}\mathbf{K}\end{aligned} \tag{23}$$

to ensure symmetry of $\mathbf{Q}$

Thus, we can model $\mathbf{Q}$ as a function of a parameter vector $\boldsymbol{\theta}$, where:

$$\boldsymbol{\theta} = [q_1, \rho, q_{43}, \sigma^2, a_{-1}, b_0, b_{-1}, b_1, c_0, c_{-1}, c_1]^{\top} \tag{24}$$

5 Estimation

This section discusses how to fit the GMRF defined by $\mathbf{Q}(\boldsymbol{\theta})$ to the observations. This task can be divided into two parts. Firstly, one needs to decide on the discrepancy measure between the observed data and the suggested GMRF. Secondly, one needs to propose a way to ensure that the parameter estimates belong to the valid parameter space $\boldsymbol{\Theta}^{+}$ which would ensure that the resulting precision matrix is symmetric positive definite (SPD).

5.1 The valid parameter space

In section 4 the precision matrix $\mathbf{Q}$ is described as a function of the parameter vector $\boldsymbol{\theta}$. In this section we discuss how to ensure that parameter estimates $\hat{\boldsymbol{\theta}}$ belong to the valid parameter space $\boldsymbol{\Theta}^{+}$ which would ensure that the resulting precision matrix $\mathbf{Q}(\hat{\boldsymbol{\theta}})$ is SPD.

Symmetry of $\mathbf{Q}$ is imposed by its construction (see Section 4). Thus, we are left with the concerns of whether the matrix is positive definite.

Unfortunately, in general it is hard to determine $\boldsymbol{\Theta}^{+}$. There are some analytical results available for precision matrices that are Toeplitz [34]. This could be used when working with homogeneous stationary GMRF, but this is not the case in this study. When there is no knowledge on $\boldsymbol{\Theta}^{+}$ available, the common practice is to consider a subset of $\boldsymbol{\Theta}^{+}$ which is given by the sufficient condition of $\mathbf{Q}$ being diagonal dominant.

Diagonal dominance is most often easy to treat analytically. On the downside, this approach becomes more and more restrictive for an increasing number of parameters. This issue is discussed in more detail in [34]. For instance, for our particular test case we could see that the assumption of diagonal dominance was too restrictive, as the estimated parameters (if no such restriction was imposed) far from fulfilled the criterion of diagonal dominant precision matrix.

If the full valid parameter space $\boldsymbol{\Theta}^{+}$ is unknown and its diagonal dominant subset is deemed as too restrictive, it is always possible to use a "brute force" approach (following terminology of [34]). This entails checking if $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}^{+}$ by direct verification of whether the resulting $\mathbf{Q}(\hat{\boldsymbol{\theta}})$ is SPD or not. This is most easily done by trying to compute the Cholesky factorization which will be successful if and only if $\mathbf{Q}$ is positive definite. The "brute force" method was the one used in this work.

However, it is worth mentioning some advantages given by the diagonal dominance approach over the "brute force" method. An important one is that if one estimates parameters while requiring the diagonal dominance, then one can be sure that if a new territory is to be included to the considered setup, there is no strict necessity (other than aiming for optimality) to re-estimate the parameters. In other words, one can be sure that the "old" parameter vector would guarantee a valid covariance structure for the enlarged lattice. This is not exactly the case for the "brute force" approach. If we want to take an additional zone into consideration, we cannot be guaranteed that the previously estimated parameters would result

in a valid covariance structure. That is, we might need to re-estimate. However, the experiments have shown a "new" set of parameters being very close to the "old" one. Thus, if we use previously estimated parameters as the initial condition for the optimization routine, then we can expect to get fast estimates of the "new" parameter vector.

5.2 Choosing an appropriate optimization criterion

When estimating $\boldsymbol{\theta}$ from the real data, one needs to decide on some discrepancy measure between the imposed GMRF and the observations.

In this work we focused on parameter estimation using maximum likelihood theory. In [35] the authors argue that maximum likelihood estimators for GMRF are not robust with respect to model errors and might result in coefficient estimates which do not describe well the global properties of the data. See [35] for more details. The authors propose a new optimization criterion which resolves this difficulty. The criterion is based on a norm distance between the estimated and the observed correlation structures. In this work we considered both the norm- based discrepancy optimization and the likelihood approach. Since estimates obtained with both approaches were consistent, further focus is on the likelihood based inference. This choice is made, since, following [35], if a GMRF describes the data adequately, then maximum likelihood-based inference is more efficient than the norm-optimization. The reader is referred to [35] for a broader discussion on the existing alternatives.

5.3 Parameter estimation using maximum likelihood

Let us focus on a single time t and recall some of the notation introduced in Section 3. The corresponding observation of the latent Gaussian field $\mathbf{x}_t = [x_{t,1}, x_{t,2}, \cdots, x_{t,n}]^\top$ is then given by the corresponding transformations of the related power measurements $y_{t,1}, y_{t,2}, \cdots, y_{t,n}$. That is,

$$\mathbf{x}_t = [\Phi^{-1}(F_{t,1}(y_{t,1})), \Phi^{-1}(F_{t,2}(y_{t,2})), \cdots, \Phi^{-1}(F_{t,n}(y_{t,n})]^\top$$

The essence of the presented methodology is based on the assumption that $\mathbf{x}_t$ follows a multivariate Gaussian distribution with zero mean and correlation matrix given by $\mathbf{Q}^{-1}$.

Then the log likelihood contribution given by $\mathbf{x}_t$ writes as:

$$l_t = -\frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln|\mathbf{Q}| - \frac{1}{2}(\mathbf{x}_t)^\top \mathbf{Q}\mathbf{x}_t \tag{25}$$

Given H realizations of the random process $\mathbf{X}$, the overall likelihood is given by

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{H} l_t = -\frac{nH}{2}\ln(2\pi) + \frac{H}{2}\ln|\mathbf{Q}(\boldsymbol{\theta})| - \frac{1}{2}\sum_{t=1}^{H}(\mathbf{x}_t)^\top \mathbf{Q}(\boldsymbol{\theta})\mathbf{x}_t \tag{26}$$

By solving $\dfrac{\partial l(\boldsymbol{\theta})}{\partial \sigma^2} = 0$ with respect to σ^2 yields the following profile maximum likelihood estimator for σ^2

$$\widehat{\sigma^2} = \frac{\sum_{t=1}^{H} \mathbf{x}_t \top \mathcal{P} \mathbf{x}_t}{Hn} \tag{27}$$

with

$$\mathcal{P} = \begin{array}{c} \text{lead time} \\ 1 \\ 2 \\ 3 \\ \vdots \\ 42 \\ 43 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & \cdots & 42 & 43 \end{array} \\ \begin{pmatrix} q_1 & & & & & \\ & \rho & & & & \\ & & \rho^2 & & & \\ & & & \ddots & & \\ & & & & \rho^{41} & \\ & & & & & q_{43} \end{pmatrix} \end{array} \tag{28}$$

Having the profile likelihood estimate for σ^2, we view $\mathbf{Q}$ as a function of the parameter vector $\boldsymbol{\theta}^-$:

$$\boldsymbol{\theta}^- = [q_1, \rho, q_{43}, a_{-1}, b_0, b_{-1}, b_1, c_0, c_{-1}, c_1]^\top \tag{29}$$

Estimate for $\boldsymbol{\theta}^-$ is obtained by a numerical optimization of the likelihood function given in eq. (26) with respect to the parameter vector $\boldsymbol{\theta}^-$.

The requirement for the resulting $\hat{\mathbf{Q}}$ to be symmetric positive definite is equivalent to requiring all eigenvalues to be positive. Similarly to [35], we approach the constrained optimization problem as an unconstrained one, adding an infinite penalty if some of the eigenvalues are negative. This approach works well in practice.

Also, $\boldsymbol{\Sigma} = \mathbf{Q}^{-1}$ is required to have a unit diagonal. In practice this is achieved by the corresponding scaling of the estimate $\hat{\mathbf{Q}}$ as suggested in [34].

6 Results

6.1 Assessing global model fit

Verification starts with examination of the global properties of the estimated dependence structure. This is done in the spirit of [35], i.e. by visually comparing the estimated covariance structure with the sample one. The estimated correlation matrix is illustrated in Fig. 8, while the sample one is shown in Fig. 3. The fit seems adequate.

The motivation for checking the global resemblance between the dependence structures in addition to the overall likelihood evaluation is given by the following. When optimizing the likelihood, the optimal fit is given by fitting the covariances within the neighbourhood exactly, while the remaining ones are determined by the inversion of the fitted precision matrix [35]. This may result in the estimates, which instead of capturing dependencies between all the variable pairs in some reasonable way, capture just some of them with a very high precision, while ignoring the others.

The fact, that the the estimated covariance matrix is in line with the the sample one, indicates that the model describes the data adequately and that the resulting joint density can be used for inferring on the global properties of the process.

6.2 Assessing predictive model performance

In this section focus is on evaluating predictive performance of the derived probabilistic forecasts. While the first year of data covering a period from the 1st of January, 2006 to the 30th of November, 2006 has been used for model estimation, validation is performed using another data subset, which is covering a period from the 30th of November, 2006 to the 24th of October, 2007.

The section starts with a presentation of the benchmark approaches. Further, scores used for the overall quality assessment are discussed. Finally, the empirical results are presented.

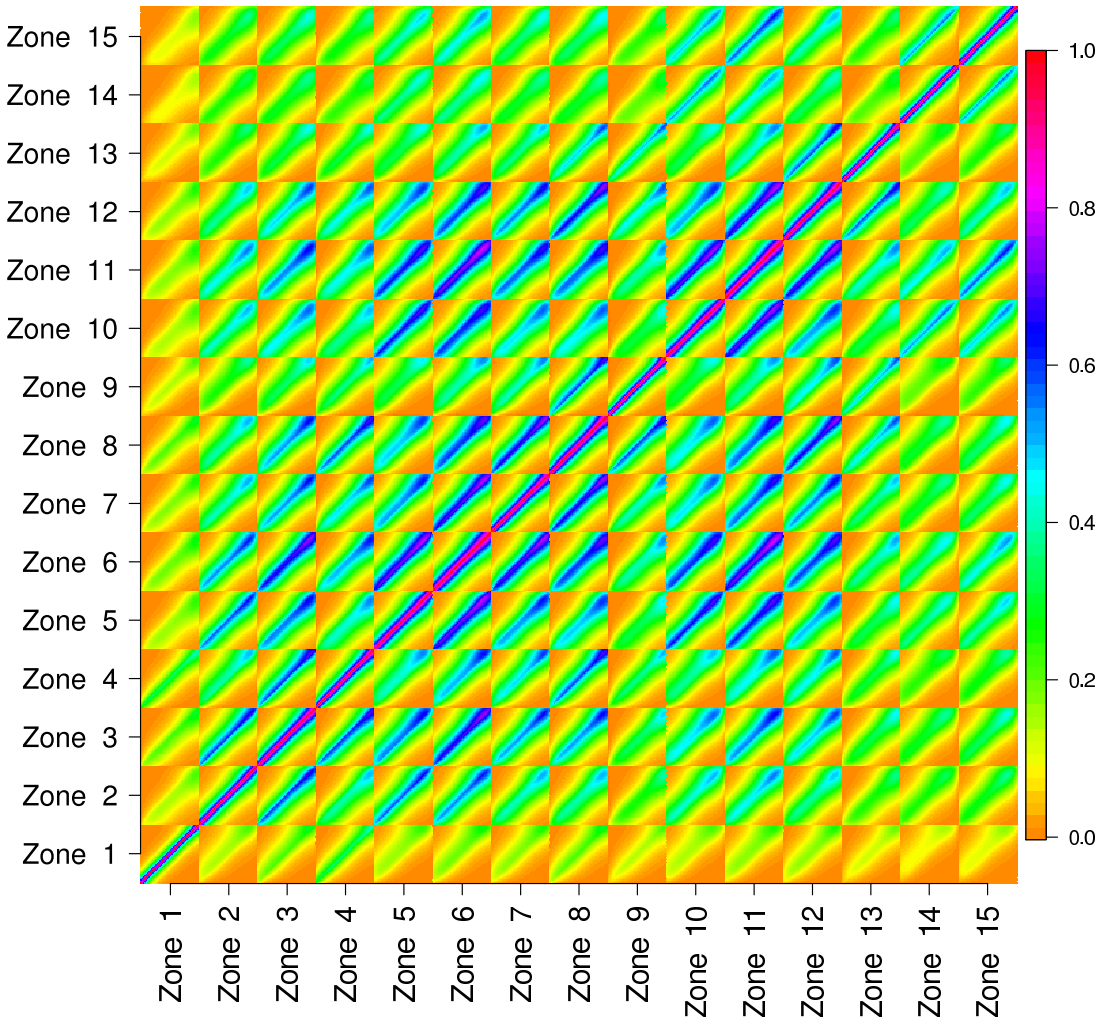


Figure 8: *Estimated correlation matrix*

6.2.1 Considered models

The following models are considered in this study:

1. *Independent*: The corresponding multivariate predictive densities are based on the assumption that the marginal densities are independent. That is:

$$F_t(y_1, y_2, \cdots, y_n) = F_{t,1}(y_1)F_{t,2}(y_2)\cdots F_{t,n}(y_n) \tag{30}$$

2. *First order time-dependence*: The corresponding multivariate densities are obtained using a Gaussian copula approach. The covariance matrix accounts only for the temporal dependencies while completely ignoring the spatial ones. This is done by constructing the precision matrix $\mathbf{Q}$ as described in Section 4.3, but setting $\rho = q_1 = q_{43} = 1$ and $b_{-1} = b_0 = b_1 = c_{-1} = c_0 = c_1 = 0$. That is, the precision matrix in this case is described by the parameters a_1 and σ^2 only. This model does not allow for any special arrangement for the boundary points. The conditional precisions are assumed to be constant. In other words, this model corresponds to a conventional stationary GMRF defined by the first order autoregressive process in time.

3. *Separable model with first order decays in time and in space* allowing for non-constant conditional precisions: The corresponding multivariate densities are obtained using a Gaussian copula approach. The precision matrix $\mathbf{Q}$ is parametrized as in Section 4.3 while setting $c_0 = b_0$, $b_1 = b_{-1} = c_1 = c_{-1} = a_1 * b_0$. That is, the precision matrix in this case is described the first order time-dependence (given by a_1) and the first order spatial dependence (given by b_0). Additionally, the model gives more flexibility compared with the conventional separable covariance structures by considering non-constant conditional precisions (modelled by ρ, q_1 and q_{43}). The model does not account for the directional influence, and that is why c_j is set to be equal to b_j with all $j = -1, \cdots, 1$

4. *Sample correlation*: The corresponding multivariate predictive densities are obtained using a Gaussian copula approach with the correlation structure given by the sample correlation matrix.

5. *Full model*: The first order model which proposed in this study. That is the precision matrix is described by the full parameter vector $\boldsymbol{\theta}$ as given in eq. (24).

6.2.2 Choosing an appropriate scoring rule for the quality evaluation

In order to evaluate and compare the overall quality of multivariate probabilistic forecasts proper scoring rules are to be employed [8, 17]. An overview of proper scoring rules used for the multivariate forecast verification is given in [18]. In this work the Logarithmic score is used as a lead score for evaluating the performance of the joint predictive densities. The logarithmic scoring rule, s, is defined as

$$s(p(\mathbf{x}), \mathbf{x}_t) = -\ln(p(\mathbf{x}_t)) \tag{31}$$

Where $p(\mathbf{x})$ stands for the predictive density, which in our case is given by $\mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})^{-1})$. $\mathbf{x}_t$ denotes the corresponding observation.

Suppose, the verification set consists of H observations, then the overall score, S, is given by the average value of the corresponding $s(p(\mathbf{x}), \mathbf{x}_t)$

$$S(p(\mathbf{x})) = -\frac{\sum_{t=1}^{H} \ln(p(\mathbf{x}_t))}{H} \tag{32}$$

That is, essentially the Logarithmic score is given by the average minus log likelihood derived from the observations. Therefore, this score is negatively orientated.

Table 1: *Quality assessment of the predictive densities in terms of the Logarithmic score (S).*

Model	Nr. of parameters	S
Independent	0	853.14
First order in time	1	409.98
Separable space-time model	6	357.84
Full model	10	318.07
Sample correlation	207690	267.96

There are several reasons for choosing the Logarithmic score as the lead evaluation criterion.

Firstly, it is consistent with the optimization criterion used when estimating the model parameters.

Secondly, allowing for some affine transformations, this is the only local proper score (see Theorem 2 in [2]). Locality means that the score depends on the predictive distribution only through the value which the predictive density attains at the observation [8]. An important advantage of using local scores when dealing with multivariate predictive densities comes with the related computational benefits. When dealing with local scores, there is no need to draw random samples from the predictive density in order to make the evaluation.

For instance, an alternative is to use the Energy score (see detailed information on this in [18]). This score is non-local and is based on the expected Euclidean distance between forecasts and the related observations. Most often, closed form expressions for such expectation are unavailable and one needs to employ Monte Carlo methods in order to estimate the score [18]. When dealing with problems of a very high dimension, Monte Carlo techniques result in computational challenges.

On the downside of local scores is their sensitivity to outliers. For instance, the Logarithmic score is infinite if the forecast assigns a vanishing probability to the event which occurs. In practice, when working with the real data, such sensitivity might be a problem.

In this work, we considered both the Energy score and the Logarithmic score for the final density evaluation. In general the results suggested by the two scores were consistent and no contradictions were observed. However, what we noticed is that the Energy score was not very sensitive to the changes in the correlation structure. That is, the changes in the Energy score when moving from the assumption of independence between the marginal predictive densities to models accounting for the dependence structure were rather small (even though they still proved statistically significant based on Diebold-Mariano test statistics [11]). This is caused by low sensitivity of the Energy score to changes in the dependence structure as argued in [33]. This is another reason to focus on the Logarithmic score further in this study.

6.2.3 Empirical results

One can appreciate the importance of accounting for the dependence structure from the fact that multivariate predictive densities derived from the independence assumption are shown to be of the lowest quality (see results in Table 1). The full model proposed in this study outperforms another two considered dependency structures: first order time-dependence as well as the separable space-time model. Statistical significance of the improvements was verified using a likelihood ratio test [27]. This confirms that letting the related conditional correlations change depending on the direction as well as allowing for non-separable space-time influence results in better quality of the multivariate probabilistic forecasts.

Predictive densities defined by the sample correlation matrix provide the best quality forecasts. This is also expected, since in this study the estimation period consisted of one year of hourly data. Large amount of data made it possible to estimate the covariance structure of the given dimension. However, the main interest in the future is to make the covariance structure dependent on meteorological conditions. In this setup, tracking sample covariance will become impossible. Thus, the proposed parametrization is crucial

for further development of the methodology as it significantly reduces the effective problem dimension.

6.3 Scenario generation

As an illustration of probabilistic forecasts obtained with the proposed approach Fig. 9 shows five scenarios describing wind power generation at zones 6 and 7 from 1 to 43 hours ahead issued on the 15th of June, 2007, at 01:00.

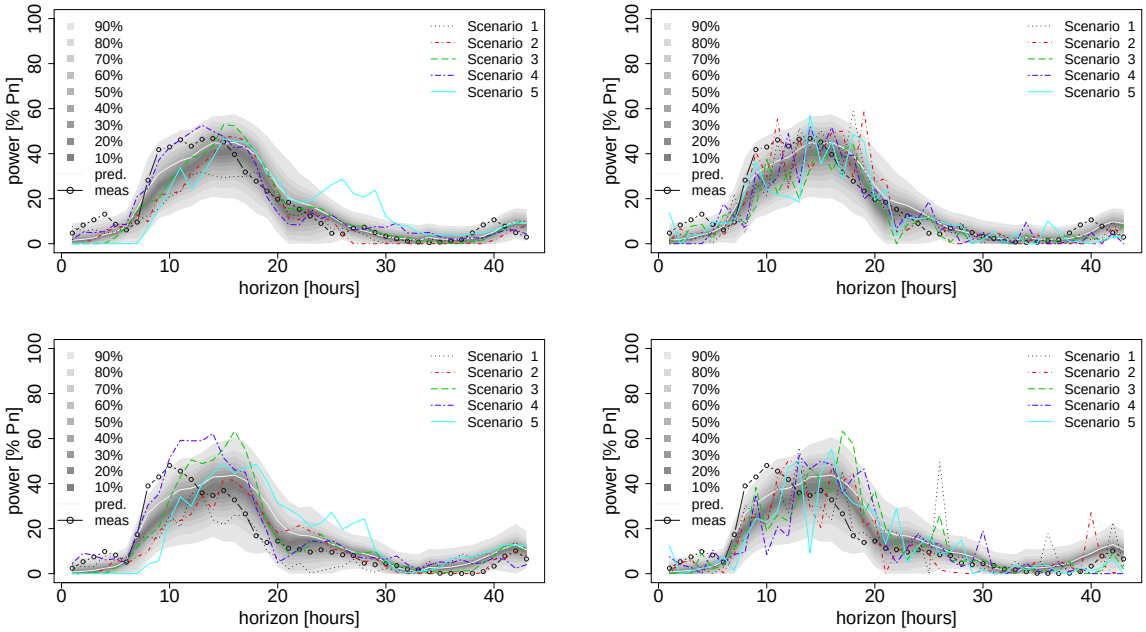


Figure 9: *Scenarios describing wind power generation at zones 6 (top) and 7 (bottom) from 1 to 43 hours ahead issued on the 15th of June, 2007, at 01:00. The scenarios given in the left column correspond to the ones obtained with the model proposed in this study, meanwhile the scenarios on the right are obtained under the assumption of independent marginals, thus, not respecting neither temporal, not spatial dependencies in the data.*

One can see that the scenarios generated using the model proposed in this study respect dependencies both in time and in space. Respecting correlations in time ensures that the corresponding scenarios evolve smoothly with time. That is, given that a scenario predicts wind power generation at time t to be far from the marginal expectation, then the power generation at time $t + 1$ is also expected to deviate a lot from its marginal expectation. As an example see scenario 5 for zone 6 for lead times from 22 to 30 hours ahead.

Respecting spatial dependency between the zones ensures that when large (small) forecast errors are observed at one zone, the errors at the other zone are also expected to be large (small). This is also visible from Fig. 9. For example, in the case of scenario 4, wind power generation deviates a lot from the expected value in both zones 6 and 7.

On another hand, one can see that the corresponding scenarios generated using the independent model do not respect neither temporal, not spatial dependencies in the data.

7 Conclusions

This study considers the problem of obtaining a joint multivariate predictive density for describing wind power generation at a number of sites over a period of time from the set of marginal predictive densities,

targeting each site and each lead time individually. A Gaussian copula approach has been employed for this purpose. The novelty of the proposed methodology consists in the proposed parametrization of the dependence structure. More specifically, instead of modelling the covariance matrix directly, focus is given to its inverse (precision matrix). This solution results in several benefits.

Firstly, the precision matrix is shown to be very sparse. This puts us in the framework of Gaussian Markov random fields and results in computational benefits due to the faster factorization algorithms available for sparse matrices.

Secondly, the proposed parametrization allows for more flexibility as one can easily obtain non separable in space and in time dependence structures following a more complex pattern than the conventional exponential decay in time (and/or space). Additionally, the study has revealed that the empirical precision matrix is given by the non-constant conditional precisions as well as by the varying conditional correlations. This puts us beyond the framework of the conventional approaches given by the homogeneous stationary Gaussian fields. We propose a way to model the changes in the conditional precisions and we permit for conditional correlations to change with the direction. Accounting for such directional influence is not only clearly necessary when looking at the data, but it is also quite intuitive, provided that wind power forecast errors propagate in time and in space under the influence of meteorological conditions.

All the empirical results were obtained by considering a test case of 15 groups of wind farms covering the territory of western Denmark. The results have shown that the joint predictive densities derived from the proposed methodology outperform the benchmark approaches in terms of the overall quality.

Additionally, the study raised a number of new questions and gave ideas for future work.

Firstly, when considering the same problem setup, the direct extension of the proposed methodology could be given by conditioning the precision matrix on the meteorological conditions. Specifically, we suggest that the precision matrix would change with the prevailing wind direction. The easiest way to account for this would be to employ a regime switching approach by allowing a neighbourhood structure to change with the wind direction. In other words, instead of distinguishing between "West-East" and "North-South" neighbourhood as we did in this study, one could then consider "Up Wind"-"Down Wind" and "Concurrent"-"Concurrent". Also, it would be interesting to investigate ways to explain the variations in the conditional precisions among the zones. Possibly some clustering techniques could be employed.

Further, an interesting challenge is to move from the lattice setup considered in this study to a fully continuous approach. Based on [23] there is a link between stochastic partial differential equations and some type of precision matrices. Thus, by understanding how the elements of the precision matrix evolve with distance between the zones and prevailing meteorological conditions, one can get a process description via stochastic partial differential equations.

Another interesting challenge comes with the verification of probabilistic forecasts of a (very) large dimension. Already when working with a dimension of 645, we have faced certain challenges when considering the different scoring rules available for multivariate probabilistic forecast verification. In this study the Logarithmic and the Energy score have been considered. Both scores are proper, thus in theory they can both be used for the forecast verification exercise. However, each of them is associated with some challenges.

The Energy score, being a non-local score, comes with associated computational challenges, since its estimation requires Monte Carlo techniques. Furthermore, following [33], this score has low sensitivity to changes in covariance structure.

The challenges associated with the likelihood-based inference are given by its sensitivity to outliers which might cause difficulties in practical applications. Moreover, following [35] the log likelihood criterion is not robust to model errors, which may result in inconsistent estimates.

As a conclusion, more research is needed in order to propose better ways (more informative, robust and computationally feasible) to evaluate probabilistic forecasts of multivariate quantities.

Acknowledgement

Acknowledgments are due to Roland Löwe for his valuable editorial comments and corrections.

References

[1] T Ackermann et al. *Wind power in power systems*, volume 140. Wiley Online Library, 2005.

[2] J. M. Bernardo. Expected information as expected utility. *The Annals of Statistics*, 7(3):686–690, 1979.

[3] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974.

[4] R. J. Bessa, M. A. Matos, I. C. Costa, L. Bremermann, I. G. Franchin, R. Pestana, N. Machado, H. P. Waldl, and C. Wichmann. Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study. *IEEE Transactions on Sustainable Energy*, 2012.

[5] R. J. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang. Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renewable Energy*, 40(1):29–39, 2012.

[6] A. Botterud, Z. Zhou, J. Wang, R.J. Bessa, H. Keko, J. Sumaili, and V. Miranda. Wind power trading under uncertainty in lmp markets. *Power Systems, IEEE Transactions on*, 27(2):894–903, 2012.

[7] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda. Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of illinois. *Sustainable Energy, IEEE Transactions on*, 4(1):250–261, 2013.

[8] J. Bröcker and L. A. Smith. Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, 22(2):382–388, 2007.

[9] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12(6):1725–1744, 2008.

[10] Guzman D. A note on the multivariate archimedian dependence structure in small wind generation sites. *Wind Energy*.

[11] F. X. Diebold and R. S Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995.

[12] Á. J. Duque, E. D. Castronuovo, I. Sánchez, and J. Usaola. Optimal operation of a pumped-storage hydro plant that compensates the imbalances of a wind power producer. *Electric Power Systems Research*, 81(9):1767–1777, 2011.

[13] Ch. Genest and A.-C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.

[14] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl. The state-of-the-art in short-term prediction of wind power–A literature overview. 2nd edition. Technical report, 2011.

[15] R. Girard and D. Allard. Spatio-temporal propagation of wind power prediction errors. *Wind Energy*, 2012.

[16] T. Gneiting. Editorial: probabilistic forecasting. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2):319–321, 2008.

[17] T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[18] T. Gneiting, L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson. Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, 17(2):211–235, 2008.

[19] S. Hagspiel, A. Papaemannouil, M. Schmid, and G. Andersson. Copula-based modeling of stochastic wind power in europe and implications for the swiss power grid. *Applied Energy*, 96:33–44, 2012.

[20] M. Hofert, M. Mächler, and A. J. Mcneil. Likelihood inference for archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, 2012.

[21] L. Jones and C. Clark. Wind integration - A survey of global views of grid operators. In *Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems*,

2011.

[22] M. Kyung and S. K. Ghosh. Maximum likelihood estimation for directional conditionally autoregressive models. *Journal of Statistical Planning and Inference*, 140(11):3160–3179, 2010.

[23] F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

[24] C. Liu, J. Wang, A. Botterud, Y. Zhou, and A. Vyas. Assessment of impacts of phev charging patterns on wind-thermal scheduling by stochastic unit commitment. *Smart Grid, IEEE Transactions on*, 3(2):675–683, 2012.

[25] X. Liu. Economic load dispatch constrained by wind power availability: A wait-and-see approach. *Smart Grid, IEEE Transactions on*, 1(3):347–355, 2010.

[26] H. Louie. Evaluation of bivariate archimedean and elliptical copulas to model wind power dependency structures. *Wind Energy*, 2012.

[27] H. Madsen and P. Thyregod. *Introduction to general and generalized linear models*. CRC Press, 2011.

[28] M. A. Ortega-Vazquez and D. S. Kirschen. Assessing the impact of wind power generation on operating costs. *Smart Grid, IEEE Transactions on*, 1(3):295–301, 2010.

[29] Pinson P. *Estimation of the uncertainty in wind power forecasting*. PhD thesis, Ecole des Mines de Paris, 2006.

[30] G. Papaefthymiou and D. Kurowicka. Using copulas for modeling stochastic dependence in power system uncertainty analysis. *Power Systems, IEEE Transactions on*, 24(1):40–49, 2009.

[31] P. Pinson, H. Madsen, H. Aa. Nielsen, G. Papaefthymiou, and B. Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009.

[32] P. Pinson, H. A. Nielsen, J. K. Mø ller, H. Madsen, and G. N. Kariniotakis. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, 10(6):497–516, 2007.

[33] P. Pinson and J. Tastu. Discrimination ability of the Energy score. *working paper*, 2013.

[34] H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*, volume 104. Chapman & Hall, 2005.

[35] H. Rue and H. Tjelmeland. Fitting gaussian markov random fields to gaussian fields. *Scandinavian Journal of Statistics*, 29(1):31–49, 2002.

[36] D. Simpson, F. Lindgren, and H. Rue. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*, 23(1):65–74, 2012.

[37] M. Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.

[38] J. Tastu, P. Pinson, and H. Madsen. Multivariate conditional parametric models for a spatiotemporal analysis of short-term wind power forecast errors. In *Scientific Proceedings of the European Wind Energy Conference, Warsaw*, pages 77–81, 2010.