

# Gaussian linear state-space model for wind fields in the North-East Atlantic

Julie Bessac<sup>1</sup>, Pierre Ailliot<sup>2</sup>, Valérie Monbet<sup>1</sup>

December 15, 2013

<sup>1</sup> *Institut de Recherche Mathématiques de Rennes, UMR 6625, Université de Rennes 1, France*

<sup>2</sup> *Laboratoire de Mathématiques de Bretagne Atlantique, UMR 6205, Université de Brest, France*

## Abstract

A space-time model for wind fields is proposed. It aims at simulating realistic wind conditions with a focus on reproducing the space-time motions of the meteorological systems. A Gaussian linear state-space model is used where the latent state may be interpreted as regional wind condition and the observation equation links regional and local scales. Parameter estimation is performed by combining a method of moment and the EM algorithm whose performances are discussed using simulation studies. The model is fitted to 6-hourly reanalysis data in the North-East Atlantic. It is shown that the fitted model is interpretable and provide a good description of important properties of the space-

time covariance function of the data, such as the non full-symmetry induced by prevailing flows in this area.

**Keywords:** Stochastic weather generators, Wind time series, Spatio-temporal modeling, State-space model, EM algorithm, Identifiability.

## 1 Introduction

Many natural phenomena and human activities depend on wind conditions. Meteorological data are often available over periods of time that are not long enough to estimate reliably probabilities of complex events. Stochastic weather generators have been developed to overcome this insufficiency by simulating sequences of meteorological variables with statistical properties resampling the ones of the observations. They have been adopted in impact studies as a computationally inexpensive tool. Wind generators have in particular been used to assess wind power production (see [BKM84, HS13]), drift of objects in the ocean (see [AFM06]) or coastal erosion (see [ST<sup>+</sup>90]).

A review of stochastic models for simulation of artificial wind time series can be found in [MAP07]. The most classical approach for modeling wind time series at a single location consists in using the Box-Jenkins methodology, where an ARIMA model is fitted to the process with transformed margins. Since wind is a non negative variable, transformations are used to obtain Gaussian like margins. The most usual transformation for wind is a power transformation (see [BKM84], [HR89], [NBS96], [KJ97]). But specific distributions are used as well, for instance Weibull [?], truncated Gaussian [GLW<sup>+</sup>06] or skew distributions [?]. The conditionnal mean (and variance) is then modelled given the wind at the previous time steps. Non-linear models have also been proposed

more recently (see [AM12] and references therein).

Several ways have been explored to generalize time series models to space time models for short term prediction of wind at several locations. Multisites wind models have to deal with temporal and spatial dependence and it is well known that they are not separable [Gne02]. Black box models like artificial neural network can be fitted. But they lead to non interpretable models and they will not be discussed here. Two other approaches have been detailed in the literature: the models based on Gaussian fields (or Gaussian vector) with a parametric non separable covariance function [HR89, RM13] and the models based on Vector Autoregressive models (VAR) [dLG05]. Both approaches allow to characterize space and time variability of the wind. In particular, the motions of the air masses are taken into account. In the Gaussian field method, these displacements are characterized by the interaction of time and space in the covariance matrix ([RM13], [HR89], [?]). One difficulty of these models is to infer the parametric form of the covariance function and standard forms are usually used as Gaussian or Matérn. In the VAR models, motions are considered via the introduction of covariates or switchings. For example, in [GLW<sup>+</sup>06], the authors propose to define regimes which describes the main weather types (westerly/easterly blowing direction) and fit a different VAR model in each regime taking into account the location of the sites to each other. Hering [?] show that similar short term predictions can be obtained by introducing the wind direction in the VAR model. In [AMP06], the VAR coefficients depend on a latent process directly describing the motion of the air masses. In [ŠBŠ11], a latent field describes the spatial structure of the AR parameters. In VAR models two scales are implicitly modelled: a regional scale representing the prevailing flows and a local one corresponding to the

locally observed wind. But VAR models may lead to over parametrization especially when sites are high correlated.

In the present paper, a new approach is investigated. The regional mean wind is explicitly introduced as a latent variable with its own autoregressive dynamic and the local wind is expressed with respect to the regional wind at different lags depending on the location of each site. The delays imitates the mean displacements of the air masses. The model is kept simple and interpretable since it is a linear Gaussian state space model and covariates can be easily added. The inference may be performed via Kalman filtering and smoothing. Such models lead to non separable and anisotropic covariance functions.

The data considered in this paper are presented in Section 2. The model is described in Section 3. Parameter estimation and fitting procedures are also discussed in this section. Validation of the model is discussed in Section 4. It is shown that the fitted model is able to reproduce the anisotropy and non-separability of the data. Various reduced models are introduced in Section 5 and conclusions are given in Section 6. Parameter identifiability and non full-symmetry are proven in Appendix A.

## 2 The wind data set

In situ data are neither available on a long time period nor on a large area offshore Brittany in France. For a reliable study we choose to use reanalysis data which are obtained by combining observations with numerical weather prediction models. It provides relevant datasets for meteorological or climatological studies. The data under study are wind speed intensities at 10

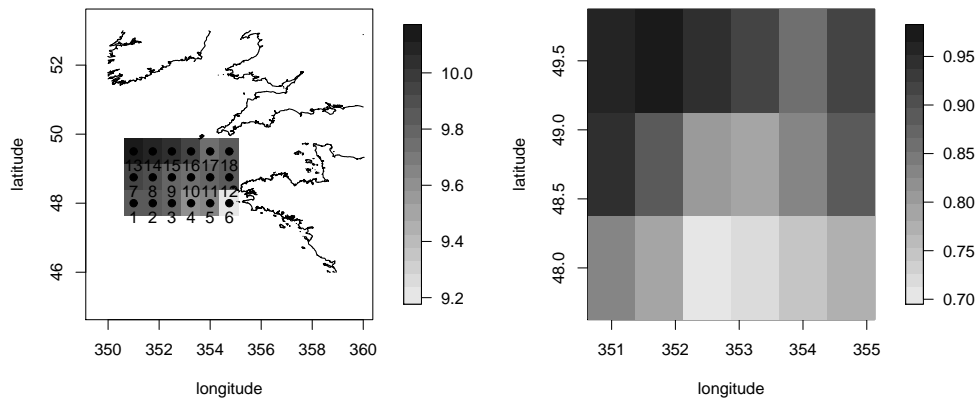


Figure 1: Left panel: mean wind speed at the 18 numbered points under study in the North-East Atlantic. Right panel: estimated values of the power in the Box-Cox method at the 18 locations.

meters above sea level extracted from the ERA Interim Full data set produced by the European Center of Medium-range Weather Forecast (ECMWF). It can be freely downloaded and used for scientific purposes at the URL <http://data.ecmwf.int/data/>.

This dataset is available on a regular space-time grid with a temporal resolution of 6 hours and a spatial resolution of  $0.75^\circ$ . The methodology introduced in this paper could however easily be adapted to handle datasets with more complicated space-time sampling such as the one obtained when considering networks of meteorological stations. We focus on 18 gridded locations between latitudes  $48^\circ\text{N}$  and  $49.5^\circ\text{N}$  and longitudes  $6.25^\circ\text{W}$  and  $9^\circ\text{W}$  (see Figure 1). The data set consists of 33 years of wind data from 1979 to 2011 and we focus on the month of January. Further, the statistical inference is based on the assumption that the 33 months of January wind data are 33 independent realizations of a common stochastic process. This assumption is not unusual for meteorological processes but it does not take into account low frequency

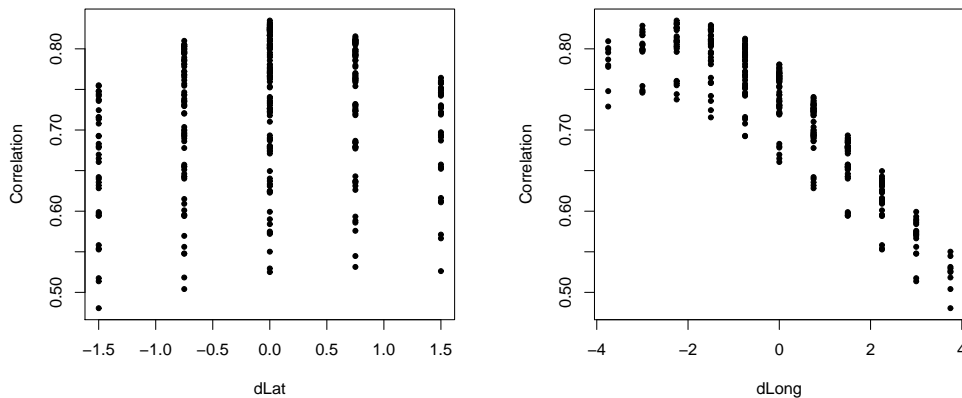


Figure 2: Lagged one cross-correlations against differences of latitude (left) and longitude (right).

variations such as the the North Atlantic Oscillation (NAO).

In the studied area prevailing air masses are generally moving eastward. It creates non-separability and non full-symmetry properties of the associated space-time covariance function (see [Gne02]). In Figure 2 lagged one cross-correlations highlight this phenomenon. The asymmetry with respect to the difference of longitude reveals that the correlation between  $Y_t(p)$  and  $Y_{t+1}(p')$  is higher when location  $p$  is more westerly than  $p'$  than when  $p$  is easterly with respect to  $p'$ . This asymmetry is less pronounced in latitude but reveals flows from north to south. Full-symmetry is rejected as for the famous data set of wind speed in Ireland studied in [Gne02], [dLG05] and [FHI07, chapter 4]. As in Figure 2 the study of correlations at lag 0 reveals some anisotropy as dependences in latitude and longitude differ.

Wind speed distribution is known to be skewed. It is generally modeled as a Weibull distribution (see e.g. [BKM84]) but other distributions such as the skew normal distribution have also been considered (see [FNAB10]). A

classical method to handle such asymmetry in time series analysis consists in applying a Box-Cox transformation in order to get a time series with approximately marginal Gaussian distribution and in fitting an ARMA model to the transformed time series. This method has been extensively used for analyzing wind time series (see e.g. [BKM84], [HR89], [NBS96], [KJ97]) and is generalized in a space-time context here. It has been proposed to use a different transformation at each location (see [RM13]) but we have chosen to use the same power transformation at all sites in order to preserve the spatial structure of the wind fields following e.g. [HR89]. More precisely, we consider

$$\begin{cases} y_{\lambda,i,t} = \frac{y_{i,t}^\lambda - 1}{\lambda} & \text{if } \lambda > 0 \\ y_{\lambda,i,t} = \log(y_{i,t}) & \text{if } \lambda = 0. \end{cases}$$

with  $y_{i,t}$  the wind speed at time  $t$  and location  $i$ . The value  $\hat{\lambda} = 0.85$  has been used in the sequel. It is derived as the mean value of the  $\hat{\lambda}_i$  depicted on Figure 1 which are obtained at individual locations with the following criterion given in [Hin77]. It is based on the empirical measure of asymmetry

$$S(\lambda_i) = \frac{\text{mean}(y_{\lambda_i,i,t}) - \text{median}(y_{\lambda_i,i,t})}{\text{var}(y_{\lambda_i,i,t})}$$

and  $\lambda_i$  is chosen as equaling to 0 this statistic. Figure 1 shows that the data close to the coast are less Gaussian than the offshore ones. The model will be fitted on this power transformed and mean corrected data. For sake of simplicity we denote  $Y_t$  the transformed observed field instead of  $Y_{\lambda,t}$  in the sequel.

### 3 A linear Gaussian state-space model for wind speed

State-space models first appeared in engineering and have then been extensively used in many domains. State-space representations bring a very flexible framework for modeling time series (see [DK12] and [BD06]) and space-time processes (see [WH10]). The model introduced in this Section is a linear Gaussian state-space model. One of the main advantages of this class of model estimation, is that forecasting and smoothing can be processed through general and efficient procedures.

#### 3.1 Model

The observed wind fields are generally smooth leading to a high correlation between the different sites. It suggests that it may be possible to explain an important part of the signal by using a common scalar process (the regional wind condition) for the different locations. This scalar process, denoted  $\{X_t\}$  in the sequel, is not directly observable and is introduced as a latent process. Due to the mean motion of air masses we expect that the wind conditions at western locations will depend more on the leading one lag  $X_{t+1}$  and  $X_t$  signals than on the lagged signal  $X_{t-1}$  with the reverse phenomenon at eastern locations. These points led us to consider the following Gaussian state-space model

$$(M) \begin{cases} X_{t+1} &= \rho X_t + \sigma \epsilon_{t+1}, \\ Y_t &= \alpha_1 X_{t+1} + \alpha_0 X_t + \alpha_{-1} X_{t-1} + \Gamma^{1/2} \eta_t \end{cases} \quad \text{for } t \geq 0.$$



$Y_t \in \mathbb{R}^K$  is the observed process, its  $K$  coordinates correspond to the mean corrected transformed wind speed at the  $K = 18$  locations.  $\{\epsilon_t\}$  and  $\{\eta_t\}$  are independent Gaussian white noise sequences with zeros mean and identity covariance matrices.  $\alpha_1$ ,  $\alpha_0$  and  $\alpha_{-1}$  are  $K$ -dimensional vectors which link the lagged values of the regional process  $\{X_t\}$  to local wind conditions. The covariance matrix  $\Gamma \in \mathbb{R}^{K \times K}$  models the spatial structure of the difference between the observed process  $\{Y_t\}$  and the local conditions  $W_t = \alpha_1 X_{t+1} + \alpha_0 X_t + \alpha_{-1} X_{t-1}$  deduced from the regional process. It may correspond to small scale fluctuations. In finance and economics this covariance matrix of error of measurement is often diagonal. In [WH10] a parametrization of this matrix or the use of a diagonal matrix are advised. As a first step this covariance matrix is not parameterized but reduced models which takes into account the spatial information are investigated in Subsection 5.1. In the sequel we denote  $\Lambda = (\alpha_1 | \alpha_0 | \alpha_{-1}) \in \mathbb{R}^{K \times 3}$  and  $\theta = (\rho, \sigma, \Lambda, \Gamma)$  the unknown parameter.

The temporal dynamics of the observed process is mainly contained in the latent process  $\{X_t\}$ . The model thus imposes the same long term temporal dynamics, corresponding to the regional scale, at each location. Under the assumption  $|\rho| < 1$ , the AR(1) process  $\{X_t\}$  is stationary and so is the process  $\{Y_t\}$ .  $\{W_t\}$  is an ARMA(1,2) process since

$$W_t - \rho W_{t-1} = \alpha_1 \epsilon_{t+1} + \alpha_0 \epsilon_t + \alpha_{-1} \epsilon_{t-1}.$$

Signs and values of  $(\alpha_1, \alpha_0, \alpha_{-1})$  can be interpreted in terms of autocovariance function of the moving average part  $\alpha_1 \epsilon_{t+1} + \alpha_0 \epsilon_t + \alpha_{-1} \epsilon_{t-1}$ .

### 3.2 Second order structure and identifiability

Identifiability is required to get sensible and reliable parameter estimation. Gaussian linear state-space models are often non-identifiable without additional constraints (see e.g. [HD88], [Lju99], [BW12], [Bor10]). Indeed the introduction of the latent process  $\{X_t\}$  is a source of non-identifiability since the unknown parameters need to be identified uniquely from the distribution of the observed  $\{Y_t\}$ .

$\{Y_t\}$  is a zero mean stationary Gaussian process which is thus characterized by its second order structure given below

$$\begin{aligned} \text{cov}_\theta(Y_t, Y_t) &= \frac{\sigma^2}{1-\rho^2} \left( \alpha_1(\alpha_1 + \rho\alpha_0 + \rho^2\alpha_{-1})^t + \alpha_0(\rho\alpha_1 + \alpha_0 + \rho\alpha_{-1})^t + \right. \\ &\quad \left. \alpha_{-1}(\rho^2\alpha_1 + \rho\alpha_0 + \alpha_{-1})^t \right) + \Gamma, \end{aligned} \quad (1)$$

$$\begin{aligned} \text{cov}_\theta(Y_t, Y_{t+1}) &= \frac{\sigma^2}{1-\rho^2} \left( \alpha_1(\rho\alpha_1 + \alpha_0 + \rho\alpha_{-1})^t + \alpha_0(\rho^2\alpha_1 + \rho\alpha_0 + \alpha_{-1})^t + \right. \\ &\quad \left. \rho\alpha_{-1}(\rho^2\alpha_1 + \rho\alpha_0 + \alpha_{-1})^t \right), \end{aligned} \quad (2)$$

$$\begin{aligned} \text{cov}_\theta(Y_t, Y_{t+k}) &= \frac{\sigma^2}{1-\rho^2} \rho^{k-2} (\alpha_1 + \rho\alpha_0 + \rho^2\alpha_{-1})(\rho^2\alpha_1 + \rho\alpha_0 + \alpha_{-1})^t, \quad (3) \\ &\text{for all } k \geq 2. \end{aligned}$$

The study of this space-time covariance function leads to the following Proposition which is proven in Appendix A.

**Proposition 1** *Assume that (M) holds. Assume further that  $\frac{\sigma^2}{1-\rho^2} = 1$  and that the vectors  $\alpha_1$ ,  $\alpha_0$  and  $\alpha_{-1}$  are linearly independent. Then the parameters can be identified from the distribution of the process  $\{Y_t\}$ .*

These identifiability constraints were always satisfied when fitting the model to the data. The first condition requires  $X$  to have a unit stationary variance, the local variability is then accounted in the vectors  $\alpha_1$ ,  $\alpha_0$  and  $\alpha_{-1}$ . The

second one is natural since in case of linear dependence between these vectors the model (M) is reduced to a sub-model dealing with one or two lagged versions of  $X$ . Identifiability of linear Gaussian state-space models have been investigated during the last decades and initially in control theory. Literature is abundant on stochastic linear systems identification ([Lju99, HD88]). To the best of our knowledge most of sufficient conditions of identifiability are structural constraints on parameters (see [PD10] for references and examples) associated with identification procedures. Identifiability is examined through different criteria such that transfer functions ([Lju99]) or likelihood criterion like in [PD10]. Identification procedures are realized through controllability and observability of several parameters in [Lju99] and via the EM algorithm in [PD10]. In econometrics the identifiability of the latent factors and the loading matrix is considered (see for example [BW12, Bor10]). In most cases structural constraints are also applied depending on the interpretability desired. However the general conditions given in [BW12] do not guarantee identifiability of the model (M) as  $X$  is scalar in (M).

Considering this function as a discretized space-time covariance function, we show in Appendix A that properties of full-symmetry and separability are not fulfilled under the identifiability constraints of Proposition 1. Other non-symmetric space-time covariance models have been proposed in the literature and some of them have been adjusted to the Irish wind dataset that exhibits non-symmetry (see [Gne02] and [FHI07, chapter 4]). Strong spatial assumptions are commonly assumed in these models such as spatial stationarity and isotropy. However due to prevailing flows a lot of meteorological data have anisotropic patterns like the dataset under study or the Irish one. A model, based on the specification of a vector autoregressive process, is proposed in

[dLG05] to capture a part of the anisotropy that is observable in the correlations of the Irish dataset. We will see in Section 4 that the proposed model enables to reproduce various above mentioned complex properties of the space-time covariance of our wind data.

### 3.3 Parameter estimation

Two methods of estimation have been implemented and compared. The first one is a method of moment based on the second order structure of the process  $\{Y_t\}$  given by (1-3). It consists in minimizing numerically the following objective function

$$\begin{aligned} \theta \rightarrow & \|\widehat{\text{cov}}(Y_t, Y_t) - \text{cov}_\theta(Y_t, Y_t)\|_2^2 + \|\widehat{\text{cov}}(Y_t, Y_{t+1}) - \text{cov}_\theta(Y_t, Y_{t+1})\|_2^2 \\ & + \|\widehat{\text{cov}}(Y_t, Y_{t+2}) - \text{cov}_\theta(Y_t, Y_{t+2})\|_2^2 + \|\widehat{\text{cov}}(Y_t, Y_{t+3}) - \text{cov}_\theta(Y_t, Y_{t+3})\|_2^2, \end{aligned} \quad (4)$$

where  $\widehat{\text{cov}}$  denotes the empirical covariance function and  $\|\cdot\|_2$  stands for the matrix Frobenius norm. This method, denoted GMM in the sequel, is a standard method in geostatistics (see e.g. [Cre91]). We have chosen to consider only the first four lags of the autocovariance function when building the objective function (4). It corresponds to the minimal number of terms needed to identify the parameters (see Appendix A). Simulation results indicate that including more lags in the objective function does not lead to more accurate estimates.

The second method performs Maximum Likelihood (ML) estimation using the Expectation-Maximization (EM) algorithm ([CMR05] and [DLR77]). For linear Gaussian state-space model efficient numerical procedures exist for both the E-step, where the Kalman recursions lead to an exact computation of

the smoothing probabilities, and the M-step with analytical expressions being available for the parameters which maximize the intermediate function of the EM algorithm. More details can be found in the supplementary materials.

Both methods are sensitive to the initial parameter value which need to be chosen carefully. We used the following procedure which exploits the properties of the second order structure of  $\{Y_t\}$ :

- $\rho = \frac{\text{cov}(Y_t, Y_{t+3})_{i,j}}{\text{cov}(Y_t, Y_{t+2})_{i,j}}$  for all  $i, j \in \{1, \dots, K\}$  is initialized as the empirical mean of  $\frac{\widehat{\text{cov}}(Y_t, Y_{t+3})_{i,j}}{\widehat{\text{cov}}(Y_t, Y_{t+2})_{i,j}}$ .

- $\Lambda$  is estimated by minimizing

$$\theta_\Lambda \rightarrow \|\widehat{\text{cov}}(Y_t, Y_{t+1}) - \text{cov}_\theta(Y_t, Y_{t+1})\|_2^2 + \|\widehat{\text{cov}}(Y_t, Y_{t+2}) - \text{cov}_\theta(Y_t, Y_{t+2})\|_2^2$$

as a function of  $\Lambda$  with  $\rho$  being fixed to the value obtained in the previous step. Note that this function does not depend on  $\Gamma$  according to (2) and (3).

- $\Gamma$  is determined by minimizing

$$\theta_\Gamma \rightarrow \|\widehat{\text{cov}}(Y_t, Y_t) - \text{cov}_\theta(Y_t, Y_t)\|_2^2$$

as a function of  $\Gamma$  with  $\rho$  and  $\Lambda$  being fixed to the value obtained in the previous steps.

These rough estimates are used as initial conditions for the numerical optimization of 4 which leads to the GMM estimates. The GMM estimates are then used to initialize the EM algorithm and get the ML estimates.

Parameters	Bias		Sd		RMSE	
	GMM	ML	GMM	ML	GMM	ML
$\rho$	0.036	0.004	0.022	0.017	0.042	0.017
$\alpha_1$	[-0.11;-0.009]	[-0.069;-0.019]	[0.065;0.108]	[0.071;0.097]	[0.067;0.149]	[0.068;0.127]
$\alpha_0$	[-0.047;-0.234]	[0.054;0.144]	[0.11;0.182]	[0.11;0.144]	[0.125;0.292]	[0.127;0.228]
$\alpha_{-1}$	[-0.080;0.022]	[-0.035;0.012]	[0.078;0.114]	[0.062;0.104]	[0.086;0.139]	[0.079;0.117]
$\Gamma$	[-0.199;0.007]	[-0.108;0.013]	[0.058;0.367]	[0.029;0.368]	[0.053;0.199]	[0.053;0.115]

Table 1: Bias, standard deviation and RMSE of parameters estimates. For the multidimensional parameters, minimal and maximal values are given in brackets.

### 3.4 Properties of the estimates

Under suitable conditions, GMM (see [NM94]) and ML (see [NM94, SS06, HD88, Cai88]) estimators are consistent and asymptotically Gaussian. A simulation study was performed in order to assess the performance of the estimates in a situation comparable to the practical application.  $N = 100$  independent sets of the size of the studied data are simulated for the parameters set estimated by ML on the wind data. Table 1 gives the bias, standard deviation and Root Mean Square Error (RMSE) of ML and GMM estimates computed from the simulations. Bias and standard deviation are low. ML generally outperforms GMM except for estimating  $\Gamma$  where both methods give comparable results. For both methods  $\alpha_1$  and  $\alpha_{-1}$  are more accurately estimated than  $\alpha_0$  and  $\Gamma$  is the less accurately estimated quantity.

## 4 Results

In order to validate the proposed model we checked its physical realism, its ability to generate realistic wind conditions and to produce accurate forecasts. GMM and ML estimates are also compared through this validation in order to check their robustness in a practical context.

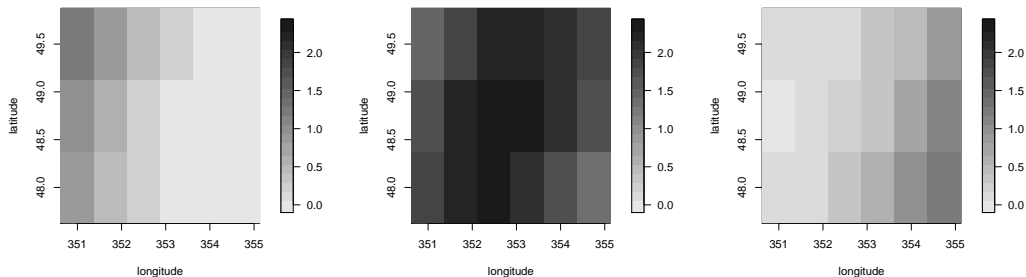


Figure 3: ML estimate of  $\alpha_1$  (left panel)  $\alpha_0$  (middle panel) and  $\alpha_{-1}$  (right panel).

## 4.1 Interpretability

The loading matrix  $\Lambda$  links the latent process to observed wind conditions. The values of  $\alpha_1$  and  $\alpha_{-1}$  shown on Figure 3 reveal the site-dependent relations with the regional mean wind. As expected western locations, which are generally the first affected when meteorological events enter in the studied region, depend more on  $X_{t+1}$  than on  $X_{t-1}$  and the reverse is true for eastern locations.

Since large scale variability is supposed to be contained in the latent process,  $\Gamma$  should contain only small scale variations due to the differences between the observed wind  $\{Y_t\}$  and the downscaled regional wind  $\{W_t\}$ . This is confirmed when comparing the spatial sill and range of  $\Gamma$  with the ones of the original covariance function of the data (see Figure 4). The shape of  $\Gamma$  has a block structure which is induced by the geometry of the domain and the numbering of the sites (see Figure 1). The level sets of the blocks, except the top right corner (and by symmetry bottom left corner), are similar to saddle point level sets: the model better explains the wind observed at the central locations of the domain than at the locations which are close to the boundary. The top right corner has elliptical level sets. This different geometry raises

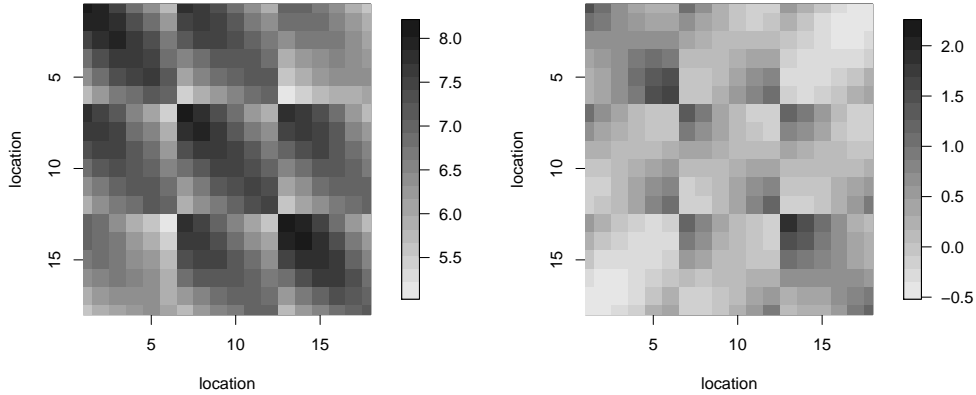


Figure 4: Covariance matrix of  $Y$  (left) and ML estimate of  $\Gamma$  (right)

problems when trying to develop simple parametric model for  $\Gamma$  (see Section 5.1).

## 4.2 Realism of simulated sequences

To further validate the model we have checked its ability to simulate realistic wind conditions. For that we have generated artificial time series from the model, and have compared statistics corresponding to the artificial sequences with those from the original data. We first looked at the marginal distributions of wind speed, both observed and fitted, at each location. According to quantile-quantile plots shown on Figure 5, the model is able to reproduce the general shape of the marginal distribution of the process at the central station 9 except for very low wind speed. Similar results were obtained at other locations.

The model assumes that the dynamics of wind speed are inherited mainly from the common latent process  $\{X_t\}$ . To check this assumption, we have compared the autocorrelation functions of the model to the data ones at the



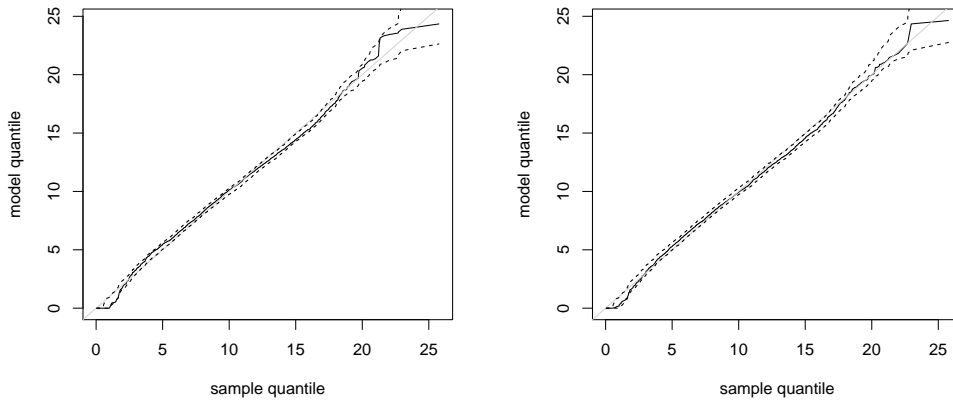


Figure 5: Quantile-Quantile plot at location 9 for the model (M) and the parameters estimated by GMM (left) and by ML (right). The dashed lines corresponds to 90% prediction intervals computed by simulation.

various locations. Typical results are shown in Figure 6. The model is able to reproduce the first coefficients of the autocorrelation function at the central location 9 and of the cross-correlation function between locations 13 and 18 which exhibits a time shift due the prevailing westerly flow. GMM estimate is designed to make coincide the first lags of the empirical autocovariance functions with the one of the fitted model. Figure 6 shows that the agreement is indeed very good. However the ML method, which takes into account the longer term dynamics in the likelihood function, is better in reproducing the correlation structure for time lags above one day. It is mainly due to the higher value of  $\rho$  (0.76 by ML and 0.70 obtained by GMM). The better performance of ML estimates is coherent with Table 1. Figure 6 also shows the important bias on the estimates of the second order structure of the process which is problematic when using the GMM approach.

Figure 7 shows the theoretical correlations at lags 0 and 1 against the empirical ones to assess the model's ability to capture the spatial depen-

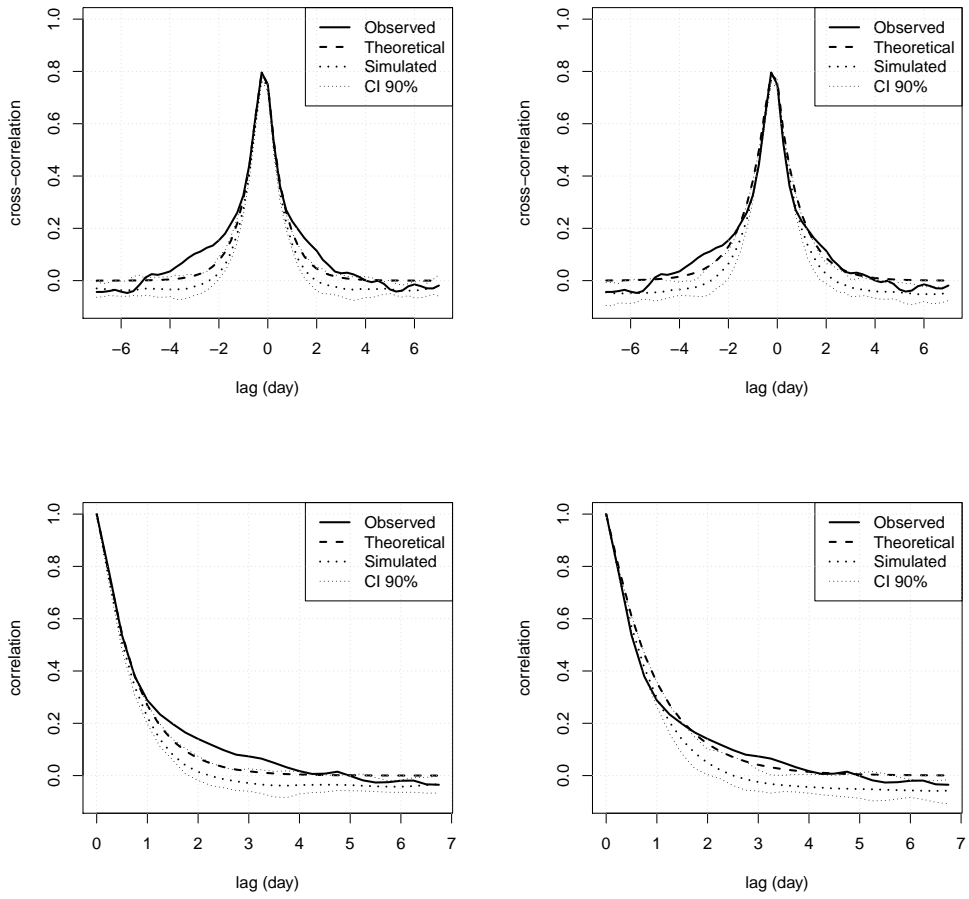


Figure 6: Observed (full lines) and theoretical (dashed lines) cross-correlations between locations 13 and 18 (upper line) and auto-correlation at location 9 (lower line) for the model (M) with parameters estimated by GMM (left) and by ML (right). 90% prediction intervals are computed from 100 independent samples of the size of the original data.

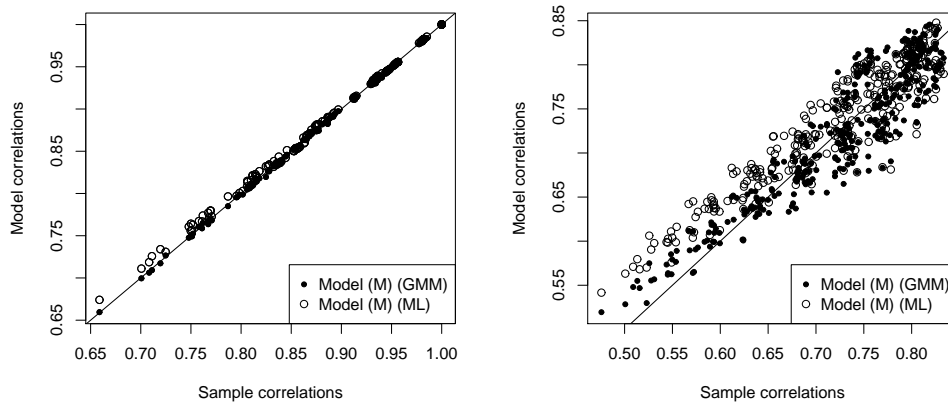


Figure 7: Theoretical correlations against observed correlations at lag 0 (left) and lag 1 (right) for the model (M) and the two methods of estimation.

dence structure. Correlations at lag 0 are very well reproduced and lagged one correlations are also generally well reproduced. As expected, the GMM method provides again a better fit. Similar figures than Figure 2 reveal that the anisotropy and non-separability observable in the correlations are partly reproduced by the model.

### 4.3 Prediction

An advantage brought by the Markovian structure of the model (M) is that short term forecast can be efficiently computed through the Kalman recursions (see [BD06, chapter 8]). The forecast performance is assessed by computing the one-step ahead forecast on the last 8 years of data (validation set) after fitting the model on the first 25 years of data (training set). In practice the forecast skill of the model at location  $i \in \{1, \dots, K\}$  is evaluated by computing the natural empirical estimate of the Mean Square Percentage Error (MSPE)

defined as

$$MSPE(i) = \frac{\text{var}(Y_t(i) - E[Y_t(i)|Y_0, \dots, Y_{t-1}])}{\text{var}(Y_t(i))}$$

where the MSE of the forecast error (the numerator) is normalized by the variance of the field at the individual locations, with  $Y_t$  the original non transformed wind.

According to Table 2, model (M) leads to significant improvements over persistence forecast (about 28%), which is the classical benchmark for wind prediction (see e.g. [ZG12]), and the forecast obtained when fitting a different ARMA(2,1) model at each site (about 17%). It illustrates the gain of using the spatial information when computing short-term probabilistic forecasts. For comparison purpose, a vector autoregressive model of order 1 (VAR(1)) was also fitted. The VAR(1) model gives slightly better results compared to model (M) with an improvement of about 9% in average over all locations. However the VAR(1) model involves 495 parameters, compared to the 208 parameters for model (M), and the VAR(1) parameters are difficult to interpret.

The spatial structure of the forecast error is shown on Figure 8. The improvement obtained with the VAR(1) and (M) models, which include the spatial information, is more pronounced at eastern locations. It is due to the prevailing westerly flow. Indeed the wind speed observed at the western locations (1, 2, 7, 8, 13, 14) at given time  $t$  brings information on the wind speed observed at the eastern locations (5, 6, 11, 12, 17, 18) at the next time  $t + 1$ . Note also that the difference between models (M) and VAR(1) is very low at the central locations 9, 10, 11 but the forecast performance of model (M) tends to deteriorate close to the boundaries of the studied region. We observed that the forecasts of model (M) exhibit less spatial variability than the observations and the forecasts obtained with the VAR(1) model. It suggests that using of

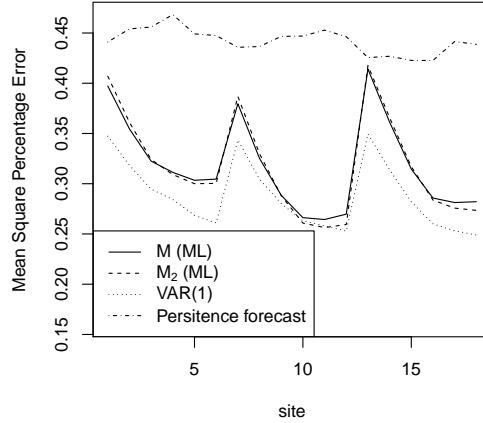


Figure 8: Mean Square Percentage Error (MSPE) between observed wind speed and back transformed predicted wind speed at each location for models (M), ( $M_2$ ), VAR(1) and persistence.

a scalar latent process  $X$  is too simplistic to catch all the complexity of the space-time structure of the data and that  $X$  is designed mainly to describe the wind conditions at central locations.

## 5 Some improvements of the model

In this section we explore reduced models for the matrices  $\Gamma$  and  $\Lambda$  and higher order models for  $\{X_t\}$ .

### 5.1 Parameterization of $\Gamma$

The spatial structure of the estimated  $\Gamma$  shown on Figures 4 and 9 suggests modeling the covariance between locations  $i$  and  $j$  in  $\{1, \dots, K\}$  as a function of the distance  $d_{i,j}$  between these locations. In the sequel we consider two

different models, one with Gaussian correlation function

$$\Gamma_{i,j} = \sigma_i \sigma_j (\exp(-\lambda_1 d_{i,j}^2) + \lambda_2 \delta_{i,j}) \text{ for } i, j \in \{1, \dots, K\},$$

and the other with wave correlation function

$$\Gamma_{i,j} = \sigma_i \sigma_j \left( \frac{\sin(\lambda_1 d_{i,j})}{\lambda_1 d_{i,j}} + \lambda_2 \delta_{i,j} \right) \text{ for } i, j \in \{1, \dots, K\},$$

where  $(\sigma_1, \dots, \sigma_K, \lambda_1, \lambda_2)$  are positive parameters and  $\delta_{i,j}$  denotes the Kronecker delta.  $\lambda_1$  and  $\lambda_2$  model respectively the range and nugget parameters, and  $\sigma_i(1 + \lambda_2)$  represents the standard deviation of the field at location  $i$ . These models are usual well defined covariance functions (see e.g. [Cre91, Abr97]). They will be denoted respectively  $(M_{\Gamma \sim \text{Gauss}})$  and  $(M_{\Gamma \sim \text{Sinus}})$  in the sequel.

The difference in dependence from latitude and longitude of  $\Gamma$  (Figure 9) suggests the use of an anisotropic distance (see [RBBP11, Has07, ŠBŠ11])

$$d_{i,j} = \sqrt{\Delta \text{Lat}(i,j)^2 + \theta_1 \Delta \text{Long}(i,j)^2 + \theta_2 \Delta \text{Lat}(i,j) \Delta \text{Long}(i,j)}$$

where  $\Delta \text{Lat}(i,j)$  and  $\Delta \text{Long}(i,j)$  denote respectively the difference in latitude and longitude between locations  $i$  and  $j$  expressed in kilometers. The constraint  $\theta_1 > \frac{\theta_2^2}{4}$  is imposed to ensure positive-definiteness of the distance.

These covariance structures have first been fitted by least square estimation to the estimated  $\Gamma$  shown on Figure 4 and the results are shown on Figure 9. The fit is globally satisfying for the wave covariance whereas the Gaussian shape can not cope with the negative correlations observed between western and eastern locations. However the covariance between the northern and southern locations are poorly reproduced (bottom left corner and top right

corner of the images of  $\Gamma$  in Figures 9 and 4). As mentioned in Section 4.1 these blocks have a particular elliptical shape which can not be reproduced by the parametric models.

Anisotropy coefficients are for the sinus structure:  $(\theta_1, \theta_2) = (0.2, 0.04)$  and for the Gaussian covariance  $(\theta_1, \theta_2) = (0.23, 0.005)$ .  $\theta_1 \leq 1$  which is reasonable as the spatial range of the coefficients of  $\Gamma$  in longitude is weaker than the one in latitude (Figure 9). The interaction between latitude and longitude is very weak and almost non existing for the Gaussian shape.

In a second step, the parameters have been re-estimated using the GMM and ML methods. A numerical optimization need to be performed in the M-step of the EM algorithm to update the values of  $(\sigma_1, \dots, \sigma_K, \lambda_1, \lambda_2)$ . Note that the function to minimize can be expressed in a compact way (see supplementary materials) which leads to an efficient numerical procedure. The models have been validated in the same way than model (M) (see Section 4). Similar results were obtained as concerns the marginal distributions and the temporal correlation functions but the description of the spatial structure is deteriorated when using a  $(M_\Gamma)$  model instead of (M) (see Figure 10). This miss-specification is also confirmed by the Bayes Information Criterion (BIC) and MSPE values given in Table 2 where  $BIC = -2 \log L + N_p \log(N_{obs})$  with  $L$  the likelihood of the model,  $N_p$  the number of parameters and  $N_{obs}$  the number of observations. The reduced models  $(M_\Gamma)$  is clearly outperformed by the full model (M). Other parametric models such as the Matérn family have been tried without more success and it seems difficult to find a simple reduced model which can reproduce all the complexity of the observation error  $\Gamma$ .

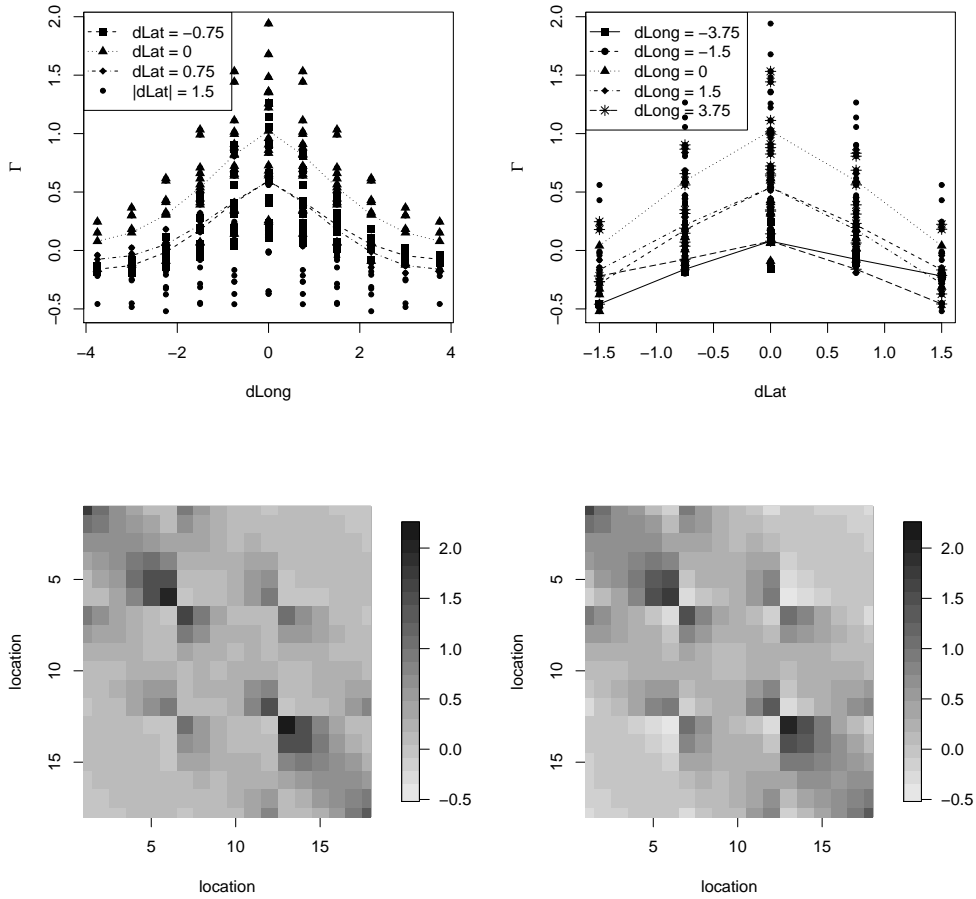


Figure 9: Top panels: estimate of  $\Gamma$  as a function of the difference in longitude (left) and latitude (right), solid lines: mean values according a given latitude or longitude. Bottom panels: image of covariance matrices fitted by least square to the matrix shown on Figure 4 (Gaussian covariance (left), wave covariance (right)).



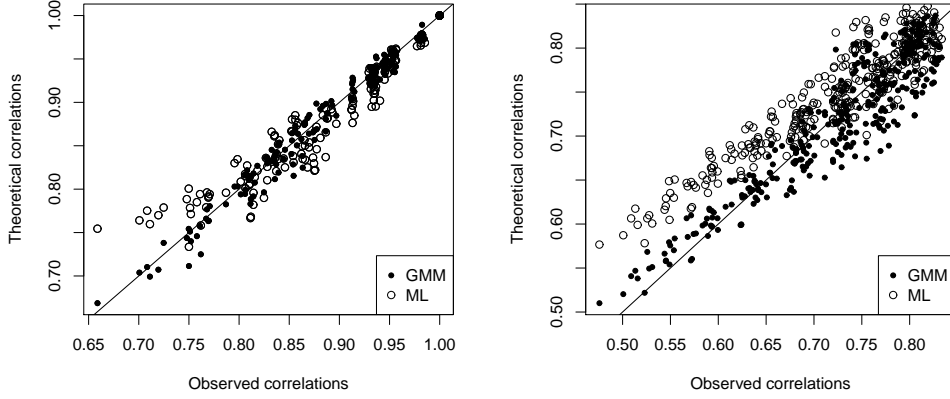


Figure 10: Theoretical correlations of the process  $Y$  for model  $(M_{\Gamma \sim \text{Sinus}})$  against observed ones at lag 0 (left) and lag 1 (right)

## 5.2 Parameterization of $\Lambda$

The structure of  $\alpha_1$ ,  $\alpha_0$  and  $\alpha_{-1}$  reveals a quadratic dependence in longitude and the dependence in latitude suggests the use of an intercept depending on latitude (see Figure 11). This following parameterization is then proposed.

$$\Lambda = \left( 1 \mid \text{Long} \mid \text{Long}^2 \right) \begin{pmatrix} \beta_1^{\text{Lat}} & \beta_4^{\text{Lat}} & \beta_7^{\text{Lat}} \\ \beta_2 & \beta_5 & \beta_8 \\ \beta_3 & \beta_6 & \beta_9 \end{pmatrix}$$

where  $\beta_i^{\text{Lat}}$  for  $i \in \{1, 4, 7\}$  takes a different value for each latitude and  $\text{Long} \in \mathbb{R}^K$  is a vector containing the longitude of the sites. Let  $(M_\Lambda)$  denote the corresponding model.  $\Lambda$  is of rank 3 if the matrix  $\begin{pmatrix} \beta_1^{\text{Lat}} & \beta_4^{\text{Lat}} & \beta_7^{\text{Lat}} \\ \beta_2 & \beta_5 & \beta_8 \\ \beta_3 & \beta_6 & \beta_9 \end{pmatrix}$  is full ranked because the matrix  $\left( 1 \mid \text{Long} \mid \text{Long}^2 \right)$  is full ranked.

The parameterization is easily handled in the GMM procedure whereas a

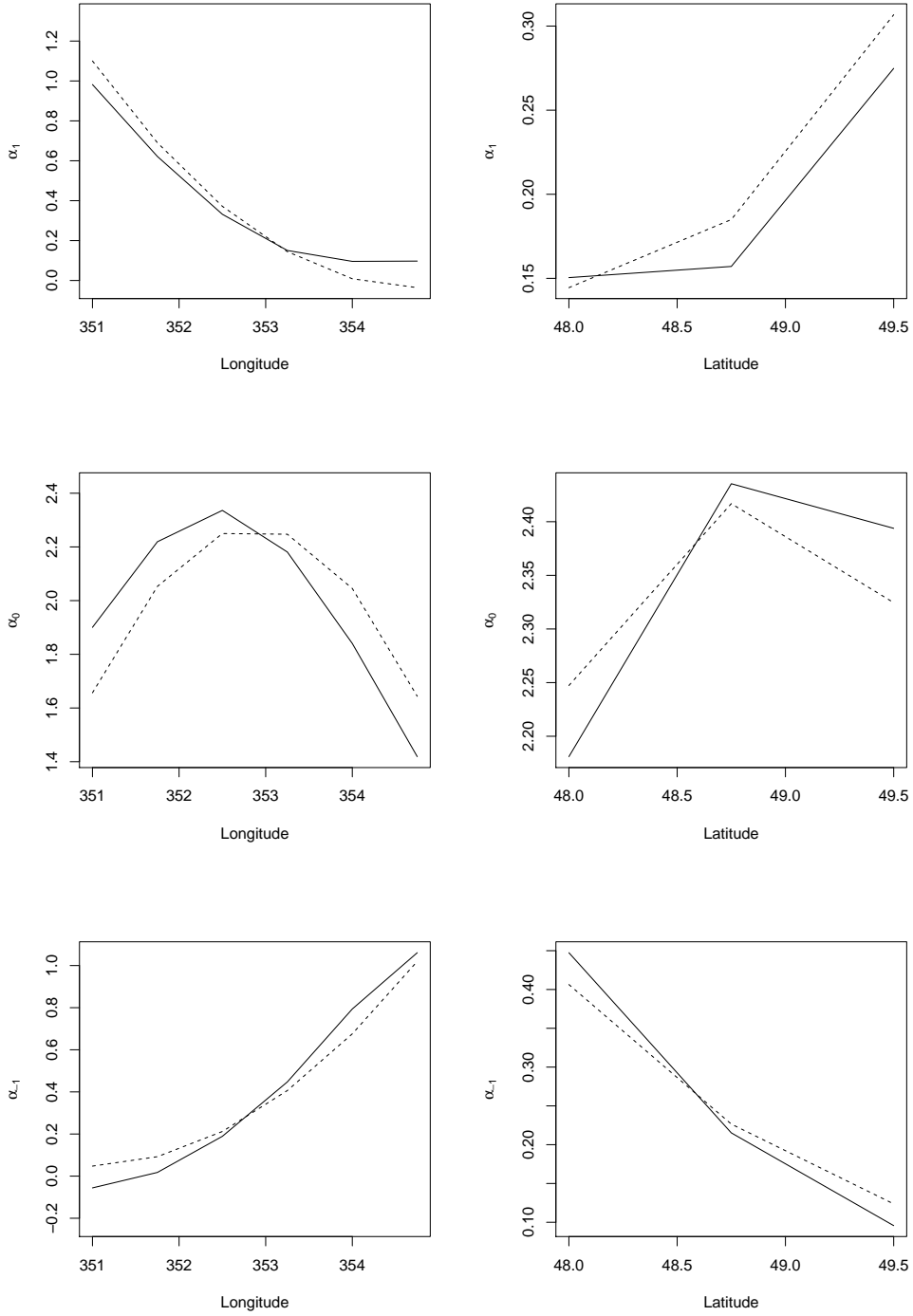


Figure 11: Estimated  $\alpha_1$  (top),  $\alpha_0$  (middle) and  $\alpha_{-1}$  (bottom) against longitude at latitude  $48^\circ$  N (left) and against latitude at longitude  $6.75^\circ$  W (right). Solid line: *ML* estimation of  $\Lambda$  for model (M), dashed line: parametric structure fitted by least square.

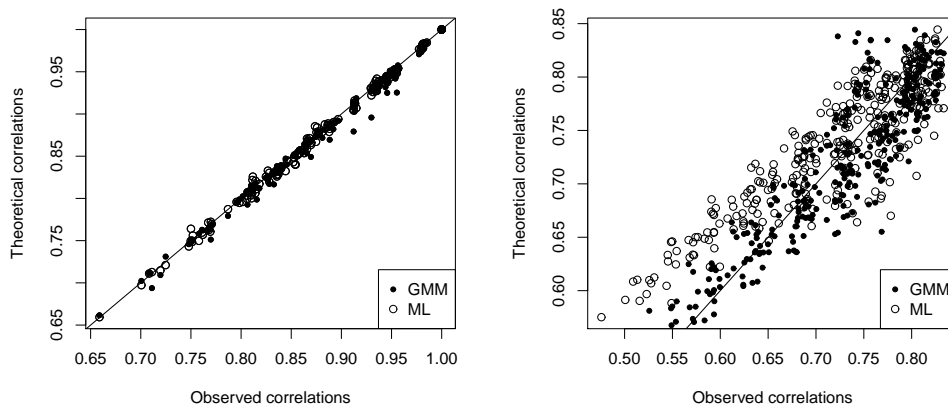


Figure 12: Theoretical correlations against observed ones at lag 0 (left) and lag 1 (right) for the model ( $M_\Lambda$ ).

numerical optimization is again needed to update  $\Lambda$  in the M-step. Moreover a joint optimization on  $\Lambda$  and  $\Gamma$  should be done since both of them are involved in the same part of log-likelihood. In order to avoid a numerical optimization in a high dimensional space, separate optimizations in  $\Lambda$  and in  $\Gamma$  have been performed leading to a so-called Generalized EM algorithm (see the supplementary materials for more details). The reduced ( $M_\Lambda$ ) and the full ( $M$ ) models give again similar results for the marginal distribution and the auto-correlation function. ( $M_\Lambda$ ) also leads to an accurate description of the spatial structure of the data (see Figure 12). Lagged one correlations are better reproduced by GMM parameters than ML parameters. The model ( $M_\Lambda$ ) is slightly inferior to the full model ( $M$ ) according to the BIC and MSPE values given in Table 2 but clearly outperforms the models ( $M_\Gamma$ ). It seems easier to find an appropriate reduced model for the loading matrix  $\Lambda$  than for the covariance matrix of the observation error  $\Gamma$ .

Model	Parameters	Log-likelihood	BIC	MSPE [min ; max]	
				GMM	ML
VAR(1)	495	-20707	46961		[ 0.249 ; 0.350 ]
(M <sub>2</sub> )	209	-24849	52040	[ 0.268 ; 0.410 ]	[ 0.256 ; 0.418 ]
(M)	208	-24954	52238	[ 0.264 ; 0.410 ]	[ 0.264 ; 0.414 ]
(M <sub>Λ</sub> )	186	-25399	52895	[ 0.277 ; 0.428 ]	[ 0.264 ; 0.417 ]
(M <sub>Γ~Gauss</sub> )	78	-29110	59082	[ 0.308 ; 0.428 ]	[ 0.274 ; 0.389 ]
(M <sub>Γ~Sinus</sub> )	78	-35615	72094	[ 0.349 ; 0.478 ]	[ 0.292 ; 0.403 ]
ARMA(2,1)	72				[ 0.366 ; 0.405 ]
Persistence forecast					[ 0.423 ; 0.468 ]

Table 2: Table of log-likelihoods and BIC indexes for the different models and Mean Square Percentage Error of one-step ahead forecasts by these models.

### 5.3 The hidden state as an AR(2) process

Higher order autoregressive models have been considered for modeling the dynamics of the hidden state. The best results have been obtained with an AR(2) model defined as

$$(M_2) \begin{cases} X_{t+1} = \rho_1 X_t + \rho_2 X_{t-1} + \sigma \epsilon_{t+1}, \\ Y_t = \alpha_1 X_{t+1} + \alpha_0 X_t + \alpha_{-1} X_{t-1} + \Gamma^{-1/2} \eta_t \end{cases} \quad \text{for } t \geq 0.$$

(M<sub>2</sub>) has been fitted using the same procedure than for model (M) and (M<sub>2</sub>). The ML estimate of  $\rho_1$  and  $\rho_2$  are respectively 0.91 and  $-0.11$ . They are close to the values obtained when fitting an AR(2) process to wind data at a single location (see [AM12]). (M<sub>2</sub>) slightly outperforms (M) according to the criteria considered in Section 4 and the values given in Table 2. According to Figure 8, the gain of using the (M<sub>2</sub>) model to produce short term forecasts is more important at central locations where (M<sub>2</sub>) and VAR(1) give similar results.

## 6 General discussion and perspectives

Several multisite models, all based on Gaussian linear state-space models, are proposed for wind speed. The main innovation with respect to the other space-time models which have been proposed for meteorological variables is the introduction of a latent process which describes regional conditions. It leads to interpretable models which can reproduce the marginal distribution of wind speed and important properties of the space-time covariance structure such as the asymmetries induced by prevailing motions of the air masses.

An important advantage of Gaussian linear state-space models is that efficient and easy to implement procedure can be used to fit the model. Two estimation procedures, one based on a method of moment (GMM) and the other on the likelihood function (ML) have been compared. GMM appears to be better when looking at the short-term space-time structure but ML seems to better capture the long-term dynamics. According to Table 2, ML estimates also generally lead to more accurate short-term forecasts except for the model (M). With this latter the southern line (points numbered 1 to 6) is better predicted by GMM parameters than by ML parameters whereas for the other models prediction is better with ML parameters at almost all stations.

According to Table 2 the rankings of the models according to BIC or MSPE, which measures the accuracy of one-step ahead forecasts on a validation set, are generally consistent. It indicates that BIC can be trusted when selecting the best model despite the very high conditioning number (ratio of the greatest eigenvalue over the smallest one) of the estimated matrix  $\Gamma$  which inverse appears in the log-likelihood function. The ranking also coincides with the complexity of the model: the quality of the model is deteriorated when the number of parameters is reduced. It highlights the difficulty to find parsimo-

nious and realistic models for describing the space-time evolution of the wind. Similar results have been obtained on the Irish wind data set considered in [HR89, Gne02] which has a different space-time sampling with daily data and stations on an irregular spatial grid.

The proposed models could be improved in several ways. In particular, we would like to introduce covariates such as large-scale climate variables in order to better understand the interannual variability (see [AM12]) or other meteorological variables which could help improving the quality of the short term forecasts (e.g. wind direction, pressure). Alternatively we could include a discrete component within the latent process (a hierarchical model) to describe the regime shifts induced by the weather types (e.g. cycloning conditions with westerly flows and anticycloning conditions with easterly flows).

Parler de l'estimation?? Modele tps continu??

## A Proof of proposition 1

Let  $\{Y_t\}$  [resp.  $\{\tilde{Y}_t\}$ ] denote a process satisfying (M) with parameters  $\theta = (\rho, \sigma, \Lambda, \Gamma)$  [resp.  $\tilde{\theta} = (\tilde{\rho}, \tilde{\sigma}, \tilde{\Lambda}, \tilde{\Gamma})$ ]. We assume that  $\frac{\sigma^2}{1-\rho^2} = 1$  and  $\Lambda$  is full ranked, with the same constraints holding true for  $\tilde{\theta}$ . We also assume that  $\{Y_t\}$  and  $\{\tilde{Y}_t\}$  have the same second order structure. We prove below that if these conditions hold true then  $\theta = \tilde{\theta}$ . The proof is based on the properties of  $C_k = cov(Y_t, Y_{t+k})$ .

- **Identification of  $\rho$  and  $\sigma$ .** According to (3), we have  $C_k = \rho^{k-2}C_2$  for  $k \geq 2$  and

$$C_2 = \frac{\sigma^2}{1-\rho^2}uv^t$$

with  $u = \alpha_1 + \rho\alpha_0 + \rho^2\alpha_{-1}$  and  $v = \rho^2\alpha_1 + \rho\alpha_0 + \alpha_{-1}$ . Since  $\alpha_{-1}$ ,  $\alpha_0$  and

$\alpha_1$  are linearly independent,  $u \neq 0$  and  $v \neq 0$  and thus  $C_2 \neq 0$ .  $\rho$  can thus be expressed as a ratio between some coefficients of  $C_3$  and  $C_2$  and we deduce that  $\rho = \tilde{\rho}$ . Using the constraint  $\frac{\sigma^2}{1-\rho^2} = 1$ , we also deduce that  $\sigma^2 = \tilde{\sigma}^2$ .

- **Identification of  $\Lambda$  when  $\rho \neq 0$ .** According to (2-3) we have  $C_2 - \rho C_1 = (1 - \rho^2)\alpha_1\alpha_{-1}^t$  and thus  $\alpha_1\alpha_{-1}^t = \tilde{\alpha}_1\tilde{\alpha}_{-1}^t$  since  $\rho^2 \neq 1$ . We deduce that there exists a real constant  $k_1 \neq 0$  such that  $\alpha_{-1} = k_1\tilde{\alpha}_{-1}$  and  $\alpha_1 = k_1^{-1}\tilde{\alpha}_1$ . We also have  $uv^t = \tilde{u}\tilde{v}^t$  where  $\tilde{u}$  and  $\tilde{v}$  are defined similarly to  $u$  and  $v$ . We deduce that there exists a real constant  $k_2 \neq 0$  such that  $\tilde{u} = k_2u$  and  $\tilde{v} = k_2^{-1}v$  and thus  $\tilde{u} - \tilde{v} = k_2u - k_2^{-1}v$  with

$$\begin{aligned}\tilde{u} - \tilde{v} &= (1 - \rho^2)\tilde{\alpha}_1 + (\rho^2 - 1)\tilde{\alpha}_{-1} \\ &= (1 - \rho^2)k_1^{-1}\alpha_1 + (\rho^2 - 1)k_1\alpha_{-1}\end{aligned}\tag{5}$$

$$k_2u - k_2^{-1}v = (k_2 - \rho^2k_2^{-1})\alpha_1 + \rho(k_2 - k_2^{-1})\alpha_0 + (k_2\rho^2 - k_2^{-1})\alpha_{-1}\tag{6}$$

Since  $\alpha_{-1}$ ,  $\alpha_0$  and  $\alpha_1$  are linearly independent, we can identify the coefficients of the linear combinations (56) and deduce, when  $\rho \neq 0$  that  $k_2 \in \{-1, 1\}$  and  $\alpha_i = k_2\tilde{\alpha}_i$  for  $i \in \{-1, 0, 1\}$ .

- **Identification of  $\Lambda$  when  $\rho = 0$ .** J'ai une preuve mais c'est un peu "bourrin"... Dans le cas où  $\frac{1}{2}\rho = 0$ ,  $Y_t$  est un MA(4). Est-ce qu'on ne peut pas juste donner une référence? Par exemple Akaike 74.

In this case,  $\{Y_t\}$  is a multivariate moving average process. Identifiability of such processes is studied in Akaike 74.

- **Identification of  $\Gamma$ .** According to (1),  $\Gamma$  can be expressed from  $C_0$  and the other parameters. We easily deduce that  $\tilde{\Gamma} = \Gamma$

Here we prove that full-symmetry can not be achieved under the chosen identifiability constraints. Separability of a space-time covariance function implies full-symmetry of this latter ([Gne02]). Full-symmetry of the space-time covariance function implies that the matrix  $C_2$  is a symmetric matrix. The symmetry of  $C_2$  implies  $uv^t = vu^t$ ,  $u$  and  $v$  are then collinear vectors which implies a collinearity between  $\alpha_1$ ,  $\alpha_0$  and  $\alpha_{-1}$ . The space-time covariance function defined by the model is not fully-symmetric and then non-separable.

## B Supplementary materials

**MLE for the model (M) and associated reduced models:** This file contains a description of the Expectation-Maximization (EM) algorithm used to fit the model (M) and the associated reduced models (supp\_estimation.pdf).

## References

- [Abr97] P. Abrahamsen. *A review of Gaussian random fields and correlation functions*. Norsk Regnesentral/Norwegian Computing Center, 1997.
- [AFM06] P. Ailliot, E. Frenod, and V. Monbet. Long term object drift forecast in the ocean with tide and wind. *Multiscale Modeling & Simulation*, 5(2):514–531, 2006.
- [AM12] P. Ailliot and V. Monbet. Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*, 30:92–101, 2012.



- [AMP06] P. Ailliot, V. Monbet, and M. Prevosto. An autoregressive model with time-varying coefficients for wind fields. *Environmetrics*, 17(2):107–117, 2006.
- [BD06] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer, New York, 2006. Reprint of the second (1991) edition.
- [BKM84] B. G. Brown, R. W. Katz, and A. H. Murphy. Time series models to simulate and forecast wind speed and wind power. *Journal of climate and applied meteorology*, 23:1184–1195, 1984.
- [Bor10] L. Bork. *Macro Factors, Monetary Policy Analysis and Affine Term Structure Models*. Aarhus School of Business, Department of Business Studies, 2010.
- [BW12] J. Bai and P. Wang. Identification and estimation of dynamic factor models. 2012.
- [Cai88] P. E. Caines. *Linear stochastic systems*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1988.
- [CMR05] O. Cappé, E. Moulines, and T. Rydén. *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York, 2005.
- [Cre91] N. A. C. Cressie. *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1991. A Wiley-Interscience Publication.

- [DK12] J. Durbin and S. J. Koopman. *Time series analysis by state space methods*, volume 38 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, second edition, 2012.
- [dLG05] X. de Luna and M.G. Genton. Predictive spatio-temporal models for spatially sparse environmental data. *Statist. Sinica*, 15(2):547–568, 2005.
- [DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38, 1977. With discussion.
- [FHI07] B. Finkenstädt, L. Held, and V. Isham, editors. *Statistical methods for spatio-temporal systems*, volume 107 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2007. Papers from the 6th Séminaire Européen de Statistique held in Bernried, December 12–18, 2004.
- [FNAB10] C. Flecher, P. Naveau, D. Allard, and N. Brisson. A stochastic daily weather generator for skewed data. *Water Resources Research*, 46(7):W07519, 2010.
- [GLW<sup>+</sup>06] T. Gneiting, K. Larson, K. Westrick, M. G. Genton, and E. Aldrich. Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space–time method. *Journal of the American Statistical Association*, 101(475):968–979, 2006.
- [Gne02] T. Gneiting. Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.*, 97(458):590–600, 2002.

- [Has07] K. A. Haskard. An anisotropic matérn spatial covariance model: Reml estimation and properties. *Ph.D. dissertation, University of Adelaide, Australia*, 2007.
- [HD88] E. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. Springer Texts in Statistics. John Wiley, New York, second edition, 1988. With 1 CD-ROM (Windows).
- [Hin77] D. Hinkley. On quick choice of power transformation. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26(1):pp. 67–69, 1977.
- [HR89] J. Haslett and A. E. Raftery. Space-time modelling with long-memory dependence: Assessing ireland’s wind power resource. *Applied Statistics*, pages 1–50, 1989.
- [HS13] M. Hofmann and I. B. Sperstad. Nowicob—a tool for reducing the maintenance costs of offshore wind farms. *Energy Procedia*, 35:177–186, 2013.
- [KJ97] L. Kamal and Y. Z. Jafri. Time series models to simulate and forecast hourly averaged wind speed in quetta, pakistan. *Solar Energy*, 61(1):23–32, 1997.
- [Lju99] L. Ljung. *System Identifiability*. Springer Texts in Statistics. Prentice Hall, New Jersey, second edition, 1999. With 1 CD-ROM (Windows).
- [MAP07] V. Monbet, P. Ailliot, and M. Prevosto. Survey of stochastic models for wind and sea state time series. *Probabilistic Engineering Mechanics*, 22(2):113–126, 2007.

- [NBS96] H. Nfaoui, J. Buret, and A. A. M. Sayigh. Stochastic simulation of hourly average wind speed sequences in tangiers (morocco). *Solar Energy*, 56(3):301–314, 1996.
- [NM94] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2111–2245. North-Holland, Amsterdam, 1994.
- [PD10] P. Papadopoulos and V. Digalakis. Identification of linear systems in canonical form through an em framework. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 4110–4113. IEEE, 2010.
- [RBBP11] A. Refice, A. Belmonte, F. Bovenga, and G. Pasquariello. On the use of anisotropic covariance models in estimating atmospheric din-sar contributions. *Geoscience and Remote Sensing Letters, IEEE*, 8(2):341–345, 2011.
- [RM13] I. Rychlik and A. Mustedanagic. A spatial-temporal model for wind speeds variability. 2013.
- [ŠBŠ11] J. Šaltytė Benth and L. Šaltytė. Spatial-temporal model for wind speed in Lithuania. *J. Appl. Stat.*, 38(6):1151–1168, 2011.
- [SS06] R. H. Shumway and D.S. Stoffer. *Time series analysis and its applications*. Springer Texts in Statistics. Springer, New York, second edition, 2006. With R examples.
- [ST<sup>+</sup>90] E. Skidmore, J. Tatarko, et al. Stochastic wind simulation for erosion modeling. *Transactions of the ASAE*, 33(6):1893–1899, 1990.

- [WH10] C. K. Wikle and M. B. Hooten. A general science-based framework for dynamical spatio-temporal models. *Test*, 19(3):417–451, 2010.
- [ZG12] X. Zhu and M. G. Genton. Short-term wind speed forecasting for power system operations. *International Statistical Review*, 80(1):2–23, 2012.