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

J. Bessac[1], P. Ailliot[2], V. Monbet [1,3]

Supplementary materials

[1] *Institut de Recherche Mathématiques de Rennes, UMR 6625, Université de Rennes 1, Rennes, France*
[2] *Laboratoire de Mathématiques de Bretagne Atlantique, UMR 6205, Université de Brest, Brest, France*
[3] *INRIA Rennes, ASPI, Rennes, France*

Second order structure and identifiability

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

$$\begin{aligned}
\mathrm{cov}_\theta(\boldsymbol{Y}_t, \boldsymbol{Y}_t) &= \frac{\sigma^2}{1-\rho^2}\Big(\boldsymbol{\alpha}_1(\boldsymbol{\alpha}_1+\rho\boldsymbol{\alpha}_0+\rho^2\boldsymbol{\alpha}_{-1})^t+\boldsymbol{\alpha}_0(\rho\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_0+\rho\boldsymbol{\alpha}_{-1})^t+\\
&\quad \boldsymbol{\alpha}_{-1}(\rho^2\boldsymbol{\alpha}_1+\rho\boldsymbol{\alpha}_0+\boldsymbol{\alpha}_{-1})^t\Big)+\boldsymbol{\Gamma}, \qquad (1)\\
\mathrm{cov}_\theta(\boldsymbol{Y}_t, \boldsymbol{Y}_{t+1}) &= \frac{\sigma^2}{1-\rho^2}\Big(\boldsymbol{\alpha}_1(\rho\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_0+\rho\boldsymbol{\alpha}_{-1})^t+\boldsymbol{\alpha}_0(\rho^2\boldsymbol{\alpha}_1+\rho\boldsymbol{\alpha}_0+\boldsymbol{\alpha}_{-1})^t+\\
&\quad \rho\boldsymbol{\alpha}_{-1}(\rho^2\boldsymbol{\alpha}_1+\rho\boldsymbol{\alpha}_0+\boldsymbol{\alpha}_{-1})^t\Big), \qquad (2)\\
\mathrm{cov}_\theta(\boldsymbol{Y}_t, \boldsymbol{Y}_{t+k}) &= \frac{\sigma^2}{1-\rho^2}\rho^{k-2}(\boldsymbol{\alpha}_1+\rho\boldsymbol{\alpha}_0+\rho^2\boldsymbol{\alpha}_{-1})(\rho^2\boldsymbol{\alpha}_1+\rho\boldsymbol{\alpha}_0+\boldsymbol{\alpha}_{-1})^t, \qquad (3)\\
&\quad \text{for all } k\geq 2.
\end{aligned}$$

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

Proposition 1 *Assume that* (M) *holds. Assume further that* $\frac{\sigma^2}{1-\rho^2}=1$ *and that the vectors* $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_0$ *and* $\boldsymbol{\alpha}_{-1}$ *are linearly independent. Then the parameters can be identified from the distribution of the process* $\{\boldsymbol{Y_t}\}$.

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

- **Identification of ρ and σ.** According to (3), we have $\boldsymbol{C}_k = \rho^{k-2}\boldsymbol{C}_2$ for $k \geq 2$ and
$$\boldsymbol{C}_2 = \frac{\sigma^2}{1-\rho^2}\boldsymbol{u}\boldsymbol{v}^t$$
with $\boldsymbol{u} = \boldsymbol{\alpha}_1 + \rho\boldsymbol{\alpha}_0 + \rho^2\boldsymbol{\alpha}_{-1}$ and $\boldsymbol{v} = \rho^2\boldsymbol{\alpha}_1 + \rho\boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_{-1}$. Since $\boldsymbol{\alpha}_{-1}$, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ are linearly independent, $\boldsymbol{u} \neq 0$ and $\boldsymbol{v} \neq 0$ and thus $\boldsymbol{C}_2 \neq 0$. ρ can thus be expressed as a ratio between some coefficients of $\boldsymbol{C}_3$ and $\boldsymbol{C}_2$ 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 $\boldsymbol{\Lambda}$ when $\rho \neq 0$.** According to (2-3) we have $\boldsymbol{C}_2 - \rho\boldsymbol{C}_1 = (1-\rho^2)\boldsymbol{\alpha}_1\boldsymbol{\alpha}_{-1}^t$ and thus $\boldsymbol{\alpha}_1\boldsymbol{\alpha}_{-1}^t = \tilde{\boldsymbol{\alpha}}_1\tilde{\boldsymbol{\alpha}}_{-1}^t$ since $\rho^2 \neq 1$. We deduce that there exists a real constant $k_1 \neq 0$ such that $\boldsymbol{\alpha}_{-1} = k_1\tilde{\boldsymbol{\alpha}}_{-1}$ and $\boldsymbol{\alpha}_1 = k_1^{-1}\tilde{\boldsymbol{\alpha}}_1$. We also have $\boldsymbol{u}\boldsymbol{v}^t = \tilde{\boldsymbol{u}}\tilde{\boldsymbol{v}}^t$ where $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{v}}$ are defined similarly to $\boldsymbol{u}$ and $\boldsymbol{v}$. We deduce that there exists a real constant $k_2 \neq 0$ such that $\tilde{\boldsymbol{u}} = k_2\boldsymbol{u}$ and $\tilde{\boldsymbol{v}} = k_2^{-1}\boldsymbol{v}$ and thus $\tilde{\boldsymbol{u}} - \tilde{\boldsymbol{v}} = k_2\boldsymbol{u} - k_2^{-1}\boldsymbol{v}$ with
$$\begin{aligned}
\tilde{\boldsymbol{u}} - \tilde{\boldsymbol{v}} &= (1-\rho^2)\tilde{\boldsymbol{\alpha}}_1 + (\rho^2-1)\tilde{\boldsymbol{\alpha}}_{-1} \\
&= (1-\rho^2)k_1^{-1}\boldsymbol{\alpha}_1 + (\rho^2-1)k_1\boldsymbol{\alpha}_{-1} && (4)\\
k_2\boldsymbol{u} - k_2^{-1}\boldsymbol{v} &= (k_2 - \rho^2 k_2^{-1})\boldsymbol{\alpha}_1 + \rho(k_2 - k_2^{-1})\boldsymbol{\alpha}_0 \\
&\quad + (k_2\rho^2 - k_2^{-1})\boldsymbol{\alpha}_{-1} && (5)
\end{aligned}$$
Since $\boldsymbol{\alpha}_{-1}$, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ are linearly independent, we can identify the coefficients of the linear combinations (4-5) and deduce, when $\rho \neq 0$ that $k_2 \in \{-1, 1\}$ and $\boldsymbol{\alpha}_i = k_2\tilde{\boldsymbol{\alpha}}_i$ for $i \in \{-1, 0, 1\}$.

- **Identification of $\boldsymbol{\Lambda}$ when $\rho = 0$.** In this case,
$$\begin{aligned}
\boldsymbol{C_1} &= \sigma^2(\boldsymbol{\alpha}_1\boldsymbol{\alpha}_0^t + \boldsymbol{\alpha}_0\boldsymbol{\alpha}_{-1}^t), && (6)\\
\boldsymbol{C_2} &= \sigma^2\boldsymbol{\alpha}_1\boldsymbol{\alpha}_{-1}^t && (7)
\end{aligned}$$
By similar reasoning as previously from (7) there exists $k_1 \neq 0$ such that $\boldsymbol{\alpha}_{-1} = k_1\tilde{\boldsymbol{\alpha}}_{-1}$ and $\boldsymbol{\alpha}_1 = k_1^{-1}\tilde{\boldsymbol{\alpha}}_1$. From (6) we deduce that $\boldsymbol{\alpha}_1(k_1\tilde{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0)^t + (\frac{\tilde{\boldsymbol{\alpha}}_0}{k_1} - \boldsymbol{\alpha}_0)\boldsymbol{\alpha}_{-1}^t = 0$.

If $k_1\tilde{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0 \neq 0$ then there exists $k_2 \neq 0$ such that $\boldsymbol{\alpha}_1 - \frac{k_2}{k_1}\tilde{\boldsymbol{\alpha}}_0 - k_2\boldsymbol{\alpha}_0 = 0$ (R_1) and $\frac{1}{k_2}\boldsymbol{\alpha}_{-1} + \boldsymbol{\alpha}_0 + k_1\tilde{\boldsymbol{\alpha}}_0 = 0$ (R_2).

Then

$$(\mathrm{R}_1) - \frac{k_2}{k_1}(\mathrm{R}_2) \quad = \quad \boldsymbol{\alpha}_1 + (k_2 + \frac{k_2}{k_1^2})\boldsymbol{\alpha}_0 + \frac{1}{k_1^2}\boldsymbol{\alpha}_{-1} = 0.$$

Since $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_{-1}$ are linearly independent we obtain $k_1 = k_2 = 0$ which is a contradiction.

If $k_1\tilde{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0 = 0$, this implies $\frac{\tilde{\boldsymbol{\alpha}}_0}{k_1} - \boldsymbol{\alpha}_0 = 0$, then $k_1 = \pm 1$. In both cases, $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_0$ and then identifiable from the covariance $\boldsymbol{C_2}$ and $\boldsymbol{C_1}$.

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

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 (Gneiting, 2002). Full-symmetry of the space-time covariance function implies that the matrix $\boldsymbol{C_2}$ is a symmetric matrix. The symmetry of $\boldsymbol{C_2}$ implies $\boldsymbol{u}\boldsymbol{v}^t = \boldsymbol{v}\boldsymbol{u}^t$, $\boldsymbol{u}$ and $\boldsymbol{v}$ are then collinear vectors which implies a collinearity between $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_{-1}$. The space-time covariance function defined by the model is not fully-symmetric and then non-separable.

Maximum Likelihood Estimation for the model (M) and associated reduced models

Maximum likelihood estimation of the parameter θ for models with latent variables consists in maximizing the incomplete likelihood function based on observed set $(\boldsymbol{y}_1, ..., \boldsymbol{y}_T)$:

$$\mathcal{L}(\theta; \boldsymbol{y}_1, ..., \boldsymbol{y}_T) = p(\boldsymbol{y}_1, ..., \boldsymbol{y}_T; \theta) = \mathcal{L}(\theta; \boldsymbol{y}_1, ..., \boldsymbol{y}_T) = p(\boldsymbol{y}_1)\prod_{t=2}^{T} p(\boldsymbol{y}_t|\boldsymbol{y}_1, ..., \boldsymbol{y}_{t-1}; \theta).$$

In the Gaussian linear case, the likelihood of the observations $(\boldsymbol{y}_1, ..., \boldsymbol{y}_T)$ can be computed easily since for all $t \in \{1, ..., T\}$ $(\boldsymbol{Y}_1, ..., \boldsymbol{Y}_t)$ is a Gaussian vector. It gives for the model (M):

$$\mathcal{L}(\boldsymbol{Y}_t|\boldsymbol{Y}_1^{t-1} = \boldsymbol{y}_1^{t-1}) = \mathcal{N}(\boldsymbol{\Lambda}\tilde{\boldsymbol{X}}_{t|t-1}, \boldsymbol{F}_{t|t-1}) \text{ where } \tilde{\boldsymbol{X}}_t = \begin{pmatrix} X_{t+1} \\ X_t \\ X_{t-1} \end{pmatrix},$$

with $\tilde{\boldsymbol{X}}_{t|t-1} = \mathrm{E}(\tilde{\boldsymbol{X}}_t|\boldsymbol{Y}_1^{t-1} = \boldsymbol{y}_1^{t-1})$ and $\boldsymbol{F}_{t|t-1} = \mathrm{Var}(\boldsymbol{Y}_t|\boldsymbol{Y}_1^{t-1} = \boldsymbol{y}_1^{t-1}) = \boldsymbol{\Lambda}\boldsymbol{P}_{t|t-1}\boldsymbol{\Lambda}^t + \boldsymbol{\Gamma}$

where $\boldsymbol{P}_{t|t-1} = \mathrm{Var}(\tilde{\boldsymbol{X}}_t|\boldsymbol{Y}_1^{t-1} = \boldsymbol{y}_1^{t-1}) = \mathrm{E}((\tilde{\boldsymbol{X}}_t - \tilde{\boldsymbol{X}}_{t|t-1})(\tilde{\boldsymbol{X}}_t - \tilde{\boldsymbol{X}}_{t|t-1})^t|\boldsymbol{Y}_1^{t-1} = \boldsymbol{y}_1^{t-1})$ with $\boldsymbol{y}_1^{t-1} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_{t-1})$. Both quantities $\tilde{\boldsymbol{X}}_{t|t-1}$ and $\boldsymbol{P}_{t|t-1}$ are computed from Kalman filter described below (see also (Shumway and Stoffer, 2006)). However no explicit expressions of the optimal parameters are available from this incomplete likelihood, a maximum likelihood estimation procedure would involve a numerical optimization of this function which is not reasonable in high dimension. A major feature of the EM algorithm (Dempster et al., 1977) is the maximization of the complete likelihood over the parameter θ.

Kalman recursions

The goal of filtering (respectively smoothing, respectively prediction) is to obtain as much as possible information about the hidden variable X_t from the observations $(\boldsymbol{y}_1, ..., \boldsymbol{y}_t)$ (respectively $(\boldsymbol{y}_1, ..., \boldsymbol{y}_T)$, respectively $(\boldsymbol{y}_1, ..., \boldsymbol{y}_{t-1})$). The solution consists in computing recursively the conditional law of X_t according to $(\boldsymbol{y}_1, ..., \boldsymbol{y}_t)$ (respectively $(\boldsymbol{y}_1, ..., \boldsymbol{y}_T)$, respectively $(\boldsymbol{y}_1, ..., \boldsymbol{y}_{t-1})$), which realizes the best approximation of X_t according to $(\boldsymbol{y}_1, ..., \boldsymbol{y}_t)$ in terms of mean square error.

Kalman prediction and filtering: $(\tilde{\boldsymbol{X}}_t, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_{t-1})$ is a Gaussian vector then the conditional distribution of $\tilde{\boldsymbol{X}}_t$ according to $(\boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1})$ is a Gaussian distribution with parameters: $\tilde{\boldsymbol{X}}_{t|t-1} = E(\tilde{\boldsymbol{X}}_t|\boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1})$ and $\boldsymbol{P}_{t|t-1} = \mathrm{Var}(\tilde{\boldsymbol{X}}_t|\boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = \mathrm{E}((\tilde{\boldsymbol{X}}_t - \tilde{\boldsymbol{X}}_{t|t-1})(\tilde{\boldsymbol{X}}_t - \tilde{\boldsymbol{X}}_{t|t-1})^t|\boldsymbol{Y}_1^{t-1} = \boldsymbol{y}_1^{t-1})$; and $\mathcal{L}(\tilde{\boldsymbol{X}}_t|\boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_t = \boldsymbol{y}_t) = \mathcal{N}(\boldsymbol{\Lambda}\tilde{\boldsymbol{X}}_{t|t}, \boldsymbol{P}_{t|t})$. Relationships between predicted and filtered quantities are the following:

$$\tilde{\boldsymbol{X}}_{t|t-1} = \tilde{\boldsymbol{\rho}}\tilde{\boldsymbol{X}}_{t-1|t-1},$$

$$\tilde{\boldsymbol{P}}_{t|t-1} = \tilde{\boldsymbol{\rho}}\tilde{\boldsymbol{P}}_{t-1|t-1}\tilde{\boldsymbol{\rho}}^t + \tilde{\boldsymbol{\sigma}},$$

$$\tilde{\boldsymbol{X}}_{t|t} = \tilde{\boldsymbol{X}}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \boldsymbol{\Lambda}\tilde{\boldsymbol{X}}_{t|t-1})$$

and

$$\tilde{\boldsymbol{P}}_{t|t} = (I - \boldsymbol{K}_t\boldsymbol{\Lambda})\tilde{\boldsymbol{P}}_{t|t-1},$$

where $\boldsymbol{K}_t = \tilde{\boldsymbol{P}}_{t|t-1}\boldsymbol{\Lambda}^t(\boldsymbol{\Lambda}\tilde{\boldsymbol{P}}_{t|t-1}\boldsymbol{\Lambda}^t + \boldsymbol{\Gamma})^{-1}$ and $\boldsymbol{K}$ is called the Kalman gain. The two first expressions are easily derived from independence of $\boldsymbol{\epsilon}_t$ and $\boldsymbol{Y}_{t-1}$ and of $(\tilde{\boldsymbol{X}}_{t-1} - \tilde{\boldsymbol{X}}_{t-1|t-1})$ and $\boldsymbol{\epsilon}_t$. The two last relations are based on properties of the Gaussian process of innovations $\boldsymbol{I}_t = \boldsymbol{Y}_t - \mathrm{E}(\boldsymbol{Y}_t|\boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1})$.

Kalman smoothing: Computation of $\tilde{\boldsymbol{X}}_{t|t}$ and $\tilde{\boldsymbol{P}}_{t|t}$ is obtained through the following backward recursions:

$$\tilde{\boldsymbol{X}}_{t|t} = \tilde{\boldsymbol{X}}_{t|t} + \boldsymbol{J}_t(\tilde{\boldsymbol{X}}_{t+1|T} - \tilde{\boldsymbol{\rho}}\tilde{\boldsymbol{X}}_{t|t}),$$

$$\tilde{\boldsymbol{P}}_{t|t} = \tilde{\boldsymbol{P}}_{t|t} + \boldsymbol{J}_t(\tilde{\boldsymbol{P}}_{t+1|T} - \tilde{\boldsymbol{P}}_{t+1|t})\boldsymbol{J}_t^t,$$
$$\boldsymbol{J}_t = \tilde{\boldsymbol{P}}_{t|t}\tilde{\boldsymbol{\rho}}^t\tilde{\boldsymbol{P}}_{t+1|t}^{-1}$$

and

$$\tilde{\boldsymbol{P}}_{t,t-1|T} = \tilde{\boldsymbol{P}}_{t|t}J_{t-1}^t + \boldsymbol{J}_t(\tilde{\boldsymbol{P}}_{t+1,t|T} - \tilde{\boldsymbol{\rho}}\tilde{\boldsymbol{P}}_{t|t})J_{t-1}^t.$$

Similar computations of the previous ones based on conditional expectation of multivariate normal distribution are used to compute these quantities.

EM algorithm

Thanks to the Markov properties and Bayes formula, the complete likelihood of the model (M) for $\theta = (\rho, \sigma, \boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ is written as:

$$\begin{aligned}
\mathcal{L}(\theta; x_0, ..., x_T, \boldsymbol{y}_1, ..., \boldsymbol{y}_T) &= \mathcal{L}(\theta; \tilde{\boldsymbol{X}}_1, ..., \tilde{\boldsymbol{X}}_{T-1}, \boldsymbol{y}_1, ..., \boldsymbol{y}_T) \\
&= p(x_0)\prod_{i=1}^{T} p(x_i|x_{i-1};\theta)\prod_{i=1}^{T} p(\boldsymbol{y}_i|\tilde{\boldsymbol{X}}_i;\theta).
\end{aligned}$$

However the set $(x_0, ..., x_T)$ is not observed, the EM-algorithm enables to approximate $\hat{\theta}$ that maximizes the quantity $\mathrm{E}(\log(p(\tilde{\boldsymbol{X}}_1, ..., \tilde{\boldsymbol{X}}_t, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_T;\theta))|\boldsymbol{Y}_1^T = \boldsymbol{y}_1^T)$. The EM-algorithm computes approximations $\hat{\theta}_n$ of $\hat{\theta}$ in a recursive way by performing the following two steps at each iteration n:

Expectation step: Computation of

$$Q(\theta, \hat{\theta}_n) = \mathrm{E}(\log(\mathcal{L}(\tilde{\boldsymbol{X}}_1, ..., \tilde{\boldsymbol{X}}_{T-1}, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_T;\theta))|\boldsymbol{Y}_1^T = \boldsymbol{y}_1^T;\hat{\theta}_n),$$

through the Kalman filtering and smoothing recursions (see (Shumway and Stoffer, 2006)).

Maximization step: Computation of $\hat{\theta}_{n+1}$ by maximization of the function $(\theta \to Q(\theta, \hat{\theta}_n))$.

Since $\boldsymbol{X}^t\boldsymbol{M}\boldsymbol{X} = \mathrm{Trace}(\boldsymbol{M}\boldsymbol{X}\boldsymbol{X}^t)$ for all K-dimensional vector $\boldsymbol{X}$ and $K\times K$-matrix $\boldsymbol{M}$, the quantity $Q(\theta, \hat{\theta}_n)$ is derived:

$$\begin{aligned}
Q(\theta, \hat{\theta}_n) = & -\frac{1}{2}\Big((T-1)(\log(2\pi) + \log(\sigma^2)) + \frac{1}{\sigma^2}\sum_{i=2}^{T}\mathrm{E}((X_i - \rho X_{i-1})^2|\boldsymbol{y}_1^T;\hat{\theta}_n) \\
& + T(K\log(2\pi) + \log(\det(\boldsymbol{\Gamma}))) \\
& + \sum_{i=1}^{T}\mathrm{Trace}(\boldsymbol{\Gamma}^{-1}\mathrm{E}((\boldsymbol{y}_i - \boldsymbol{\Lambda}\tilde{\boldsymbol{X}}_i)(\boldsymbol{y}_i - \boldsymbol{\Lambda}\tilde{\boldsymbol{X}}_i)^t|\boldsymbol{y}_1^T;\hat{\theta}_n))\Big).
\end{aligned}$$

Then the following quantities $\hat{x}_i = \mathrm{E}(X_i|\boldsymbol{y}_1^T;\hat{\theta}_n)$, $\hat{x}_{i,i-1} = \mathrm{E}(X_iX_{i-1}{}^t|\boldsymbol{y}_1^T;\hat{\theta}_n)$, $\hat{\tilde{\boldsymbol{X}}}_i = \mathrm{E}(\tilde{\boldsymbol{X}}_i|\boldsymbol{y}_1^T;\hat{\theta}_n)$ and $\hat{\tilde{\boldsymbol{X}}}_{i,i} = \mathrm{E}(\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i{}^t|\boldsymbol{y}_1^T;\hat{\theta}_n)$ are needed for all $i \in$

$\{1, ..., T\}$ and derived from the Kalman filter and smoother. At each M-step, analytical expressions of the estimates of the parameters can be derived:

$$\rho_n = \frac{\sum_{i=2}^{T} \hat{x}_{i,i-1}}{\sum_{i=1}^{T} \hat{x}_{i,i}},$$

$$\boldsymbol{\Lambda}_n = \Big(\sum_{i=1}^{T} \boldsymbol{y}_i \hat{\tilde{\boldsymbol{X}}}_i{}^t\Big)\Big(\sum_{i=1}^{T} \hat{\tilde{\boldsymbol{X}}}_{i,i}\Big)^{-1} \text{ and } \boldsymbol{\Gamma}_n = \frac{1}{T}\sum_{i=1}^{T}(\boldsymbol{y}_i \boldsymbol{y}_i{}^t - \boldsymbol{\Lambda}_n \hat{\tilde{\boldsymbol{X}}}_i{}^t \boldsymbol{y}_i{}^t).$$

The estimation of $\boldsymbol{\Gamma}_n$ in models (M_Γ) and of $\boldsymbol{\Lambda}_n$ in the model (M_Λ) are processed by numerical optimization of the associated part of the log-likelihood. For the model (M_Γ), $\boldsymbol{\Lambda}_n$ is determined by its analytical expression and injected in the associated part of the likelihood which is optimized numerically to determine the parameters that structure $\boldsymbol{\Gamma}_n$. $\boldsymbol{\Gamma}_n$ is the maximizer of:

$$\begin{aligned}(\sigma_1, ..., \sigma_K, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2) \to \quad & T(K\log(2\pi) + \log(\det(\boldsymbol{\Gamma}_{par}))) \\ & + \sum_{i=1}^{T} \mathrm{Trace}(\boldsymbol{\Gamma}_{par}^{-1} \mathrm{E}((\boldsymbol{y}_i - \boldsymbol{\Lambda}_n \tilde{\boldsymbol{X}}_i)(\boldsymbol{y}_i - \boldsymbol{\Lambda}_n \tilde{\boldsymbol{X}}_i)^t | \boldsymbol{y}_1^T; \hat{\theta}_n)).\end{aligned}$$

Where $\boldsymbol{\Gamma}_{par}$ is the parametric covariance defined by $(\sigma_1, ..., \sigma_K, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2)$. Initial conditions of the parameters of the structure of $\boldsymbol{\Gamma}$ are determined empirically. In the estimation procedure associated with (M_Λ), $\boldsymbol{\Lambda}_n$ is determined as the maximizer of the function:

$$\begin{aligned}(\beta_1^{\mathrm{Lat}}, ..., \beta_9) \to \quad & T(K\log(2\pi) + \log(\det(\boldsymbol{\Gamma}_{n-1}))) \\ & + \sum_{i=1}^{T} \mathrm{Trace}(\boldsymbol{\Gamma}_{n-1}^{-1} \mathrm{E}((\boldsymbol{y}_i - \boldsymbol{\Lambda}_{par} \tilde{\boldsymbol{X}}_i)(\boldsymbol{y}_i - \boldsymbol{\Lambda}_{par} \tilde{\boldsymbol{X}}_i)^t | \boldsymbol{y}_1^T; \hat{\theta}_n)),\end{aligned}$$

with $\boldsymbol{\Lambda}_{par} = \begin{pmatrix} 1 & | & \mathrm{Long} & | & \mathrm{Long}^2 \end{pmatrix} \begin{pmatrix} \beta_1^{\mathrm{Lat}} & \beta_4^{\mathrm{Lat}} & \beta_7^{\mathrm{Lat}} \\ \beta_2 & \beta_5 & \beta_8 \\ \beta_3 & \beta_6 & \beta_9 \end{pmatrix}$. Initial conditions of this optimization are determined by a least square estimation between $\hat{\boldsymbol{\Lambda}}$, the output of the EM processes for the model (M), and $\boldsymbol{\Lambda}_{par}$. $\boldsymbol{\Gamma}_n$ is then determined as the maximizer of:

$$\begin{aligned}\boldsymbol{\Gamma} \to \quad & T(K\log(2\pi) + \log(\det(\boldsymbol{\Gamma}))) \\ & + \sum_{i=1}^{T} \mathrm{Trace}(\boldsymbol{\Gamma}^{-1} \mathrm{E}((\boldsymbol{y}_i - \boldsymbol{\Lambda}_n \tilde{\boldsymbol{X}}_i)(\boldsymbol{y}_i - \boldsymbol{\Lambda}_n \tilde{\boldsymbol{X}}_i)^t | \boldsymbol{y}_1^T; \hat{\theta}_n)).\end{aligned}$$

The splitting of optimization in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ into the EM algorithm refers to a Generalized Expectation-Maximization algorithm in which at each M-step only an improvement of the approximated incomplete likelihood is required.

Prediction as a validation tool

The time-step of the data makes unrealistic the use of the proposed model as a forecasting tool. Nevertheless, forecasting is used here a classical statistical tool for validation. Indeed it enables to evaluate many features linked to statistical modeling and it can, for instance, help to detect overfitting. The Markovian structure of the model (M) is such that the short-term forecast can be efficiently computed through the Kalman recursions (see (Brockwell and Davis, 2006, chapter 8)). The forecast is performed 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 skills of the model at location $i \in \{1, ..., K\}$ is evaluated by computing the natural empirical estimate of the Mean Square Percentage Error (MSPE) defined as

$$\text{MSPE}(i) = \frac{\text{Var}(\boldsymbol{Y_t}(i) - \text{E}[\boldsymbol{Y_t}(i)|\boldsymbol{Y}_0, ..., \boldsymbol{Y}_{t-1}])}{\text{Var}(\boldsymbol{Y_t}(i))}$$

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

For comparison purpose, a vector autoregressive model of order 1 (VAR(1)) was also fitted on the multivariate process $\boldsymbol{Y}$ of transformed mean-corrected wind speed. Such a high-dimension response vector may lead to a model VAR which suffers from over-parameterization and to a difficult interpretation of the parameters. Note that the BIC and MSPE criteria lead to coherent results.

Model	Parameters	Log-likelihood	BIC	MSPE [min ; max]	
				GMM	ML
VAR(1)	495	-20707	46961		[0.249 ; 0.350]
(M_2)	209	-24849	52040	[0.268 ; 0.410]	[0.256 ; 0.418]
(M)	208	-24954	52238	[0.264 ; 0.410]	[0.264 ; 0.414]
(M_Λ)	186	-25399	52895	[0.277 ; 0.428]	[0.264 ; 0.417]
($M_{\Gamma\sim Gauss}$)	78	-29110	59082	[0.308 ; 0.428]	[0.274 ; 0.389]
($M_{\Gamma\sim Sinus}$)	78	-35615	72094	[0.349 ; 0.478]	[0.292 ; 0.403]
Persistence forecast					[0.423 ; 0.468]

Table 1: Table of log-likelihoods and BIC indexes for the different models and Mean Square Percentage Error of one-step ahead forecasts by these models.

References

Brockwell, P. J. and Davis, R. A. (2006). *Time series: theory and methods.* Springer Series in Statistics. Springer, New York. Reprint of the second (1991) edition.

Dempster, A. P., M., L. N., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39(1):1–38.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.*, 97(458):590–600.

Shumway, R. H. and Stoffer, D. S. (2006). *Time series analysis and its applications*. Springer Texts in Statistics. Springer, New York, second edition. With R examples.