Markov-Switching AutoRegressive models for Cartesian component of wind fields in the North-East Atlantic

Julie BESSAC - Université de Rennes 1
with Pierre AILLIOT and Valérie MONBET

September 2014

Plan

Motivations for the use of weather generators

Many natural phenomena and human activities depend on wind conditions

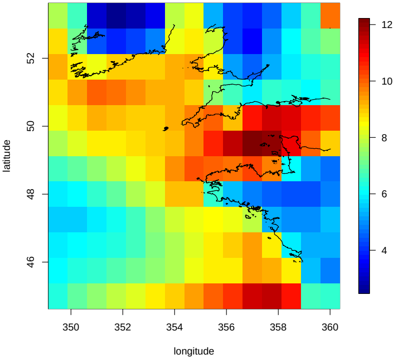
- Production of electricity by wind turbines
- Evolution of a coast line
- Maritime transport
- Drift of objects in the ocean

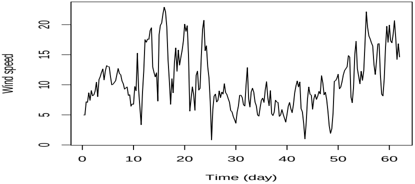
Wind data generally available on short periods of time

- 50 years of data maximum
- Not enough to compute reliable estimates of the probability of complex events

$\rightarrow$ Stochastic model used to simulate unlimited numbers of artificial wind sequences

Plan





Goals:

- to propose a stochastic generator for (u, v)-wind fields
- to account for the regime-switching induced by synoptic conditions
- to compare several regime-switching models

Plan

Plan

Reanalysis data ERA Interim from ECMWF

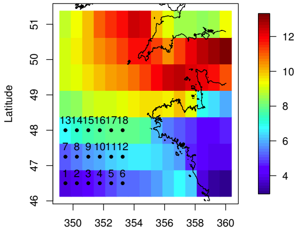
- Zonal and meridional components: u and v
- Wind speed: U
- Wind direction: Φ

Available with regular sampling: $\Delta x = 0.75^o$, $\Delta t = 6h$

Study of months of January from 1979 to 2011

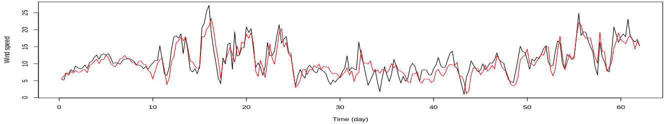
Transformation with $\alpha > 1$ facilitates the modeling of the bi-modal distributions of u and v

$$\begin{cases} \tilde{u}_t & = & U_t^\alpha \cos(\Phi_t) \\ \tilde{v}_t & = & U_t^\alpha \sin(\Phi_t) \end{cases}$$



Proposed models fitted on the 5 locations (1, 6, 10, 13, 18) to avoid over-parameterization

One of the main goals is: reproducing space-time motions of meteorological systems and the associated alternation in intensity and temporal variability



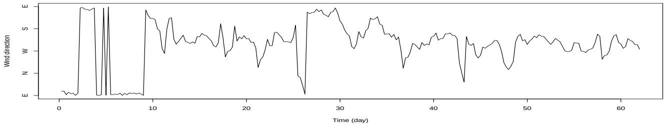


Figure : Top: time series of wind speed, black: western location, red: eastern location Bottom: time series of wind direction

Some statistics computed on data

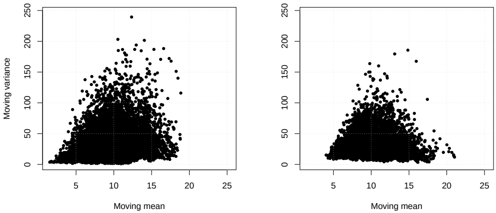


Figure : Moving variance over 9 time steps against of the value U against its moving mean at location 10. Left: data, right: simulation from the VAR(2)

To account for the alternation of temporal variability $\rightarrow$ Vector AutoRegressive models with regime-switching

Plan

- $\{S_t\}$: Markov chain valued in $\{1, ..., M\}$ describes the current weather type

- $\{S_t\}$ can be latent or observed

- Given the value of S_t, the observation $\mathbf{Y}_t$ is written as:

$$\mathbf{Y}_t = \mathbf{A}_0^{(S_t)} + \mathbf{A}_1^{(S_t)}\mathbf{Y}_{t-1} + \mathbf{A}_2^{(S_t)}\mathbf{Y}_{t-2} + ... + \mathbf{A}_p^{(S_t)}\mathbf{Y}_{t-p} + (\mathbf{\Sigma}^{(S_t)})^{-1/2}\epsilon_t, \quad (1)$$

$\mathbf{Y}$: observed power-transformed K-dimensional process
For $i \in \{1, ..., M\}$, $\mathbf{A}_0^{(i)} \in \mathbb{R}^K$, $\mathbf{A}_1^{(i)}, ..., \mathbf{A}_p^{(i)}, \mathbf{\Sigma}^{(i)} \in \mathcal{M}_{K,K}$
ϵ is a Gaussian white noise of dimension K

Conditional independences between S and $\mathbf{Y}$ for $p = 1$:

$$\begin{tikzcd}
\cdots \arrow[r] \arrow[d] & S_{t-1} \arrow[r] \arrow[d] & S_t \arrow[r] \arrow[d] & S_{t+1} \arrow[r] \arrow[d] & \cdots \arrow[d] \\
\cdots \arrow[r] & \mathbf{Y}_{t-1} \arrow[r] & \mathbf{Y}_t \arrow[r] & \mathbf{Y}_{t+1} \arrow[r] & \cdots
\end{tikzcd}$$

The regime S can be latent or observed:

· The regime is said to be observed when regimes are identified separately from the conditional model
Clustering methods are run on extra-variables, such as descriptors of atmospheric circulation or from local variables

· The regime is said to be latent when it is introduced as a hidden variable in the model
More complex framework from a statistical point of view

→ Propose several regime-switching models to reproduce the various temporal dynamics and scales present in the wind data

→ Discuss the computation of relevant observed clustering

→ Compare observed and latent regime-switching models

MLE for observed MS-VAR models

The complete set $(\mathbf{y}_1, ...\mathbf{y}_T, s_1, ...s_T)$ is available, the complete log-likelihood :

$$\begin{aligned}\log(\mathcal{L}(\theta; \mathbf{y}_1, ...\mathbf{y}_T, s_1, ...s_T|\mathbf{y}_{-1}, \mathbf{y}_0)) &= \log(\mathcal{L}(\theta^{(\mathbf{Y})}; \mathbf{y}_1^T|\mathbf{y}_{-1}, \mathbf{y}_0, s_1^T)) \\ &+ \log(\mathcal{L}(\theta^{(S)}; s_1^T|\mathbf{y}_{-1}, \mathbf{y}_0)),\end{aligned}$$

- $\log(\mathcal{L}(\theta^{(S)}; s_1, ..., s_T|\mathbf{y}_{-1}, \mathbf{y}_0)) \rightarrow$ usual MLE of a Markov chain parameters

- $\log(\mathcal{L}(\theta^{(\mathbf{Y})}; \mathbf{y}_1, ..., \mathbf{y}_T|\mathbf{y}_{-1}, \mathbf{y}_0, s_1^T)) = \sum_{i=1}^{M} \sum_{t \in \{t|s_t=i\}} \log(p(\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, s_t)),$

for each $i \in \{1, ..., M\}$, maximization of each function :

$$\theta^{(\mathbf{Y},i)} \rightarrow n_i(-\frac{d}{2}\log(2\pi) - \frac{1}{2}\log(\det(\mathbf{\Sigma}^{(i)})) - \sum_{t \in \{t|s_t=i\}} \frac{1}{2}\mathbf{e}_t'(\mathbf{\Sigma}^{(i)})^{-1}\mathbf{e}_t,$$

where $\mathbf{e}_t = (\mathbf{y}_t - \mathbf{A}_0^{(i)} - \mathbf{A}_1^{(i)}\mathbf{y}_{t-1} - \mathbf{A}_2^{(i)}\mathbf{y}_{t-2}) \rightarrow$ usual MLE for VAR models

EM for hidden MS-VAR models

Only $(\mathbf{y}_1, \ldots \mathbf{y}_T)$ is available $\rightarrow$ maximization of

$$\theta \rightarrow \mathrm{E}_\theta(\log(\mathcal{L}(\theta; \mathbf{Y}_1, \ldots, \mathbf{Y}_T, S_1, \ldots, S_T))|\mathbf{Y}_{-1}^T = \mathbf{y}_{-1}^T).$$

via the Expectation-Maximization algorithm ([Cappé et al., 2005]):

E-step: Computation of the probabilities $P(S_t|\mathbf{Y}_1^T = \mathbf{y}_1^T)$ through Forward-Backward recursions to derive the incomplete likelihood,

M-step: Explicit forms of the parameters

In the following:

- a hidden MS-VAR model is fitted on the data with M=3 regimes and order of AR p=2

- 3 observed MS-VAR models are built and compared:
  - 1 model with regimes extracted from a large-scale variable
  - 2 models with regimes extracted from the wind data

One is selected and compared to the hidden MS-VAR

Plan

Derivation of regimes from extra-variables

· 4 regimes obtained over the North-Atlantic / European sector by a kmeans-clustering of 500 mb geopotential anomalies, provided by Julien Cattiaux, CNRM-GAME

· In winter, four weather regimes are identified and described in various references [Michelangeli et al., 1995, Cassou, 2008, Najac, 2008]: North-Atlantic phases: NAO+, NAO-, Blocage: BL and Atlantic Ridge: AR $\rightarrow$ correspond to characteristic patterns of atmospheric circulation

· Associated privileged flows: south-western flows (NAO+), western flows (NAO-), southern or eastern stable flows (BL) and northern flows (AR)

Clustering denoted C_{Z500}, *associated MS-VAR model* $AP\text{-}MS\text{-}VAR_{C_{Z500}}$

Derivation of regimes from the local variables

Clustering with 3 clusters via a Hidden Markov Model with Gaussian probability of emission:

- the time series associated to the first Empirical Orthogonal Functions (EOF) of the anomalies (mean-corrected fields) of $\{\mathbf{u}_t, \mathbf{v}_t\}$,
denoted $C_{EOF-(u,v)}$, *associated MS-VAR model* $AP\text{-}MS\text{-}VAR_{C_{EOF-(u,v)}}$

- the bivariate process $\{\mathbf{u}_t - \mathbf{u}_{t-1}, \mathbf{v}_t - \mathbf{v}_{t-1}\}$,
denoted $C_{Diff(u,v)}$, *associated MS-VAR model* $AP\text{-}MS\text{-}VAR_{C_{Diff(u,v)}}$

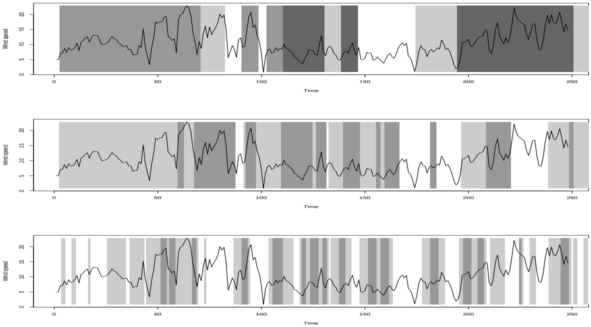


Figure : Time series of wind speed and *a priori* regimes extracted from the proposed methods above. The darker is the grey, the smaller is the determinant of $\Sigma^{(i)}$. From top to bottom: C_{Z500}, $\mathrm{C}_{EOF-(u,v)}$ and $\mathrm{C}_{Diff(u,v)}$.

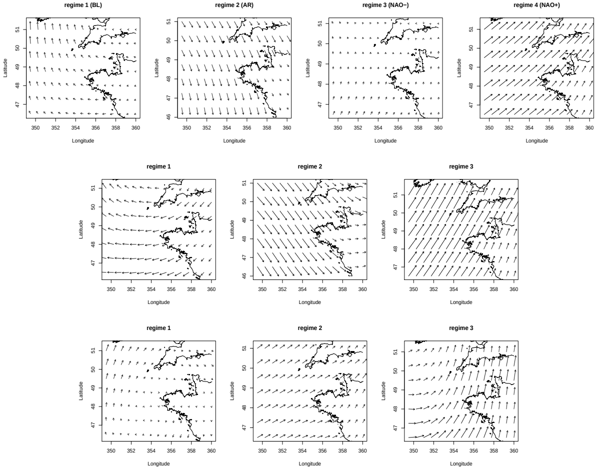


Figure : Average fields of $\{\mathbf{u}_t, \mathbf{v}_t\}$ in each regime of the proposed clusterings, from top to bottom: C_{Z500}, $\mathrm{C}_{EOF-(u,v)}$, $\mathrm{C}_{Diff(u,v)}$

Link between large-scale weather regimes and the other regimes

Explore the joint occurrences of large-scale weather regimes and the local regimes provided by the proposed clustering and by the model H-MS-VAR.

	$C_{EOF-(u,v)}$					$C_{Diff(u,v)}$					H-MS-VAR				
	BL	AR	NA0 -	NAO+	Total	BL	AR	NA0 -	NAO+	Total	BL	AR	NA0 -	NAO+	Total
R1	0.17	0.06	0.08	0.01	0.32	0.15	0.10	0.07	0.13	0.45	0.13	0.09	0.07	0.14	0.43
R2	0.04	0.10	0.05	0.08	0.27	0.09	0.06	0.09	0.16	0.40	0.10	0.06	0.09	0.15	0.41
R3	0.07	0.02	0.07	0.26	0.42	0.03	0.02	0.04	0.06	0.15	0.04	0.02	0.05	0.06	0.16
Total	0.28	0.18	0.20	0.35	1	0.27	0.18	0.20	0.35	1	0.27	0.17	0.21	0.35	1

Table : Joint probability of occurrence of the three regimes identified by the proposed models in lines and the large-scale regimes in columns

$\rightarrow$ small-scale regimes seem to appear in privileged large-scale weather regimes.

BIC indexes and log-likelihood

Select the clustering that is the most physically meaningful and appropriate in terms of conditional autoregressive models

$$\mathrm{BIC} = -2\log \mathrm{L} + \mathrm{N}_p \log(\mathrm{N}_{obs}) \text{ with L the likelihood}$$

Model	BIC	log-$\mathcal{L}$ of S	log-$\mathcal{L}$ of $\mathbf{Y}$	N_p
Unconditional VAR	542640	-	-269825	265
$\mathrm{AP\text{-}MS\text{-}VAR}_{\mathrm{C}_{Z500}}$	542730	-1510	-263808	1072
$\mathrm{AP\text{-}MS\text{-}VAR}_{\mathrm{C}_{EOF-(u,v)}}$	545730	-2331	-266015	801
$\mathrm{AP\text{-}MS\text{-}VAR}_{\mathrm{C}_{Diff(u,v)}}$	520759	-4762	-251099	801
H-MS-VAR	459458	-	-229616	801

Table : N_p the number of parameters. Values are computed from models fitted on $\{\mathbf{u}_t, \mathbf{v}_t\}$ at the 5 locations (1,6,10,13,18).

In the following, H-MS-VAR and $\mathrm{AP\text{-}MS\text{-}VAR}_{\mathrm{C}_{Diff(u,v)}}$ are compared

Plan

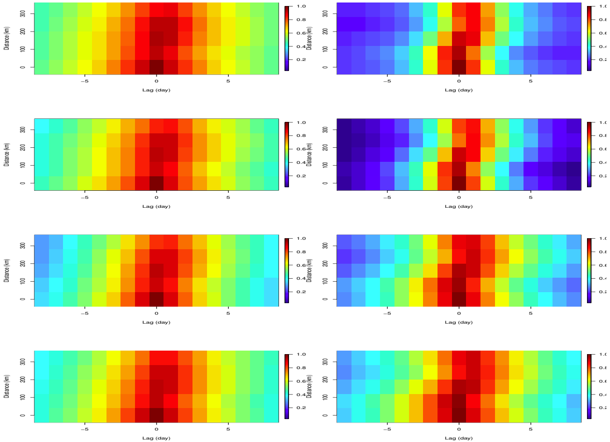


Figure : Correlation of between $\{u_t\}$ at site 1 and $\{u_t\}$ (left and similar quantities for $\{v_t\}$ on the right) at the other locations at various time-lag. From top to bottom: data, simulation from VAR(2), $\mathrm{AP\text{-}MS\text{-}VAR}_{C_{Diff(u,v)}}$ and $\mathrm{H\text{-}MS\text{-}VAR}$.

$\rightarrow$ average space-time motions are in part reproduced by all the models

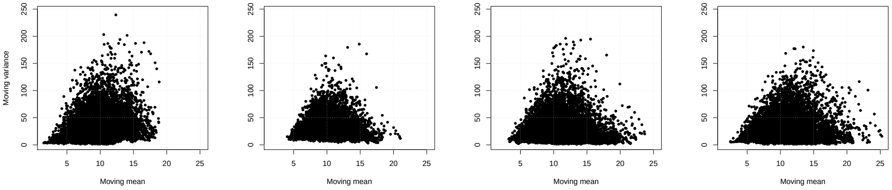


Figure : Moving variance against of the value $\{U_t\}$ against its moving mean at location 10. From left to right and top to bottom: data, simulation from the VAR(2), $\mathrm{AP\text{-}MS\text{-}VAR}_{C_{Diff(u,v)}}$ and $\mathrm{H\text{-}MS\text{-}VAR}$

$\rightarrow$ better description by the MS-VAR models and especially by the hidden MS-VAR model

Plan

Discussion and perspectives

- both types of models have related advantages

- compromise between meteorological consistency of the clustering and a good description of the conditional distribution by a VAR framework

- account for spatial information into the regime ?
$\rightarrow$ develop a test procedure to decide the relevance of a regional or site-specific regime

- develop parameterization of autoregressive parameters

Cappé, O., Moulines, E., and T., R. (2005).
Inference in hidden Markov models.
Springer-Verlag, New York.

Cassou, C. (2008).
Intraseasonal interaction between the madden–julian oscillation and the north atlantic oscillation.
Nature, 455(7212):523–527.

Michelangeli, P. A., Vautard, R., and Legras, B. (1995).
Weather regimes: recurrence and quasi stationarity.
Journal of the Atmospheric Sciences, 52(8):1237–1256.

Najac, J. (2008).
Impacts du changement climatique sur le potentiel éolien en France: une étude de régionalisation.
PhD thesis, Université Paul Sabatier-Toulouse III.