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

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

SWGEN - 2012

# 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
- ...

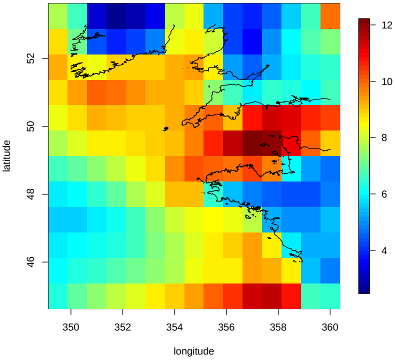
Wind data generally available on short periods of time

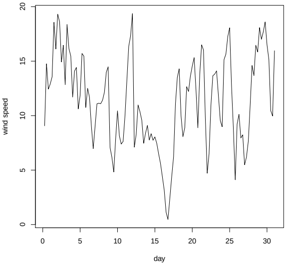
- 50 years of data maximum
- Not enough to compute reliable estimates of the probability of complex events

Stochastic model used to simulate artificial wind conditions

- Monte-Carlo methods

# Plan



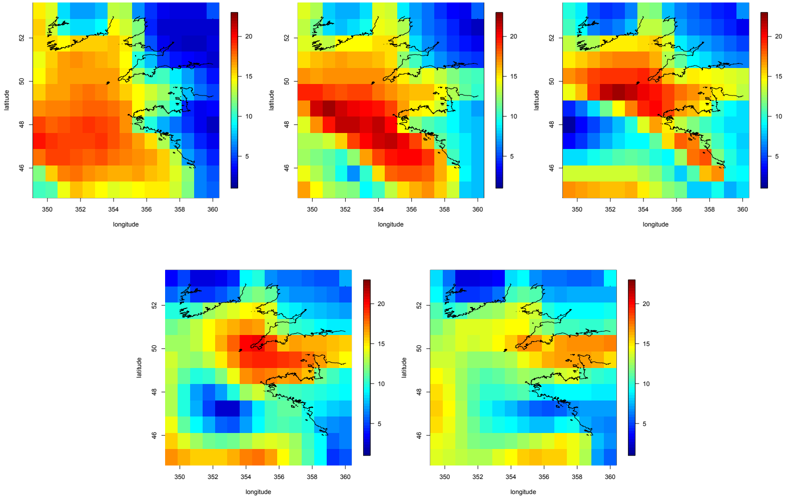

**Available data:**

- Reanalysis data from ECMWF: wind speed intensity

**Goals:**

- to propose a stochastic state-space model for wind fields
- to generate realistic wind conditions

**One of the main difficulties and goals is:** reproducing space-time motions of meteorological systems *e.g.* propagation of low pressure systems.



# Plan

# Plan

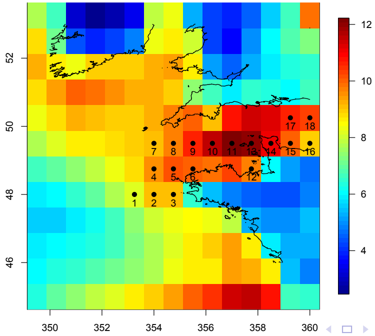
Data under study: reanalysis data ERA Interim from ECMWF

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

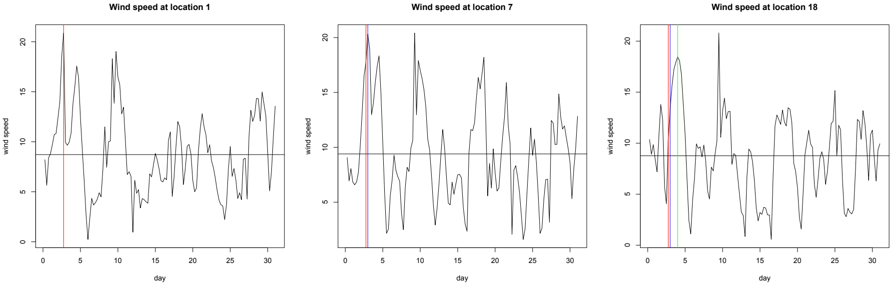
To eliminate seasonality: study of months of January from 1979 to 2011

Wind speed are squared root transformed to reduce skewness

18 gridded points under study in the North-East Atlantic channel

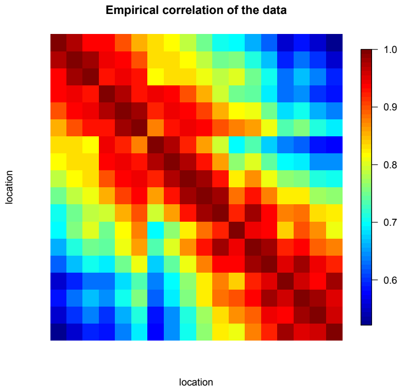


# Some time series




Time shifts are observable on time series

# Some statistics computed on data



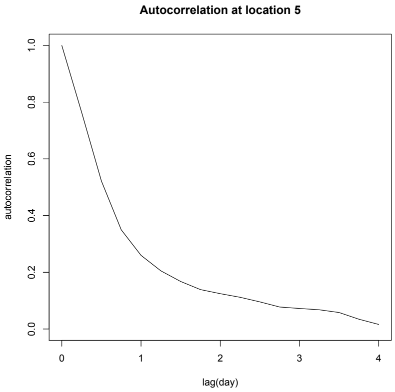




Figure: Spatial correlation at temporal lag 0 and autocorrelation function at location 5

# Plan

Modelling by a Gaussian state-space model [DK01], [CMR]

$$\left\{\begin{array}{lcl} X_t & = & bX_{t-1} + \epsilon_t \\ Y_t(1) & = & Z(1)X_t + \eta_t(1) \\ \vdots & & \vdots \\ Y_t(K) & = & Z(K)X_t + \eta_t(K) \end{array}\right. \quad \text{for } t \geq 0,$$

$Y_t \in \mathbb{R}^K$: vector of observations, $K$: number of locations studied
$X_t \in \mathbb{R}$: scalar hidden state,
$\epsilon$ and $\eta$ are serially independent Gaussian white noises
$\epsilon_t \sim \mathcal{N}(0, Q)$ and $\eta_t \sim \mathcal{N}(0, R)$

**Parameters and equations interpretation:**

X : · mean squared root wind conditions at regional scale,
· temporal dynamic is contained in the state equation,

Y : mean corrected squared root wind speed at local scale,

Z : links regional and local scales,

R : contains a part of the spatial covariance between studied locations.

$$\begin{cases} X_t & = & bX_{t-1} + \epsilon_t \\ Y_t(1) & = & Z(1)X_t + \eta_t(1) \\ \vdots & & \vdots \\ Y_t(K) & = & Z(K)X_t + \eta_t(K) \end{cases} \quad \text{for } t \geq 0,$$

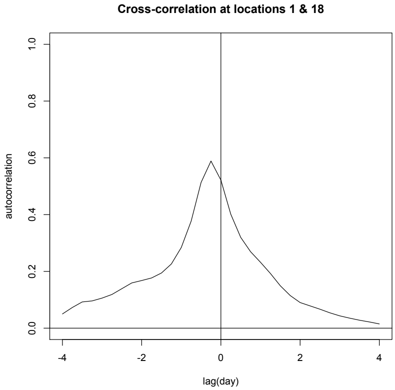
$\epsilon_t \sim \mathcal{N}(0, Q)$ and $\eta_t \sim \mathcal{N}(0, R)$

**Stationarity:**
Under the assumption $|b| < 1$, the autoregressive process $X$ is stationary and so is the process $Y$ [BD].

**Identifiability:**
$Q$ is fixed in order to ensure identifiability of the model (study of the second order structure of the Gaussian process $Y$ [TA10]).

# Cross-correlation

The estimate of cross-correlation shows that some locations have temporal advance on others:




To capture this phenomenon, we add more flexibility to the model

# More flexible dependance from regional mean wind speed

To get a more flexible temporal dependance of observations from the regional mean wind speed, we consider the followed observation equation:

$$Y_t = Z_0 X_t + Z_1 X_{t-1} + \eta_t,$$

Taking $\tilde{X}_t = \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix}$, the system can be rewritten as:

$$\left\{ \begin{array}{rcl} \tilde{X}_t & = & \tilde{B}\tilde{X}_{t-1} + \tilde{\epsilon}_t \\ Y_t & = & \tilde{Z}\tilde{X}_t + \eta_t \end{array} \right. \text{for } t \geq 0,$$

with $\tilde{B} = \begin{pmatrix} b & 0 \\ 1 & 0 \end{pmatrix}$, $\tilde{Z} = \begin{pmatrix} Z_0 & Z_1 \end{pmatrix}$ and $\tilde{\epsilon}_t \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} \right)$.

This model is still a linear Gaussian state-space model.

**Identifiability:** This model brings new problems of identifiability....

**Covariance structure:** [Cre], [HR89], [Abr]

(Observation equation: $Y_t = Z_0 X_t + Z_1 X_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, R)$)

For parsimony purpose we choose to fit the covariance observation error matrix to a parametric family of spatial covariance:

Assume that $R$ depends on the distance between locations:

$$R_{i,j} = \left(\sigma_i \sigma_j (\exp(-\lambda_1 d_{i,j}^2) + \lambda_2 \delta_{i,j})\right) \quad \text{for} \quad i, j \in \{1, ..., K\},$$

with $(\sigma_1, ..., \sigma_K, \lambda_1, \lambda_2)$ to be estimated and strictly positive.

**Advantage:** it reduces the number of parameters to estimate

# Maximum Likelihood Estimation

Generalized Expectation-Maximization algorithm is used [CMR]:

**E-step:** Linearity and Gaussianity of the model enable to use Kalman filter and smoother to derive the incomplete likelihood,

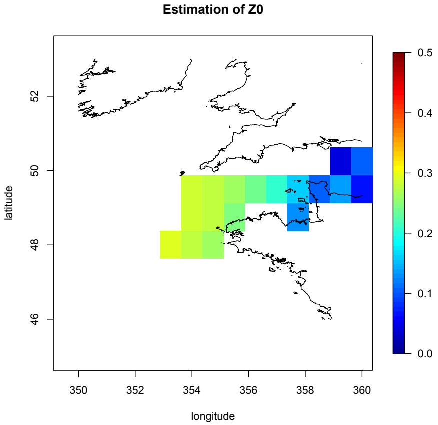
**M-step:**
- Explicit forms of the parameters: $\tilde{B}$ and $\tilde{Z}$,
- $R$ is estimated by numerical optimization of a part of the incomplete likelihood.

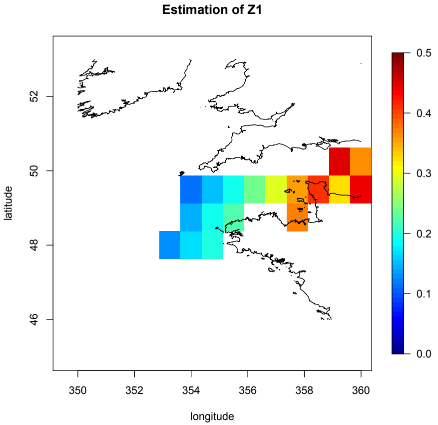
**Starting point:** Least square estimation on second order structure.

# Plan

**Estimate of $Z$:**

Observation equation: $Y_t = Z_0 X_t + Z_1 X_{t-1} + \eta_t$,

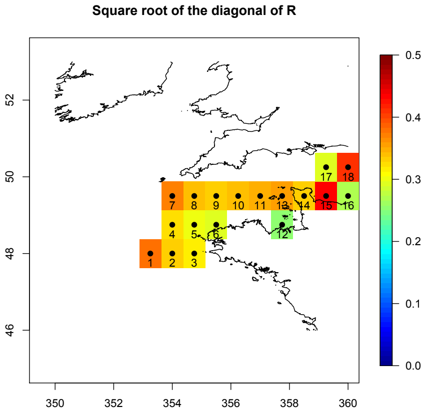



Western locations depend more on $X_t$ than on $X_{t-1}$ and conversely for the eastern locations.

## Estimate of $R$:

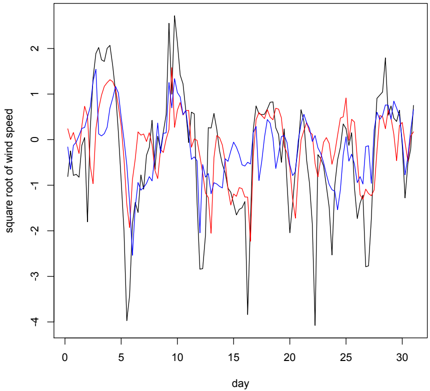
Observation equation: $Y_t = Z_0 X_t + Z_1 X_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, R)$




Location 12 has the weakest error measurement, the most westerly and easterly locations have the largest ones.

Comparison of the smoothed hidden state $\hat{X}$ to the squared root of wind speed $Y$ at locations 1 and 18, with

$$\hat{X}_t = E(X_t | Y_1, ..., Y_T)$$




–Smoothed hidden state

–Squared root wind speed at location 1

–Squared root wind speed at location 18

$X$ is a good compromise between square root of wind speed at each location.

# Marginal Distributions

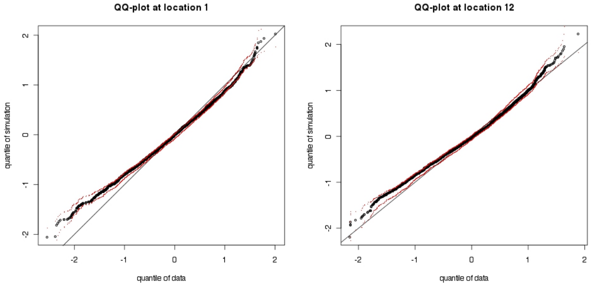
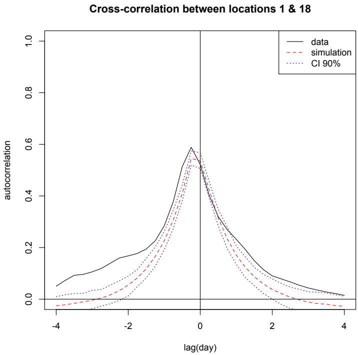
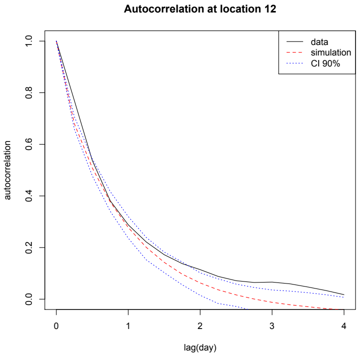



Figure: Locations 1 and 12

Weak values are overestimated and high values are not correctly reproduced. Location 12 distribution matches more with the data distribution.

# Temporal dynamic



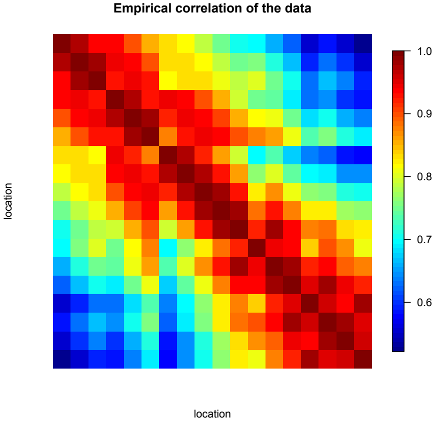



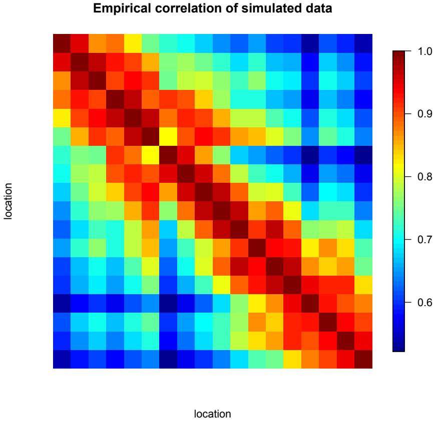

◦ Time shifts are not entirely reproduced.

◦ For about one day time-lag, temporal dynamic is reproduced in some locations and underestimated for greater time-lags.

 Validation of the model

# Spatial structure




Spatial covariance of the observed process $Y$ is underestimated but the shape is in part reproduced.

# Spatial structure

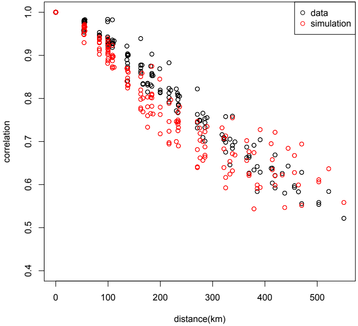



Figure: Empirical correlation of $Y$ at lag 0 against distance

Correlation is underestimated for distances smaller than 300km and variability is larger for distances greater than 400 km.

# Comparison of different spatial structures

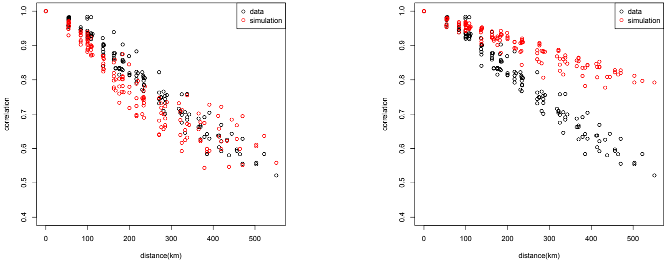



Figure: Corrrelation against distance for Gaussian variogram model and exponential variogram model

Spatial structure and others properties are not reproduced with the exponential model which overestimates some parameters and underestimates others.

**Comparison of log-likelihood:**

Exponential variogram model 36779

Gaussian variogram model 43286

# Plan

# Conclusion

**Advantages of the model:**

- Ease of interpretation of the parameters,
- Linearity and Gaussian properties: ease of implementation of estimation procedure
- Some statistics are well reproduced by the model,

**Drawbacks:**

- Lack of flexibility of the model: it only catches average behaviors of wind speed in the area, does not account for the weather type,
- Linearity of the model may not be appropriate to catch possible local effects,

- Improving the temporal dynamic of this model,

- Adding Markovian regime switching to account for the weather type and to give more flexibility to the model,

- Modelling wind speed and direction at the same time will be more relevant: the dependance is different according the direction in which the wind is blowing, and we know that according the weather type the wind has prevailing direction.

P. Abrahamsen.
A review of gaussian random fields and correlation functions.

Peter J. Brockwell and Richard A. Davis.
*Introduction to time series and forecasting*.
Springer Texts in Statistics. Second edition.

Olivier Cappé, Eric Moulines, and Tobias Rydén.
*Inference in hidden Markov models*.
Springer Series in Statistics.

Noel A. C. Cressie.
*Statistics for spatial data*.
Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics.

J. Durbin and S. J. Koopman.
*Time series analysis by state space methods*, volume 24 of *Oxford Statistical Science Series*.
Oxford University Press, Oxford, 2001.

J. Haslett and A.E. Raftery.
Space-time modelling with long-memory dependence: Assessing ireland's wind power ressource.
*Applied Statistics*, 1989.

Ailliot P. Tandeo, P. and E. Autret.
Linear gaussian state-space model with irregular sampling: application to sea surface temperature.
*Stoch Environ Res Risk Assess*, 2010.

diagonale $R$ simulée:
0.57 0.56 0.56 0.55 0.56 0.56 0.55 0.55 0.57 0.59 0.62 0.57 0.63 0.63 0.65 0.59 0.59 0.63
diagonale $R$ observée:
0.66 0.66 0.64 0.65 0.65 0.63 0.64 0.64 0.64 0.65 0.65 0.60 0.64 0.64 0.65 0.63 0.61 0.66