Multivariate spatio-temporal stochastic weather generators based on HMM models

A. Cuzol, LMBA - Université de Bretagne Sud
DESIRES project / D. Allard, V. Monbet, M. Thakur

Workshop "Modèles spatio-temporels en météorologie et océanographie"
Rennes, 28 novembre 2018

Stochastic weather generators

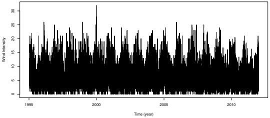
- Stochastic tools for generating sequences of meteorological variables.
- Used in impact studies when climate is involved.

Synthetic weather conditions → System → Distribution of the response

- Most usual applications : hydrology, agriculture.

Historical data

10 years of meteorological data (rainfall, temperature, solar radiation, humidity, wind speed,...) on several locations



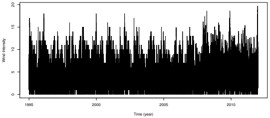
Stochastic model

SWGEN calibrated on historical data

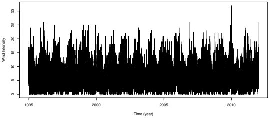
⟹

Synthetic data

Large number of synthetic sequences with statistical properties similar to original data



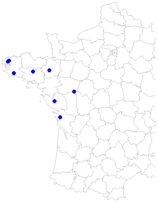
⋮



Framework / objective

West France data :

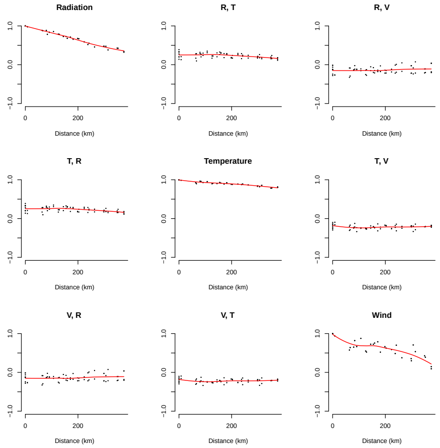
- 10 years of daily data
- 3 variables : Radiation, Temperature, Wind intensity
- Summer months : June-July-August
- Focus on West area in this talk : 8 stations



⇒ Objective : construct a multivariate spatio-temporal generator for existing or new stations.

What do we want to reproduce ?

⇒ Spatial correlations :



- Black points : observed correlations.
- Red line : LOESS approximation.
- Long range correlation for the temperature.
- Shorter range for radiation and wind.

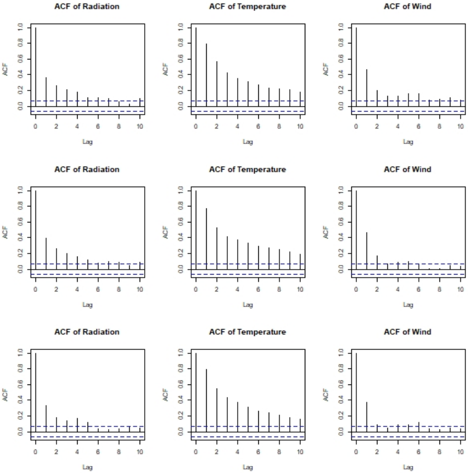
What do we want to reproduce ?

⇒ Temporal correlations :

Plomelin :

Saint-Jacques-de-la-Lande :

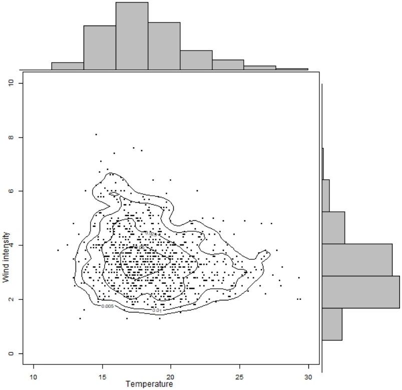
La Roche-sur-Yon :



Variables have different temporal behaviors.

What do we want to reproduce ?

⇒ Relations between meteorological variables :



Temperature and wind for one location (La Roche-sur-Yon)

Existing methods

- Non parametric methods (based on data : bootstrap, analogs)
  - statistical properties are reproduced by construction ;
  - unable to create unobserved meteorological situations ;
  - multisite, unable to simulate new locations.
- Parametric models
  - $\Rightarrow$ Allard, Bourotte (2016), Ailliot, Monbet (2017), Monbet's talk (METMA 2018)
  - multivariate, multisite or spatial ($\Rightarrow$ new locations) ;
  - can simulate non observed situations ;
  - weather types can be introduced.

$\Rightarrow$ Within the family of parametric models, we propose a multivariate spatio-temporal generator, including weather types through an HMM model.

First HMM model

$$\begin{array}{lccccccccc} \textit{Weather type} & \cdots & \rightarrow & X_{t-1} & \rightarrow & X_t & \rightarrow & X_{t+1} & \rightarrow & \cdots \\ & & & \downarrow & & \downarrow & & \downarrow & & \\ \textit{Meteo variables} & \cdots & & \mathbf{Y}_{t-1} & & \mathbf{Y}_t & & \mathbf{Y}_{t+1} & & \cdots \end{array}$$

where :

- X is the hidden weather type modeled as a first order Markov chain, with homogeneous transitions.
- $\mathbf{Y}$ contains the meteorological variables.
- $\forall t \quad \mathbf{Y}_t$ is Gaussian conditionally to the weather type X_t :

$$\mathbf{Y}_t | X_t = x \sim \mathcal{N}(\mathbf{m}_x, \mathbf{\Sigma}_x)$$

Extension toward a joint model

$$
\begin{array}{lccccccccc}
\textit{Weather type} & \cdots & \rightarrow & \begin{pmatrix} X_{t-1} \\ X_t \end{pmatrix} & \rightarrow & \begin{pmatrix} X_t \\ X_{t+1} \end{pmatrix} & \rightarrow & \begin{pmatrix} X_{t+1} \\ X_{t+2} \end{pmatrix} & \rightarrow & \cdots \\
 & & & \downarrow & & \downarrow & & \downarrow & & \\
\textit{Meteo variables} & \cdots & & \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{Y}_t \end{pmatrix} & & \begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t+1} \end{pmatrix} & & \begin{pmatrix} \mathbf{Y}_{t+1} \\ \mathbf{Y}_{t+2} \end{pmatrix} & & \cdots
\end{array}
$$

where $\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t+1} \end{pmatrix}$ is Gaussian conditionally to $\begin{pmatrix} X_t \\ X_{t+1} \end{pmatrix}$:

$$
\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t+1} \end{pmatrix} \Bigg| \begin{pmatrix} \mathbf{X}_t = x \\ \mathbf{X}_{t+1} = x' \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{m}_x \\ \mathbf{m}_{x'} \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_x & \mathbf{\Sigma}_{x,x'} \\ \mathbf{\Sigma}'_{x,x'} & \mathbf{\Sigma}_{x'} \end{pmatrix} \right)
$$

Resulting dependence graph

This joint model implies :

$$\mathbf{Y}_{t+1}|\mathbf{Y}_t = y, X_t = x, X_{t+1} = x' \sim \mathcal{N}\left(\mathbf{m}_{x'} + \mathbf{\Sigma}_{x,x'}\mathbf{\Sigma}_x^{-1}(y - \mathbf{m}_x), \mathbf{\Sigma}_{x'} - \mathbf{\Sigma}_{x,x'}\mathbf{\Sigma}_x^{-1}\mathbf{\Sigma}'_{x,x'}\right)$$

⇒ The following temporal relationships between weather types and meteorological states are described :

$$\begin{array}{lccccccccc}
\textit{Weather type} & \cdots & \rightarrow & X_{t-1} & \rightarrow & X_t & \rightarrow & X_{t+1} & \rightarrow & \cdots \\
 & & \searrow & \downarrow & \searrow & \downarrow & \searrow & \downarrow & \searrow & \\
\textit{Meteo variables} & \cdots & \rightarrow & \mathbf{Y}_{t-1} & \rightarrow & \mathbf{Y}_t & \rightarrow & \mathbf{Y}_{t+1} & \rightarrow & \cdots
\end{array}$$

Covariance model

- We use the multivariate Gneiting-Matérn space-time model (Bourotte et al. 2016)

$$Cov\left(Y_i(s,t), Y_j(s+\mathbf{h}, t+u)\right) = \frac{\sigma_i\sigma_j}{\psi(u^2)}\rho_{ij}\mathcal{M}\left(\frac{\mathbf{h}}{\psi(u^2)^{1/2}}; r_{ij}, \nu_{ij}\right)$$

  where we need the case $u = 0$ and $u = 1$.
- It is a valid matrix-valued covariance function if, for all $i, j = 1, \cdots, p$,

$$\begin{aligned} r_{ij} &= ((r_i^2 + r_j^2)/2)^{1/2} \\ \nu_{ij} &= (\nu_i + \nu_j)/2 \\ \rho_{ij} &= \beta_{ij}\frac{\Gamma(\nu_{ij})}{\Gamma(\nu_i)^{1/2}\Gamma(\nu_j)^{1/2}}\frac{r_i^{\nu_i} r_j^{\nu_j}}{r_{ij}^{2\nu_{ij}}} \end{aligned}$$

  where σ_i, ν_i, r_i are positive for all $i = 1, \cdots, p$, $\beta = (\beta_{ij})_{i,j=1}^{p}$ is a correlation matrix $\psi(t)$, $t \geq 0$, a positive function with a completely monotone derivative.
- This model is non separable with parameters constant in time and space.
- We allow ψ to vary with the variable.
- All parameters vary with the weather regimes.

Estimation

Unknown parameters :

- For p variables and S sites we have to estimate for each regime : $\mathbf{m} \in \mathbb{R}^{p*S}$ and $(\sigma_i, r_i, \nu_i, \beta_i, \psi_i) \quad \forall i = 1, \ldots, p$.
- Initial states probabilities and matrix of transitions :
  If $N = 2$ (weather states "1" and "2") :

	11	12	21	22
11	p_{111}	p_{112}	0	0
12	0	0	p_{121}	p_{122}
21	p_{211}	p_{212}	0	0
22	0	0	p_{221}	p_{222}

Estimation steps :

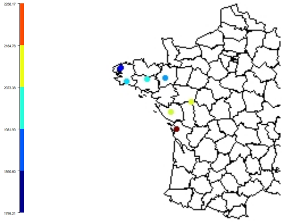
- Initialization : k-means (N^2 classes) + gaussian mixture model.
- Initial values of $\mathbf{m}$ and $(\sigma_i, r_i, \nu_i, \beta_i, \psi_i) \quad \forall i = 1, \ldots, p$ are then obtained from the resulting means and covariances in each regime.
- EM algorithm.

Prediction problem

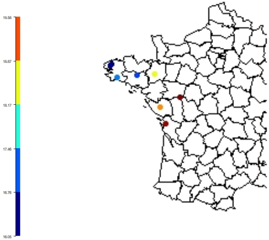
In order to predict new sites at unobserved locations, a spatial trend is adjusted for the mean of each variable in each regime :

$$\hat{m}_i(Lon, Lat) = \hat{\alpha}_0^i + \hat{\alpha}_1^i Lon + \hat{\alpha}_2^i Lat \quad \forall i = 1, \dots, p$$

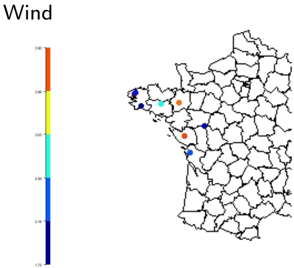
Radiation



Temperature

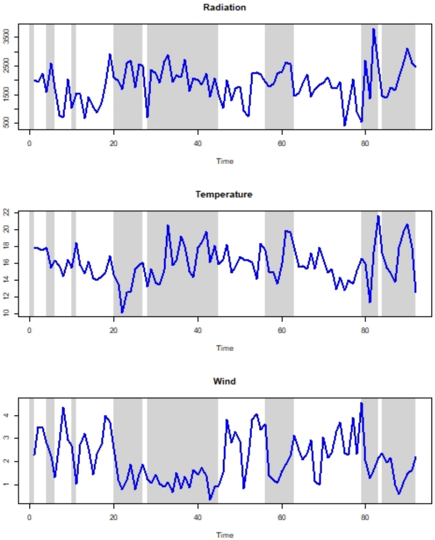


... but this trend hypothesis is not satisfying for the wind :



Weather states and transitions

For instance, for one summer in Plomelin :

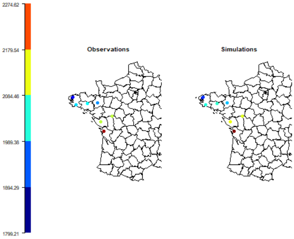


- State 1 (grey) :
  $\bar{R} = 2234$, $\bar{T} = 19$, $\bar{W} = 1.8$
  $\Rightarrow$ more sun, "hotter", less wind
- State 2 (white) :
  $\bar{R} = 1880$, $\bar{T} = 17$, $\bar{W} = 2.8$
  $\Rightarrow$ less sun, "colder", more wind
- Covariance parameters :
  - ☐ σ_W higher in state 2
  - ☐ Spatial and temporal ranges higher in state 1
- Transition probabilities :

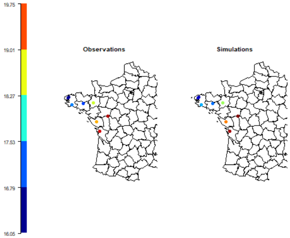
	11	12	21	22
11	0.83	0.17	0	0
12	0	0	0.18	0.82
21	0.77	0.23	0	0
22	0	0	0.22	0.78

Spatial means

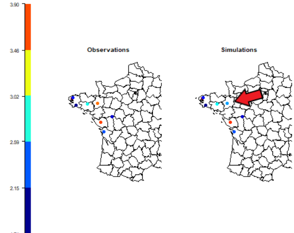
Radiation



Temperature



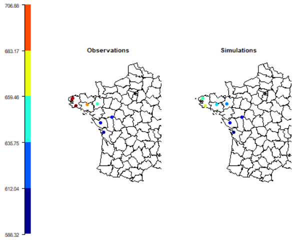
Wind intensity



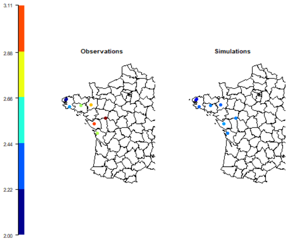
- NB time series of 10 years of summer months have been simulated ($NB = 20$).
- Simulated means of all variables are good for observed stations.
- Saint-Jacques-de-La-Lande () is not observed : hard to predict a local unknown behavior for the wind.

Spatial standard deviations

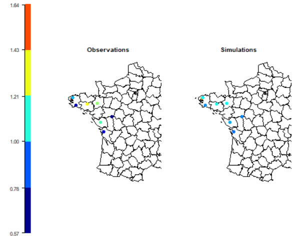
Radiation



Temperature

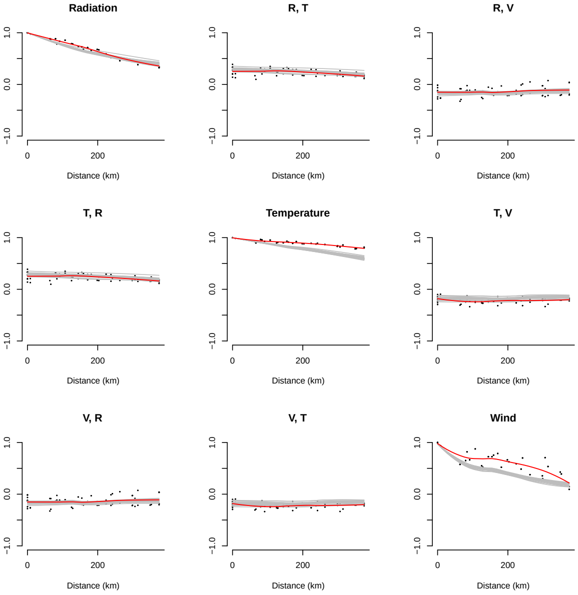


Wind intensity



Standard deviations are smoothed, local behaviors are not taken into account by the model (σ_i for each variable i does not depend on the location).

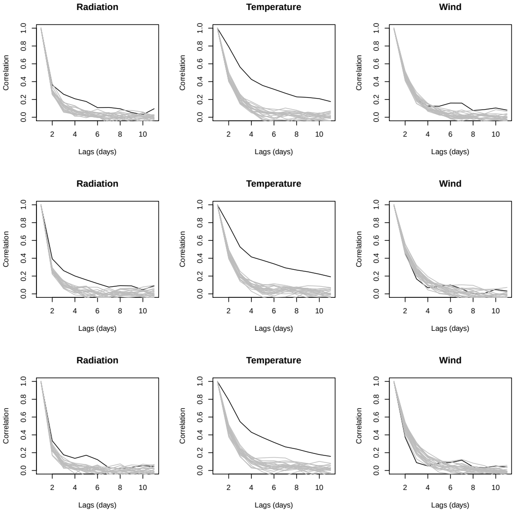
Spatial cross-correlations



- Black points : observed correlations.
- Red line : LOESS approximation.
- Grey lines : LOESS approximation for $NB = 20$ simulations of 10 years.
- Temperature : the correlation is under-estimated for long distances.
- Wind : under-estimation of correlation except for very short distances.

Temporal correlations

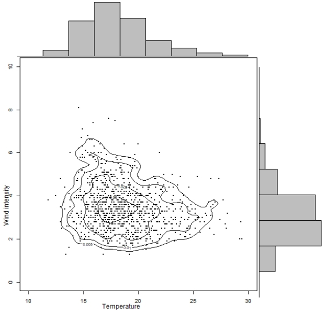
For Plomelin, Saint-Jacques-de-la-Lande and La Roche-sur-Yon :

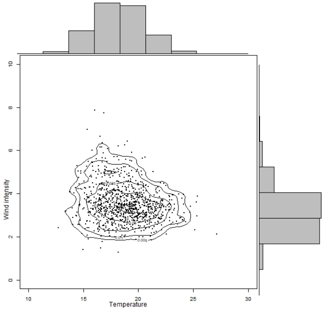


- Black lines : observed correlations.
- Grey lines : correlations for $NB = 20$ simulations of 10 years.
- Temporal correlations of temperatures are underestimated.
- Ψ is allowed to vary with i in the model but is not able to describe the specific behavior of the temperature.

Relations between meteorological variables

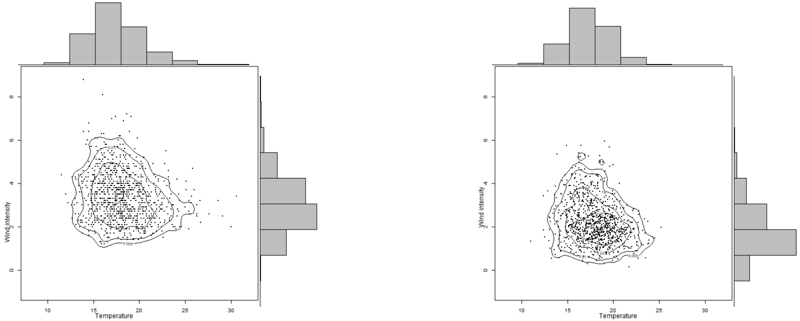
Temperature and Wind in one station (La Roche-sur-Yon) : observations / 1 simulation (10 years)





Relations between meteorological variables

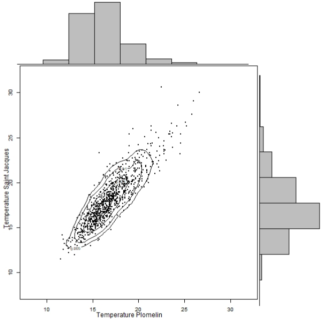
Temperature and Wind in one station (Saint-Jacques, predicted) : observations / 1 simulation (10 years)

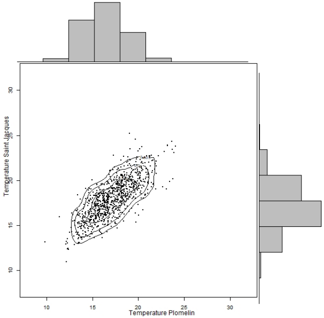


⇒ The wind is under-estimated but the relation between the two variables is well reproduced.

Relations between sites

Temperature in two stations (Plomelin and Saint-Jacques) : observations / 1 simulation (10 years)





⇒ Distributions are well reproduced.

What is the benefit of coupling ?

HMM model without coupling :

$$\begin{array}{lccccccccc}
\textit{Weather type} & \cdots & \rightarrow & X_{t-1} & \rightarrow & X_t & \rightarrow & X_{t+1} & \rightarrow & \cdots \\
 & & & \downarrow & & \downarrow & & \downarrow & & \\
\textit{Meteo variables} & \cdots & & \mathbf{Y}_{t-1} & & \mathbf{Y}_t & & \mathbf{Y}_{t+1} & & \cdots
\end{array}$$

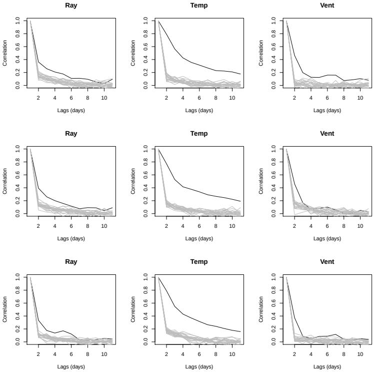
HMM model with coupling :

$$\begin{array}{lccccccccc}
\textit{Weather type} & \cdots & \rightarrow & X_{t-1} & \rightarrow & X_t & \rightarrow & X_{t+1} & \rightarrow & \cdots \\
 & & \searrow & \downarrow & \searrow & \downarrow & \searrow & \downarrow & \searrow & \\
\textit{Meteo variables} & \cdots & \rightarrow & \mathbf{Y}_{t-1} & \rightarrow & \mathbf{Y}_t & \rightarrow & \mathbf{Y}_{t+1} & \rightarrow & \cdots
\end{array}$$

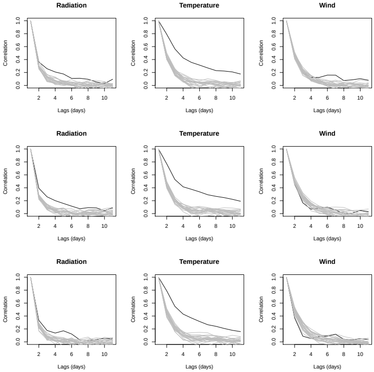
What is the benefit of coupling ?

Temporal correlations for Plomelin, Saint-Jacques-de-la-Lande and La Roche-sur-Yon :

Without coupling

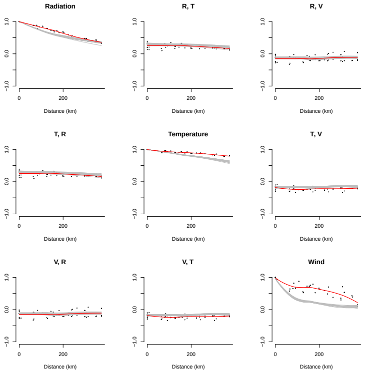


With coupling

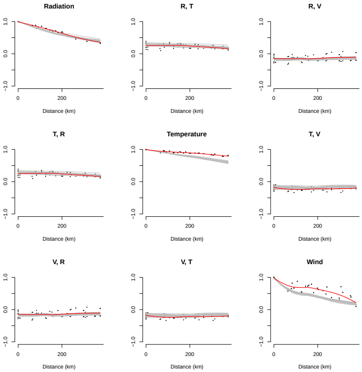


What is the benefit of coupling ?

Spatial correlations **without** coupling



Spatial correlations **with** coupling



Conclusion and perspectives

- Hard to predict local phenomena for unobserved locations : need of a better prediction of local means and standard deviations.
- Bad estimation of Ψ_i when variables have different temporal behaviors.
- Has to be tested with a higher number of locations $\Rightarrow$ use of composite likelihood.
- The specific problem of precipitation has to be studied.
- This approach has to be compared with parametric MSAR.