Stochastic weather generators with non-homogeneous hidden Markov switching Application to temperature series

Valérie Monbet [1], Pierre Ailliot[2]

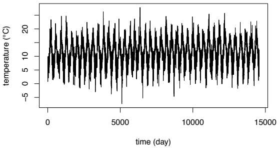
[1]IRMAR, Rennes 1
[2]LMBA, UBO

Stochastic weather generators?

- Stochastic tools for generating sequences of meteorological variables

Historical data

40 years of daily mean temperature at Brest



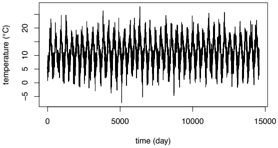
Stochastic model

SWGEN calibrated on historical data

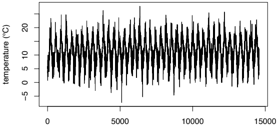
→

Synthetic data

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



⋮

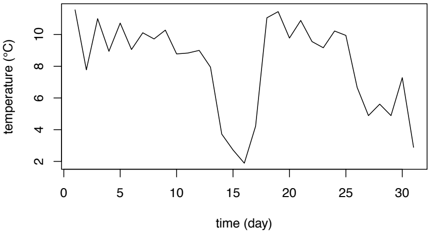


Stochastic weather generators?

- Stochastic tools for generating sequences of meteorological variables
- Used in impact studies when climate is involved and not enough data to estimate the quantities of interest

Synthetic weather conditions	→	System	→	Distribution of the response

- Most usual applications: hydrology and agriculture
  - Meteorological variables: rainfall, temperature, solar radiation, humidity, wind speed,...
- In this talk, firstly focus on daily mean temperature at a single location
  - Brest, 40 years, $\Delta t = 1$ day
  - Preprocessing step to remove seasonal components (+ trends, daily components)?
    In this talk, data blocked by month, results for January



Some existing stochastic weather generators

- Data oriented (non parametric) : random sampling with replacement from the original data also called Bootstraping
  - \+ Marginal distribution of the data reproduced by construction
  - \- Cannot create unobserved meteorological situations
  - \- It may be difficult to restore the dependance structure (especially for long events)

  Ref : Rajagopolan and Lall (1999)

- Model oriented (parametric)
  - \- Difficult to build realistic multi-site and multi-parameter generators
  - \+ Can create unrecorded situations
  - \+ Interpretable

Transformed Gaussian processes

- Classical approach for modeling time series
- Simulation of the stationary time series $\{Y_t\}$ in three steps:
  - First step: find a deterministic monotonic function g such that
    $$X_t = g(Y_t)$$
    has (approximatively) Gaussian margins
  - Second step: $\{X_t\}$ is a stationary process with (approximately) Gaussian margins. Further assume that $\{X_t\}$ is a Gaussian process and simulate this process.
    - Exact simulation
    - ARMA models
  - Third step: apply the inverse transform g^{-1} to the simulated Gaussian sequence

- **Temperature data**
  - Transformation to Gaussian margins : Normal scores transformation $Y_t^{(transf)} = \Phi^{-1}(\hat{F}(Y_t))$
  - Time series model AR(2) : $Y_t^{transf} = 0.57 + 0.83 Y_{t-1}^{transf} - 0.07 Y_{t-2}^{transf} + 0.41\epsilon_t$
  - Inverse transformation $\hat{Y}_t = \hat{F}^{-1}(\Phi(Y_t^{(transf)}))$

Transformed Gaussian processes

+ Easy to implement
+ Can be used for multivariate data (several sites and/or variables)
- May be hard to interpret
- Can not reproduce some non-linearities

- Alternatives?
  - More general transformation $X_t = g(Y_{h(t)})$ with g and h (stochastic?) transformations such that $\{X_t\}$ is a Gaussian process
  - Non-linear time series models

- Weather type models
  - Weather type = typical weather situation
    - e.g. dry/wet days, convective/frontal rainfall, cyclonic/anticyclonic conditions,...
    - Generally between 2 and 10 weather types are used; enough to summarize climate
  - Weather type models characterized by
    - A model for the weather type sequence
    - A model for the meteorological variables conditionally to the weather type
  - widely used for meteorological variables
- Various strategies available for defining the weather type
  - Weather type = typical weather situation
    - e.g. dry/wet days, convective/frontal rainfall, cyclonic/anticyclonic conditions,...
    - Generally between 2 and 10 weather types are used; enough to summarize the climate
  - Weather type models characterized by
    - A model for the weather type sequence
    - A model for the meteorological variables conditionally to the weather type

Weather type models

- An example : the chain dependent model (Richardson, 1981)
  - Weather type S_t={Dry,Wet}
    - First order Markov chain with 2 states
  - Rainfall amount R_t conditionally independent in time given the weather type sequence
  - Temperature, solar radiation, wind speed Y_t
    - Residual correlation between successive observations modeled as an AR process with time varying coefficients

$$Y_t = \beta_0^{(S_t)} + \beta_1^{(S_t)} Y_{t-1} + \sigma^{(S_t)} \epsilon_t$$

$(\beta_i^{(s)})$, $(\sigma^{(s)})$ unknown parameters and $\{\epsilon_t\}$ a Gaussian white noise

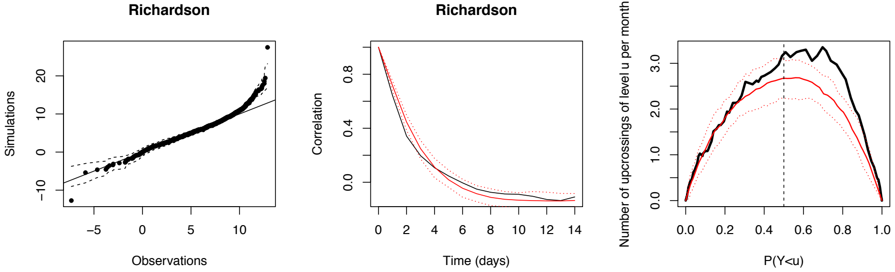
$$\begin{array}{lccccccccc}
\textit{Rainfall} & \cdots & & R_{t-1} & & R_t & & R_{t+1} & & \cdots \\
 & & & \uparrow & & \uparrow & & \uparrow & & \\
\textit{Weather type} & \cdots & \rightarrow & S_{t-1} & \rightarrow & S_t & \rightarrow & S_{t+1} & \rightarrow & \cdots \\
 & & & \downarrow & & \downarrow & & \downarrow & & \\
\textit{Temperature, (wind, ...)} & \cdots & \rightarrow & Y_{t-1} & \rightarrow & Y_t & \rightarrow & Y_{t+1} & \rightarrow & \cdots
\end{array}$$

  - Temperature dynamics modeled as a mixture of two AR(2) processes
  - Probably not the optimal clustering for describing temperature conditions!
- More sophisticated models for temperature time series discussed in the sequel

- The selected model is not easily interpretable (results for January at Brest) : the regimes are close to each other

$$Y_t = \begin{cases} 1.26 + 0.88\,Y_{t-1} - 0.08\,Y_{t-2} + 1.96\epsilon_t & (S_{t-1} = dry), \quad \mu^{(1)} = 6.41 \\ 1.28 + 1.01\,Y_{t-1} - 0.18\,Y_{t-2} + 2.10\epsilon_t & (S_{t-1} = wet), \quad \mu^{(2)} = 7.39 \end{cases}$$

Transition matrix : $\begin{bmatrix} 0.55 & 0.45 \\ 0.33 & 0.77 \end{bmatrix}$, stationary distribution $\begin{bmatrix} 0.24 \\ 0.76 \end{bmatrix}$



Richardson's model

- The dry/wet classification is not convenient for the temperature in Brest's climate.
- The non linearities are not reproduces by the model.
- Weather types can be obtained from other variables.

- Other variables (pressure, wind at the same site) could be considered for the clustering, but results are not improved

A SWGEN model for the temperature

- Introduce the weather type as a latent (hidden) variable : "Markov Switching AutoRegressive" model (MS-AR)

$$\begin{array}{lccccccccc} \textit{Weather type} & \cdots & \rightarrow & S_{t-1} & \rightarrow & S_t & \rightarrow & S_{t+1} & \rightarrow & \cdots \\ & & & \downarrow & & \downarrow & & \downarrow & & \\ \textit{Temperature} & \cdots & \rightarrow & Y_{t-1} & \rightarrow & Y_t & \rightarrow & Y_{t+1} & \rightarrow & \cdots \end{array}$$

  + Estimation procedure will find the "optimal" weather type
  - Estimation procedure more complicated, simpler models are needed
  - Hidden weather type modeled as a first order Markov chain
  - Linear Gaussian AR(p) model for the temperature evolution conditionally to the weather type

$$Y_t = \beta_0^{(S_t)} + \beta_1^{(S_t)} Y_{t-1} + \ldots \beta_p^{(S_t)} Y_{t-p} + \sigma^{(S_t)} \epsilon_t$$

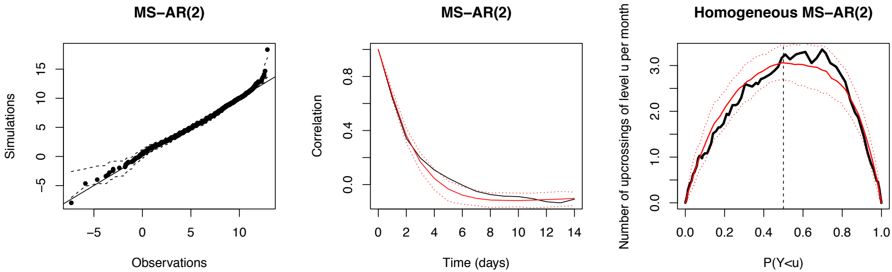
    - $(\beta_i^{(s)})$, $(\sigma^{(s)})$ unknown parameters and $\{\epsilon_t\}$ iid $\mathcal{N}(0, 1)$ sequence

- Maximum likelihood estimates (MLE) can be computed with the EM algorithm
- The selected model is interpretable (results for January at Brest)

$$Y_t = \begin{cases} 1.10 + 0.86 Y_{t-1} - 0.09 Y_{t-2} + 1.83 \epsilon_t & (S_{t-1} = 1), \quad \mu^{(1)} = 4.64 \\ 7.20 + 0.32 Y_{t-1} - 0.03 Y_{t-2} + 1.02 \epsilon_t & (S_{t-1} = 2), \quad \mu^{(2)} = 10.07 \end{cases}$$

Transition matrix : $\begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}$, stationary distribution $\begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix}$

- **Regime 1** : lower temperature with slow variations (anticyclonic)
- **Regime 2** : higher temperature with low and quick variation around the mean (cyclonic)

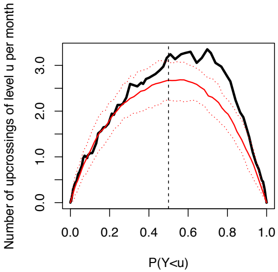


model is able to reproduce some important statistical properties of the data.

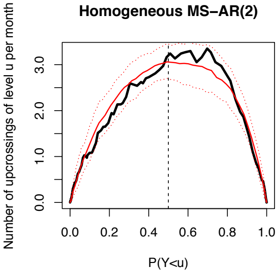
A MS-AR model for the temperature

- MS-AR models much better reproduce the observed asymmetry in up-crossing rate
  - No enough variability around middle levels?

A priori classification



MS-AR model



- In MS-AR models the regime switchings are independent of past temperature conditions
  - The probability of staying in the cold regime at time t is higher if the temperature is low at time $t - 1$?
  - Adding this in the model could create more asymmetry?

$$
\begin{array}{lccccccccc}
\textbf{Hidden Regime} & \cdots & \rightarrow & S_{t-1} & \rightarrow & S_t & \rightarrow & S_{t+1} & \rightarrow & \cdots \\
 & & & \downarrow & \nearrow & \downarrow & \nearrow & \downarrow & & \\
\textbf{Temperature} & \cdots & \rightarrow & Y_{t-1} & \rightarrow & Y_t & \rightarrow & Y_{t+1} & \rightarrow & \cdots
\end{array}
$$

- Logistic link function for the switching probabilities. For $s \in \{1, 2\}$,

$$P(S_t = s | S_{t-1} = s, Y_{t-1} = y_{t-1}) = \pi_-^{(s)} + \frac{1 - \pi_-^{(s)} - \pi_+^{(s)}}{1 + \exp\left(\lambda_0^{(s)} + \lambda_1^{(s)} y_{t-1}\right)}$$

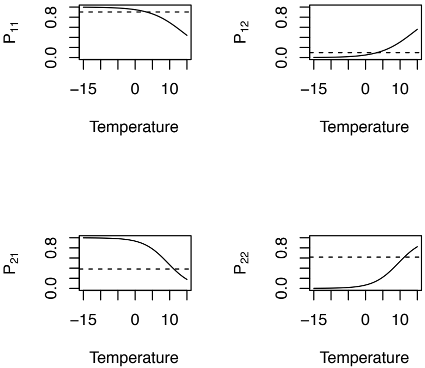
- Linear Gaussian AR(p) model for the temperature conditionally to the weather type

$$Y_t = \beta_0^{(S_t)} + \beta_1^{(S_t)} Y_{t-1} + ...\beta_p^{(S_t)} Y_{t-p} + \sigma^{(S_t)} \epsilon_t$$

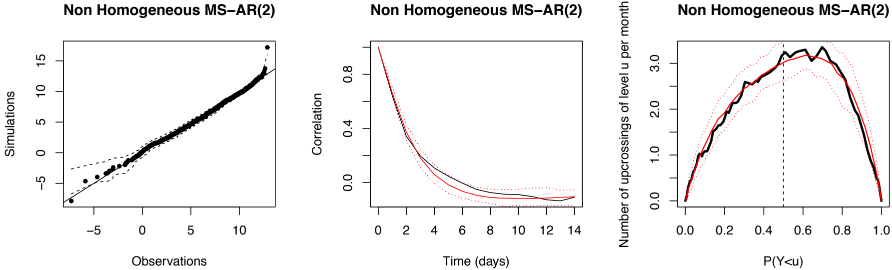
  - $(\beta_i^{(s)})$, $(\sigma^{(s)})$ unknown parameters and (ϵ_t) iid sequence of $\mathcal{N}(0, 1)$ r.v.

A non-homogeneous MS-AR models for the temperature

- Model fitted with the EM algorithm
- Similar interpretation for the regimes
  - **Regime 1** : lower temperature with slow variations (anticyclonic)
  - **Regime 2** : higher temperature with low and quick variation around the mean (cyclonic)
- Higher probability of staying in regime 1 if low temperature



A non-homogeneous MS-AR model for the temperature

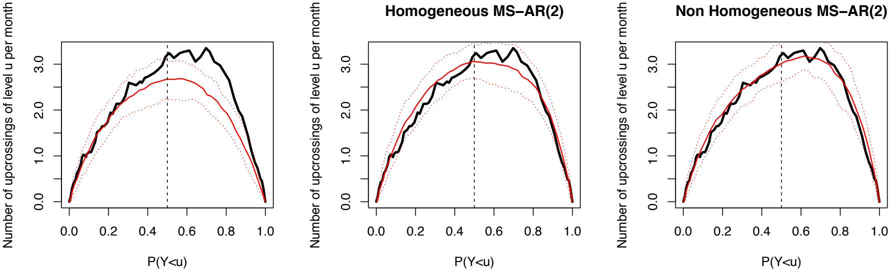


data, simulation

- The model is able to reproduce asymetries more accurately.

A non-homogeneous MS-AR models for the temperatures

- Durations in the regimes are no more geometric (mixture of geometric)
- Create more asymmetry in the dynamics



- Theoretical framework flexible enough to include models with exogenous covariates
- For example non-homogeneous HMMs used for statistical downscaling

$$
\begin{array}{lccccccccc}
\textbf{Covariates} & \cdots & \rightarrow & Z_{k-1} & \rightarrow & Z_k & \rightarrow & Z_{k+1} & \rightarrow & \cdots \\
 & & & & \searrow & & \searrow & & \searrow & \\
\textbf{Hidden Regime} & \cdots & \rightarrow & S_{k-1} & \rightarrow & S_k & \rightarrow & S_{k+1} & \rightarrow & \cdots \\
 & & & \downarrow & & \downarrow & & \downarrow & & \\
\textbf{Output time series} & \cdots & & Y_{k-1} & & Y_k & & Y_{k+1} & & \cdots
\end{array}
$$

- Covariates = large scale information
  - Assumed to be an ergodic Markov process
- Output time series = local weather conditions
  - (rainfall,temperature,...) at a meteorological station

Conclusions

- Weather type models provide a flexible and interpretable family of models for meteorological time series
- Models have been developed/validated for generating
  - (rainfall, temperature, solar radiation, humidity, wind speed) simultaneously at a single location with a priori clustering
  - Wind speed, wind direction, temperature independently at a single location with latent clustering
  - Rainfall and temperature independently at several locations simultaneously with latent clustering
- More works needed for
  - Developing multi-site and multi-parameters generators
  - Improving some aspects of existing models
    - Interannual variability underestimated ("overdispersion" phenomenon)
    - Probability of long "events" (dry spell, heat wave,...) generally underestimated but can be improved by introducing large scale covariate
- R package under development
- References

  - Ailliot P., Monbet V., (2012), Markov-switching autoregressive models for wind time series. Environmental Modelling & Software, 30, pp 92-101.
  - Ailliot P., Pène F. (2013), Consistency of the maximum likelihood estimate for Non-homogeneous Markov-switching models. Submitted.