

Markov-switching autoregressive models for wind time series

Pierre Ailliot¹ Valérie Monbet²

¹ Université de Brest
² Université de Rennes

Outline

- 1 Data
- 2 MS-AR models for monthly data
- 3 Seasonal model
- 4 Model with interannual component
- 5 Conclusions/perspectives

Plan

- 1 Data
- 2 MS-AR models for monthly data
- 3 Seasonal model
- 4 Model with interannual component
- 5 Conclusions/perspectives

Data

- Wind speed in France (Ouessant, $48^{\circ}27'$ N, $5^{\circ}6'$ W)

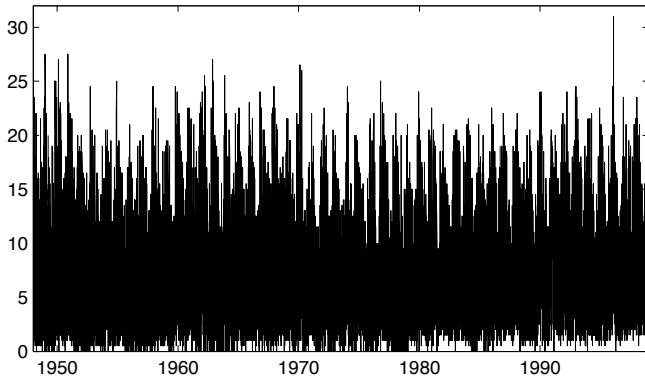


FIG.: Wind speed in ms^{-1} (y-axis) at Ouessant between 1948 and 1998.

Data

- Tests performed on other (shorter) time series
- Pretreatment
 - Filter hourly data to get 6 hours 'mean' wind speed
 - Remove short term fluctuations ("synoptic wind")
 - Decreases the number of missing data/outliers
 - Remove years with "too many" missing data (2 years)
- Characteristics ?
 - Values in $(0, +\infty)$
 - Strong dependence between successive observations
 - Non-stationary components : daily, seasonal, interannual (?)

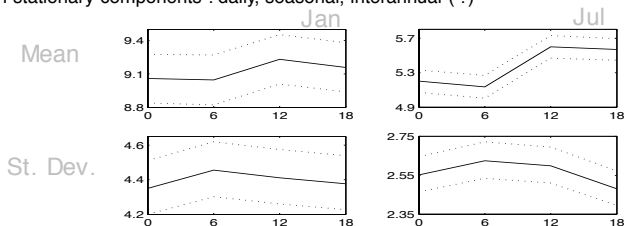


FIG.: Daily variations for the mean (top) and standard deviation (bottom) of the wind speed in January (left) and July (right). The x-axis represents the time in the day. The dotted lines correspond to 95% confidence interval computed using the (unrealistic) assumption that the observations comes from an i.i.d. Gaussian sample to help interpretation.

Data

- Tests performed on other (shorter) time series
- Pretreatment
 - Filter hourly data to get 6 hours 'mean' wind speed
 - Remove short term fluctuations ("synoptic wind")
 - Decreases the number of missing data/outliers
 - Remove years with "too many" missing data (2 years)
- Characteristics ?
 - Values in $(0, +\infty)$
 - Strong dependence between successive observations
 - Non-stationary components : daily, seasonal, interannual (?)
 - Existence of different weather regimes (e.g. cyclonic/anticyclonic)

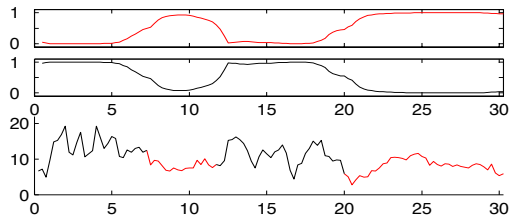


FIG.: One month of wind speed (bottom) with corresponding smoothing probabilities.

Plan

- 1 Data
- 2 MS-AR models for monthly data
- 3 Seasonal model
- 4 Model with interannual component
- 5 Conclusions/perspectives

Monthly model

- Focus on 1 month, $S_t \in \{1, \dots, M\}$ weather type, $Y_t \geq 0$ wind speed
- Assumptions : Markov-Switching AutoRegressive (MS-AR) model

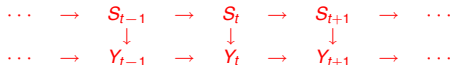
- S_t is a hidden Markov chain

$$P(S_t = s_t | S_0 = s_0, \dots, S_{t-1} = s_{t-1}, Y_0 = y_0, \dots, Y_{t-1} = y_{t-1}) = P(S_t | S_{t-1} = s_{t-1})$$

- Y_t is an AR(p) process with time-evolving coefficients

$$P(Y_t | S_0 = s_0, \dots, S_t = s_t, Y_0 = y_0, \dots, Y_{t-1} = y_{t-1}) = P(Y_t | S_t = s_t, Y_{t-p} = y_{t-p}, \dots, Y_{t-1} = y_{t-1})$$

- DAG ($p = 1$)



- Paramatrization of the AR(p) models

$$Y_t = a_0(t)^{(S_t)} + a_1^{(S_t)} Y_{t-1} + \dots + a_p^{(S_t)} Y_{t-p} + \sigma^{(S_t)} \epsilon_t$$

with $\{\epsilon_t\}$ i.i.d. $\mathcal{N}(0, 1)$ and

$$a_0^{(s)}(t) = \alpha_0^{(s)} + \alpha_1^{(s)} \cos\left(\frac{2\pi}{T_d}(t - \alpha_2^{(s)})\right)$$

The "amplitude" $\alpha_1^{(s)}$ and "peak" hour $\alpha_2^{(s)}$ of the daily component depend on the weather type s
 Usual MS-AR model (Hamilton, 1989) : $a_0^{(s_t)}(t) = a_0^{(s_t)}$

Discussion

- Other parametrizations have been tried

- Allow daily fluctuations for other parameters (e.g. $\sigma^{(s_t)}$)
- Replace Gaussian by Gamma distribution to ensure $Y_t \geq 0$
 - Minor improvement as concerns the marginal distribution ; Gaussian model quite good !
 - Requires unrealistic constraints (such as $a_2^{(s_t)} > 0$) , increases CPU time

- Comparison with other models

- Brown's model (1984) (\sim Box-Jenkins methodology)

$$\left(\frac{Y_t - \mu(t)}{\sigma(t)} \right)^\beta = Z_t$$

- $\{Z_t\}$ AR(2) process, $\mu(t)$ and $\sigma(t)$ periodic function (period 1 day)
- A power distribution is applied to get approximative marginal normality
- Only one regime
- Parlange and Katz (2000) (extension of Richardson's model to include wind)
 - Daily data
 - $S_t \in \{0, 1\}$ (0 if dry day and 1 if wet day) is a Markov chain, **not hidden**
 - $Y_t^{1/2} = \mu^{(S_t)} + \sigma^{(S_t)} Z_t$ with $Z_t = aZ_{t-1} + \epsilon_t$
 - $\mu^{(s)}$ and $\sigma^{(s)}$ related to the mean and standard deviation of the wind speed conditionally on dry/wet

$$Y_t^{1/2} = \mu^{(S_t)} + \sigma^{(S_t)} \left(a \frac{Y_{t-1}^{1/2} - \mu^{(S_{t-1})}}{\sigma^{(S_{t-1})}} \right) + \epsilon'_t$$

- $\{S_t, (Y_t)^{1/2}\}$ is not a MS-AR model but dynamics in each regime is an AR model !

Results

- The model has been fitted on data sets with different climatology (various locations including Ouessant) and length (from a few years to a few decades)
Inference : Expectation-Maximization algorithm.
Initialization : we use the inclusion of the model.
- BIC generally selects a model with 3 regimes (sometimes 2 regimes) and order $p = 2$, with no daily component in January and a daily component in July.

January											
M		1	2	3	4	5	1	2	3	4	5
MC	AR	$p = 1$					$p = 2$				
H	H	29452	28898	28921	28844	28890	29120	28518	28524	28546	28577
H	N	29464	28925	28964	29026	28977	29132	28545	28569	28631	28688
July											
M		1	2	3	4	5	1	2	3	4	5
MC	AR	$p = 1$					$p = 2$				
H	H	23572	23299	23295	23367	23408	23380	23100	23109	23190	23233
H	N	23446	23142	23135	23192	23279	23258	22952	22952	23028	23110

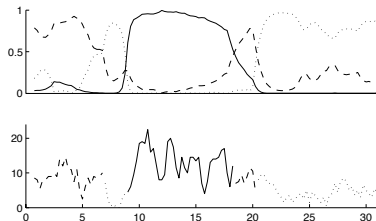
TAB.: BIC values for the various MS-AR models fitted for the months of January and July. The second column indicates if the AR models are homogeneous (H) or non-homogeneous (HN).

Results

- The selected models are interpretable
- Main difference between the different regimes : volatility

Transition matrix					AR models				
S_{t-1}	S_t			$\pi^{(s)}$	Coefficients				$\mu^{(s)}$
	1	2	3		$a_0^{(s)}$	$a_1^{(s)}$	$a_2^{(s)}$	$\sigma^{(s)}$	
1	0.92 [0.026]	0.07 [0.050]	0.01 [0.032]	0.35	1.13 [0.121]	0.96 [0.012]	-0.13 [0.008]	1.65 [0.051]	6.63
2	0.07 [0.023]	0.91 [0.015]	0.02 [0.023]	0.35	2.83 [0.346]	0.86 [0.023]	-0.19 [0.012]	2.66 [0.073]	8.77
3	0.01 [0.023]	0.03 [0.046]	0.96 [0.038]	0.30	6.36 [0.213]	0.69 [0.035]	-0.20 [0.024]	3.44 [0.165]	12.32

TAB.: Estimated parameters for the homogeneous model with $M = 3$ regimes and autoregressive models of order $p = 2$. Asymptotic standard errors are given in brackets. Results for January.



Results

- When used for simulation purpose, the models are able to reproduce observed statistics (marginal distribution, dynamics, daily component)

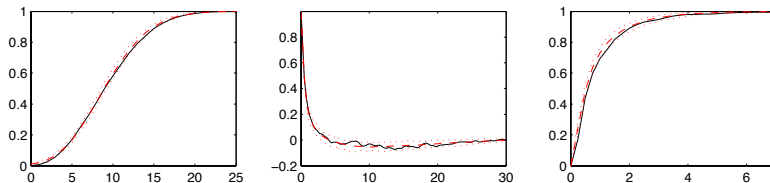


FIG.: Left : cumulative distribution function of the marginal distribution. Middle : autocorrelation function. Right : cumulative distribution function of the time duration of the sojourns below the threshold 6 ms^{-1} . The full line corresponds to the sample functions and the dashed line to the fitted model with a 95% prediction intervals (dotted line). The distributions for the fitted model was obtained by simulation. Results for January/July at Ouessant.

Plan

- 1 Data
- 2 MS-AR models for monthly data
- 3 Seasonal model**
- 4 Model with interannual component
- 5 Conclusions/perspectives

Seasonal model

- A practical motivation : problems at the end/beginning of the month when simulating
- Classical approaches in the literature ?
 - Apply transformation (scaling using seasonal mean and variance) to "stationarize" data
 - Assume that some of the coefficients of the model are periodic function
- Seasonal evolution of the parameters (monthly models with 3 regimes, **Ouessant**)
 - Smooth seasonal evolution of all the parameters, good picture of the climatology

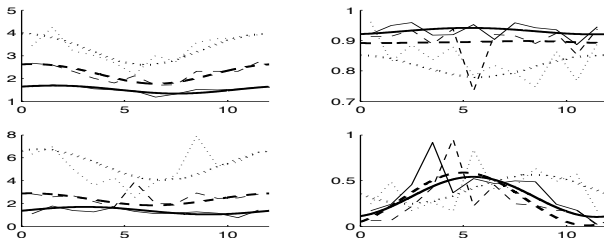


FIG.: Seasonal variations of $\sigma^{(s)}$ (top left), $q_{s,s}$ (top right), $\alpha_0^{(s)}$ (bottom left) and $\alpha_1^{(s)}$ (bottom right). The full line corresponds to regime $s = 1$, the dashed line to $s = 2$ and the dotted line to $s = 3$. The thin line corresponds to the values obtained when fitting separately the models to monthly data whereas the thick line corresponds to the values obtained after fitting the seasonal model on yearly data. The x-axis represents the time in month.

Seasonal model

- In the seasonal MS-AR model, coefficients (transition matrix and emission probabilities) are assumed to be simple periodic functions

$$f(t) = f_0 + f_1 \cos\left(\frac{2\pi}{T_y}(t - f_2)\right)$$

- f_0, f_1, f_2 unknown parameters, T_y is the number of observations per year
- Constraints are added for some parameters (e.g. transition probabilities must be positive and sum to one)
- Both the frequency and the characteristics of the weather types are allowed to change with the season
- EM algorithm initialized with parameters values obtained from fitted monthly models
 - Save CPU time and avoid convergence problems

Results

- Model validated on a monthly basis : similar results than with the monthly models
- ...but underestimation of the interannual variability ("overdispersion")

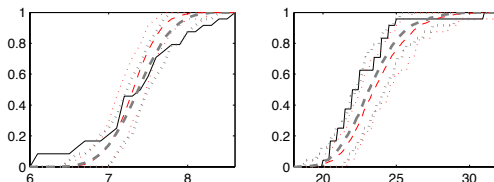


FIG.: Distribution function of the annual mean (left) and annual maxima (right). The full line corresponds to the sample function and the dashed line to the fitted models with a 95% prediction intervals (dotted line). The thin lines correspond to the **seasonal model without trend** and the thick lines to the **model with trend** (explained hereafter). The distribution for the fitted model was obtained by simulation. Results for the time period 1973-1998.

- According to Katz and Parlange (1999), two possible explanations :
 - Inadequate modelling of high frequency fluctuations (seems ok here !)
 - Existence of low frequency fluctuations (investigated in the next part)
- What about stochastic seasonality ?

Plan

- 1 Data
- 2 MS-AR models for monthly data
- 3 Seasonal model
- 4 Model with interannual component
- 5 Conclusions/perspectives

Model with interannual component

- How should we include long-term fluctuations ?
 - Frequency and/or characteristics of the weather types ?
- $E[S_t | Y_1 = y_1, \dots, Y_T = y_T]$ gives information on the distribution of weather types
 - Related to the volatility at time t
 - High values corresponds to high variability (cyclonic conditions)
 - Low values corresponds to low variability (anticyclonic conditions)

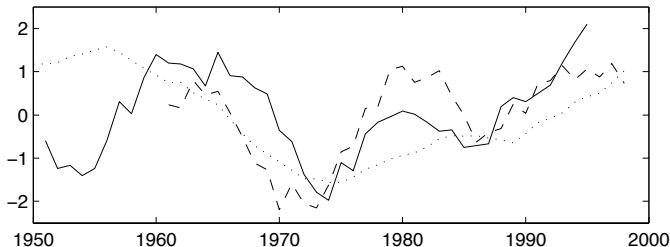


FIG.: 7-year running mean of the smoothing expectation $E[S_t | Y_{1-p} = y_{1-p}, \dots, Y_T = y_T]$ for the seasonal model together with the AMO index (dotted line). The full line corresponds to the model fitted on in-situ data whereas the dashed line corresponds to the model fitted on ERA40 data at the same location. The three time series have been scaled. Results for Ouessant.

Model with interannual component

- Focus on 1973-1998 (almost linear trend ?)
- A linear trend was added in the transition probabilities

$$P(S_t = s' | S_{t-1} = s) \propto q_{s,s'} \exp \left(\kappa_{s'} \cos \left(\frac{2\pi}{T_y} (t - \phi_{s'}) \right) + \lambda_{s'} t \right)$$

- Assumption : interannual components only impact the frequency of the weather type ; not their characteristics
- EM algorithm used to fit the model (preliminary estimation with the seasonal model)
- Better description of the interannual variability

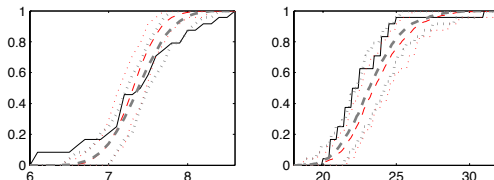


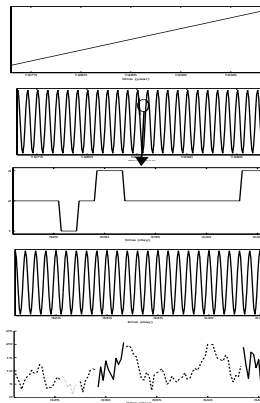
FIG.: Distribution function of the annual mean (left) and annual maxima (right). The full line corresponds to the sample function and the dashed line to the fitted models with a 95% prediction intervals (dotted line). The thin lines correspond to the **seasonal model without trend** and the thick lines to the **model with trend**. The distribution for the fitted model was obtained by simulation. Results for the time period 1973-1998.

Plan

- 1 Data
- 2 MS-AR models for monthly data
- 3 Seasonal model
- 4 Model with interannual component
- 5 Conclusions/perspectives**

Summary

Time scale	Model
Year	Deterministic Linear trend
Season	Deterministic Trigonometric function
Week	Stochastic weather type Markov chain
Day	Deterministic Trigonometric function
Hour	Stochastic AR(2) model



Conclusions/perspectives

- Good features
 - Flexible enough to reproduce complex properties of the data (skewed marginal distribution, non-linearities in the dynamics,...)
 - Remain interpretable : allow to understand the deficiencies and build extensions (seasonal model, trends,...)
- Disadvantage ?
 - Parameter estimation more "technical" than with other existing stochastic weather generator models
- Perspectives ?
 - Space-time model → Julie's talk.
 - Wind direction : AR models for circular variables ?
 - Other parameters (rain, temperature,...)
 - Stochastic seasonality ?