

# An autoregressive model with time-varying coefficients for wind fields

Pierre Ailliot, Valérie Monbet, Marc Prevosto

IFREMER/Metocean, Brest, France

UBS/SABRES, Vannes, France

pierre.ailliot@univ-ubs.fr, valerie.monbet@univ-ubs.fr, mpre@ifremer.fr

12th April 2005

## Abstract

**In this paper, an original Markov-switching autoregressive model is proposed to describe the space-time evolution of wind fields. At first, a non-observable process is introduced in order to model the motion of the meteorological structures. Then, conditionally to this process, the evolution of the wind fields is described by using autoregressive models with time varying coefficients. The proposed model is calibrated and validated on data in North Atlantic.**

***Keywords:* Space-time model, Wind fields, Markov-switching autoregressive model.**

## 1 Introduction

Wind conditions have a determining influence on many human activities and physical phenomena. For instance, wind time series permit to evaluate the power values produced by wind turbines (Brown, 1984; Castino 1998), to predict the propagation of an oil-spill, to assess the profitability of a maritime line (Ailliot, 2003) or to study coastal erosion. In order to understand and predict the evolution of these phenomena, it is useful to develop stochastic models for wind time series. Such models can be used to make short-term forecasts, to perform Monte-Carlo studies (Ailliot, 2003) or to assimilate in-situ measurements in numerical forecasts (Malmberg, 2005).

For some applications, it is sufficient to have a model which describes the evolution of the wind in a fixed location, and different models have been proposed (see Brown *et al.* (1984), Toll *et al.* (1997), Ailliot *et al.* (2004), Monbet *et al.* (2005)). For other applications, it is necessary to know

the wind conditions in several points simultaneously a space-time model is useful. The few models which have been proposed in the literature are based on Lagrangian or Eulerian methods. In the first approach, one will follow the space-time evolution of some entities, like the storms for example. It has been used for example by Cason (1998) to predict extreme wind speeds in hurricanes or by Boukhanovsky (2003b) for storms in the Barents sea. In the second approach, one will model directly the evolution of the process at a fixed set of locations. For most of the applications, it is necessary to know the wind conditions on a fixed spatial grid and the eulerian approach is then more appropriate. We will focus on this approach in this paper.

Boukhanovsky (2003a) and Malmberg (2004) proposed linear autoregressive models to describe time series of wind-fields. In this two articles, a principal component analysis is first used in order to reduce number dimension of the observations. However, this type of model can not reproduce some features of the wind fields, and in particular the motions of the meteorological structures. In this paper, we propose an original model, in which these motions are introduced as a hidden Markov chain. Conditionally to this hidden process, the evolution of the wind fields is modeled using autoregressive models with time-varying coefficients.

In the first part, we present the data and the model. Then, in the second part, we briefly discuss the problem of parameter estimation, and finally, the last part is devoted to model validation.

## 2 Model description

### 2.1 The data

In this study, we have used hindcast data produced by ECMWF (European Center for Medium range Weather Forecast). This data set describes the wind conditions at 10m above the sea-level for the period 1992-2003, on a regular grid with a resolution of 1.125 degree both in latitude and longitude and a time-step of 6 hours. This dataset is available everywhere on the earth and we have extracted an area which is located between latitude 45N and 50N and longitude 23W and 31W (see Figure 1). Its dimension is about 600km per 600km, and it contains  $N = 35$  points of the grid of the hindcast model. These points will be denoted by  $R_0 = (r_1, \dots, r_N)$ . They have been ordered in the following way  $r_i = (x_i, y_i) < r_j = (x_j, y_j)$  if  $x_i < x_j$  or  $x_i = x_j$  and  $y_i < y_j$ . With this convention, the point  $r_1$  is located at the South-West of the area  $R_0$ ,  $r_5$  at the North-West,  $r_{31}$  at the South-East and  $r_{35}$  at the North-East.

We will denote  $Z_t(R_0) = (u_t(r_1), u_t(r_2), \dots, u_t(r_N), v_t(r_1), v_t(r_2), \dots, v_t(r_N))$  the wind field on the area  $R_0$  at time  $t$ , with  $u_t(r_i)$  (*resp.*  $v_t(r_i)$ ) the zonal (*resp.* meridional) component at point  $r_i$  and time  $t$ . More generally, if  $R$  represents an arbitrary set of points,  $Z_t(R)$  will denote the

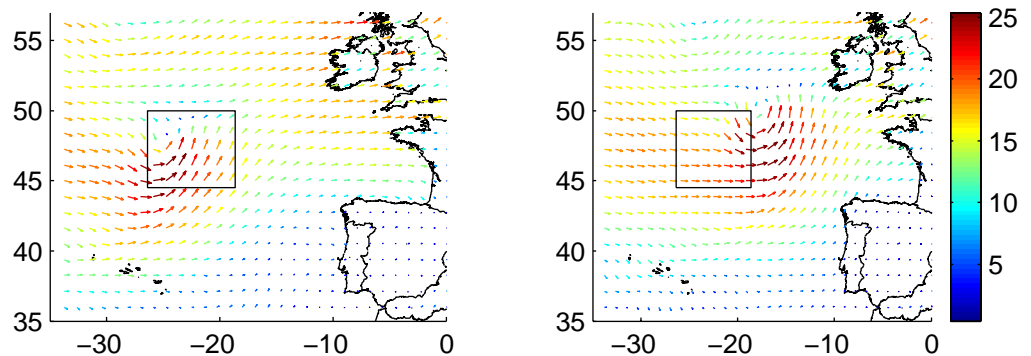


Figure 1: Wind fields on the 4th of January 1997: 12:00 on the left and 18:00 on the right.

wind field on this area. In the subsection 2.2 we propose a model for the multivariate time series  $\{Z_t(R_0)\}_{t \in \mathbf{N}}$  with values in  $\mathbf{R}^{70}$ .

It is well known that meteorological time series are generally non-stationary in time, with seasonal and daily components, and eventually a trend. In this article, we will only consider the data corresponding to the month of January. We did not find any daily components nor trend in this time series, and we will assume that the 11 months of January, which are available in our dataset are independent realizations of a stationary and ergodic process.

## 2.2 Model description

Wind conditions in the studied area are closely linked to the position and to the movement of the principal air masses which govern the climatology of the region (vast low-pressure or high-pressure systems). Depending on these positions, there may be systems of smaller amplitude (fronts or cyclones for example) that go through the considered area. An example, which corresponds to the 4th of January 1997, is given on Figure 1. At this date low crossed the area  $R_0$  between 12:00 and 18:00.

In Boukhanovsky (2003a) and Malmberg (2004), classical AR models are used in order to describe the space-time evolution of the wind fields. These models can successfully describe the motions of objects that translate at a constant speed. However, the motions of the meteorological structures depend on the position of the principal air masses. As a consequence, these motions evolve in time and AR models are not adapted. In this paper, we propose a more elaborate model, in which the motions are introduced explicitly. For this, let us denote  $S_t$  the translation of the wind field  $Z_{t-1}(R_0)$  between the dates  $t-1$  and  $t$  and  $\mathbf{S} \subset \mathbf{R}^2$  the state-space of this process. In order this translation to be well defined, it is necessary to assume that all the points of the area  $R_0$  move with the same speed, and in particular that the meteorological structures do not rotate

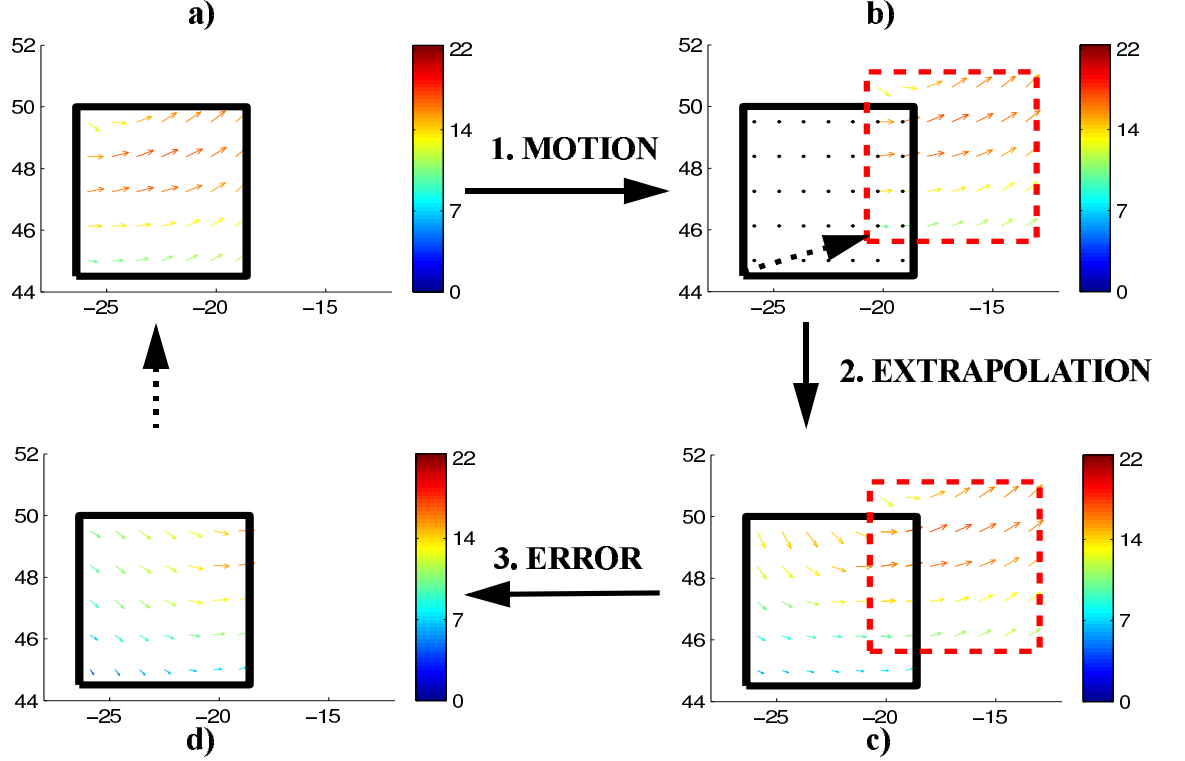


Figure 2: **a)** Wind fields on the 4th of January 1997, 12.00. **b)** Translated field (dashed box) and initial area  $R_0$  (solid box). **c)** Translated field (dashed box) and predicted field  $A^{(S_t)}Z_{t-1}(R_0) + B^{(S_t)}$  on  $R_0$  (solid box) **d)** Wind fields on the 4th of January 1997, 18.00.

and do not change their sizes. This assumption seems realistic according to the size of  $R_0$  and the model proposed in this paper could be adapted in order to describe more complex motions than pure translations. In Aberg et al. (2004), the evolution of the wind fields are modeled using image warping and a detailed discussion on the motions of wind fields can be found in this paper.

The translations  $\{S_t\}$  can not be observed directly, and thus are introduced as a hidden process. Then, in order to describe the evolution of the wind fields  $\{Z_t(R_0)\}$  conditionally to this hidden process, we use linear autoregressive models with time-varying coefficients. More precisely, it is assumed that

$$Z_t(R_0) = A^{(S_t)}Z_{t-1}(R_0) + B^{(S_t)} + H^{(S_t)}E_t \quad (1)$$

where  $\{E_t\}$  represents a standard Gaussian noise with 0 mean and identity covariance matrix, and  $A^{(s)} \in \mathbf{R}^{2N \times 2N}$ ,  $B^{(s)} \in \mathbf{R}^{2N \times 1}$  and  $H^{(s)} \in \mathbf{R}^{2N \times 2N}$  are matrices that depend on a reduced number of parameters  $\theta \in \Theta$ . In equation (1),  $A^{(S_t)}Z_{t-1}(R_0) + B^{(S_t)}$  represents a forecast of the wind field on the area  $R_0$  at time  $t$  conditionally to the field on  $R_0$  at time  $t-1$  and to the translation  $S_t$  and  $H^{(S_t)}E_t$  represents the conditional error. This is illustrated on the figure 2.

Let us now give more details on the modeling of the evolution of the hidden process  $\{S_t\}$  and

on the parametrization of the matrices  $A^s$ ,  $B^s$  and  $H^s$ . In subsection 3.1, we briefly explain how we can check the realism of the different assumptions made hereafter .

### Parametrization of the evolution of $\{S_t\}$

First, for the sake of simplicity, we assume that the state space  $\mathbf{S}$  of the hidden process  $\{S_t\}$  is a finite subset  $\{\mathbf{s}_1, \dots, \mathbf{s}_M\}$  of  $R^2$  which corresponds to displacements on the grid of the hindcast model at a speed lower than  $40ms^{-1}$ . With this choice, the cardinal  $M$  of  $\mathbf{S}$  is about 300. This assumption makes it possible to compute the maximum likelihood estimates using standard optimization algorithms (see subsection 3.2), and does not seem too much restrictive in practice. Indeed, the motions are large compared to the distance between two grid points of the numerical model, so that the error induced by the discretization of the state-space is low.

Then, we assume that the hidden process is a first order Markov chain and  $Q$  will denote its transition matrix. It is assumed to be a discretized Gaussian kernel: for all  $i, j \in \{1, \dots, M\}$ ,

$$Q(i, j) = P(S_t = \mathbf{s}_j | S_{t-1} = \mathbf{s}_i) \sim \exp(-\|\mathbf{s}_i - \mathbf{s}_j\|^2 / \sigma^2) \exp((\mathbf{s}_j - \mathbf{s}_0)' \Gamma (\mathbf{s}_j - \mathbf{s}_0)) \quad (2)$$

with unknown parameters  $\sigma > 0$ ,  $\mathbf{s}_0 \in \mathbf{R}^2$  and  $\Gamma \in \mathbf{R}^{2 \times 2}$  a positive definite symmetric matrix, and  $\|\cdot\|$  the Euclidean distance on  $\mathbf{R}^2$ .

### Parametrization of $A^{(s)}$

The matrices  $A^{(s)}$  are chosen such that they perform linear extrapolation of the field on the area  $R_0$  from the field on  $R_0 + s$ . This choice seems natural since, by definition of  $S_t$ , we know that if  $S_t = s$  then

$$Z_{t-1}(R_0) = Z_t(R_0 + s) + \delta_t$$

with  $\delta_t$  the deformation of the wind field  $Z_{t-1}(R_0)$  between the dates  $t - 1$  and  $t$ .

More precisely, if  $X$  and  $Y$  are two random vectors, let us denote by  $cov(X, Y) = E[XY'] - E[X]E[Y']$  the covariance matrix of  $X$  and  $Y$ . The matrices  $A^{(s)}$  are then defined as:

$$A^{(s)} = cov(Z_t(R_0), Z_t(R_0 + s))(cov(Z_t(R_0), Z_t(R_0)) + D)^{-1} \quad (3)$$

where  $D = cov(\delta_t, \delta_t)$  represents the covariance matrix of the deformation, which is assumed to be independent of  $\{S_t\}$ . Equation (3) is inspired by classical least-square estimates in linear regression models. In practice, the spatial covariance matrices  $cov(Z_t(R_0), Z_t(R_0 + s))$  and  $cov(Z_t(R_0), Z_t(R_0))$  are estimated directly from the data, thanks to our assumption of temporal

ergodicity. For example, the estimate of  $cov(Z_t(R_0), Z_t(R_0 + s))$  we have used is:

$$\hat{cov}(Z_t(R_0), Z_t(R_0 + s)) = \frac{1}{T} \sum_{t=1}^T z_t(R_0) z_t(R_0 + s)' - \frac{1}{T} \sum_{t=1}^T z_t(R_0) \frac{1}{T} \sum_{t=1}^T z_t(R_0 + s)'$$

where  $\{z_t\}_{t \in \{1 \dots T\}}$  denotes the time series of observed wind fields. The estimation of the matrix  $D$  is discussed in section 3.1.

### Parametrization of $\mathbf{B}^{(s)}$

We can check that the spectral radius of the matrices  $A^{(s)}$  defined above are lower than 1. It implies that the AR(1) model defined as

$$Y_t = A^{(s)} Y_{t-1} + B^{(s)} + H^{(s)} E_t$$

and which corresponds to the regime  $s \in \mathbf{S}$ , admits a unique stationary distribution. Let us denote  $M^{(s)} = (I - A^{(s)})^{-1} B^{(s)}$  the mean of this stationary distribution. It strongly depends on  $s$  since the mean wind in a meteorological structure is related to its speed and direction of displacement. To model this relation, we use the linear model described hereafter. Let us denote  $M^{(s)} = (m_u^{(s)}(1), \dots, m_u^{(s)}(N), m_v^{(s)}(1), \dots, m_v^{(s)}(N))'$  with  $m_u^{(s)}(i)$  (*resp.*  $m_v^{(s)}(i)$ ) the mean of the zonal (*resp.* meridional) component at point  $r_i$ . We assume that

1.  $m_u^{(s)}(i) = m_u^{(s)}$  and  $m_v^{(s)}(i) = m_v^{(s)}$  for all  $s \in \mathbf{S}$  and  $i \in \{1 \dots N\}$ . In other terms, it is assumed that the mean of the stationary distribution is the same at the different locations in  $R_0$ .
2.  $m^{(s)} = F s + G$  for all  $s \in \mathbf{S}$  with  $m^{(s)} = (m_u^{(s)}, m_v^{(s)})'$ ,  $F \in \mathbf{R}^{2 \times 2}$  and  $G \in \mathbf{R}^{2 \times 1}$ . Here,  $F$  and  $G$  denote matrices of unknown parameters.

### Parametrization of $\Sigma^{(s)} = \mathbf{H}^{(s)} \times (\mathbf{H}^{(s)})'$

The forecast error  $Z_t(R_0) - A^{(S_t)} Z_{t-1}(R_0) - B^{(S_t)}$  is mainly due to two terms:

- the deformation of the wind field  $Z_{t-1}(R_0)$  between dates  $t - 1$  and  $t$
- the extrapolation error. This error is linked to the existence of new meteorological structures coming in the area  $R_0$ .

In order to model these two terms, we make the assumptions listed below on the covariance matrices of the error  $\Sigma^{(s)}$ .

1. **Diagonal of the matrix.** To model the diagonal of  $\Sigma^{(s)}$ , we suppose that

$$\begin{aligned}\Sigma^{(s)}(i, i) &= f_u(\text{dist}(r_i, R_0 + s)) \\ \Sigma^{(s)}(i + N, i + N) &= f_v(\text{dist}(r_i, R_0 + s))\end{aligned}\quad (4)$$

for all  $s \in \mathbf{S}$  and  $i \in \{1 \dots N\}$ . In other terms, it is assumed that the variance of the error on the zonal (*resp. meridional*) component at a point  $r_i$ , namely  $\Sigma^{(s)}(i, i)$  (*resp.*  $\Sigma^{(s)}(i + N, i + N)$ ), only depends on the Euclidean distance  $\text{dist}(r_i, R_0 + s)$  between this point and the translated area  $R_0 + s$ . When this distance is null, the error is only due to the deformation of the wind fields, and then the error increases with the distance and the extrapolation error. Several parametric forms can be chosen for  $f_u$  and  $f_v$ . In this article, we use  $f_u(d) = \beta_u + \alpha_u \exp(-d/d_u)$  and  $f_v(d) = \beta_v + \alpha_v \exp(-d/d_v)$  where  $\beta_u, \alpha_u, d_u, \beta_v, \alpha_v$  and  $d_v$  are unknown positive parameters.

2. **Correlation matrix.** Then, it remains to model the correlation matrix of the error, defined as

$$R^{(s)}(i, j) = \Sigma^{(s)}(i, j) / \sqrt{\Sigma^{(s)}(i, i) \times \Sigma^{(s)}(j, j)}$$

For this, we assume that, for all  $s \in \mathbf{S}$ ,

$$\begin{aligned}R^{(s)}(i, j) &= g_u(\|r_i - r_j\|) \quad \forall (i, j) \in \{1, \dots, N\}^2 \\ R^{(s)}(i + N, j + N) &= g_v(\|r_i - r_j\|) \quad \forall (i, j) \in \{1, \dots, N\}^2 \\ R^{(s)}(i, j) &= 0 \quad \forall (i, j) \in \{1, \dots, N\} \times \{N + 1, \dots, 2N\}\end{aligned}\quad (5)$$

It means that the correlation between the errors on the zonal (*resp. meridional*) components at two points  $r_i$  and  $r_j$  only depends on the Euclidean distance  $\|r_i - r_j\|$  between these points, and that there is no correlation between the error on the zonal and meridional components. The functions  $g_u$  and  $g_v$  must satisfy  $g_u(0) = 1$  and  $g_v(0) = 1$  and be decreasing. In practice, we assume that  $g_u(d) = \exp(-d/l_u)$  and  $g_v(d) = \exp(-d/l_v)$  where  $l_u$  and  $l_v$  are unknown positive parameters. With this parametrization, the matrices  $R^{(s)}$  and thus  $\Sigma^{(s)}$  are positive definite symmetric (see Cressie (1993) pp 85).

The proposed model belongs to the class of the Markov Switching AutoRegressive models (denoted MS-AR hereafter). This kind of model has first been introduced by Hamilton in 1989 (see Hamilton (1989)) to analyse the rate of growth of USA GNP and then used in different fields of applications to model time series subject to regime shifts. The proposed MS-AR model has 20 parameters: 6 for the matrices  $B^{(s)}$ , 8 for the matrices  $\Sigma^{(s)}$  and 6 for the matrix  $Q$ . We will denote

by  $\theta \in \Theta \subset R^{20}$  the set of unknown parameters. The estimation of these parameters is described in next section.

### 3 Estimation

The most usual approach to estimate the parameters of a MS-AR model is to compute the Maximum Likelihood Estimates (MLE). The numerical computation of the MLE in models with hidden variables has been addressed by many authors. The most popular method is probably the EM algorithm which has first been introduced by Baum *et al.* (1970) for Hidden Markov Models (HMM) and then generalized to other models with missing variables by Dempster *et al.* (1977). However, this algorithm has several well-known limitations, one of them being the possible convergence to local maxima of the likelihood function. It is then important to use a good guess of the MLE as starting point of the algorithm.

In order to get a first estimate of the unknown parameters, a natural method consists in estimating the values taken by the hidden process  $\{S_t\}$  and then use this estimated sequence in order to compute empirical estimates of the quantities of interest. We have used such an approach for the wind fields and the proposed method is described in subsection 3.1. Then, in subsection 3.2, we briefly describe the Generalized EM algorithm which has been used to compute the MLE and we discuss some statistical properties of these estimates.

#### 3.1 Preliminary estimation

Let us first describe how we have computed estimates of the hidden translations  $\{S_t\}$ . For this, we have used an extra-information which is available in the dataset. Indeed, this dataset describes the wind conditions not only on  $R_0$ , but everywhere on the earth, so that it is possible to compute

$$(\hat{s}_2, \dots, \hat{s}_T) = \underset{(s_2, \dots, s_T) \in \mathbf{S}^{T-1}}{\operatorname{argmax}} \{f(s_2, \dots, s_T)\}$$

with

$$f(s_2, \dots, s_T) = \prod_{t=2}^T \exp\left(-\frac{\|z_{t-1}(R_0) - z_t(R_0 + s_t)\|^2}{d_0^2}\right) \times Q(s_{t-1}, s_t) \quad (6)$$

In equation 6,  $Q$  denotes the transition matrix defined in equation (2) and  $d_0$  a fixed real number. The first term in equation (6) constrains the estimated sequence of displacements  $\{\hat{s}_t\}$  to be such that  $z_{t-1}(R_0) \approx z_t(R_0 + \hat{s}_t)$ , and the second term will penalize sequences which evolve too quickly in time and thus act as a smoothing term. The function  $f$  can be interpreted as the likelihood function of a HMM, and thus the sequence  $(\hat{s}_2, \dots, \hat{s}_T)$  can be computed quickly using the Viterbi



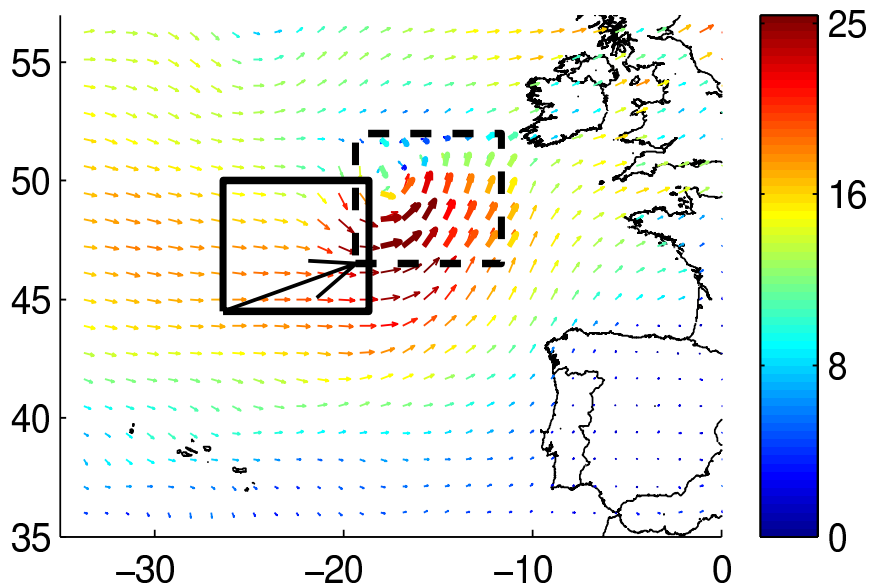


Figure 3: Situation on the 4th of January 1997 at 18:00. Solid box:  $R_0$ , dashed box:  $R_0 + s$ , thin arrows: wind field at 18:00, thick arrows: wind field observed at 12:00 on  $R_0$ .

algorithm (a recent review on HMM can be found in Ephraim *et al.* (2002)).

Off course, the choice of the function  $f$  is arbitrary, and there exists no objective criterion to validate this choice. In order to verify the meteorological realism of the estimated sequences, we performed visual verification. An example is given on Figure 3. It shows that the method succeeds to estimate the movement of the low pressure which go through the studied area on the 4th of January 1997. In particular, the two fields represented in the dashed box, and which represent respectively  $z_{t-1}(R_0)$  (thick arrows) and  $z_t(R_0 + \hat{s}_t)$  (thin arrows) are close to each other. In practice, we have used the numerical values  $d_0 = 3ms^{-1}$  and  $\Gamma = 0$ .

Once these displacements have been estimated, we can compute empirical estimates of the covariance matrix  $D$  of the deformation which is used to define the matrices  $A^{(s)}$  (see equation (3)). For this, we have computed

$$\hat{D} = \frac{1}{T-1} \sum_{t=2}^T [z_{t-1}(R_0) - z_t(R_0 + \hat{s}_t)][z_{t-1}(R_0) - z_t(R_0 + \hat{s}_t)]'$$

In the same way, we can use the sequence  $\{\hat{s}_t\}$  in order to estimate the conditional means  $M^{(s)}$ , the covariance matrices of the error  $\Sigma^{(s)}$ , and the transitions matrix  $Q$ . These empirical estimates can first be used in the analysis part, in order to guess parsimonious and realistic parametric form for these matrices and make it possible to check the realism of the different assumptions which are

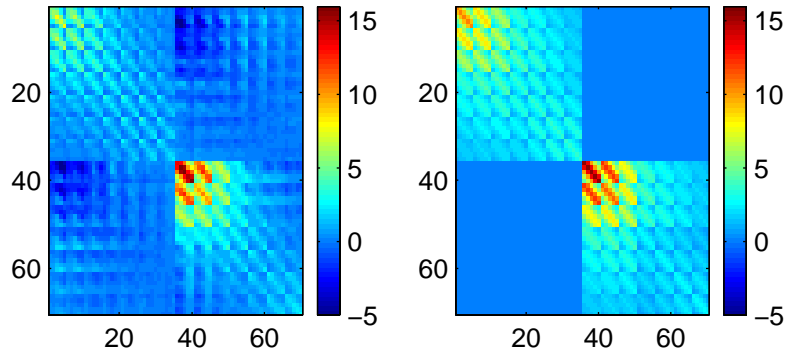


Figure 4: Empirical covariance matrix of the error (displacement of 3 grid steps through the east). Empirical version estimated from the sequence  $\{\hat{s}_t\}$  on the left and parametric version on the right.

made in section 2.2. They can also be used to fit the chosen parametric models using, for example, least square estimates.

As an example, we have plotted on Figure 4 the empirical covariance matrix of the error corresponding to a displacement  $s$  of 3 grid steps to the east and its parametric version. The shapes of these two matrices are in a good agreement, and thus the chosen parametrization seems to be adapted to our data.

### 3.2 Maximum likelihood estimates

The estimates obtained in the previous subsection have then been used as a starting point to an optimization algorithm. Let us briefly describe this algorithm which belongs to the family of the Generalized EM algorithms. If  $\theta_{n-1}$  denotes the value of the parameters after  $n-1$  iterations, the  $n$ th iteration consists in two main steps:

- **E-step.** It consists in calculating the same auxiliary function as in the classical EM algorithm (see Demspster et al, 1977). This function is denoted by  $R(\theta, \theta_{n-1})$ . When the hidden state-space  $\mathbf{S}$  is finite, it can be computed quickly thanks to the Forward-Backward algorithm which has been first proposed by Chang *et al.* (1966). Otherwise, one can use Monte-Carlo approximations of this function (see Douc *et al.* (2004)) and references therein).
- **M-step.** In the usual EM algorithm, this step consists in maximizing the function  $R(\theta, \theta_{n-1})$ , and update the current value of the parameters as  $\theta_n = \operatorname{argmax}_{\theta \in \Theta} R(\theta, \theta_{n-1})$ . For the model proposed in section 2, we did not find any analytical solution to this optimization problem, so that we had to use a numerical optimization algorithm. When the iterations stop with a non optimal value  $\theta_n$ , which is such that  $R(\theta_n, \theta_{n-1}) > R(\theta_{n-1}, \theta_{n-1})$ , the algorithm is called Generalized EM algorithm. This algorithm has first been proposed by Demspster *et*

*al.* (1977) and general conditions which imply the convergence of this algorithm to a local maximum of the likelihood function can be found in Wu (1982). In practice, we have used some iterations of a quasi-Newton algorithm to compute  $\theta_n$ . Each iteration of this algorithm is time consuming because of the complexity of the objective function and the number of parameters. Thus, a careful choice of this number of iterations, denoted by  $k_n$ , has to be done. According to the numerical tests that we have done, the best strategy seems to chose a small number of iterations for small values of  $n$  and then to increase progressively this number of iterations when the current values of the parameters get closer to a maximum of the likelihood function. In practice, we have used  $k_n = n$ .

Whereas there exists a considerable amount of literature about the numerical computation of MLE, the statistical issues concerning the asymptotic properties of these estimators have been addressed only recently. Explicit conditions which warrant the consistency of the MLE in MS-AR models have first been established simultaneously in 1998 by Krishnamurthy *et al.* and Francq *et al.* and assumptions which warrant consistency and asymptotic normality of MLE in MS-AR models with compact, not necessary finite, hidden state space  $\mathbf{S}$  can be found in Douc *et al.* (2004). It would be beyond the scope of this paper to show that the MS-AR model proposed in section 2 verifies these assumptions.

## 4 Model validation

In order to validate the MS-AR model, we have checked its capacity to perform short-term forecasts, and the obtained results have been compared to those of an AR(1) model.

Let us briefly explain how we have computed the best one-step ahead predictor for the MS-AR model. It is given by:

$$\begin{aligned}
 & E[Z_t(R_0)|Z_1(R_0), \dots, Z_{t-1}(R_0)] \\
 = & \sum_{s \in \mathbf{S}} P[S_t = s|Z_1(R_0), \dots, Z_{t-1}(R_0)] E[Z_t(R_0)|Z_{t-1}(R_0), S_t = s] \\
 = & \sum_{s \in \mathbf{S}} P[S_t = s|Z_1(R_0), \dots, Z_{t-1}(R_0)] (A^{(s)} Z_{t-1}(R_0) + B^{(s)})
 \end{aligned} \tag{7}$$

The forecast probabilities  $P[S_t = s|Z_1(R_0), \dots, Z_{t-1}(R_0)]$  which appears in equation (7) can be computed recursively (Forward algorithm). The generalization to other forecast horizons is straightforward. More details can be found in Krolzig (1997).

On Figure 5, we have plotted the variance of the one-step ahead forecast error on the zonal and meridional component at the different points of  $R_0$ , for the MS-AR model and AR(1) model. The

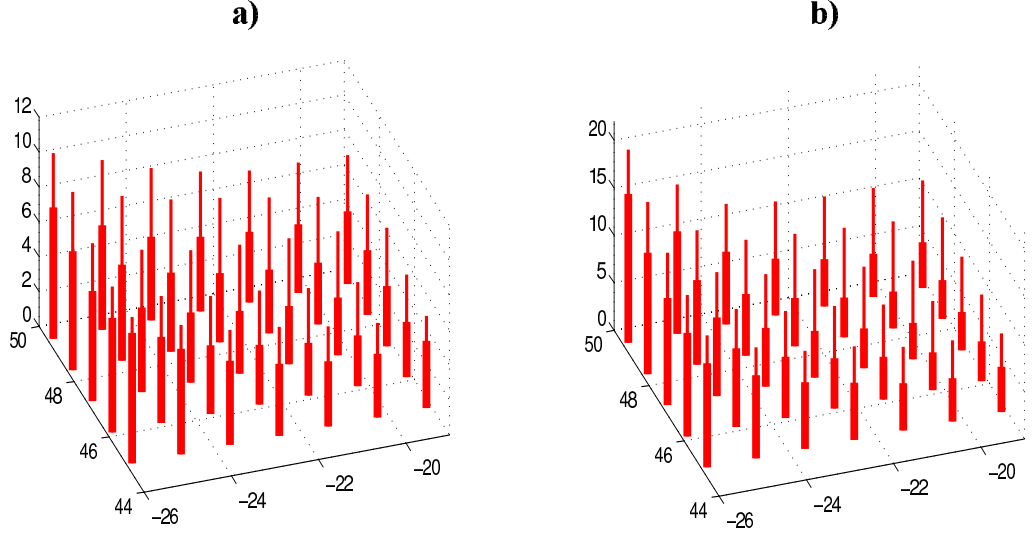


Figure 5: Variance of the one-step ahead forecast error on the zonal (a) and meridional (b) components at the different points of  $R_0$  (thick line: MS-AR model, thin line: AR(1) model)

$h$	1	2	3	4	5
$MS - AR$	151	538	952	1263	1498
$AR$	300	981	1719	2257	2560

Table 1: Evolution of the Perron-Frobenius norms with the forecast horizon  $h$

MS-AR model gives significantly better forecasts in all points of the domain  $R_0$ , the improvement being more important at the points located at the east of the domain. It is due to the fact that in the considered area the air-masses generally moves toward the east, and these points are generally closer to the translated area  $R_0 + S_t$ . Thus, in these points, the extrapolation error is low and the error is mainly due to the deformation of the wind fields. On the opposite, the errors at the points located at the west are more important because the "new" meteorological structures come in the area  $R_0$  from the west.

We have also computed the variance of the error corresponding to other time-step prediction. In order to compare the variance-covariance matrices of the forecast error corresponding to the AR(1) model and the MS-AR model, we have calculated their Perron-Frobenius norms. It is defined for any matrix  $\Sigma$  as  $\|\Sigma\|_{PF} = \sqrt{\text{tr}(\Sigma'\Sigma)}$ . The results are given in Table 1. The MS-AR model clearly improves the results obtained with the AR(1) model.

## 5 Conclusion

In the present paper, an original Markov Switching Autoregressive models is proposed for modeling the space-time evolution of the wind. The main originality of this model consists in introducing the movement of the air masses as a hidden variable. The hidden process governs the evolution of the short-term evolution of the wind fields. The model is fitted on hindcast data in North Atlantic and it is shown that it provides better short-term forecasts than AR(1) model.

This model has been developed specifically for wind fields, but similar methodology could be used to describe other space time processes in which motions are present. In the proposed model, the state space of the hidden process is assumed to be finite. This assumption seems realistic for wind fields, but restrictive for other space-time processes. The model could be refined by allowing continuous translations or more complex motions, such as rotations for instance.

## References

- Aberg, S. (2002). Modelling and prediction of wind fields using Gaussien Markov random fields and warping. *Master's Thesis, Lund University*
- Ailliot, P., Prevosto, M., Soukissian, T., Diamanti, C., Theodoulides, A. and Politis C. (2003). Simulation of sea state parameters process to study the profitability of a maritime line. *Proc. of ISOPE Conf.*
- Ailliot, P. (2004). *Modèles autorégressifs à changement de régimes markovien - Application à la modélisation de séries temporelles de vent.* PhD Thesis. Université de Rennes 1.
- Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Statistic.*, **41**, 164-171.
- Baxevani, A., Rychlik, I., Wilson, R.J. (2005). Modelling space variability of Hs in the North Atlantic. *Submitted to Extremes*
- Boukhanovsky A.V., Krogstad H.E., Lopatoukhin L.J. , Rozhkov V.A. (2003a). Stochastic simulation of inhomogeneous metocean fields. Part I: annual variability. *International Conference on Computational Science* , Springer Verlag, pp 213-222.
- Boukhanovsky A.V., Krogstad H.E., Lopatoukhin L.J., Rozhkov V.A. Athanassoulis G.A., Stephanakos C.N. (2003b). Stochastic simulation of inhomogeneous metocean fields. Part II: syn-

- optic variability and rare events. *International Conference on Computational Science*, Springer Verlag, pp 213-222.
- Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis, forecasting and control*. (revised edn.) Holden-Day, San Francisco.
- Brown, B.G., Katz, R.W. and Murphy, A.H. (1984). Time series models to simulate and forecast wind speed and wind power. *J. of clim. and appl. meteor.* **23**, 1184-1195.
- Casson E., Coles S. (1998). Extreme hurricane wind speeds: estimation, extrapolation and spatial smoothing. *J. of Wind Engineering and Industrial Aerodynamics*, **74-76**, 131-140.
- Castino, F., Festa, R. and Ratto, C.F. (1998). Stochastic modelling of wind velocities time series. *Journ. Wind Eng.*, **74-76**, 141-151.
- Cressie N. (1993). *Statistics for spatial data*. Wiley.
- Dempster, A.P., Laird, N.M. and Rubin, D.B., (1977). Maximum likelihood from incomplete data, *J. R. Stat. Soc. B*, **39** 1-38.
- Douc, R., Moulines, E. and Ryden, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, *Ann. Statis.*, **32;5**, 2254-2304.
- Ephraim, Y. and Merhav N. (2002). Hidden markov process. *IEE Transactions on Information Theory*, **48;6**, 1518-1569.
- Francq, C. and Roussignol, M. (1998). Ergodicity of autoregressive processes with Markov switching and consistency of the maximum-likelihood estimator. *Ann. Statis.*, **27;2** 1-38.
- Hamilton, J.D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, **57**, 357-384.
- Krishnamurthy, V. and Ryden, T. (1998). Consistent estimation of linear and non-linear autoregressive models with markov regime. *Journal of time series analysis*. **19(3)**, 291-307.
- Krolzig, H.M. (1997). *Markov-switching vector Autoregressions. Modelling, statistical inference and applications to business cycle analysis*, Lecture notes in economics and mathematical systems 454. Springer-Verlag, Berlin.
- Malmberg, A., Holst, U., Holst, J. (2005). Forecasting near-surface ocean winds with Kalman Filter techniques, *Ocean Engineering*.

- Monbet, V., Ailliot, P., Prevosto, M. (2005). Survey of stochastic models for wind and sea state time series, *Submitted to Probabilistic Engineering Mechanics*, **32**, 273-291.
- Toll, R.S.J. (1997). Autoregressive conditional heteroscedasticity in daily wind speed measurements, *Theor Appl Climatol*, **56**, 113-122.
- Wu, C.F.J. (1983). On the convergence properties of the EM algorithm, *Ann. Stat*, **11**, 95-103.