Journal of Statistical Planning and Inference 160 (2015) 75–88

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Non-homogeneous hidden Markov-switching models for wind time series

Pierre Ailliot [a,*], Julie Bessac [b], Valérie Monbet [b], Françoise Pène [a]

[a] Laboratoire de Mathématiques de Bretagne Atlantique, UMR 6205, Université de Brest, France

[b] Institut de Recherche Mathématiques de Rennes, UMR 6625, Université de Rennes 1, France

ARTICLE INFO

Article history:
Received 25 August 2014
Received in revised form 11 December 2014
Accepted 13 December 2014
Available online 24 December 2014

Keywords:
Wind time series
Markov-switching autoregressive process
Non-homogeneous hidden Markov process
Linear–circular time series
Consistency

ABSTRACT

In this paper we propose various Markov-switching autoregressive models for bivariate time series which describe wind conditions at a single location. The main originality of the proposed models is that the hidden Markov chain is not homogeneous, its evolution depending on past wind conditions. It is shown that they have good theoretical properties and permit to reproduce complex features of wind time series such as non-linear dynamics and multimodal marginal distributions. This is illustrated on a wind time series for a location off the French Atlantic coast using the R package NHMSAR.

1. Introduction

Meteorological time series are a key input in many risk forecasting and impact studies applications and historical data are often available over periods of time that are not long enough to get reliable estimates of the quantities of interest. Stochastic weather generators have been developed to overcome this insufficiency by simulating artificial sequences of unlimited length of meteorological variables with statistical properties similar to those of the observations (see Srikanthan and McMahon (1999) and references therein). These generators can also be useful for downscaling global climate models (see e.g. Maraun et al. (2010) and references therein) or infilling missing values by conditional simulation. In this paper we focus on wind time series. Various approaches have been proposed in the literature for modeling time series of wind speed (see e.g. Monbet et al. (2007) for a review). In comparison, there exist only very few models for circular time series of wind direction or for bivariate time series describing simultaneously the evolution of the wind speed and the wind direction. There is thus a need for models which can reproduce the specificities of such time series and this paper aims at filling this gap.

In this work we consider a wind time series for a location off the French Brittany coast in which bivariate marginal distribution has complex features (see Fig. 1 and Section 4 for a description of the data). In particular it clearly exhibits two modes, each one corresponding to a different meteorological regime or 'weather type'. The prevailing mode (westerly winds) corresponds to cyclonic conditions with low pressure systems coming from the North-Atlantic ocean, whereas the second mode (easterly winds) corresponds to anticyclonic conditions which can temporarily deviate or block the westerly flow. The alternation of such weather regimes is a well-known characteristic of the North-Atlantic/European area, and after Vautard (1990) brought some evidence for quasi-stationary solutions in the equations of the atmospheric flow, thus giving a physical meaning to the statistically-derived regimes, they have been broadly used in climate studies. More generally, the presence

* Corresponding author.
E-mail address: pierre.ailliot@univ-brest.fr (P. Ailliot).

http://dx.doi.org/10.1016/j.jspi.2014.12.005
0378-3758/

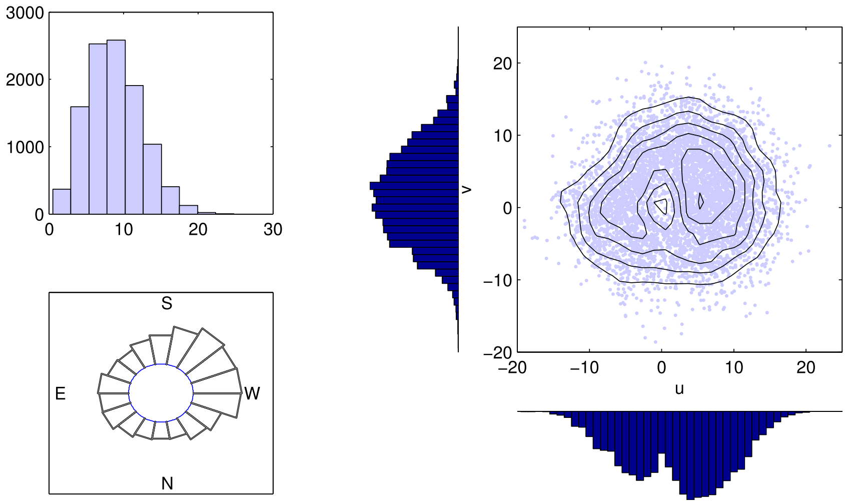

Fig. 1. Histogram of $\{U_t\}$ (top left), rose plot of $\{\Phi_t\}$ (bottom left), histograms of $\{u_t\}$ and $\{v_t\}$ and joint distribution of $\{u_t, v_t\}$ (right). The lines on the scatter plots are levels of a non-parametric kernel estimate of the bivariate density. Results for the months of January.

of regimes with distinct weather conditions is a usual feature of meteorological time series and a classical approach for modeling these meteorological regimes consists in introducing a hidden (or latent) variable. This idea goes back to Zucchini and Guttorp (1991) where Hidden Markov Models (HMMs) were proposed for modeling the space–time evolution of daily rainfall. HMMs have also been proposed for modeling time series of wind direction in Zucchini and MacDonald (2009). However HMMs assume that the successive observations are conditionally independent given the latent weather type and cannot reproduce the strong relationship which exists between the wind conditions at successive time steps for the dataset considered in this work. Markov-Switching AutoRegressive (MS-AR) have been proposed in this context to model time series of wind speed in Monbet et al. (2007); Pinson et al. (2008); Ailliot and Monbet (2012). MS-AR models extend both the usual HMMs, by adding dynamics in the regimes, and AR models, which are often used to model wind time series (see e.g. Brown et al. (1984)), by introducing regime switchings through a latent variable.

In HMMs or MS-AR models, the evolution of the weather type is independent of the past weather conditions. For our particular example, it would imply for example that the probability of switching from the cyclonic conditions to the anticyclonic conditions between time t and time $t+1$ does not depend on the wind conditions observed at time t whereas we know that these switchings generally occur when the wind is blowing from the north and is very unlikely to occur when the wind is blowing from the south. One originality of the models proposed in this paper is that the evolution of the latent weather type depends on past wind direction leading to non-homogeneous MS-AR (NHMS-AR) models. We show that NHMS-AR models lead to a better description of important characteristics of the data considered in this work, such as multimodality and non-linear dynamics, compared to MS-AR models.

The wind condition at a single location at time t can be described using the polar coordinates $\{U_t, \Phi_t\}$, where U_t denotes the wind speed with values in $\mathbb{R}^+$ and Φ_t the wind direction with values in $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ or the Cartesian coordinates $\{u_t, v_t\}$ where u_t and v_t denote respectively the zonal and meridional components with values in $\mathbb{R}$. The polar coordinates are generally used by meteorologists, probably because they are easier to interpret. However, from a statistical point of view, it is probably more straightforward to model the time series of Cartesian components since many models, such as Gaussian vector AR models, have been proposed for bivariate time series with values in $\mathbb{R}^2$ whereas the process $\{U_t, \Phi_t\}$ is a linear–circular process with values in $\mathbb{R}^+ \times \mathbb{T}$ and very few models have been proposed for such variables. Both representations are considered in this paper and a discussion of their respective advantages is given.

The paper is organized as follows. NHMS-AR models are introduced in Section 2 with specific parametrizations proposed when considering Cartesian and polar coordinates. In Section 3, we briefly describe the EM algorithm which has been used to maximize the likelihood function and discuss the asymptotic properties of the maximum likelihood estimates (MLE). Then the performances of the models are discussed and compared in Section 4 using simulations. The data used in this paper are also introduced at the beginning of this section. At last, we make a synthesis of the obtained results and we give some perspectives in Section 5. The R codes based on the package HMSAR and the data which have been used in this study are available as supplementary material (see Appendix A).

2. Models

2.1. Non-homogeneous Markov-switching autoregressive models

Let $X_t \in \{1, \ldots, M\}$ represent the latent weather type and Y_t denote the observed wind conditions at time t. In this paper $\{Y_t\}$ will represent successively the bivariate process of Cartesian coordinates of wind in Section 2.3, the wind