Non linear statistical models to improve wind power forecasts

M. Pilar Muñoz[1], Josep A. Sánchez-Espigares[1], M. Dolores Márquez[2]

[1]Barcelona-Tech. Universitat Politècnica de Catalunya, Barcelona, Spain
[2]Universitat Autònoma de Barcelona, Bellaterra, Spain
Corresponding autor: M. Pilar Muñoz, e-mail: pilar.munyoz@upc.edu

Given the need to control energy financial markets, forecasting wind power generation has become an important subject of research. One of the most common and important components of wind power forecasting models is wind behavior. Usually, data show different behaviors depending on wind direction. This is the case of the data from the Gibraltar Strait, which is presented in this work; it shows two clearly different wind directions. In order to better fit the wind direction, deterministic (Threshold) and stochastic (Markov Switching) models for circular variables with a Von-Mises conditional distribution are implemented and tested with this data. Results show better agreement between observed and modeled data when considering a stochastic mechanism that governs the change between regimes. The study includes a code implemented in the R package MSwM that deals with the proposed models (https://cran.r-project.org/web/packages/MSwM/index.html). This improved prediction of wind direction enhances the capability of forecasting energy generation.

Keywords: Circular variables, forecasting, Von-Mises distribution, wind direction, wind power

1. Introduction

Knowing and forecasting wind direction is very important in a lot of fields that cover situations related with meteorology, air pollution, climate, migratory patterns of birds, ozone concentrations and wind power, among many other applications. The main problem is that it is not easy to predict the wind because of its characteristics. All of this has generated the need for obtaining statistical models that allow the analysis and accurate prediction of wind components, in particular wind direction, which is very hard to obtain and predict. Methodologically, wind direction is considered to be a circular variable, and one of the most recognized probability distributions for analyzing circular variables is the von Mises distribution, also known as the circular normal distribution, originally proposed by von Mises (1918). In fact, classical models can be extended to deal with circular variables by using their specific distribution instead of the Gaussian assumption for linear variables. Holzman et al. (2006) introduced a hidden Markov model for bivariate circular time series, generating the new hidden Markov models: von Mises-HMM; wrapped normal-HMM; and wrapped Cauchy-HMM, in which von Mises, wrapped normal and wrapped Cauchy are circular distributions. Also, these models are applied to wind direction, among other examples. Carta et al. (2008) model wind speed and direction by using mixtures of von Mises distributions. The use of mixtures gives flexibility to the model and, therefore, forecasts have to be more accurate. Artes and Toloi (2009) also apply an AR circular model with von Mises random errors in estimating wind direction. Recently, Kato (2010) suggests a Markov model for circular data, applying it to estimating wind direction. It differs from the AR model proposed by Fisher and Lee (1994), in the sense that the angle error follows a wrapped Cauchy distribution and not a von Mises, as in the Fisher and Lee (1994) proposal. Erdem and Shi (2011a) proposed ARIMA models for forecasting wind speed and direction together while Erdem and Shi (2011b) fit a bivariate statistical model, also for wind speed and direction, with the intention of increasing the effectiveness of wind power generation. Their approach is based on the combination of several distributions of probability for wind speed and direction. Finally, a novel idea is to estimate the circular variable, like direction, via Bayesian techniques. Nuñez-Antonio et al. (2011) fit a Bayesian regression model for estimating wind direction from other covariates such as humidity and temperature. Modlin et al. (2012) estimate spatial wind direction under a Bayesian point of view also. All this makes us think that there is room for promising research in this area.

Data used in this study is provided by the HIRLAM model and we have concentrated on 36º 9'N 5º 42'W. This geographical point has been selected because it is very close to a wind generation farm. Knowing the behavior of the direction at this point can be very useful in forecasting the wind generation on the farm. This point is located in the south of Spain, close to the Strait of Gibraltar. The wind direction in this zone is not easy to estimate because of the geographical situation of this area:

- There is a topographic barrier formed by two mountain lines: One along the coastline of the Iberian Peninsula and another formed by the foothills of the Atlas Mountains in Morocco. These topographic features cause the air to flow like a tube in an east-west direction.
- The temperature differences on the two sides of the strait can produce changes in wind direction. The Mediterranean is generally 2-5 degrees warmer than the Atlantic, which causes a difference in pressure that produces these changes.
- Sea breezes are produced by the thermal contrast temperatures because of the temperature differences between the sea and the coast.

This means that the wind direction in the Strait of Gibraltar area depends mainly on the surface pressure gradient between the two sides of the strait.

The period analyzed covers January 1st 2009 to December 31st 2010; the frequency of data is every 3 hours (temporal reference system is UTC); it starts at 00.00 (daily analysis) and forecasting is at 3.00, 6.00, 9.00, 12.00, 15.00, 18.00 and 21.00. The wind direction is recorded every day, eight times a day, so we have a time series for this variable. As the whole data set expands over two years, it will permit detecting changes due to meteorological seasons and will also permit analysis of diurnal and nocturnal patterns. The meteorological direction of the wind (angle between the unit vector and geographical north) is calculated as tangent arc (v_x,v_y), where v_x and v_y are the horizontal and vertical components provided by the HIRLAM model. Information about temperature and atmospheric pressure is also available.

2. Methodology

In the context of circular data like wind direction, several approaches have been developed in order to forecast time series. ARMA-type models can be fitted by considering a conditional distribution for circular data like von Mises, wrapped normal or Cauchy distributions. But sometimes the series has a non-linear behavior, like when considering a mixture of distributions. In this work, we compare two different approaches for dealing with nonlinearity when fitting Circular Autoregressive models. The mechanism behind the two techniques is similar: the series changes between different alternative models according to some condition. The condition that governs the change of regime can be a threshold defined by a deterministic mechanism (Logistic Mixture Autoregressive Model, LMARX) or a non-observable Markovian Process (Markov Switching Model, MSM). The two models are presented and algorithms for fitting the models by using the maximum likelihood criterion have been developed. In the case of the MSM model for circular data, the EM algorithm is implemented and applied to fit the model as a part of an R package for dealing with MSM for different types of data.

Let us consider a circular random variable, Θ, that takes the values $\mathbb{R}/2\pi\mathbb{Z}$. Fisher and Lee (1944) developed a method for transforming circular processes into linear processes by means of a link function. Let g be the link function, a strictly monotonic (therefore invertible) function that maps the real support into the open interval ($-\pi, \pi$). So, if X_t is a stationary process and $\mu \in [0,2\pi]$, then $\Theta = g(X) + \mu \bmod 2\pi$ would be called the linked circular process and will also be stationary. Conversely, for a stationary circular process Θ, the transformation by using the link function $X = g^{-1}(\Theta - \mu)$ is a stationary linear process. Some examples of link function in this context are: $g(x) = 2\tan^{-1}(x - \mu)$ and $g(x) = 2\pi(\Phi(x) - 0.5) + \mu$. For this approach to be useful, it is convenient for the series to have low dispersion. The von Mises distribution was also introduced by Fisher and Lee (1994) as a distribution for circular time series with an AR structure in a similar way as the linear AR(p) process. The expression of a von Mises density with parameters mu and kappa is given by:

$$f(\theta) = \frac{e^{\kappa\cos(x-\mu)}}{2\pi I_o(\kappa)}$$

where $I_0(\kappa)$ is the modified Bessel function of order 0.

A process (Θ) is called a circular autoregressive of order p, CAR(p), with link function g if Θ_t, given $\Theta_{t-1} = \theta_{t-1}, \dots, \Theta_1 = \theta_1$ is distributed as a VM(μ, κ) for t=1…p and VM(μ_t, κ) for t>p, where:

$$\mu_t \coloneqq \mu + g(\phi_1 g^{-1}(\theta_{t-1} - \mu) + \cdots + \phi_p g^{-1}(\theta_{t-p} - \mu))$$

Sometimes there are structural changes in the series not necessarily related to the presence of outlier data. Usually, some partition of the time series can be found in which the behavior in each subset is homogeneous and can be described by a stationary process. However, the series as a whole is a mixture of different sub-models. The definition of this partition can be made in several ways. In this work, we consider two possibilities: the LMARX model and MSM for circular data.

LMARX model

This model is a kind of regime-switching model where the state is determined by the value of some available observations from the series and other exogenous variables. If the threshold is applied to the series itself, then the model is called SETAR (Self-exciting Threshold Autoregressive). If some exogenous variables are used instead, the model is the TARSO model (Open-loop Threshold autoregressive), see Tong (1990). Another extension is to consider a logistic function with both the series itself and a set of exogenous variables as explanatory variables to determine which model of the mixture is to be applied. This kind of model is known as LMARX (Logistic Mixture Autoregressive), see Wong and Li (2001). The regimes are settled by a certain number of thresholds, $l_0,\ldots,l_m$ that divide the space of the [0,1] interval into m subsets. The location of the value of the response variable from a logistic model in one of these subsets determines which model governs the dynamic of the series at that moment. Specification of the model m-regimes LMARX(p) --for circular data with the state defined by the logistic model and one exogenous variable-- is as follows:

$$\theta_t|\theta_{t-1},\ldots,\theta_1 \sim VM\left(\mu_t,\kappa^{(S_t)}\right)$$

$$g^{-1}(\mu_t-\mu^{(S_t)})=\sum_{i=1}^{p_1}\phi_i^{(S_t)}g^{-1}\left(\theta_{t-i}-\mu^{(S_{t-i})}\right)+\sum_{j=1}^{q_1}\delta_j^{(S_t)}X_{t-j}$$

$$log\left(\frac{\pi_t}{1-\pi_t}\right)=\beta_0+\sum_{i=1}^{p_2}\beta_i g^{-1}(\theta_{t-i})+\sum_{j=1}^{q_2}\beta_{p_2+j}X_{t-j}$$

$$S_t=\begin{cases}1 & 0=l_0\le\pi_t\le l_1\\ \ldots & \ldots\\ m & l_{m-1}\le\pi_t\le l_m=1\end{cases}$$

There is a set of parameters $(\mu^{(S_t)},\phi_1^{(S_t)},\ldots,\phi_{p_1}^{(S_t)},\delta_0^{(S_t)},\ldots,\delta_{q_1}^{(S_t)},\kappa^{(S_t)})$ for each sub-model. Another set of coefficients $(\beta_0,\ldots,\beta_{p_2+q_2})$ is related to the linear predictor of the logistic function determining the active sub-model. There are p_2 parameters related to previous observations of the circular variable and q_2 parameters for the exogenous variable. Algorithms for estimating this model in the linear case can be found in Wong and Li (2001). In our case, we include only one lag for the wind direction and also for the two exogenous variables (temperature and atmospheric pressure) in the linear predictor, and a purely autoregressive model for the circular variable in all the sub-models from the mixture. We have implemented an adaptation of the EM algorithm to deal with the likelihood of the circular variable.

MSA for circular data

A Markov-Switching Autoregressive model for circular data includes an unobserved process that determines the changes in regime. Usually, this process is Markovian of the first order, and is parameterized by a transition probability matrix. Each element of this matrix represents the probability of change between the correspondent states. For each regime there is a current AR(p) model that reflects how the present observations are linearly related with past observations of the linked scale. MSA models with an underlying Markov process $(S_t)_{t\ge1}$ can be expressed as follows:

$$\theta_t|\theta_{t-1},\ldots,\theta_1,S_t \sim VM\left(\mu_t,\kappa^{(S_t)}\right)$$

$$g^{-1}(\mu_t-\mu^{(S_t)})=\sum_{i=1}^{p_1}\phi_i^{(S_t)}g^{-1}\left(\theta_{t-i}-\mu^{(S_{t-i})}\right)+\sum_{j=1}^{q_1}\delta_j^{(S_t)}X_{t-j}$$

$$P(S_{t-1} = i, S_t = j) = p_{ij} \quad i = 1 \ldots m, j = 1 \ldots m$$

The process defined by $(S_t)_{t\geq 1}$is an irreducible and homogeneous Markov chain with transition probability matrix $\Pi = (p_{ij})_{i,j=1,\ldots,m}$. Again, the model contains a set of sub-models that depend on the current regime. The expression of these sub-models is the same as the one used in the LMARX model. The parameters for the MSA model include the probability matrix Π and the coefficients for the model under each state. Estimation of parameters under the maximum likelihood criterion has to take into account the existence of the underlying process S_t and, due to the fact that this is not observable, has to build the marginal likelihood for the circular data. The expression of this marginal likelihood is the joint density of the data. After integration, the unobserved process is:

$$f(\theta_1, \ldots, \theta_T) = \sum_{i_1=1}^{m} \ldots \sum_{i_T=1}^{m} P(\theta_1, \ldots, \theta_T | S_1 = i_1, \ldots, S_T = i_T) P(S_1 = i_1, \ldots, S_T = i_T)$$

In this case, by using the Markovian property of the S_t process and the conditional distribution of the observations, given the previous ones and the current state, the joint density can be re-expressed as follows:

$$f(\theta_1, \ldots, \theta_T) = \sum_{i_1=1}^{m} \ldots \sum_{i_T=1}^{m} \prod_{t=1}^{p} P(\theta_t | S_t = i_t) P(S_t = i_t) \prod_{t=p+1}^{T} P(\theta_t | \theta_{t-1}, \ldots \theta_1, S_t = i_t) P(S_{t-1} = i_{t-1}, S_t = i_t)$$

The first p observations are included with an unconditional density, similar to the classical AR(p) processes. For the rest of the observations, the conditional distribution is von Mises and the probabilities for the transition between states correspond to the elements of the transition matrix Π. Therefore, the expression for the likelihood can be expressed in terms of the set of parameters and a nonlinear optimization algorithm can be applied to find their estimates. But the dimension of the parameter space can be huge when considering a high number of states and the estimation procedure can become unstable. In these cases, an alternative based on the EM algorithm can be used to avoid this problem.

3. Conclusions

In the MSA model, the changes between states have a stochastic pattern instead of the deterministic criteria defined in the LMARX model. So, the main difference between the two approaches is that in the LMARX model the state is determined at time *t-1*, whereas in the MSA model, the state is assigned to time t after obtaining the value for S_t.

Short-term forecasting with a LMARX model implies the use of a specific sub-model, depending on the last observation value. On the other hand, forecasts from an MSA model come from a linear combination of predictions from each sub-model.

References

Artes, R. and Toloi, C. M. C. (2009). An Autoregressive model for time series of circular data. *Communication in Statistics- Theory and Methods*. 39, 186-194.

Carta, J.A., Bueno, C. and Ramírez, P. (2008). Statistical modelling of directional wind speeds using mixtures of von Mises distribution: Case study. *Energy Conversion & Management*, 49, 897-907.

Erdem, E. and Shi, J. (2011a). ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy* , 88(4): 1405-1414

Erdem, E. and Shi, J. (2011b) Comparison of bivariate distribution construction approaches for analysing wind speed and direction data. *Wind Energy*, 14(1): 27–41.

Fisher, N. I. and Lee, A. J. (1994) "Time Series Analysis of Circular Data". Journal of the Royal Statistical Society. Series B (Methodological), 56(2), 327-339

Holzmann, H., Munk, A., Suster, M. and Zucchini W. (2006) Hidden Markov models for circular and linear-circular time series, *Environmental and Ecological Statistics*, 13(3), 325-347.

Kato, S. (2010). A Markov process for circular data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72 (5): 655-672.

Modlin, D., Fuentes, M. and Reich, B. (2012), Circular conditional autoregressive modeling of vector fields. *Environmetrics*, 23(1): 46–53.

Nuñez-Antonio, G., Gutierrez-Peña, E. and Escarela, G. (2011). A Bayesian regression model for circular data based on the projected normal distribution. *Statistical Modelling*, 11: 185-201.

Tong, H. (1990) *Non-linear time series: a dynamic system approach*. Oxford. University Press.

Von Mises, R. (1918). UJber die 'Ganzzahligheit' der Atomgewichte und vermandte Fragen. *Phys. Z.* 19, 490-500.

Wong, C. S. and Li, W. K. (2001) "On logistic mixture autoregressive model", *Biometrika*, 88(3), 833-846