Analysis and modeling of wind directions time series

Salvatore Basile, Riccardo Burlon, Davide Gurrera
Dipartimento Energia, Ingegneria dell'Informazione e Modelli Matematici
Università degli Studi di Palermo
Viale delle Scienze, Edificio 6, I-90128 Palermo, Italy
salvatore.basile@unipa.it

***Abstract* — This work aims at studying some aspects of wind directions in Italy and supplying appropriate models. A comparison is presented between independent mixture and Hidden Markov models, which seem to be appropriate as far as the series we studied.**

Keywords: directional data, wind direction time series

I. INTRODUCTION

The investigation of wind speed and direction and the development of predicting models are extremely important in evaluating the characteristics of the wind regime of a particular region. Wind influences the rate at which pollutants are dispersed or diffused and the propagation of forest fires. It may prove dangerous for many human activities, such as air freight, and its local properties should always be analyzed prior to the construction of structure like long suspension bridges or skyscrapers. In addition, it provides green energy and the production of electricity by wind energy is fast-increasing.

Although the motion of air masses is effectively described and daily predicted by hydrodynamic models, this deterministic approach may sometimes prove insufficient because of the great complexity of the underlying dynamics. In order to fill this void, one can turn to statistical methods and the models thus obtained may provide an effective support and in certain cases a more appropriate description [1-7].

Wind direction has been considered as an important aspect in the evaluation of wind energy because it can give more information about wind speed in order to exploit better its potential energy. It is well-known that in order to maximize the amount of captured energy, we need to analyze and predict the wind direction and consequently the parameters of the wind turbine [4].

However, despite of the progresses made, the investigation of the statistical properties of the wind is still far from being complete.

The present work aims at developing appropriate statistical models for wind direction time series. The empirical data have been collected in Sicily, Italy. The area is characterized by a complex morphology, but the analysis and the obtained results may prove useful for other areas as well.

More specifically, a comparison between independent mixtures and Markov models is presented. Despite the serial dependence of the data would suggest the choice of time-dependent models, the independent model approach yields the more accurate description for all the investigated time series, with a smaller number of parameters as a fringe benefit.

The paper is organized as follows. Section II presents the models used in the statistical analysis. Section III gives some details on the data sets. Results are presented with details in Section IV. Conclusions and final remarks are in Section V.

II. MODELS STRUCTURE

A. *General considerations*

Directional data are rather special, but they do arise in many different contexts. The difficulty in their statistical analysis results from the difference of topology between the circle and the straight line. Directions near to the opposite end-points are close neighbors in a metric which respects the topology of the circle, but maximally distant in a linear metric. Thus, many standard statistical techniques are inappropriate for modeling circular data and different solutions have been proposed for distributional models and stochastic processes on the circle [8]. The von Mises distribution is a natural first choice as a model for unimodal continuous distribution [9]. Its probability density function (p.d.f.) with parameters $\mu \in (-\pi, \pi]$ (location) and $\kappa > 0$ (concentration) is:

$$f(\theta) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)} \tag{1}$$

for $\theta \in (-\pi, \pi]$ and with I_0 denoting the modified Bessel function of the first kind of order zero, given by

$$I_0(z) = \frac{1}{\pi}\int_0^{\pi} e^{z\cos(\theta)} d\theta = \sum_{k=0}^{\infty} \frac{1}{(k!)^2}\left(\frac{z}{2}\right)^{2k} \tag{2}$$

In the following analysis, we make use of (1) for describing time series of directional data by means of two different models whose details are presented below [10].

B. *Independent mixture models (IMM)*

One method of dealing with data characterized by a bimodal or multimodal distribution, such as those this work is dealing with, is to use a mixture model.

978-1-4799-1464-7/13/$31.00 ©2013 IEEE

An independent mixture distribution consists of a finite number, say m, of component distributions, either discrete or continuous, and a "parameter process" which selects from these components. If the parameter process is a series of independent random variables, the values generated by the model will be independent too. More specifically, let $\delta_1,...,\delta_m$ be the probabilities, summing to 1, assigned to the different components whose p.d.f.s are $p_1,...,p_m$. If Θ is the random variable which is supposed to have the mixture distribution, the p.d.f. is given by

$$p(\theta)=\sum_{i=1}^{m}\delta_i p_i(\theta,\boldsymbol{\eta}) \quad (3)$$

The estimation of the m components of the parameters vector $\boldsymbol{\eta}$ of the component distributions as well as of the mixing parameters can be obtained by numerical maximization of the likelihood

$$L(\eta_1,...,\eta_m;\delta_1,...,\delta_m \mid \theta_1,...,\theta_N)=\prod_{j=1}^{N}\sum_{i=1}^{m}\delta_i p_i(\boldsymbol{\theta},\boldsymbol{\eta}) \quad (4)$$

where $\theta_1,...,\theta_N$ are the observations.

C. Hidden Markov Models (HMM)

Adjacent observations in a time series usually exhibit a mutual dependence whose nature is of considerable practical interest. One natural way to allow for serial dependence, while employing a mixture model of the kind discussed above, is to assume that the parameter process is a Markov chain, the obtained model being known as a Hidden Markov Model. Indicating by $\{C_t : t=1,2,...\}$ the unobserved parameter process satisfying the Markov property and by $\{\Theta_t : t=1,2,...\}$.the time-dependent process, the distribution of Θ_t is given by:

$$p(\Theta_t=\theta)=\sum_{i=1}^{m}\Pr(C_t=i)p(\Theta_t=\theta \mid C_t=i)= \sum_{i=1}^{m}\Pr(C_t=i)p_i(\theta,\boldsymbol{\eta}) \quad (5)$$

m being the number of component distributions of the end model (and therefore the number of states of the Markov process) and $p_1,...,p_m$ their p.d.f.s. By making use of the Markov property, (5) can be rewritten as

$$p(\Theta_t=\theta)=\boldsymbol{\delta}\boldsymbol{\Gamma}^{t-1}\mathbf{p}(\theta)\mathbf{1}' \quad (6)$$

where $\boldsymbol{\delta}=\left(\Pr(C_1=1),\Pr(C_1=2),...,\Pr(C_1=m)\right)$ is the initial distribution of the Markov chain, $\boldsymbol{\Gamma}$ is the one-step transition probability matrix (with row sums equal to 1), $\mathbf{p}(\theta)$ denotes the diagonal matrix with i-th diagonal element $p_i(\theta)$ and $\mathbf{1}'$ is a column vector with all elements equal to 1.

Also in this case the estimation of the unknown parameters in (6) can be performed by numerical maximization of the likelihood, now reading:

$$L(\boldsymbol{\delta};\boldsymbol{\Gamma};\boldsymbol{\eta} \mid \theta_1,...,\theta_N)=\boldsymbol{\delta}\mathbf{p}(\theta_1)\boldsymbol{\Gamma}\boldsymbol{\delta}\mathbf{p}(\theta_2)\boldsymbol{\Gamma}\mathbf{p}(\theta_3).\boldsymbol{\Gamma}\mathbf{p}(\theta_N)\mathbf{1}' \quad (7)$$

If θ_t is a missing observation, then the diagonal matrix $\mathbf{p}(\theta_t)$ is replaced by the identity matrix.

III. Data Description

Wind speed and direction, as well other meteorological data are collected in Sicily, Italy, by the Servizio Informativo Agrometeorologico Siciliano (SIAS), on a network of 29 stations. Four years of data (2003−2006) were made available and this paper attempts to model the wind direction data, consisting of hourly averaged wind directions at 10 meters above ground level. The heterogeneity of local conditions for the different sites provides us with samples suitable for investigating the effectiveness of the two above described approaches to wind direction modeling. There are 35065 observations for each site, although some of them are "not available" because they are either missing or calms. For the sake of brevity, we shall confine ourselves to report detailed results for four sites, selected on the basis of their representativeness. Palermo, near the Tyrrhenian Sea and partly surrounded by mountains, lies almost at sea level. Its position determines a complex frequency distribution of the wind directions, affected by a land and sea breeze cycle (which dominates the wind pattern for most of the year) and characterized by 3 or 4 modes. Monreale, looking down on Palermo, is located at about 300 m above sea level and its exposed position shapes a fair bimodal distribution. At an altitude of 50 meters, Pachino is on the opposite side of Sicily, in the South-East corner of the island, facing the Mediterranean Sea and the Ionian Sea; also in this case, only two main wind directions are dominant. Finally, Leni station is located at the altitude of 200 meters on the small island of Salina, North-East of Sicily. Its winds mostly blow from two directions, but the distribution is not as clear-cut as those of Monreale and Pachino.

IV. Analysis and Results

The first studies on hourly averaged wind data forecasting were based upon a Monte Carlo simulation, assuming a given probability distribution. The predictions obtained using this method showed the weakness of not considering the correlation between consecutive observations, which is characteristic of such data.

In the following analysis we want to evaluate the effectiveness of a stochastic model including autocorrelation, versus a simpler independent mixture model. The comparison is made on the basis of the yielded probability distributions. The structure of these models, both using the von Mises p.d.f., has been described in Section II.

The first reported results concern the wind direction data sets for Palermo and Pachino. In Figs. 1 and 2 we show the polar plots of the empirical p.d.f.s for Palermo and Pachino stations along with their modeling by a 3 distribution von Mises HMM and a three component von Mises IMM. In these figures, as well in the next ones, the empirical probability densities are computed dividing the relative frequencies by the bin size.

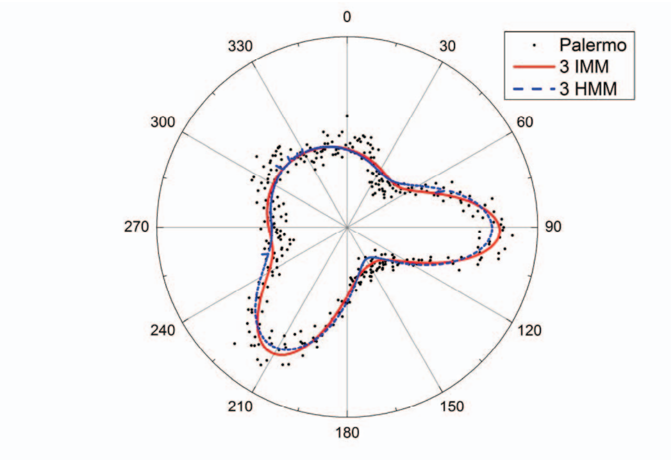

Figure 1. Empirical probability densities of wind direction for Palermo station (black dots) along with two models, namely a 3 state von Mises-HMM (blue dashed line) and a 3 component von Mises-IMM (red line).

Something not obvious is that in both cases the HMM performs worse, as indicated by the values of the root mean square error (RMSE) shown in Table I. Although here we are not showing, for sake of brevity, the graphical results with only 2 HMM/IMM, the significant amelioration when using 3 HMM/IMM is clearly evident from Table I. This suggests that it would be advisable to make use of a 3 component model even in the case of a bimodal distribution.

Fig. 3 shows empirical data and modeling for the Monreale station, with a 3 state HMM and a 3 component IMM.

The above results keep these main features for the all set of the investigated data and can be summarized as follows.

An IMM always performs better than the corresponding opponent with the same number of component distributions, although in some cases differences are very small. This has been verified up to 3 components, since a model with as many as 25 parameters, as those required by a 4 state HMM, would be quite meaningless. Secondly, a 3 component von Mises IMM was found to yield a suitable description for most of the analyzed series, although in some cases a considerable refinement may be achieved by adding an extra component. This is shown in Fig. 4 for Leni station data (see also Table I).

V. CONCLUSIONS

According to the results, this work provides a reliable model for dealing with wind direction time series, namely a von Mises independent mixture model with 3 or, at most, 4 components. Since a first-order Markov chain does not allow for long serial dependence, the cause of the failure of the Hidden Markov Model (HMM) should be traced back to its parameter process. More specifically, the findings suggest that any model for the hourly wind data should allow for higher-order dependence and, in particular, should be able to account for the daily period typical of such data [2, 7]. Daily and annual effects could be added to the HMM, but that would require an increasing in the number of the related parameters. Therefore, as far as the intended use of the pursued model is to provide an adequate probability model, the results of this study suggest employing an independent model for describing the data under investigation.

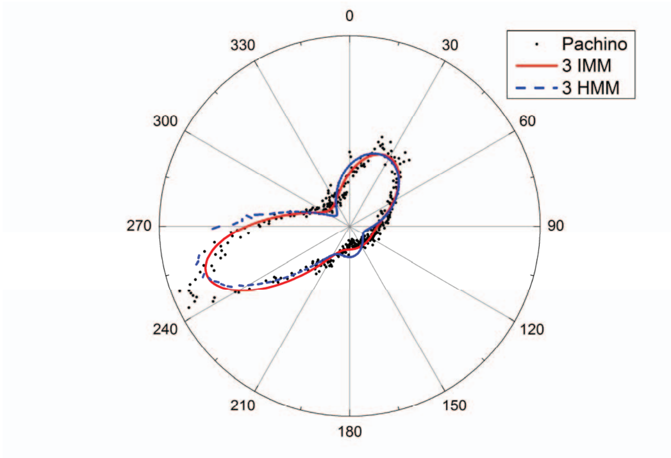

Figure 2. The same as in Fig. 1, for the Pachino station.

TABLE I. SUMMARY OF THE MODELS USED FOR THE FOUR STATIONS, ALONG WITH THE NUMBER OF INVOLVED PARAMETERS AND THE ROOT MEAN SQUARE ERRORS.

	Model	*m*	*RMSE*
Palermo	2-comp. IMM	5	0.0430
	2-state HMM	7	0.0488
	3-comp. IMM	8	0.0196
	3-state HMM	14	0.0210
Pachino	2-comp. IMM	5	0.0197
	2-state HMM	7	0.0562
	3-comp. IMM	8	0.0366
	3-state HMM	14	0.0489
Monreale	3-comp. IMM	8	0.0326
	3-state HMM	14	0.0408
Leni	3-comp. IMM	8	0.0564
	3-state HMM	14	0.0561
	4-comp. IMM	11	0.0344

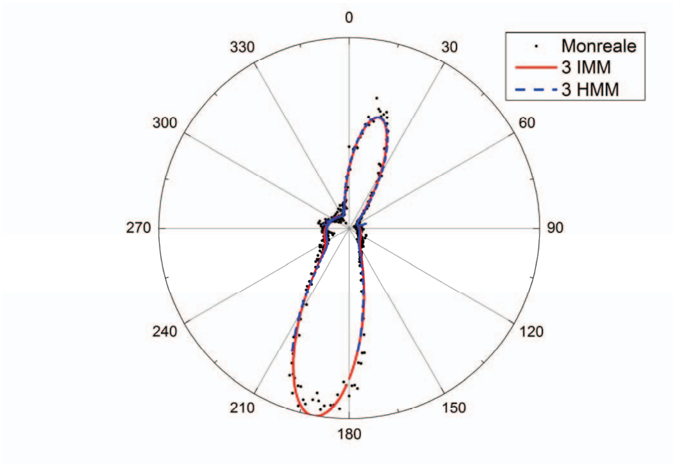

Figure 3. The same as in Fig. 1, for Monreale station.

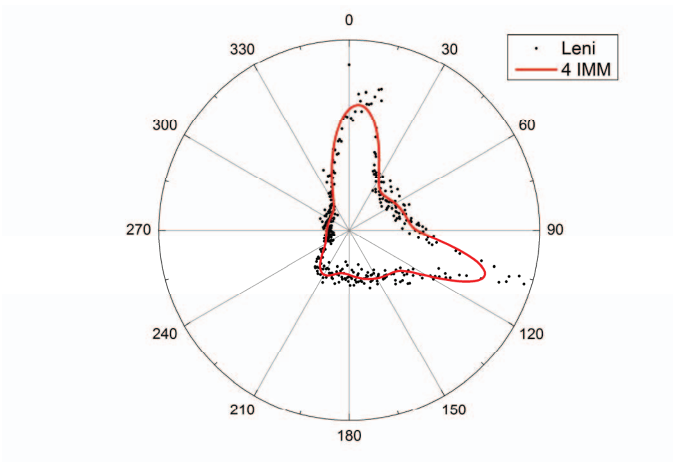

Figure 4. Empirical probability densities of wind direction for Leni station (black dots) along with results from the 4 component von Mises-IMM (red line).

ACKNOWLEDGMENTS

This research has been partially supported by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca. Authors acknowledge the collaboration of the Servizio Informativo Agrometeorologico Siciliano for providing the data.

REFERENCES

[1] D. S. Wilks, Statistical Methods in the Atmospheric Sciences, 2nd ed., Elsevier, 2006.

[2] J. Breckling, The Analysis of Directional Time Series: Applications to Wind Speed and Direction, Lecture Notes in Statistics, 61, Springer-Verlag, 1989.

[3] H. Kantz, D. Holstein, M. Ragwitz, and N. K. Vitanov, "Markov chain model for turbulent wind speed data", Physica A, vol. 342, pp. 315–321, 2004.

[4] J. A. Carta, C. Bueno, and P. Ramírez, "Statistical modelling of directional wind speeds using mixtures of von Mises distributions: Case study", Energ. Convers. Manage., vol. 49, pp. 897–907, May 2008.

[5] S. Bivona, G. Bonanno, R. Burlon, D. Gurrera, and C. Leone, "Taxonomy of correlations of wind velocity – an application to the Sicilian area", Physica A, vol. 387, pp. 5910–5915, 2008.

[6] S. Bivona, G. Bonanno, R. Burlon, D. Gurrera, and C. Leone, "Univariate and multivariate properties of wind veocity time series", J. Stat. Mech., P02026, 2009.

[7] S. Bivona, G. Bonanno, R. Burlon, D. Gurrera, and C. Leone, "Stochastic models for wind speed time series: a case study", Acta Phys. Pol. B, vol. 41, pp. 1083–1092, May 2010.

[8] S. Coles, "Inference for circular distributions and processes", Stat. Comput., vol. 8, pp. 105–113, June 1998.

[9] N. I. Fisher, Statistical Analysis of Circular Data, Cambridge University Press, 1993.

[10] W. Zucchini and I. L. MacDonald, Hidden Markov Models for Time Series: An Introduction Using R, Chapman and Hall/CRC, 2009.