



HAL
open science

Statistical bivariate modelling of wind using first-order Markov chain and Weibull distribution

Fatiha Youcef-Ettoumi, Henri Sauvageot, A.-E.-H. Adane

► **To cite this version:**

Fatiha Youcef-Ettoumi, Henri Sauvageot, A.-E.-H. Adane. Statistical bivariate modelling of wind using first-order Markov chain and Weibull distribution. *Renewable Energy*, 2003, 28, pp.1787-1802. 10.1016/S0960-1481(03)00019-3 . hal-00137536

HAL Id: hal-00137536

<https://hal.science/hal-00137536>

Submitted on 21 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Statistical bivariate modelling of wind using first-order Markov chain and Weibull distribution

F. Youcef Ettoumi ^a, H. Sauvageot ^{b,*}, A.-E.-H. Adane ^a

^a *Université de Sciences et Technologie Houari Boumediène (U.S.T.H.B.), Faculté de Génie Electrique, Laboratoire Traitement d'Images et Rayonnement, BP. 32 El Alia, Bab Ezzouar, Alger, Algeria*

^b *Université Paul Sabatier, Observatoire Midi Pyrénées, Laboratoire d'Aérodynamique, 14 avenue Edouard Belin, 31400, Toulouse, France*

This paper studies the statistical features of the wind at Oran (Algeria). The data used are the wind speed and wind direction measurements collected every 3 h at the meteorological station of Es Senia (Oran), during the 1982/92 period. The eight directions of the compass card have been considered to build the frequency distribution of the wind speed for each month of the year and each direction. The three-hourly wind data have been modelled by means of Markov chains. First-order nine-state Markov chains are found to fit well the wind direction data, whereas the related wind speed data are well fitted by first-order three-state Markov chains. The Weibull probability distribution function has also been considered and found to fit the monthly frequency distributions of wind speed measurements. Two methods of wind data retrieval are thus made available. In fact, two models of chronological bi-series are obtained describing wind speed and wind direction.

Keywords: Wind modelling; Stochastic Processes; Markov chain; Weibull distribution; Long-term probability

* Corresponding author. Tel.: +33-5-62-40-61-02; fax: +33-5-62-40-61-01.
E-mail address: sauh@aero.obs-mip.fr (H. Sauvageot).

1. Introduction

Wind contributes significantly to the thermal exchange processes between earth surface and atmosphere. It favours water evaporation at ground surface and carries various particles in suspension in the air. It also represents an appreciable source of energy. To account for the effect of wind erosion at the ground, knowledge of wind speed and direction is important. The wind can be considered as a short-time scale, random, non-stationary process. It is thus of great interest to assess its climato-statistical features. These characteristics can be studied in space and time by using the wind speed and direction measurements collected near the ground in various meteorological stations.

To describe the statistical features of the wind speed at a given location, the usual method is to fit the experimental wind data with probability distributions [1–6]. In previous studies [7,8], wind speed and direction data have been mapped for Algeria on a yearly basis. For a given location, the Weibull probability distribution function was shown to fit the annual frequency distribution of the three-hourly wind speed. The Weibull distributions obtained for the various meteorological stations of Algeria were classified by considering either the Ward method or the K-means algorithm [9–11]. In more than 50% of the meteorological stations, mean wind speed was found to be lower than 3 m/s. For about 30% of the stations, the corresponding wind speed values ranged between 3 and 5 m/s. In the rest of the stations, wind speed was higher than 5 m/s. Thanks to this classification, the wind energy potential in Algeria can be estimated with a good accuracy.

Another way of estimating the wind statistical features consists in using autoregressive processes or Markov chains [12]. This kind of modelling has the advantage of accounting for the time variations of wind speed and direction. Since the latter can easily be expressed in term of discrete states, it can be described efficiently by Markov chains or autoregressive models.

The goal of this paper is to present the results of the modelling, on a monthly basis, of the three-hourly wind speed and direction data recorded at the site of Es Senia (Oran) during the 1982/92 period using both Markov chains and Weibull probability distributions.

2. Data

The station of Es Senia is situated in Algeria at $0^{\circ} 37'$ W, $35^{\circ} 38'$ N and 90 m of altitude. At this station, the wind is measured by an anemometer and weathercock label, placed at 10 m above the ground. Wind direction angles are measured with respect to the North with an accuracy of 0.1° .

Wind speed and direction were averaged over three-hour periods. These measurements are available from 1982-1992. The resulting sequences of three-hourly wind data consist of fast random fluctuations superimposed on trend variations. The random fluctuations and trend variations are caused by atmospheric turbulence and geostrophical circulation respectively. Since the trend varies slowly, a month is a suf-

ficiently short period to represent the three-hour wind variations by a stationary random variable [13]. Hence, the obtained database has been divided into 12 subsets, corresponding to the 12 months of the year. A given monthly subset brings together all the 1982/92 three-hour wind data collected during the same months of the different years.

In meteorological applications dealing with wind measurements, the circle described by the weathercock in the horizontal plane is usually divided into eight sectors having the same spread of 45° . Every wind direction is assigned to one of the eight sectors by computing the sector corresponding to the observed direction.

3. Markov modelling

3.1. Physical aspects

The wind can blow in the same direction for a relatively long time. Then it suddenly changes from one direction to another, continues to blow in the new direction and so on. The resulting sequence of wind direction data is thus made of time series of stable states and abrupt changes between them. These states depend on the wind strength and can take all the weathercock directions. Wind speed and direction are both erratic and persistent. This means that wind variations in time can be described by first-order Markov chains [14].

According to the division into eight sectors presented in section 2, wind direction measurements, expressed in degree, were classified following nine different states. These states were labelled by numbers 0–8 corresponding to no wind, north (N), north-east (NE), east (E), south-east (SE), south (S), south-west (SW), west (W) and north-west wind (NW) respectively. The nine states thus defined are those to be used for modelling wind direction data by first-order, nine-state Markov chains.

When analysing the experimental data, three states of wind speed can be distinguished, namely weak, mean or strong. The numbers 0, 1 and 2 labels these states respectively. They can be defined by considering the relative frequency distributions of wind speed obtained in a given site, for each month of the year and each of the eight sectors. Three typical states can be identified from the histograms of three-hourly wind speed, which respectively correspond to: $0 \leq V(i) < 3$ m/s (weak wind), $3 \text{ m/s} \leq V(i) < 8$ m/s (mean wind) and $V(i) \geq 8$ m/s (strong wind). Once these states are known, wind speed data can be modelled by first-order, three-state Markov chains.

3.2. First-order Markov model

A series of wind data collected every three hours, is considered, the number of intervals of 3 h being k at time t . X_k is a random variable that may take one of the values 0, 1, ..., n , ..., $N-1$ at time t [15,16]. These values related to the N possible states of the phenomenon under study, are respectively denoted $\Phi_0, \dots, \Phi_n, \dots$,

Φ_{N-1} . Since the probability of observing Φ_n state after 3 k h, is $P(X_k = n)$, a probability vector can be defined. It is written as:

$$Q = [P(X_k = 0). \dots P(X_k = n) \dots P(X_k = N-1)], \quad (1)$$

with

$$\sum_n P(X_k = n) = 1. \quad (2)$$

For the first-order Markov process, the probability of having a transition between the states occurring for two successive observations, is:

$$P_{ij} = P(X_{k+1} = j / X_k = i), \quad (3)$$

with i or $j = 0, \dots, n, \dots, N-1$.

Considering all the possible transitions between the Φ_n states, the Markov transition matrix:

$$M = [P_{ij}], \quad (4)$$

is obtained with

$$\sum_j P_{ij} = 1. \quad (5)$$

If n_{ij} is the number of transitions from Φ_i to Φ_j observed in the sequence of wind data, the maximum likelihood estimate of the transition probabilities is:

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}. \quad (6)$$

If Q_k and Q_{k+1} , are the probability vectors observed after 3 k and 3 $(k+1)$ h respectively, the transition from Q_k to Q_{k+1} is expressed as:

$$Q_{k+1} = Q_k M. \quad (7)$$

According to the ergodic properties of Markov chains, the probability vector Q , obtained after a long period of time, becomes independent from the initial state, and relation (7) can be written:

$$QM = Q. \quad (8)$$

The solutions to this equation are the long-term Markov probabilities $P(X_k = n)$ which form the probability vector (3). These probabilities can also be estimated by calculating the corresponding a priori probabilities $P_n = n_n / N_M$, where n_n is the number of Φ_n states observed over a month, and N_M the whole number of observations in the month. To appraise the persistence of the phenomenon under study, the correlation between two three-hourly observations has to be calculated.

Thus, the autocorrelation coefficient corresponding to a lag of 3 q h, is given by:

$$C_q = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i j p_{ij}^q P(X_j = i) - \left[\sum_{i=0}^{N-1} i P(X_j = i) \right]^2}{\sum_{i=0}^{N-1} i^2 P(X_j = i) - \left[\sum_{i=0}^{N-1} i P(X_j = i) \right]^2} \quad (9)$$

where P_{ij}^q is the element (ij) of the matrix M^q .

The discrete model described above, is equivalent to a coding system which produces messages by combining N different symbols. The relative entropy of this system is, by definition:

$$H = \frac{1}{\log_2(1/N)} \sum_{i=0}^{N-1} P_i \log_2 P_i \quad (10)$$

The loss of information caused by the classification of wind data into N stable states, is then given by:

$$L = 1 - H. \quad (11)$$

3.3. Wind direction modelling

The wind direction data collected at Oran during the 1982/92 period, have been modelled by first-order, nine-state Markov chains, following two different approaches. The first one is based on the use of homogeneous Markov chains to describe the three-hourly wind direction fluctuations for each month of the year. This means that all the Markov parameters defined above are constant with time over the month.

To illustrate the resulting model, the Markov transition probabilities of wind direction data obtained in January, April, July and October are plotted against the various directions in Fig. 1. As shown in Fig. 2, the long-term Markov probability of having one of the nine states of the wind direction, calculated for each of the 12 months, is found to be almost equal to the corresponding a priori probability. More generally, homogeneous Markov models are seen to yield a good representation of the seasonal variations.

The second approach consists in setting up Markov models with time-varying parameters. The use of inhomogeneous Markov processes leads to improved models, which account for the daily variations. To obtain this kind of model, each of the 12 monthly subsets of three-hourly wind direction data has been divided into four samples, rather than considering only one subset per month. For each month of the year, these samples denoted S_1 , S_2 , S_3 and S_4 , are made of all the measurements of the 1982/92 period performed in the intervals running from 0–6 h, 6–12 h, 12–18 h, and 18–24 h respectively. Then, each of the four samples has been modelled by first-order, nine-state Markov chains. The corresponding transition matrices are denoted M_1 , M_2 , M_3 and M_4 .

Fig. 3 gives, as an example, the results of the calculation of the Markov transition probabilities characterising the four January samples of wind direction data. The

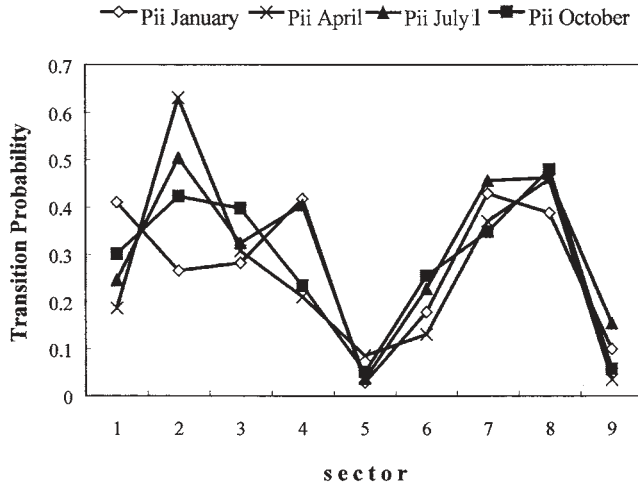


Fig. 1. Transition probability P_{ii} for the 3 h wind direction data collected at Oran and modelled by homogeneous first-order nine-state Markov chains.

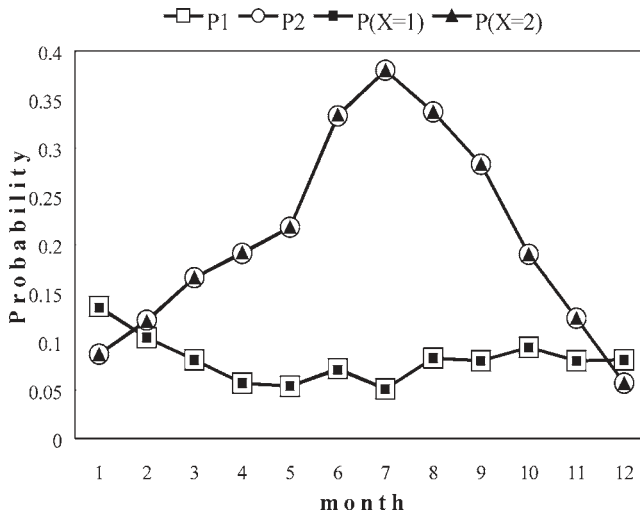


Fig. 2. Representation of long-term and a priori probability for hourly wind direction data. P_1 is the a priori probability corresponding to the sector 0. P_2 is the a priori probability corresponding to the sector 1. $P(X = 1)$ is the long-term probability corresponding to the sector 0. $P(X = 2)$ is the long-term probability corresponding to the sector 1.

product of these matrices has been performed, and the resulting matrix ($M_1 M_2 M_3 M_4$) has been compared with the transition matrix (M) obtained from the approach based on homogeneous Markov chains. For each of the 12 months, these matrices have been tested for homogeneity using the χ^2 distribution with eight degrees of freedom. The level of confidence has been fixed at 1%. In this case, all the P_{ij} values

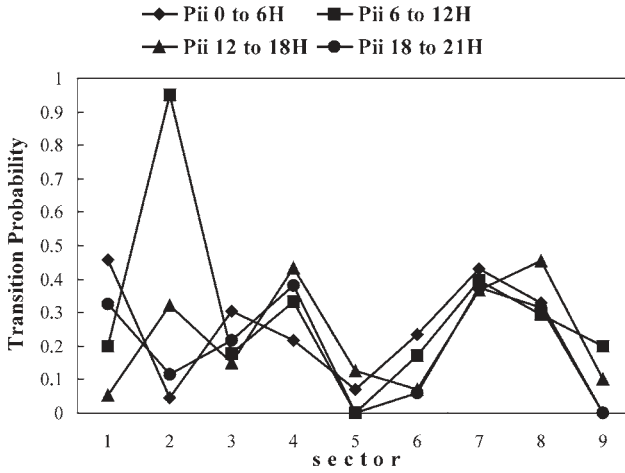


Fig. 3. Transition probability P_{ii} for 3 h wind direction data collected at Oran in January and modelled by inhomogeneous first-order nine-state Markov chains.

of these matrices are found to be falling within a 0.99 confidence range. This proves that the $(M_1 M_2 M_3 M_4)$ and (M) matrices are identical.

The autocorrelation coefficient has been calculated using (9). The values of C_q estimated for each month of the year and three different lags, namely 3, 6 and 9 h, are plotted in Fig. 4. These plots show that the autocorrelation coefficient decreases quickly when the lag (q) increases. When computing C_q for various lags, the autocor-

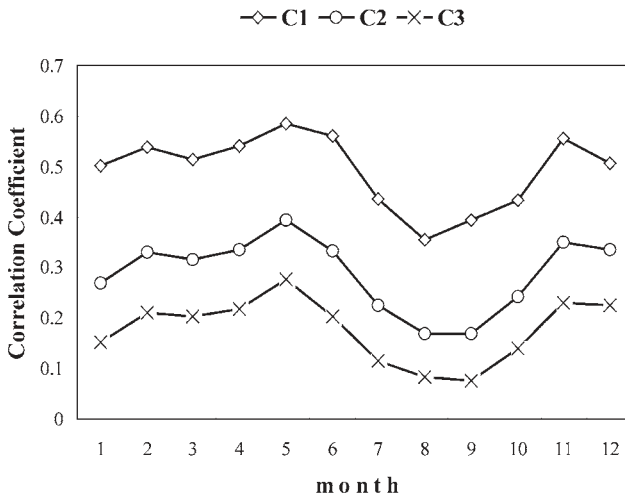


Fig. 4. Yearly variations of autocorrelation coefficients C_1 , C_2 and C_3 for the wind direction, C_1 , C_2 and C_3 are the correlation coefficients when two data are separated by three, six or nine h, respectively.

relation function has been found to be approximately equal to C_1 , where C_1 is the one-lag autocorrelation coefficient. This result means that the change in wind direction strongly depends on the immediate past. Hence, it justifies the limitation of the Markov chains used in this study to the first order.

The relative entropy was also computed for each of the 12 months. The results show that the loss of information due to the use of homogeneous first-order nine-state Markov model, only represents a few percentage points. The yearly variations of entropy for the inhomogeneous first-order nine-state Markov models are illustrated in Fig. 5. The mean monthly value of information loss is found not to exceed 11%. Hence, it may be concluded that such a model describes the wind direction variations for Oran with a reasonably good accuracy.

3.4. Wind speed modelling

First-order three-state Markov chains have been used to model the wind speed data collected every 3 h at Oran during the 1982/92 period. It has to be reminded that the three states are “weak wind” (state Φ_1), “mean wind” (state Φ_2) and “strong wind” (state Φ_3).

To illustrate the calculation of the Markov transition probabilities for wind speed, the plots of P_{11} , P_{22} , and P_{33} variations over the year are presented in Fig. 6. These diagrams show that the wind is weak at Oran almost all the time, and that strong winds are rare. This means that the wind energy potential is significant, but not very important in this region.

For each month of the year, the long-term Markov probability of having one of the three states of wind speed has been compared with the corresponding a priori probability. The results are plotted in Fig. 7 where it can be seen that both types of

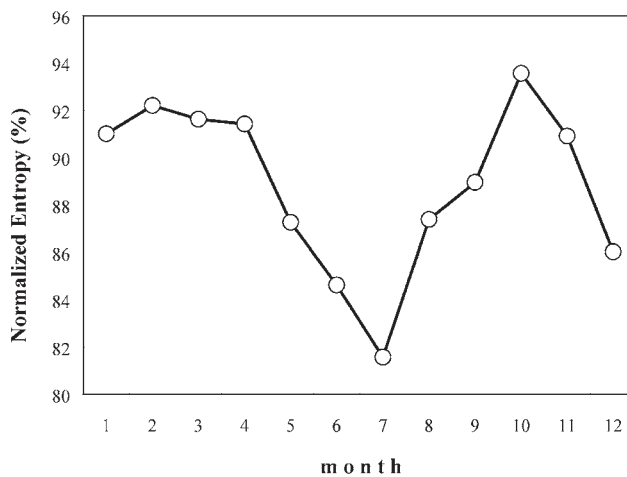


Fig. 5. Yearly variations of relative entropy for 3 h wind direction data.

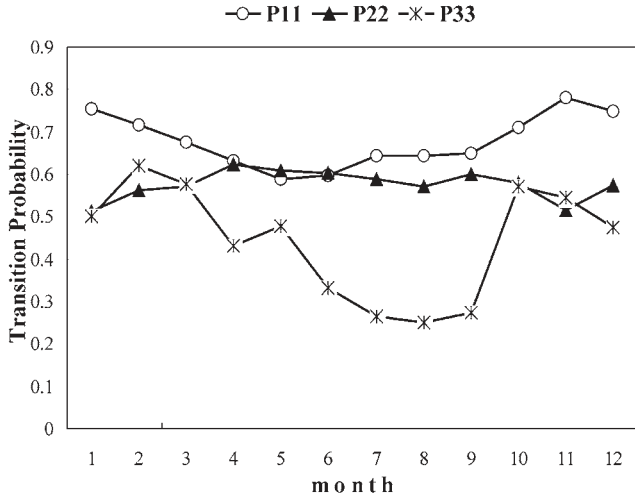


Fig. 6. P_{11} , P_{22} and P_{33} transition probabilities for the 3 h monthly wind speed data collected in Oran and modelled by first-order three-state Markov chains.

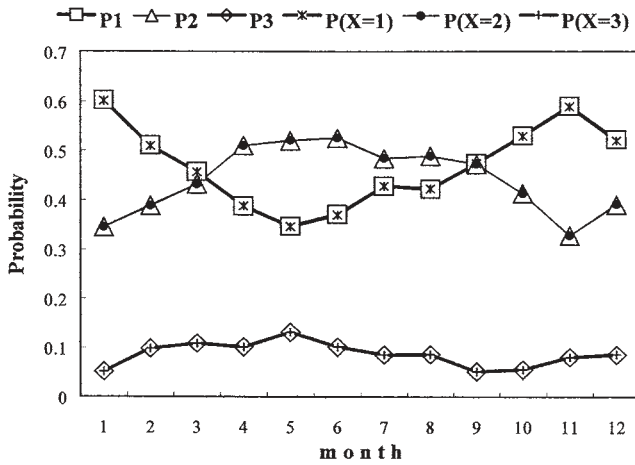


Fig. 7. A priori and Long-term probabilities for the 3 h monthly wind speed, noted P_i and $P(X = i)$ respectively with $i = 1, 2$ or 3 . i represents the sector.

probability are identical for each of the three states. This result proves that the first-order three-state Markov chains correctly fit the three-hourly wind speed variations.

The autocorrelation coefficient has been calculated for $q = 1, 2$ and 3 (i.e., lag equalling 3, 6, and 9 h respectively). The diagrams representing the variations of

C_1 , C_2 , and C_3 over the year are given in Fig. 8. The mean value of C_1 is close to 0.5, showing the strong correlation between two successive states. C_2 varies from 0.17-0.33, and C_3 is smaller than C_1 and C_2 . However, it sometimes reaches 0.22. More generally, the autocorrelation function is found to vary slowly. This means that the wind state observed at hour (h) is mainly influenced by the wind speed measured three hours before, i.e. $h - 3$, but the measurements performed at the preceding hours, $h - 6$, $h - 9, \dots$, also bring non negligible information.

Such a result is confirmed by the values of relative entropy obtained over the year. Fig. 9 indicates that the yearly variations of the relative entropy are weak. Entropy is minimum in January, when it equals 75%, and maximum in May, when it reaches 89%. Hence, the resulting loss of information varies from 21-25%.

In spite of the equality of the a priori and long-term Markov probabilities mentioned above, the strong values of information loss imply that the first-order Markov model cannot account accurately for the wind speed fluctuations. To improve this kind of modelling, second and third-order transitions must be considered, since they are as important as the first-order ones.

4. Weibull distribution-based model

The implementation of second or third-order Markov models requires tedious computation. Then, instead of such approaches, we have considered the Weibull distribution to model the 3 h wind speed measurements performed at Oran [17–21]. The Weibull probability density function is given by:

$$f(v) = \frac{k}{A} \left(\frac{v}{A}\right)^{k-1} \exp\left(-\left(\frac{v}{A}\right)^k\right) \quad (12)$$

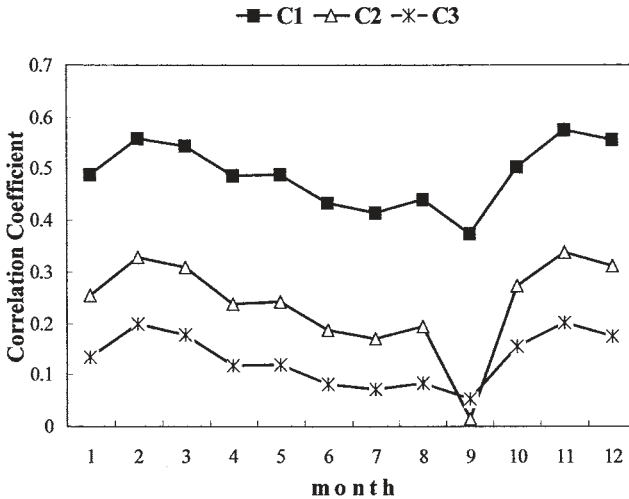


Fig. 8. Yearly variations of autocorrelation coefficients for the 3 h monthly wind speed. C_1 , C_2 and C_3 are the correlation coefficients, when two data are separated by three, six or nine hours respectively.

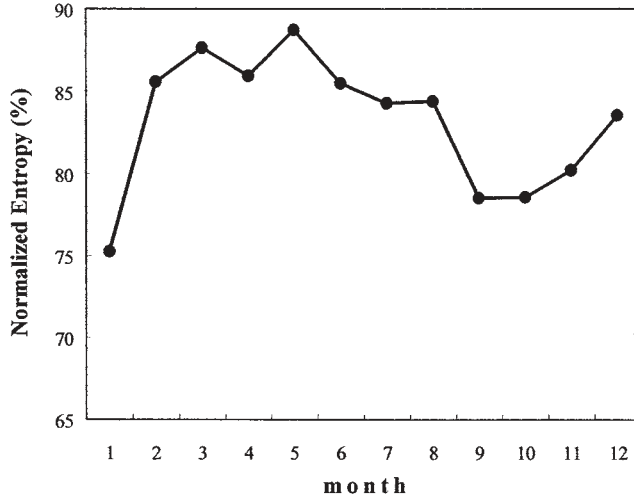


Fig. 9. Yearly variations of relative entropy for the 3 h monthly wind speed.

where, v is the wind speed in m s^{-1} , k is a dimensionless shape factor, and A is a scale factor with the dimension of speed. The Weibull cumulative distribution function is:

$$F(V) = \int_0^V f(v)dv = 1 - \exp\left(-\left(\frac{V}{A}\right)^k\right) \quad (13)$$

where $F(V)$ is the probability for the wind speed to be equal or lower than V . This function can be transformed into a linear relation of the form $y = a x + b$ by calculating:

$$y = \text{Ln}\{-\text{Ln}[1 - F(V)]\} = k \text{Ln}V - k \text{Ln}A. \quad (14)$$

The double logarithmic transformation has been applied to each of the monthly cumulative frequency distributions of the three-hourly wind speed data. The experimental data thus transformed have been smoothed using a least square method. The a and b parameters of the regression equation have been computed. Taking into account that $a = k$ and $b = -k \text{Ln}A$, the shape and scale factors of the corresponding Weibull distribution are obtained.

The plots of Figs 10 and 11 give the k and A values obtained at Oran for each of the eight directions and each of the 12 months. Fig. 10 shows that wind direction from west, north and south-west are dominant, with peak values in February, July, and June. For all sectors taken together, the peak is in April.

Fig. 11 also shows that the shape factor is always higher than 1. This means that the shape of the distribution is not exponential. However, Fig. 10 shows that, for the monthly curves, k is sometimes lower than 1.

For all the regression equations obtained in this study, the correlation coefficient is higher than 0.9 (Figs 12 and 13). For the south, south-west and west directions,

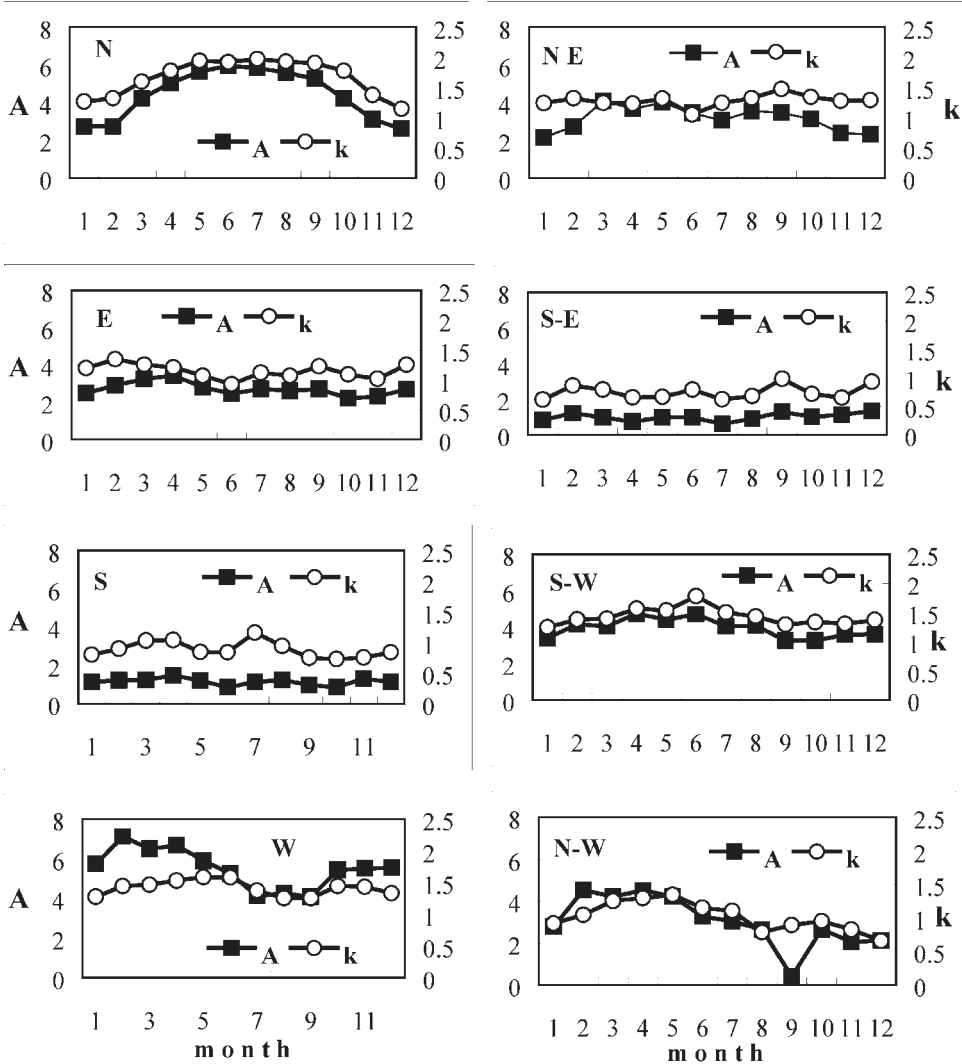


Fig. 10. Shape (k) and scale (A) factors of Weibull distribution estimated at Oran for each of the eight directions and for months running from January to December.

the correlation coefficient is higher than 0.97. For all the sectors together, it is higher than 0.96.

For a given month, each of the eight frequency distributions of three-hourly wind speed measurements is smoothed by a Weibull probability distribution function. All the results obtained with this kind of modelling, are in good agreement with those arising from experimentation. The results of the implementation of the usual hypothesis tests confirm that the Weibull probability distribution very well describes the monthly frequency distributions of hourly wind speed data.

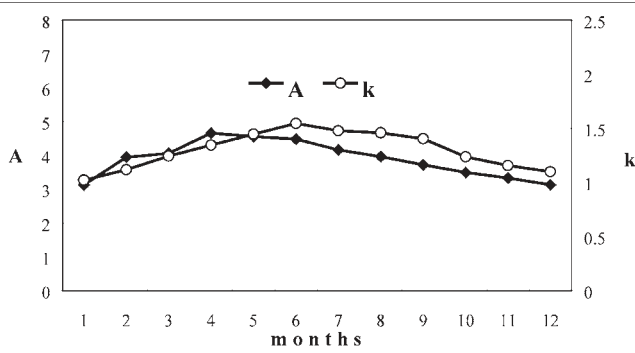


Fig. 11. Shape (k) and scale (A) factors of monthly Weibull distribution estimated at Oran for all sectors taken together.

5. Conclusion

This study confirms that wind is a Markov process, in which the main contribution arises from the first-order transitions. Two bivariate models are obtained. The first one is based on the first-order nine-state Markov chains for wind direction, and first-order three-state Markov chains for wind speed. The second one is based on the first-order nine-state Markov chains for wind direction, and Weibull distribution, for wind speed.

The memory effect that characterises the Markov processes is well mirrored by the autocorrelation obtained for the two used variables.

The bivariate model consisting in the combination of only two Markov chains, correctly describes the three-hourly wind data. It generates two time series corresponding to two independent variables. However, this model leads to non negligible loss of information.

The three-hourly wind data are better described with the Weibull probability distributions associated to Markov chains. The tight link observed between experimental and retrieved series (95%) confirms this result. Moreover, this kind of bivariate model improves the usual Weibull-based wind modelling.

Using the Markov model of wind direction obtained in section 3.3, the Weibull model of wind speed generates joint times series. The addition of the wind direction model is useful since it enables to retrieve bivariate time series of the wind. This can be used for the reconstruction, simulation, and prediction of the wind.

As mentioned in section 3.4, the wind energy potential in the region of Oran is moderate, but sufficient to be used as an alternative energy source. However, the results of section 4 show that for some periods of the year and for some directions, the wind energy can be important.

Recently, both bivariate models have also been applied to hourly wind measurements collected in the meteorological station of Chasseral (Switzerland) from 1990-2000. This station is situated in the Jura mountains at 1600 m of altitude. The measurements were performed at 40 m above ground level. The obtained results are

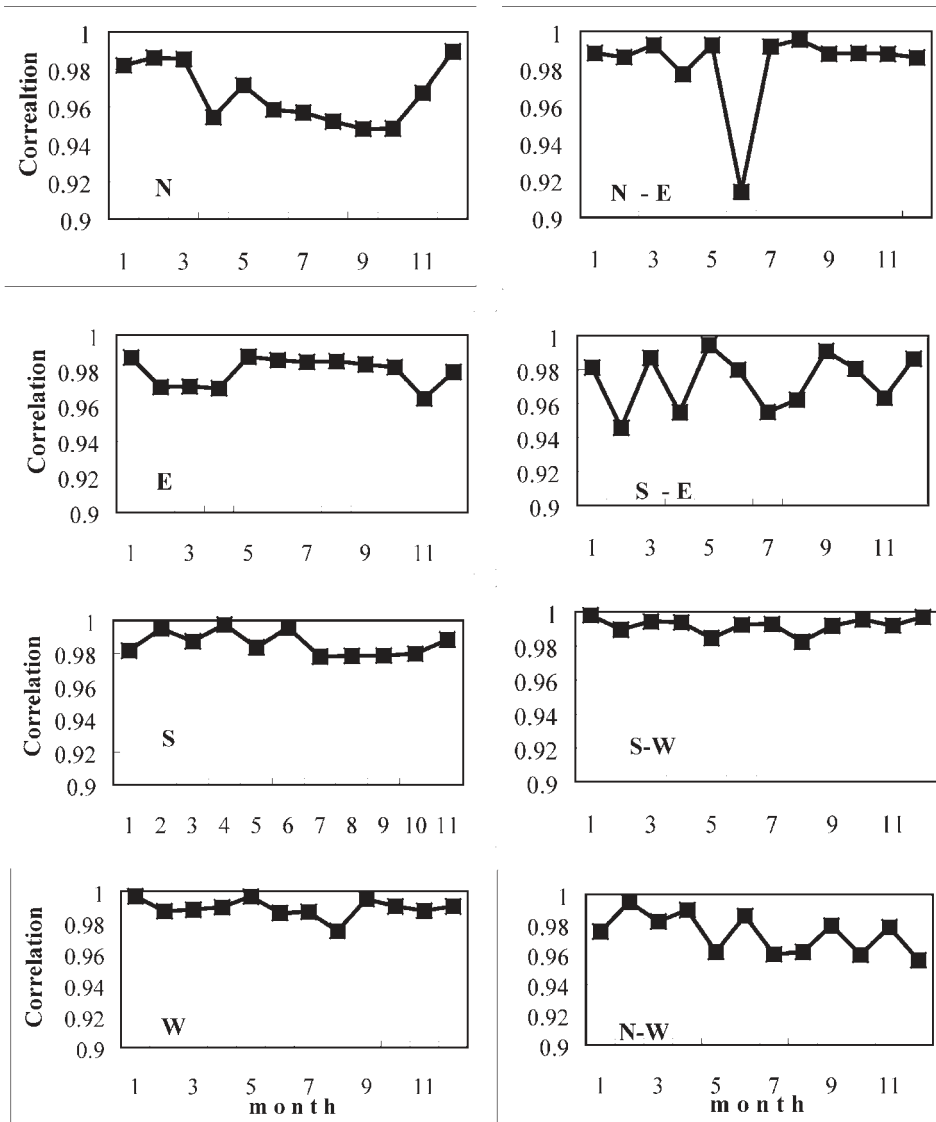


Fig. 12. Correlation coefficient between the experimental wind speed distribution and the corresponding values obtained from fitting the Weibull distribution for each sector.

in good agreement with present ones. They confirm that the combination of Weibull distributions and Markov chains yields the best representation of wind data.

The wind data generated by the improved bivariate model could be useful to size the wind power generators to be installed in any location. These data are also usable for the evaluation of the performance of such generators.

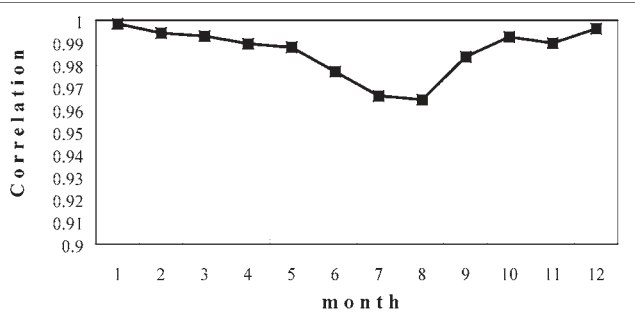


Fig. 13. Correlation coefficient between the experimental wind speed distribution and the corresponding values obtained from fitting the Weibull distribution for all sectors together.

Acknowledgements

The authors wish to thank the National Office of Meteorology of Algeria for providing the data used in this study.

References

- [1] Weibull W. A statistical distribution function of wide applicability. *J Appl Mech* 1951;18:293–7.
- [2] Khogali A, Albar OF, Yousif B. Wind, solar energy potential in Makkah (Saudi Arabia). Comparison with Red Sea coastal sites. *Renewable Energy* 1991;1(3/4):435–40.
- [3] Hasan A. Wind energy in West Bank and Gaza Strip. *Renewable Energy* 1992;2(6):637–40.
- [4] Rehman S, Halawani TO. Statistical characteristics of wind in Saudi Arabia. *Renewable Energy* 1994;4(8):949–56.
- [5] Morgan VT. Statistical distributions of wind parameters at Sydney, Australia. *Renewable Energy* 1995;6(1):39–48.
- [6] Garcia A, Torres JL, Prieto E, De Francisco A. Fitting wind speed distributions: a case study. *Solar Energy* 1998;62(2):139–44.
- [7] Youcef Ettoumi F, Adane A, Hammouche R. Modelling and simulation of hourly wind in Algeria. In *Proceedings of ISES (In Search on the Sun) Solar World Congress, Harare, Zimbabwe, 1995*.
- [8] Youcef Ettoumi F, Adane A. Optimum classification of wind energy potential in Algeria. In *Proceedings of World Renewable Energy Congress IV, Denver, CO, 1996*.
- [9] Ward JH. Hierarchical grouping to optimize an objective functions. *J Amer Stat Assoc* 1963;58:236–44.
- [10] Kalkstein LS, Tan G, Skindlov JA. An evaluation of three clustering procedures for use in synoptic climatological classification. *J Climate and Appl Meteor* 1987;26(6):717–30.
- [11] Dixon WJ. *BMDP Statistical software*. London, Los Angeles, CA: University of California Press, 1985.
- [12] Castini F, Festa R, Ratto CF. Stochastic modelling of wind velocities time series. *J Wind Eng Ind Aerodynamics* 1998;74-76:141–51.
- [13] Nfaoui H, Buret J, Sayigh AAM. Stochastic simulation of hourly average wind speed sequences in Tangers (Morocco). *Solar Energy* 1996;56:301–14.
- [14] Daniel AR, Chen AA. Stochastic simulation and forecasting of hourly average wind speed sequences in Jamaica. *Solar Energy* 1991;46:1–11.
- [15] Ross SM. *Stochastic processes*. J Wiley & Son, 1982.
- [16] Stewart WJ. *Introduction to the numerical analysis of Markov chains*. NJ: Princeton, 1994.

- [17] Hennesey Jr. JP. Some aspects of wind power statistics. *J Appl Meteor* 1977;16:119–28.
- [18] Tackle ES, Brown JM. Note on the use of Weibull statistics to characterise wind-speed data. *J Appl Meteor* 1977;17:556–9.
- [19] Justus CG, Hargraves WR, Mikhail A, Graber D. Methods for estimating wind speed frequency distributions. *J Appl Meteorol* 1978;17:350–3.
- [20] Darwish ASK, Sayigh AAM. Wind energy potential in Irak. *Solar and Wind Technol* 1988;5:215–22.
- [21] Løvseth J. The renewable energy potential of Norway and strategies for its development. *Renewable energy* 1995;6(3):207–14.