Mausam, (1983), 34, 4, 395-398

551.509.336 : 551.58 (46)

Application of a Markovian model to the study of sunshine in four cities in the north of Spain

P. L. FERNANDEZ, L. S. QUINDOS, J. SOTO and E. VILLAR

Deparatamento de Fisica Fundamental, Facultad de Ciencias, Universidad de Santander

(*Received 4 June 1982*)

सार — किसी खास दिन और उसके अगले दिन का मौसम सांख्यिकीय दृष्टि से परस्पर स्वतन्त्र नहीं होते बल्कि एक प्रकार का सह-सम्बन्ध प्रदर्शित करते हैं। इस शोधपत्र में अच्छे (एफ) एवं खराब (बी) मौसम की स्थितियों को लेकर मौसम विज्ञान सम्बन्धी एक सरलीकृत निदर्श के ढांचे में स्पेन के उत्तरी भाग के कई शहरों के लिए स्वसमाश्रयण ज्ञात किए गए हैं। इन स्थितियों को आपेक्षिक दैनिक धूप के फलन के रूप में परिभाषित किया है। निदर्श संतोषजनक है क्योंकि अधिकांश सैद्धांतिक परिणाम प्रयोगिक परिणामों से मेल खाते है। इनमें अच्छे एवं खराब दिनों के मासिक प्रतिशत और वर्ष के प्रत्येक माह में विभिन्न अनुक्रमों की औसत अवधियां विशेष रूप से निदर्श के अनुसार पाई गई।

ABSTRACT. The weather of a given day and the following one are not statistically independent, but show a kind of correlation. In this work, we have determined the autocorrelation parameters for several cities in the north of Spain in the framework of a simplified meteorological model with two states : (F) fine and (B) bad weather. These states are defined as a function of the relative daily sunshine. The model is satisfactory since most of the theoretical results are in agreement with the experimental ones, specially the monthly percentages of fine and bad days and the mean lengths of the different sequences in each month of the year.

1. Introduction

Several authors have studied the correlation in weather beween consecutive days, using the statistical interdependence of past, present and future weather situations, taking into account either the overall level of sunshine (Klein 1977) or the amount of rainfall (Levett 1976). In this work we have applied a simplified meteorological model according to Ph. Bois (Bois 1979). This model assumes that any sequence of consecutive days behaves as a Markovian chain with two states : (F) fine weather and (B) bad weather.

We apply this model to the particular cases of four cities in the north of Spain (Gijon, Santander, Bilbao and San Sebastian), whose locations are shown in Fig. 1, from 1 January 1970 to 1 January 1980, in order to know the influence that the correlation between consecutive days has on weather forecasting. This is the main point to consider in calculating potential use and storage of solar energy, a field of current research in our Department. The criterion used for discriminating between the two states mentioned above was whether the daily integrated sunshine was greater or less than $0.5\ I_0$, where I_0 represents the maximum possible amount of sunshine for that day, given a complete absence of clouds.

2. Description of the model

We will use a first order Markov chain with two states : (F) fine weather and (B) bad weather. The probabilities of the four possible transitions are designated by a (F-F), b (F-B), c (B-F) and d (B-B), respectively. These four probabilities define the stochastic P matrix :

$$P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

The main properties of this matrix are :

(a) The probability that the weather will be fine (X) or bad (Y) after $j+1$ days, given that the first day is either fine or bad is :

$$\begin{pmatrix} X \\ Y \end{pmatrix} = (P)^j \begin{pmatrix} F \\ B \end{pmatrix}$$

when (F, B) is (1, 0) or (0, 1) according to whether the first day is fine or bad.

(b) Since a, b, c and d are positive numbers (no transition can be eliminated as they are not zero values), the limit of P^j when $j \to \infty$ exists. Therefore :

$$\lim_{j \to \infty} P^j = \begin{pmatrix} c/b+c & b/b+c \\ c/b+c & b/b+c \end{pmatrix} \tag{1}$$

This limit can be easily deduced provided that P is the solution of the equation $P^\infty .\ P = P$. This result means that there is a probability of $c/b+c$ ($b/b+c$ resp.) that it will be fine (bad resp.) after a long period, whatever the weather on the first day.

TABLE 1

Monthly weighted mean values of probabilities a, d for the period 1970-79

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
						Santander						
$\bar{a}$	.44±.11	.52±.11	.55±.11	.53±.11	.57±.11	.60±.10	.64±.09	.53±.11	.66±.09	.60±.10	.59±.10	.53±.11
$\bar{d}$	.79±.08	.74±.09	.74±.09	.73±.09	.73±.09	.67±.09	.64±.10	.64±.10	.71±.09	.75±.08	.75±.08	.77±.08
						San Sebastian						
$\bar{a}$	.46±.11	.52±.11	.57±.11	.54±.11	.59±.10	.62±.10	.64±.09	.58±.10	.65±.09	.61±.10	.60±.10	.55±.11
$\bar{d}$	.77±.08	.75±.08	.71±.09	.71±.09	.68±.10	.63±.10	.63±.10	.61±.10	.67±.09	.70±.09	.72±.09	.75±.08
						Bilbao						
$\bar{a}$	.41±.12	.46±.11	.49±.11	.49±.11	.54±.11	.56±.11	.63±.10	.55±.11	.60±.10	.58±.10	.55±.11	.50±.11
$\bar{d}$	.83±.08	.79±.08	.77±.08	.77±.08	.75±.08	.70±.09	.63±.09	.63±.10	.71±.09	.75±.08	.77±.08	.81±.08
						Gijon						
$\bar{a}$	.43±.12	.48±.11	.53±.11	.51±.11	.55±.11	.60±.10	.63±.10	.57±.11	.63±.10	.59±.10	.57±.11	.50±.11
$\bar{d}$	.78±.08	.73±.09	.71±.09	.70±.09	.68±.09	.65±.10	.62±.10	.60±.10	.69±.09	.71±.09	.73±.09	.76±.08

TABLE 2

Theoretical (T) and experimental (E) probabilities P (F)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
						Santander						
T	.27±.09	.35±.09	.37±.09	.37±.08	.39±.08	.45±.08	.50±.08	.43±.08	.46±.08	.39±.08	.35±.09	.33±.09
E	.27	.32	.34	.39	.38	.43	.50	.44	.46	.39	.35	.32
						San Sebastian						
T	.30±.09	.34±.08	.40±.08	.39±.08	.44±.08	.49±.08	.51±.08	.48±.08	.49±.08	.44±.08	.41±.08	.36±.09
E	.28	.31	.37	.40	.43	.46	.50	.47	.47	.42	.39	.34
						Bilbao						
T	.22±.09	.28±.09	.31±.09	.31±.09	.35±.09	.40±.08	.50±.08	.45±.08	.42±.08	.37+.09	.34±.09	.28±.09
E	.19	.26	.29	.28	.32	.39	.47	.42	.42	.34	.31	.27
						Gijon						
T	.28±.09	.34±.09	.38±.08	.38±.08	.42±.08	.48±.08	.51±.08	.48±.08	.46±.08	.41±.08	.39±.08	.32±.09
E	.29	.32	.37	.36	.40	.45	.50	.46	.47	.40	.36	.31

TABLE 3

Theoretical (T) and experimental (E) monthly mean length of sequǝces of F days

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
						Santander						
T	1.8±0.3	21.0±.4	2.2±0.4	2.1±0.4	2.3±0.4	2.5±0.5	2.8±0.5	2.1±0.4	2.9±0.5	2.5±0.5	2.4±0.4	2.1±0.4
E	1.7	1.8	2.1	1.9	2.1	2.3	2.7	1.9	2.6	2.2	2.2	1.9
						San Sebastian						
T	1.8±0.3	2.1±0.4	2.3±0.4	2.2±0.4	2.4±0.4	2.6±0.5	2.8±0.5	2.4±0.4	2.9±0.5	2.6±0.5	2.5±0.5	2.2±0.4
E	1.7	1.8	2.1	2.0	2.2	2.4	2.8	2.4	2.6	2.4	2.2	1.9
						Bilbao						
T	1.7±0.3	1.8±0.3	2.0±0.4	2.0±0.4	2.2±0.4	2.3±0.4	2.7±0.5	2.2±0.4	2.5±0.5	2.4±0.4	2.2±0.4	2.0±0.4
E	1.4	1.6	1.8	1.7	1.9	2.1	2.4	2.0	2.3	2.1	2.0	1.8
						Gijon						
T	1.8±0.3	1.9±0.3	2.1±0.4	2.0±0.4	2.2±0.4	2.5±0.5	2.7±0.5	2.3±0.4	2.7±0.5	2.4±0.4	2.3+0.4	2.0±0.4
E	1.6	1.8	1.9	1.8	2.0	2.4	2.7	2.2	2.5	2.2	2.0	1.8

TABLE 4

Theoretical (T) and experimental (E) monthly mean lengths of sequences of B days

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
						Santander						
T	4.8±1.0	3.8±0.8	3.8±0.8	3.7±0.8	3.7±0.8	3.0±0.7	2.8±0.6	2.8±0.6	3.4±0.7	4.0±0.8	4.5±0.9	4.3±0.9
E	4.5	3.7	3.6	3.4	3.4	2.9	2.6	2.5	3.3	3 7	4.2	4.2
						San Sebastian						
T	4.4±0.9	4.0±0.8	3.4±0.7	3.4±0.7	3.1±0.7	2.7±0.6	2.7±0.6	2.6±0.6	3.0±0.7	3.3±0.7	3.6±0.8	4.0±0.8
E	4.2	3.7	3.3	3.2	2.9	2.6	2.5	2.3	2.7	3.1	3.5	4.0
						Bilbao						
T	5.9±1.2	4.8±1.0	4.4±0.9	4.4±0.9	4.0±0.8	3.3±0.7	2.7±0.6	2.7±0.6	3.4±0.7	4.0±0.8	4.4±0.9	5.3±1.1
E	5.7	4.6	4.3	4.1	3.8	3.1	2.7	2.7	3.2	3.7	4.2	5.2
						Gijon						
T	4.5±0.9	3.7±0.8	3.4±0.7	3.3±0.7	3.1±0.7	2.9±0.6	2.6±0.6	2.5±0.6	3.2+0.7	3.4±0.7	3.7±0.8	4.2±0.9
E	4.3	3.6	3.1	3.3	3.0	2.7	2.5	2.3	2.9	3.2	3.6	4.1

(c) The correlation between any two days separated by an interval of j days is defined by means of the correlation coefficient R_j, the value of which can be calculated through the formula (Brarucha-Reid 1960)

$$R_j = \frac{<Z(i).Z(i+j)> - <Z(i)> . <Z(i+j)>}{\sigma_{Z(i)} \cdot \sigma_{Z(i+j)}} \tag{2}$$

where $< >$ and σ denote their mean value and the standard deviation, respectively, of the random variable $Z(i)$, which is 1 or 0 depending on whether the i-th day is fine or bad.

The value obtained for R_j is :

$$R_j = (a - c)^j \tag{3}$$

If $a > c$ then R_j is positive for all j values. However, R_j will naturally decrease when j increases because a and b are both less than or equal to 1.

(d) The probability that any sequence beginning with a F day will not deviate from this state during a period of I days is $b.\ a^{I-1}$. The mean length of this kind of sequence is $1/b$. In the same way, the probability that any sequence beginning with a B day will persist for I days is $c.\ d^{I-1}$ and its mean length is $1/c$.

Starting from any defined sequence of $j+1$ days, among the 2^{j+1} possible ones, the probability that such a chain will happen is :

$$p(k, l, m, n) = a^k.b^l.c^m.d^n = a^k.(1-a)^l.(1-d)^m.d^n$$
$$(k+l+m+n = j) \tag{4}$$

where k, l, m and n are the number of a, b, c and d type transitions, respectively.

From the probability p (k, l, m, n) given above, we can evaluate the stochastic P matrix, using the maximum likelihood method which is based in determining the four probabilities (a, b, c, d) which maximise the p (k, l, m, n) probability and deriving this from Eqn. (4). Applied to the P matrix we obtain the following result :

$$P = \begin{pmatrix} k/l+k & l/l+k \\ m/m+n & n/m+n \end{pmatrix} \tag{5}$$

Also we can easily calculate the errors in these values from the standard deviation of a and d in the probability distribution p (k, l, m, n) :

$$\sigma_a^2 = \sigma_b^2 = \frac{(k+1).(l+1)}{(k+l+2)^2.(k+l+3)} \tag{6}$$

$$\sigma_c^2 = \sigma_d^2 = \frac{(m+1).(n+1)}{(m+n+2)^2.(m+n+3)} \tag{7}$$

3. Application of the model

In the framework of the model we have described, and analysed the daily relative sunshine data recorded by the Centro Meteorologico del Cantabrico for the cities of Gijon, Santander, Bilbao and San Sebastian, from 1 January 1970 to 1 January 1980. These cities have a temperate climate and within this denomination, the breton's one, Cfb, according to the Koppen classification (Koppen 1948). Their mean annual temperature oscillates between 13.9 and 14.2° C and the average total annual precipitation is about 1000 mm.

The monthly distributions of daily relative sunshine show high concentrations of extreme values, specially at the lower end of the scale. We have also recorded significant minima in the central region around 0.5, except in the spring and summer months, in which the weather was not so steady from a statistical point of view. For that reason, we have chosen the relative sunshine value of 0.5 as a threshold in order to discriminate the two kinds of weather : (F) fine and (B) bad.

The monthly values of a_i and d_i corresponding to each year of the period under study have been obtained by the maximum likelihood method. From these, we determine the weighted mean values $\bar{a}$ and $\bar{d}$ defined as :

$$\bar{a} = \frac{\sum_{i=1}^{10} k_i.a_i}{\sum_{i=1}^{10} k_i} \tag{8}$$

$$\bar{d} = \frac{\sum_{i=1}^{10} n_i.d_i}{\sum_{i=1}^{10} n_i} \tag{9}$$

Fig. 1. Location of the cities of Gijon, Santander, Bilbao and San Sebastian in a map of Europe

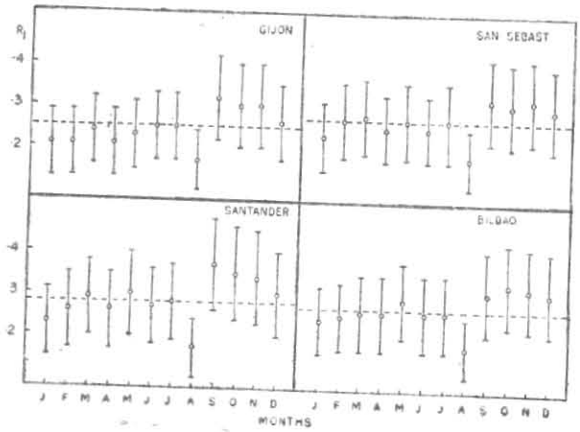

Fig. 2. Annual variation of the correlation coefficient R_1 in the four stations

where k_i and n_i are the number of a and d type transitions for each month of the i-th year, respectively. Table 1 shows the monthly a and d probabilities and their associated errors for the four stations we have studied. The seasonal variation is evident for a with a characteristic minimum in January. However, calculated values of a have anomalous values in August, less than those for July and September, whereas the expected ones would be similar to or greater than these. This fact is related to the instability of the climate during this month, so long as the maximal polar frontal activity occurs in this period. Such activity gives rise to isolated periods, one or two days of high cloudiness with the appearance sometimes of rains.

The calculated values of d are higher than those of a for all months, except July when the two are equal. But these values are of similar order throughout the whole year, except in summer, when they are slightly lower by about 0.1.

Fig. 2 shows the annual variation of the correlation coefficient R_1 for the four stations. Only the value for August is lower than the corresponding average annual one. This fact again confirms the special climate features of this month. The mean annual values of R_1, we have found for the four cities under study, fluctuate between 0.25 and 0.28. They are slightly lower than that observed (Lestienne 1978) in places of similar weather conditions.

In Table 2 we compare, for the four locations, the experimental probabilities that the weather will be fine or bad with $P(F)=c/b+c$ and $P(B)=b/b+c$, deduced from the model, for each month of the year. Also, Tables 3 and 4 compare the experimental mean monthly lengths of sequences of fine and bad days, respectively, and the theoretical ones, F and B, predicted in accordance with the model. Results confirm the validity of the model.

4. Summary

We have applied a first order Markov chain to data on sunshine in four cities in the north of Spain for the period 1970-1979. The goodness of the fitting shows the utility of this model in two aspects: (1) as a complementary method of construction of models in weather predictions and (2) in the knowledge of the frequency and sequences of fine and bad days.

References

Brarucha-Reid, A.I., 1960, *Markov processes and their applications*, Ed. McGraw Hill.

Bois, Ph., 1979, Analyse temporelle et cartographique de la matrice stochastique pour le modele markovien dans le midi de la France, *La Meteorologie*, **17**, 83-122.

Klein, S.A., 1977, "Calculation of monthly average insolation on tilted surfaces", *Solar Energy*, **19**, 325-329.

Koppen, W., 1948, "Con un estudio de los climas de la Tierra", *Climatologia*, FCE, Mexico.

Lestienne, R., 1978, "Modele markovien simplifie de meteorologie a deux etats., L'exemple d'Odeillo", *La Meteorologie*, **19**, 53-64.

Levett. R., 1976, "Rainfall reliability in Ethiopia", *Weather* (GB), **31**, 11, 417-424.