Comparison of Models for Estimating the Joint Probability of a Weather Event

JOSIP JURAS

Federal Hydrometeorological Institute, Zagreb, Yugoslavia

18 August 1981 and 26 July 1982

1. Introduction

Lund and Grantham (1979, 1980), in order to estimate spatial joint probabilities of the occurrence of a weather event, use the equation which originates from McAllister's (1969) equation, later modified by Gringorten (1971). Gringorten's equation makes it possible to obtain the joint probabilities of two correlated and normalized variables in a very simple way. However, it is important to point out that Gringorten's equation is valid only in the case of equal unconditional probabilities. Lund and Grantham adapted this equation for the cases where climatic probabilities of the event in certain places differ significantly. Therefore, in a general case, Lund and Grantham's model represents a poor approximation of the bivariate normal distribution (BND) model.

This article presents a very simple model, which serves as a substitute for the BND model. Its purpose is to emphasize that the existing range of models does not enhance our knowledge of the temporal and spatial persistence of weather events. A more extensive application of the BND model, appearing the most promising at the moment, would give the possibility of comparing the results obtained in various investigations.

2. The additional model

As Lund and Grantham obtained quite satisfactory joint probability estimates, I was encouraged to take the heretical way, starting from the equation developed by Sheppard (1898),

$$P_{ij} = \tfrac{1}{4} + (\tfrac{1}{2\pi}) \arcsin\rho. \tag{1}$$

By applying this equation, it is possible to determine the joint probabilities P_{ij} of BND for any correlation coefficient ρ, but only in the case where the marginal probabilities P_i and P_j are equal to 0.5. For the cases when P_i and P_j differ from 0.5, the unallowed generalization of Eq. (1) expressed in the following form was applied:

$$\left.\begin{aligned} \hat{P}_{ij} &= P_iP_j + P(1 - P)(2/\pi) \arcsin\rho, \\ \text{where} \quad P &= \begin{cases} \min(P_i, P_j) & \text{if} \quad P_i + P_j \leqslant 1, \\ \max(P_i, P_j) & \text{if} \quad P_i + P_j \geqslant 1. \end{cases} \end{aligned}\right\} \tag{2}$$

Sometimes, the calculated value of the $\hat{P}_{ij}$ according to Eq. (2) exceeds that of the marginal probabilities. In such a case, the joint probability $\hat{P}_{ij}$ is replaced by a smaller marginal probability.

In order to check the validity of Eq. (2) as an approximation of BND, the differences between the $\hat{P}_{ij}$ value and the corresponding BND value were calculated for various combinations of P_i, P_j and ρ. The greatest error was found for the following combination: $P_i = 0.35$, $P_j = 0.65$, $\rho = 0.75$. The calculated value of $\hat{P}_{ij}$ is 0.3503. Therefore, the smaller value of the marginal probabilities was taken (i.e., $\hat{P}_{ij} = P_i = 0.35$). In comparison, BND gives the value 0.330. The fact that the error exceeded 0.01 in only 2% of the examined combinations justifies the use of Eq. (2) as an approximation of BND. The limitation of its usefulness becomes evident when it is used in calculating conditional probabilities $\hat{P}_{j|i}$ from the relation

$$\hat{P}_{j|i} = \hat{P}_{ij}/P_i. \tag{3}$$

For small values of unconditional probabilities ($P_i < 0.1$), errors in estimating $\hat{P}_{j|i}$ even greater than 0.1 were obtained.

The suitability of Eq. (2) for estimating the probability that a weather event would occur jointly at two locations has been examined by using Lund and Grantham's data (1979, Table 4; 1980, Table 5).

It is assumed that correlation ρ depends on distance d (in statute miles) between two corresponding locations in accordance with the expression

$$\rho = \rho_2^{d/25}, \tag{4}$$

where ρ_2 represents correlation parameter (cp), which is used in the model when joint probabilities for pairs of locations are estimated. The accuracy of the model defined by Eqs. (2) and (4) has been compared with the accuracy obtained by the application of Lund and Grantham's (1980) Eq. (3) as well as of the BND model. McCabe (1968) and Gringorten (1971) described the various methods of computation for the BND model. A comparison of the results is given in Table 1.

The cp's of the model were determined for each region separately because of the intention to examine primarily the relative validity of the models, rather

0021-8952/82/121926-03$04.75

TABLE 1. The correlation parameters for various models in the estimation of the joint probabilities for the event ≥ 0.8 sky cover in winter. Root-mean-square-errors (rmse) and percent-reductions-in-error relative to independence R found by solving Eqs. (12) and (13) of Lund and Grantham (1980), respectively.

	Eastern United States stations			Central United States stations		
Model	ρ_2	rmse	R	ρ_2	rmse	R
Lund and Grantham (1980),						
Eq. (3)	0.965	0.006	95.5	0.963	0.008	93.3
Eq. (2)	0.957	0.005	96.8	0.949	0.009	92.4
BND	0.957	0.005	96.8	0.949	0.008	92.6

than to find out if the same *cp* could be used in various regions.

All the models are equally good. The BND model does not show any significant advantage compared with the other two models. The relative frequencies of the event in the examples under consideration differ slightly, which makes Lund and Grantham's model convenient for the application. All the frequencies are ~50%, which is convenient for the application of Eq. (2).

However, I would like to point out the similarities and differences between the *cp*'s in various models. My impression is that *cp* in Lund and Grantham's model does not reflect only the spatial persistence of the weather event. It is also under the strong influence of the range of climatic frequencies, which the corresponding weather event features in the area under consideration. For example, the very high value ρ_2 (0.982) for the visibility category ≥ 10 mi (Lund and Grantham, 1980, Table 1) is probably not the consequence of the strong persistence of this weather event. This value indicates that the model requires an unrealistically large *cp* when applied to the area where the climatic frequencies for various places differ a great deal.

3. Modeling the correlation field

By examining the differences between the joint probabilities estimated by the models and observed relative frequencies (see Lund and Grantham, 1979, Fig. 4; 1980, Table 5), the systematic positive errors for pairs of very near and very distant locations become obvious. This indicates that the assumption that correlation decreases with distance as formulated in Eq. (4) is only valid as the first approximation.

An attempt was made to describe the field of correlation by the relation

$$\rho = \rho_2^{(d+d_1)/25} \cos(\pi d/2d_0). \quad (5)$$

This presumes that correlation on some distance d_0 drops to zero, while parameter d_1 could be explained as observational error. Such a simple model of the correlation field yields much more accurate joint probability estimates, as seen in Table 2. For the approximate estimates of the additional parameters in Eq. (5), the values taken for the east coast of the United States are: $d_0 = 1200$, $d_1 = 14$ and for the central United States: $d_0 = 900$, $d_1 = 0$.

Very small rmse for the BND model proves not only its validity, but also the excellent quality of the basic data.

TABLE 2. As in Table 1, except the values of correlation ρ were determined by Eq. (5).

	Eastern United States stations			Central United States stations		
Model	ρ_2	rmse	R	ρ_2	rmse	R
Lund and Grantham (1980),						
Eq. (3)	0.972	0.004	97.2	0.973	0.004	96.5
Eq. (2)	0.965	0.002	98.6	0.961	0.004	96.8
BND	0.965	0.002	98.6	0.960	0.003	96.9

4. Joint occurrence probabilities at more than two locations

The probability that weather event E will occur simultaneously at four places, for example, could be determined by the multiplication rule;

$$P_{ijkl} = P(E_i)P(E_j|E_i)P(E_k|E_j, E_i)P(E_l|E_k, E_j, E_i), \quad (6)$$

where $P(E_i)$ denotes the probability of event E at the location i, $P(E_j|E_i)$ the conditional probability of E at location j, given E at location i, etc.

If, as a simplification, it is assumed that the conditional probability of the occurrence of the weather event E at some location depends only on whether this event was observed in the nearest neighboring location, the underestimated value of conditional probability is given. If the conditional probabilities in (6) are determined by (3), while the more distant locations are neglected, it is necessary to operate with some larger *cp* values than those used for pairs of location as is indicated in following equation:

$$\hat{P}_{ijkl} = P(E_i)\hat{P}(E_j|E_i; \rho_2)\hat{P}(E_k|E_j; \rho_3)\hat{P}(E_l|E_k; \rho_4). \quad (7)$$

Surprisingly, Eq. (11) of Lund and Grantham (1980), used by them for a somewhat different purpose, turned out to be suitable for determining increased *cp* values which depend on the number of locations. An attempt was made to determine the joint probability of the occurrence of some weather event at four locations by the conditional probabilities for pairs of neighboring locations with increased *cp* values. The computation procedure of joint probabilities for a combination of stations COU–STL–BLV–EVV is shown in Table 3.

The joint probability of the simultaneous occurrence of the event at four locations was obtained according to (7) by multiplying the climatic frequency

TABLE 3. An example of the computation of conditional probability estimates $P(E_n|E_{n-1})$ for the event > 0.8 sky cover in winter. The values of unconditional probabilities of the event $P(E_n)$ and distances d are taken from Table 5 of Lund of Grantham (1980); ρ_n are determined in accordance with Eq. (11) of Lund and Grantham (1980); correlation ρ for corresponding pairs of locations by Eq. (5).

n	Station*	$P(E_n)$	ρ_n	d	ρ	$P_{n\|n-1}$
1	COU	0.532				
2	STL	0.566	0.961	108	0.827	0.852
3	BLV	0.564	0.966	32	0.955	0.915
4	EVV	0.598	0.970	130	0.831	0.864

* Columbia, MO; St. Louis, MO; Belleville, IL; Evansville, IN.

of the first location with the conditional probabilities in the last column in the table. The obtained value, 0.359, is lower than the observed value by 0.014. This is the greatest error in computing 18 combinations in Table 6 of Lund and Grantham (1980).

Some drawbacks of the method presented in this section and that proposed by Lund and Grantham became more evident in the following hypothetical example. Let us assume the existence of a triplet of locations arranged as the vertices of an equilateral triangle, and another one in which the locations are distributed along a straight line. The distances between neighboring locations in both triplets are the same. Let us assume further that the climatic frequencies of a weather event at all locations are the same and equal to 0.5. Both methods yield equal estimates of joint probabilities for each triplet because they fail to take into account the fact that the locations are distributed in a different way. Since two different patterns of distribution are dealt with, the same joint probabilities for both triplets can hardly be expected.

It seems logical that a better solution of this problem could be possible through the application of some approximate formula for the multinormal distribution. This should take into account correlations of all possible pairs of locations in a given combination and not only correlations between nearest neighboring locations.

5. Conclusion

The values of the BND model in the research of spatial and temporal persistence of various weather events have been presented in a number of papers. Satisfactory results are obtainable by applying other models which may be simpler for computation. By comparing the parameters of the various models, it is impossible to get a clear picture of the persistence of a particular weather event and its dependence on climate, season, time of day, etc. The main purpose of these comments is to emphasize the need to adopt the BND model in the examination of the persistence of weather events. This would allow comparison of results obtained in different parts of the world. The BND model seems to be more promising, if for no other reason than because its title includes the word "normal".

REFERENCES

Gringorten, I. I., 1971: Modeling conditional probability. *J. Appl. Meteor.*, **10**, 646–657.

Lund, I. A., and D. D. Grantham, 1979: Estimating the joint probability of a weather event at two locations. *J. Appl. Meteor.*, **18**, 27–33.

——, and ——, 1980: Estimating the joint probability of a weather event at more than two locations. *J. Appl. Meteor.*, **19**, 1091–1100.

McAllister, C. R., 1969: Cloud-cover recurrence and diurnal variation. *J. Appl. Meteor.*, **8**, 769–777.

McCabe, J. T., 1968: Estimating conditional probability of events having marked diurnal variability. *Preprints, First Statistical Meteorological Conf.*, Hartford, Amer. Meteor. Soc., 163–172.

Sheppard, W. F., 1898: On the application of the theory of error to cases of normal correlation. *Phil. Trans. Roy. Soc. London*, **A192**, 101–167.