Mausam. (1995). **46**, 1. 35-40

551.509.22 : 551.553.21 : 551.577

Skills in prediction of 5-day monsoon rainfall by using self-analogues

ASHWINI KULKARNI and S. V. SINGH
Indian Institute of Tropical Meteorology. Pune
(Received 13 April 1992. 3 September 1993)

सार — भारत में मानसून के दौरान होने वाली 5 दिवसीय वर्षा के पूर्वानुमान में स्व-एनालॉग की क्षमता की जांच की गई है। एक वस्तुनिष्ठ कंप्यूटरीकृत पद्दति की मदद से, 1901 से 1980 तक, 80 वर्षों में जुलाई-अगस्त के दौरान, 5 दिवसीय वर्षा के कुल 960 मानचित्रों में से वर्षा के एनालॉग ढूंढने का प्रयास किया गया। देश के पश्चिमी भागों के लिए उच्चतर मानों सहित न्यून परन्तु धनात्मक दक्षताओं (स्किल्ज) का पता चला है। एनालॉगों के समूह के प्रयोग से दक्षताओं में मामूली सुधार दृष्टिगोचर होता है।

स्व-एनांलॉग तकनीक का पूर्वानुमान प्रदान करने की गुणवत्ता के आकलन हेतु, इस तकनीक का प्रयोग 1987 के सूखा वर्ष और 1988 के अच्छे मानसून वर्ष के अध्ययन के लिए किया गया है।

ABSTRACT. Potential of rainfall self analogues in forecasting 5-day monsoon rainfall over India is examined. 5-day rainfall maps during July-August of 80 years (1901-1980) (total 960) are searched for the analogues by an objective computerised procedure. Low but positive skills with higher values over western parts of the country have been obtained. Use of group of analogues is found to improve the skills though only marginally.

The self analogue technique has been applied for 1987 (drought year) and 1988 (good monsoon year) to assess its predictive value.

Key words — Pentad Rainfall. Self-analogues. Skill scores.

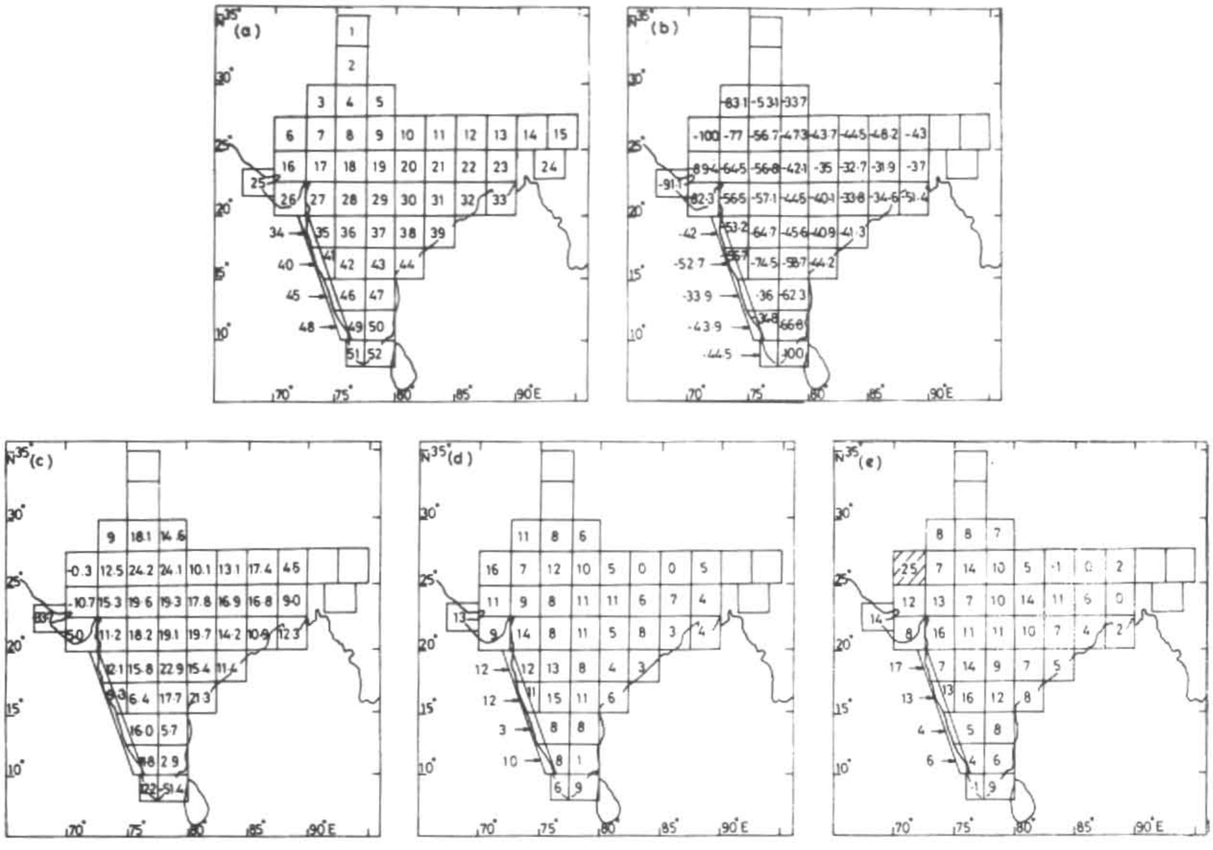
1. Introduction

The Indian monsoon rainfall shows large interannual and intra-seasonal variability, which has well publicised effect on the Indian economy. For the purpose of agricultural planning, therefore, the forecasts of weather on medium range scale (4-10 days) are very important.

The forecasts on medium range scale involve prediction of not only the intensity and movement of the current weather systems (*i.e.*, weather systems observed on the synoptic chart) but even the growth and movement of the second generation systems. Success in prediction of weather by the numerical methods has been limited to 2-3 days in the tropics. Therefore empirical techniques have been developed by many scientists for prediction of rainfall on 5-day scale. These techniques include the effect of persistence, systematic evolution and other physical factors implicitly. They, automatically, include the effect of the geographical factors. A rather simple technique is that of analogues, which has been employed by several researchers and operational meteorologists. This technique is simple but requires large data sets. We are examining here the skills of this analogue method in forecasting 5-day rainfall. The basic principles involved in analogue prediction are described in section 2 and the data used in section 3. The method followed is illustrated in section 4 and the results have been presented in section 5. Section 6 assesses the predictive value of the self-analogue technique. The technique is applied for recent contrasting monsoon years 1987, a drought year and 1988 and 1990, the good monsoon years. Finally in section 7, important conclusions are given.

2. Principles and use of the analogue method

The word analogue refers to two meteorological states resembling closely. The assumption involved in analogue prediction is that similar initial situations evolve similarly and give rise to similar weather. The dynamic solution of the initial state weather condition is implicit in analogue prediction. The method depends on how the nature itself has integrated its initial states in the historical records. A major difficulty is that the identical situations seldom occur and similar situations



Figs. 1 (a-e). (a) Location of 52 blocks (No. 1, 2, 14, 15 & 24 not used), (b) lower tercile limit (47 blocks), (c) upper tercile limits (47 blocks), (d) skill scores achieved using single best analogue and (e) skill scores achieved using 12 best analogues [actual value = figure value × 10^{-2} for (d) & (e)]

result in widely different situations. The analogue can be selected on the basis of circulation indices or fields at one or more levels but the more features we consider the more difficult it may become to select good analogues from a limited data set, necessitating longer data sets.

Csima *et al.* (1991), Rimba and Boroneant (1991) have used analogue technique for forecasting. Lorenz (1969) used the analogues for examining predictability. The experience has shown that even the best analogue may not result in similar weather, because it differs from the current situations in terms of several unaccounted factors and that the weather is a stochastic process. Also it may not be possible to get a good or satisfactory analogue in a large number of cases. To encounter these problems use of group of analogues has been suggested by Gruza *et al.* (1976). According to these authors a group of 10 or 20 analogues provides probabilistic and more skilled forecasts than the single analogue. While for predictions on daily scale a few good analogues may be found, for predictions on monthly scale the data may be inadequate to provide for more than one good analogue. Paul *et al.* (1982) have used the analogue technique for forecasting daily monsoon rainfall distribution over various parts of the country. Datta *et al.* (1981) have used it to forecast the tracks of tropical cyclone.

3. Data

The daily rainfall data of 365 stations uniformly spread over the country are obtained for a period of 80 years (1901-1980) from the India Meteorological Department, Pune. From these station data the spatial averages have been prepared for 52, 2.5° lat./long. blocks Fig. 1 (a) covering contiguous country. The density of rainfall stations is about 7 stations per block. To remove high frequency oscillations, 5-day averages are prepared for these 52 blocks. We have considered the data for 12 pentads for July and August only. Using these pentad data for blocks, Percentage Departure from Normal (*i.e.*, long term average) (PDN) or standardized departures have been determined. Block numbers 1, 2, 14, 15 & 24 (extreme eastern blocks) have not been used in the analysis. So, here we deal with only 47 blocks.

4. Method

In analogue prediction we usually search for analogues in the circulation fields and/or some other physical factors like surface boundary conditions and the weather following the analogue is considered as forecast weather. For fine tuning the circulation at many levels and past history may also be considered. The analogue forecast procedure assumes that the relationship between circulation and rainfall is perfect. which is far from the reality. Therefore. in an analogue procedure there are 3 factors responsible for the loss of predictive skill :

(*i*) perfect or good analogue may not be found.

(*ii*) apparently perfect analogue may not evolve similarly. and

(*iii*) the relationship between circulation and rainfall is far from perfect.

The chances of finding good analogue increase with the length of historical data. Although the rainfall is more spatially variable than the circulation quantities. still we hope that reasonably good analogues will be obtained because of large length of data. Therefore. we look for the analogues in the rainfall field itself and call them as self-analogues. This method of self-analogues automatically eliminates the last step in analogue forecasting. namely. the interpretation of the analogue in terms of rainfall.

For the purpose of finding the analogues the basic rainfall data has been expressed in terms of PDN for each block. This removes the seasonal cycle and singularity of any kind and puts each pentad on equal footing. A variety of similarity measures may be adopted like map-to-map correlation. distance between principal components (Koizumi *et al.* 1990) or some other indices of circulation (Paul *et al.* 1982) for the search of analogues. Here we try to get self-analogues of rainfall field by using map-to-map correlation method.

4.1. *Map-to-map correlation method*

We have 960 maps (12 pentads × 80 years) each consisting of 47 values corresponding to blocks. These 47 values of every map are correlated with 47 values of every other map. Thus. we get the matrix. $R = (r_{ij})$. $i = 1, \ldots, 960$; $j = 1, \ldots, 960$ of map-to-map correlations.

For any map *i*. the best analogue is the map which has maximum correlation with map *i*. The second best analogue is that which has second maximum correlation with map *i* and so on. The forecast. then. consists of the PDN values for the pentad following the single best analogue or the average PDN values of the pentads next to the selected number of ranked analogues. The forecasts are made only for 11 pentads in each year and the analogues are also searched from only these 11 pentads excluding the 12th pentad since it will require the rainfall for the 13th pentad for verification. Thus the observed pentad runs from 1 to 11 and the verification pentads from 2 to 12. The forecasts are verified against the observed PDN values in the next pentad through a 3 × 3 contingency table. The three classes of the contingency table are equiprobable and the tercile limites. *i.e.*. the lower and upper limit of the middle class are determined by arranging 960 (12 pentads × 80 years) PDN values for every block in ascending order. Then the values of this ordered array below which 33% of the values lie (*i.e.*. 320th value) gives the lower limit of the middle class and the value above which 33% of the values lie (*i.e.*. 640th value) gives the upper limit for the middle class. These limits are shown in Figs. 1 (b & c) respectively. There are a total 11 (pentads) × 80 (years) = 880 forecasts made. As mentioned above the observed and predicted PDN are classified in a 3 × 3 contingency table. We then compute the skill scores (*SS*) from these tables defined as follows :

$$SS = \frac{C - E}{T - E}$$

where.

C — The number of forecasts in correct category.

E — The number of forecasts expected to be correct by chance alone.

In the present case. since the marginal totals are 33%. these will be 33%.

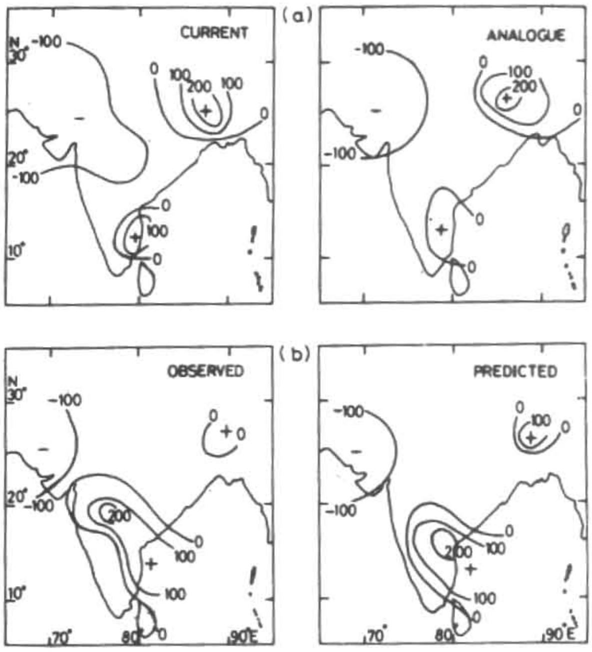
T — The total number of forecasts.

The value of skill score (*SS*) is 0 if $C = E$. *i.e.*. the number of correct forecasts are equal to those expected by chance alone. It is 1 when $C = T$. *i.e.*. when all forecasts are correct.

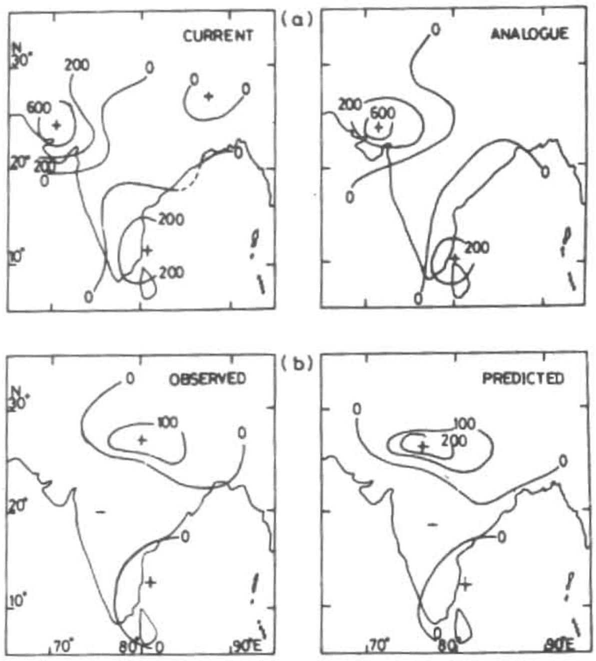
We have used single best analogue as well as combination of 3. 5. 7. 10. 12. 15 ranked analogues for making forecasts.

5. Results and discussion

The skill scores achieved by using single best analogue are shown in Fig. 1 (d). It can be seen that the skills are positive but not so satisfactory and.



Figs. 2 (a & b). (a) Best correlated current and analogue pentad patterns and (b) corresponding observed and predicted pentad patterns for 1987



Figs. 3 (a & b). (a) Best correlated current and analogue pentad patterns and (b) corresponding observed and predicted pentad patterns for 1988

perhaps, not usable. Spatially the skills are higher over the western parts of India. This distribution of skills is consistent with the skills achieved through regression equation (Singh *et al.* 1978) or through extrapolation of the rainfall fields by empirical orthogonal function analysis (Singh and Kripalani 1986). It should be added that the low frequency oscillations, namely, the Madden-Julian oscillations and persistence of 5-day rainfall is strongest over this region. This might have been responsible for the higher skills over the western parts of the country. Another reason for higher skills over this region may be our use of PDN maps whose patterns may be predominantly controlled by the dry regions. There is slight improvement in skills for group of analogues. About 80% of the blocks keep on improving the skills as the group size is enlarged with saturation in skills reaching for 12 best analogues. The *SS* achievable due to 12 best analogues are shown in Fig. 1 (e). The maximum *SS* as achieved by using 12 best analogues is 0.25 for the block number 6 [Hatched block in Fig. 1 (e)]. According to the formula of *SS*, a *SS* of 0.25 means that about 50% of the forecasts are in correct category.

We have reasoned into the possible cause of the low skills. One possible suspect is the level of resemblance of the analogues. The frequency distribution between the actual and the best analogues rainfall charts is given in Table 1.

TABLE 1

The frequency distribution between the actual and the best analogues rainfall charts

Correlation coefficients (CC)	Frequency	Percentage
0.4 - 0.5	2	0.23
0.5 - 0.6	47	5.34
0.6 - 0.7	228	25.91
0.7 - 0.8	362	41.14
0.8 - 0.9	212	24.09
0.9 - 1.0	29	3.29
Total	880	100

We can see from Table 1 that over 95% of the analogues charts have correlation coefficients (CC) more than 0.6 which is quite high and the poor skills are not because of the level of similarity but because of the possibility of lack of organised evolution of the rainfall fields having median correlations as high as 0.6. We recall that, the PDN fields are tuned to the lower rainfall regions. Therefore, as an alternate, we consider the charts of standardised rainfall anomalies. Use of standardised anomalies show marginally higher skills than those obtained through charts of PDN. In this case we obtain the highest skill of 0.18 as against 0.16 observed by using PDN for the case of single best analogue. The areas of high skills remain similar.

6. Assessment of self-analogue technique

In order to assess the predictive value of self-anlogue technique, it is applied for recent contrasting years, *i.e.*, good monsoon (1988) and bad monsoon (1987) years.

TABLE 2

Correlation coefficients between the current & analogue patterns and observed & predicted patterns

Current		Analogue		C.C.	Observed		Predicted		C.C.
Year	Pentad	Year	Pentad		Year	Pentad	Year	Pentad	
1987	7	1905	5	0.75	1987	8	1905	6	0.61
1988	8	1916	11	0.75	1988	9	1916	12	0.47

For 1987 the level of correlation for the best analogue is as high as 0.75. Pentad No. 7 has maximum correlation with its analogue, *i.e.*, pentad no. 5 of 1905 [Fig. 2 (a)]. The observed and predicted patterns [Fig. 2 (b)] also match very well with correlation 0.61. Pentad no. 7 of 1987 shows negative departures all over the country except northeast India and southeastern tip part. Its analogue is also similar with less magnitudes of PDN values. The observed pattern (pentad no. 8 of 1987) has positive departures in the northeast corner and southeastern tip of the country. The predicted pattern (pentad no. 6 of 1905) shows similar pattern.

For recent good monsoon year 1988, the level of best correlation is not as high as that for bad monsoon year 1987. The maximum correlation between pentad and its analogue is achieved for pentad no. 8 of 1988. The analogue is pentad no. 11 of 1916 [Fig. 3 (a)]. The correlation coefficient is 0.70. The correlation between observed pentad (pentad no. 9 of 1988) and predicted pentad (pentad no. 12 of 1916) is 0.47 [Fig. 3 (b)]. For pentad no. 8 of 1988 the departures are positive over northeastern parts of the country. This corresponds to mid-tropospheric cyclonic flow. Due to western disturbances, in next pentad (observed) positive departures get shifted eastward to the foothills of Himalaya, resulting in break monsoon situation over the country. The analogue pattern is similar. In the observed pattern, positive departures are there only in northwest corner and southern tip of India. The predicted pattern is almost similar to the observed one.

Thus, it is seen that the analogue technique is quite satisfactory technique for prediction purposes.

7. Conclusions

From the above results we see that :

(*i*) Low but positive skills are achievable in prediction of 5-day rainfall by analogue technique. This suggests that the rainfall fields possess some systematic evolutionary signal.

(*ii*) The skills are somewhat higher in the western parts of the country. It is over this region, the intensity of the Madden-Julian low frequency (30-60 day period) oscillations is highest. Therefore, the skill of these self-analogues seems to arrive from Madden-Julian oscillations. The magnitude of the skills is comparable to that arrived by extrapolating the Madden-Julian oscillations.

(*iii*) There is some improvement in skill when group of analogue is considered. A group of 12 analogues gives maximum skill.

We note, however, that the skills presented here may not be the best skills achievable through analogue prediction. Analogues in circulation fields may give better skills but this will need large data of circulation fields. The rainfall fields could also be treated differently. One could consider the EOFs of the rainfall and the Euclidean distance between the charts in terms of principal component scores. Alternatively, the spatial patterns of rainfall fields could be smoothed by considering the original values corresponding to the terciles of rainfall. It may also be of interest to investigate if the skills are different and better for certain types of current situations. It is hoped that investigations into this direction may improve the skills in prediction. Such skills may be used as essential elements of an operational forecast set up and the forecasts prepared from other means may be suitably combined with the forecasts prepared by the analogue method. The analogue forecasts can also be used as a basis against which skill of the forecasts prepared from more involved methods could be verified.

Acknowledgements

The authors are grateful to Shri D. R. Sikka, Director and Dr. S. S. Singh, Head, Forecasting Research Division for the encouragement and the facilities provided. The original rainfall data were obtained from the India Meteorological Department and the spatial averages for the blocks were prepared by Dr. R. H. Kripalani of the Indian Institute of Tropical Meteorology. Authors thank him for his kind cooperation.

References

Csima, G., Kala, M. and Mallor, A. J., 1991, Status and prospects of long-range forecasting at Hungarian Meteorological Service. Proc. ICTP/WMO Intern. Tech. Conf. on Long-Range Weather Forecasting Research, Trieste, Italy, 8-12 April 1991, WMO/TD No. 395, pp. 65-66.

Datta, R. K., Bansal, R. K. and Bindra, M. M. S., 1981, "Verification of forecast of movement of cyclones by analogue methods," *Vayu Mandal*, **11**, 1-2, pp. 14-18.

Gruza, G. V., Rankova, E., Ya and Esterie, G. E., 1976, A scheme of the statistical prognosis and utilisation of group of analogues,Gidrometeorizdat, Trans. No. 13, pp. 5-14.

Livezey, R. E. and Barnston, A. G., 1988, "An operational multifield analogue/anti-analogue prediction system for United States seasonal temperatures-I : System design and winter experiments., *J. Geophy. Res.*, **93**, D9, pp. 10953-10974.

Lorenz, E. N., 1969, "Atmospheric predictability as revealed by naturally occurring analogues", *J. Atmos. Sci.*, **26**, pp. 636-646.

Koizumi, K., Kurihara, K. and Maeda, S., 1990, Probability forecast of monthly mean temperatures using analog/anti-analog method. Proceedings of the WMO Training Workshop on Diagnosis and Prediction of Monthly and Seasonal Atmospheric Variations, Nanjing, China, 18-19 Oct., 1990, Long-Range Forecasting Res. Rep. No. 13 WMO/TD No. 363, pp. 87-90.

Paul, D. K., Majumdar, V. R. and Sikka, D. R., 1982, "Analogue in the southwest monsoon," *Mausam*, **33**, 1, pp. 121-130.

Rimba, N. and Boroneant, C., 1991, "Typifying the evolution in the Meteorological parameters by using analog method". Proc. of the ICTP/WMO Inter. Tech. Conf. on Long-range Weather Forecasting Research, Trieste, Italy, 8-12 April 1991, WMO/TD No. 395, pp. 229-230.

Singh, S. V., Mooley, D. A. and Prasad, K. D., 1978, "Prediction of 10-day summer monsoon rainfall over India", *Arch. Met. Geoph. Biokl.*, Ser. A., **27**, pp. 317-331.

Singh, S. V. and Kripalani, R. H., 1986, "Application of Extended EOF analysis to inter-relationships and sequential evolution of monsoon fields," *Mon. Weath. Rev.*, **114**, pp. 1603-1610.