Nordic Hydrology, 32 (2), 2001, 85-98

Spatial Rainfall Pattern Identification by Optimum Interpolation Technique and Application for Turkey

Zekai Şen
Technical University of Istanbul, Maslak 80626, Turkey

Zeyad Habib
University of Al-Fateh, 13488 Tripoli, Libya

The spatial interpolation of hydrometeorological data aims at estimating the value of rainfall at a given site based on the observation at neighboring stations. This is a problem of operational hydrometeorology that is regularly encountered in the spatial estimation procedures. In this paper, an optimum interpolation method is validated for the analysis of monthly rainfall amounts over Turkey. Hence, a two dimensional statistically optimum interpolation scheme is generated for the analysis of monthly rainfall amounts over Turkey. Monthly spatial correlation functions are used for the analysis and modelling of the spatial monthly rainfall variability.

Introduction

One of the fundamental purposes of objective analysis is the transformation of hydrometeorological observations from irregularly distributed sites to gridded networks for use in the numerical prediction schemes. Another purpose is to assign a prediction value to any point without measurement through a suitable optimum interpolation technique. Hence, in practice, there are two-fold purposes in describing the regional variability of hydrometeorological fields by objective analysis, i) values at regular grid nodes serve as the initial data for numerical forecast models; and ii) coordinates of a rather great number of isoline points help to construct equal-value lines such as isohyetal maps.

There are various methods for data interpolation from measurement stations to any desired point (Schlatter1988). These are:

i) Surface fitting methods: The first objective analysis method in meteorology was the surface fitting type devised by Panofsky (1949). In this method, the analysis value is represented as a continuous mathematical function which fits irregularly spaced observations. Among these methods are the polynomial interpolation (Panofsky 1949); orthogonal polynomials (Dixon 1969, 1976); splines (Fritsch 1971); and finally, spectral approaches (Flattery 1970).
ii) Empirical linear interpolations: The value of any variable at a particular location is estimated as a weighted sum of observations. Among such interpolation techniques are the iterative successive correction method (Cressman 1959); and Barnes analysis (Barnes 1964).
iii) Statistical objective analysis: These are estimation methods at any desired point where spatial correlation structure determines the weights applicable to each observation. The major approaches in this category are the optimal interpolation (Gandin 1963); the covariance models for atmospheric variable (Buell 1958); adaptive filtering (Kalman 1960; Şen 1980, 1983); and recently, the cumulative semivariogram method (Şen 1997).
iv) Variational techniques: These include more mathematical abstraction than other methods and two of them are, the incorporation of dynamic constraints (Sasaki 1958); and the fitting models to data at different times (Ledimet and Talagrand 1986).

On the other hand, geostatistical methods were shown as superior to these methods by Tabios and Salas (1985). In contrast, Hevesi *et al.*(1992) presented an application of multivariate geostatistics to problems of estimating areal average precipitation. Their method is based on the classical semivariogram and kriging techniques application. However, Şen and Habib (1998) presented point cumulative semivariogram method of spatial precipitation assessments in mountainous areas.

It is the main purpose of this paper to present the application of optimum interpolation technique by considering spatial correlation functions obtained from monthly rainfall data set scattered all over Turkey.

Rainfall Data in the Study Area

Turkey is the region considered in this study and its location is shown in Fig.1. The stations cover almost all the area with comparatively denser concentration in the north-west of Turkey including the European part. The grid for this analysis is on a polar stereographic projection oriented along the 35° E longitude line (Habib 1993). The data employed in this study are the mean monthly precipitation records collected from the statistics published by Turkish State Meteorological Department (DMI).

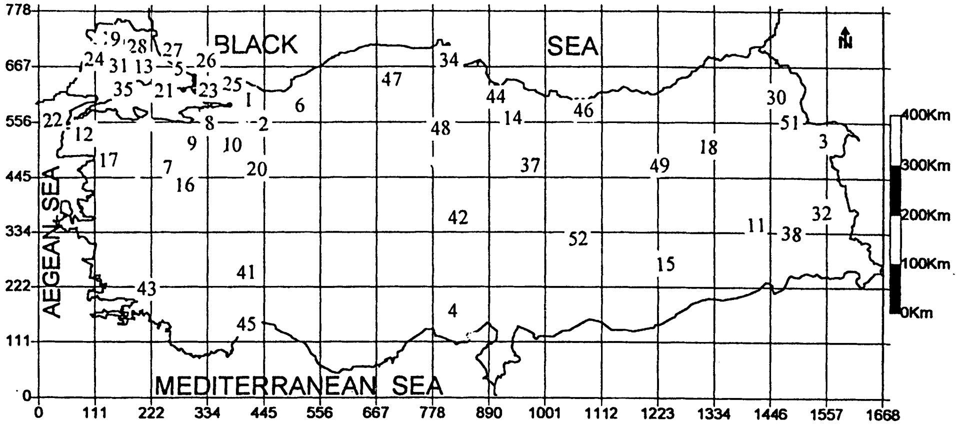


Fig. 1. Grid points and station locations in Turkey.

Fifty-two stations for the 30-year period of 1956-1985 were selected for this study which amounts to 18,720 monthly observations. Although these stations might not seem to represent sufficiently the whole of Turkey, an area of almost 780,000 km^2, they give precipitation variations on a rather coarse scale. However, smaller scale precipitation variations can be obtained by the same method as more stations are considered in the future.

In general, there are 7 different climatological regions in Turkey. Four of these regions include coastal areas, namely, along the Black Sea in the north, Aegean Sea in the west, Mediterranean Sea in the south and a transition zone between the Aegean and the Black seas as the Marmara Sea region in the northwest. Except the Black sea zone, all the other regions are similar to the Mediterranean type of climate with modifications towards the north. On the other hand, three inland regions of climate are the central Anatolian part with steppe type of features, transitional semi desertic region in the southwest and finally, the very cool region in the summer with severely cold climate during the winter in the eastern part of Turkey where average elevations above the mean sea level reaches 3,000 m.

In general, temporal rainfall occurrences show temporal and spatial correlation structures. In any spatial prediction, the problems are with spatial (areal and elevation) rainfall modeling and with the transfer of information from available irregular measurement sites to regular grid nodes or to any desired point of interest. In general, rainfall amount at any location bears some relationship to nearby locations. The strength of the relationship normally decreases as the distance between the location increases. Accurate objective analysis of rainfall is extremely difficult because of the following reasons as, i) precipitation appears as a discontinuous random field; and ii) rainfall observations are reported at irregularly scattered sites.

The optimum interpolation method for the objective analysis of rainfall produces the optimum solution in the sense that the interpolation error is minimized on the average. This method allows for the extraction of as much useful information as possible from the measurements. The problems associated with optimum interpolation analysis of rainfall are that, i) it requires a knowledge of covariance which is often not known and thus an estimate is necessary from the available data. Establishing such an estimate is often fraught with difficulty as a host of local factors are involved, and ii) essentially, one must determine the priority about which station measurements are significantly correlated with the value at the point of interpolation *i.e.*, one must determine a region of influence around the interpolation point.

Theoretical Optimum Interpolation Approach

In the practical applications of the optimum interpolation such as the analysis of rainfall, one uses the climatological mean as the first guess value. Hence, the following expression is considered for the interpolation point value (Habib 1999)

$$Z_k^a = \bar{Z}_k + \sum_{i=1}^{n} W_i \, (Z_i^0 - \bar{Z}_i) \tag{1}$$

where Z^a_k and Z^0_i are the calculated and observed rainfall values corresponding to the arithmetic averages of Z_k and Z_i, respectively, at interpolation point k and observation stations, $i = 1,2,...n$; and W_i's are the interpolation weights. Here, k is the index of the location for which the interpolation is to be done. In order to calculate these weights, the interpolation formula can be obtained by multiplying both sides of Eq.(1) by $(Z^0_j\text{-}Z_j)$ and taking the expectation of both sides leads to

$$\sum_{i=1}^{n} W_i \rho_{ij} = \rho_{kj} \tag{2}$$

where ρ_{ij} is the spatial correlation coefficients between stations i and j, and ρ_{kj} between stations k and j. This equation is valid in the case when $E(Z_k) = E(Z_k{}^a)$ which implies unbiased estimator. In short, interpolation weights W_i, are dependent on the statistical structure of the spatial correlation function (SCF) of the rainfall records at irregular sites. Once the SCFs are obtained from the available data, then the value of ρ_{kj} can be read from these function depending on the distance between k and j and consequently, the only unknowns in Eq. (2) are the weights which can be calculated from the set of n linear equations. The expected analysis error, ε_{kj}, at grid point k that results from the introduction by using information at location j can be expressed as (Habib 1999)

$$\varepsilon_{kj} = 1 - \sum_{i=1}^{n} W_i \rho_{ki} = 1 - \rho_{kj} \tag{3}$$

Most often in practice $0 < \rho_{kj} < 1$ and therefore $0 \le \varepsilon_{kj} \le 1$. It is obvious that the expected error does not depend directly on the observed values but again on the spatial statistical structure of the rainfall amounts. In the light of the aforementioned discussion, the following optimum interpolation procedure steps are presented:

i) specify the geographical locations (longitude and latitude) of interpolation points and observations stations,
ii) compute the background error correlations which correspond to the differences between observed and average values,
iv) find a suitable model for the background error SCF,
v) select the measurement sites that will influence the interpolation point,
vi) solve the system in Eq. (2), and obtain the interpolation weights,
vii) compute the interpolation point value by using Eq. (1),
viii) calculate the expected analysis error at the interpolation point by using Eq. (3),
iv) repeat steps (v) to (viii) for all desired interpolation points.

Application

The size of the correlation matrix in Eq. (2) is directly related to the number of measurement sites that influence each interpolation point. The choice of a search strategy for the number of influencing stations that should be included in the interpolation procedure is an important consideration in any approach of objective analysis methods. The most common approach in choosing the stations that contribute to the interpolation is to define a search neighborhood within which all the available stations are used. In this study, a simple search strategy is adopted using all stations within a circular search neighborhood with a limited radius of influence.

During the course of this study, it is noted that the distance based SCFs do not yield good predictions. This is due to the fact that such a procedure gives rise to unnecessary smoothing in the final predictions causing large prediction errors. The distance based SCFs are not reliable especially for moderate or large distances. Instead of distance based SCF, nearest station number based SCF approach is preferred in sparsely covered areas. It is, therefore more convenient to consider the nearest station numbers that cause the least prediction error. On the other hand, the SCF changes between months and seasons.

In order to fix this idea with the data at hand, the change of expected mean square error is plotted *versus* the number of neighboring stations for each month. However, it appeared that such graphs are very similar to each other. Therefore, Fig. 2 shows such a relationship for month January only. It is seen from this figure that on the average the number of influencing stations is equal to 4. The expected analysis error is a by-product of the optimum interpolation analysis procedure as already stated in Eq. (3). The presence of any data type which affects the analysis at a interpolation point results in a reduction of the analysis error at that point. If no data are present,

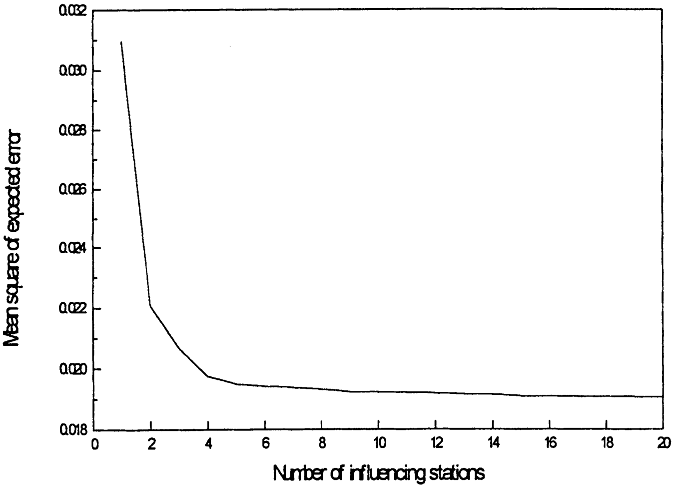


Fig. 2. Expected mean square error-number of influencing stations relationship.

the expected analysis error remains unchanged from its initial value. The calculation and modeling of the various forecast and observational errors result in an analysis that is not necessarily optimum (Lorenc 1981). The analysis error is thus a measure of what the analyzer "Thinks" the error is (Schlatter and Branstator 1987).

Fig. 3 provides maps of the estimated expected error of rainfall calculated from Eq. (3) for each month. It is obvious that there are a number of areas where expected error changes quite rapidly. For instance, there is a bull's-eye feature centered

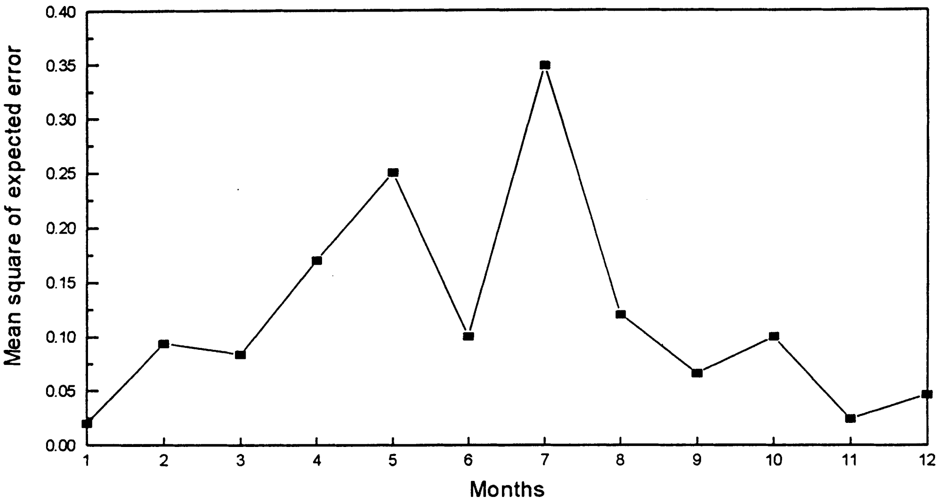


Fig. 3. Monthly expected analysis error.

over the northwestern part of the study area in all months which is enclosed by the minimum expected error. The contours surrounding this region represent the change in forecast error from the data-spare to surrounding plentiful data areas. There is also a similar area of less circular contours surrounding the region at the north center of the study area near the Black Sea coast. It is possible to depict that contours occur less densely over terrestrial Turkey during September-February winter period. However, they are very dense from March to October. The reason for such a digression is due to the fact, that the rainfall occurrences are comparatively more sporadic (*i.e.*, regionally random) during March-October. This is also because of the convective rainfall occurrences in the area which are almost independent of each other. However, areal continuity of winter rainfall is a signature of cyclonic weather movements. Such continuity results in comparatively very small error amounts, say for instance, in January (Fig. 3) where errors vary between 0.05-0.15. On the contrary, in July error band varies between 0.20 and 0.70. In order to compare monthly error amounts over Turkey areal error averages are calculated and presented in Table 1. It is possible to note that the maximum mean square expected error appears in summer months (for example July is 0.3480) because the precipitation phenomenon is more discontinuous as explained above. On the other hand, the minimum mean square expected error is in winter months (for example January is 0.0193), because the precipitation in this season is areally extensive and more continuous.

Table 1 – Monthly mean square errors.

Month	MSE	Month	MSE
January	0.0193	July	0.3480
February	0.0935	August	0.1200
March	0.0834	September	0.0657
April	0.1700	October	0.0994
May	0.2500	November	0.0238
June	0.1000	December	0.0463
		Average	0.1141

Cross Validation

The accuracy of the optimum interpolation method that is used in objective analysis of this study was investigated with the help of numerical experiments from the surrounding stations and in comparison to the interpolated values with the observed ones (Gandin 1965). Similarly cross validation techniques allow us to compare estimated and true values using only the information available in the measurement stations (Isaaks and Srivastava 1989).

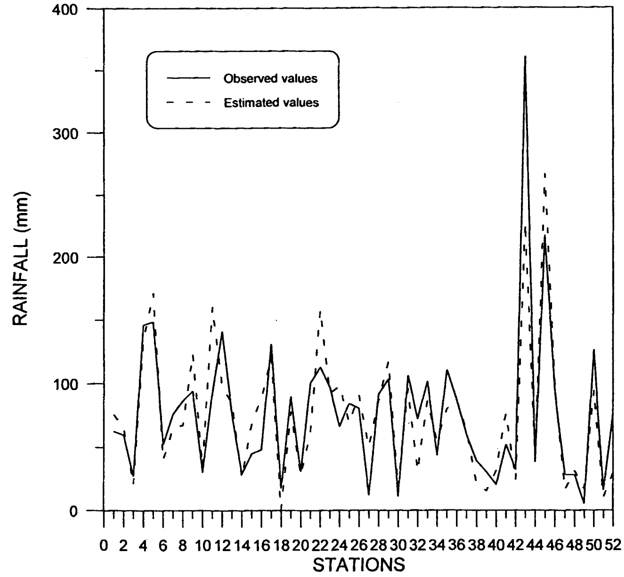


Fig. 4. Observed and estimated values for January 1984.

In a cross validation technique, the interpolation method is tested at the observation station by discarding temporarily the observed value at this station. In other words, this station is regarded as an interpolation point. The value at the same station is then estimated using the aforementioned optimum interpolation technique. Once the estimate is calculated, it is possible to compare it with the true value that was initially removed from the station. This procedure is repeated for all the available stations. Fig.4 provides the observed an estimated values for each station with 52 observation stations in the study area for January 1984. The choice of 1984 is just for the presentation of results spatially and such a choice does not have any criterion and similar conclusions and figures can be obtained for any desired year. It is to be noted from this figure that there is an overall good match, but some significant deviations occur at stations which lie within the less dense areas such as stations 12, 43 and 46.

Table 2 – Univariate statistics of optimum interpolation analysis compared to the observed values for January 1984.

	Observed value	estimated value
N	52	52
Mean	76.6	74.70
Minimum	5	0
Maximum	361	266.9
Range	356	267
Std.Dev.	58.37	53.90

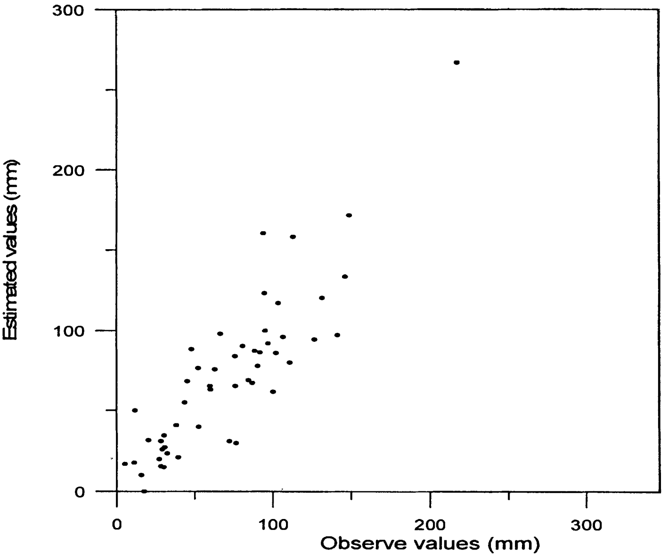


Fig. 5. Observed and estimated rainfall scatter diagram (January 1984).

However, in this case the analysis method has a slight tendency toward overestimation as seen from Table 2. Additionally, comparisons of the standard deviation values indicate that the estimated values have less variability than the observed ones. However, this reduced variability in the estimated values is often referred to as smoothing which is the feature of the method.

A scatter plot of observed *versus* predicted values provides additional evidence on how well an estimation method has been performed. Fig. 5 shows scatter plot of observed *versus* estimated values for January 1984, which indicates that, there is a good relationship between estimated and observed values especially for less than 200 mm rainfall amounts. The correlation coefficient is a good index for summarizing how close the points on a scatter plot come to fall around a straight line. This is also supported by the root mean square error which is 29.9 and the mean percentage error is –9.07%. Furthermore, the correlation between observed and predicted values is 0.87.

In order to see the correspondence matching between observed and predicted monthly rainfall amounts, the monthly statistical parameter variations are presented in Figs. 6 and 7. It is seen that on the basis of averages, observed and predicted monthly values are very close to each other, in fact, with less than 1% error. Although there are more discrepancies for monthly standard deviations reaching up to almost 25%, they appear in November and December months only. The difference in the monthly standard deviations is due to the smoothing of the interpolation which can also be seen in cross validation procedure.

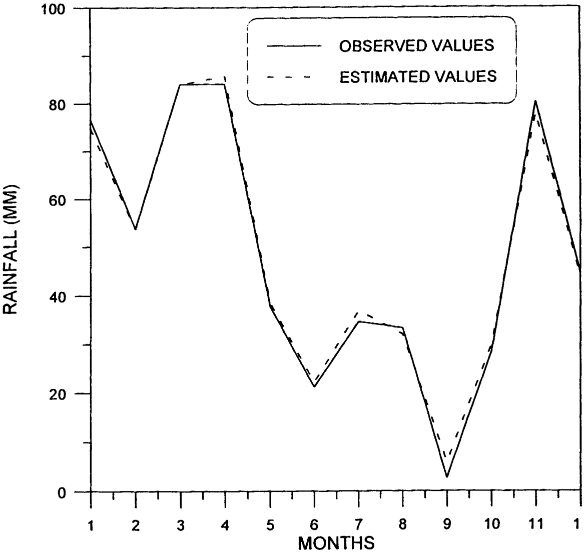


Fig. 6. Arithmetic mean of monthly rainfall.

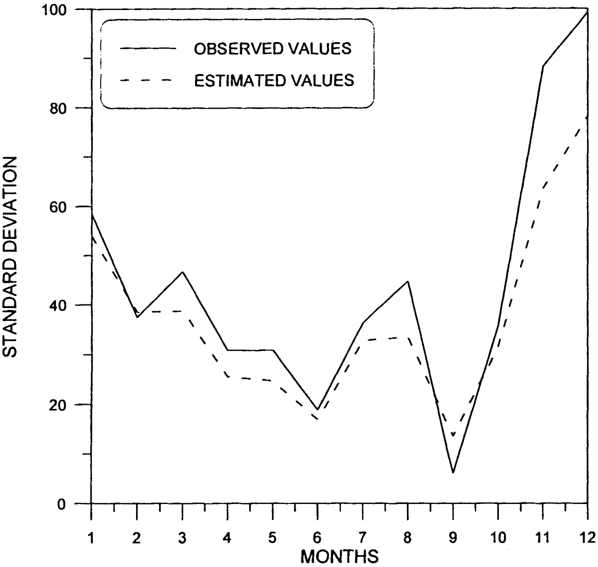


Fig. 7. Standard deviation of monthly rainfall.

Spatial Rainfall Maps and Interpretations

There are significant differences between the annual and monthly rainfall amounts in different parts of Turkey. Deviations from the annual or monthly areal average rainfall amounts are very much dependent on the location, season, continentiality, distance from the sea coast and elevation (Kadioglu 1997; Kadioglu *et al.* 1999).

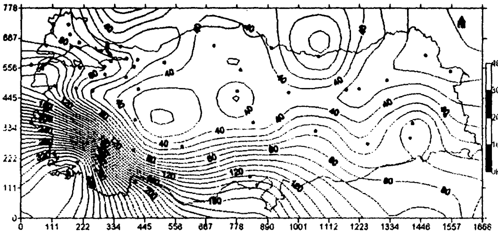
Optimum interpolation techniques yield average regional rainfall amounts, and therefore, their temporal variations can be visualized by monthly isohyetal maps as shown in Fig. 8. The following general features emerge from these isohyetal maps.

1) Comparison of these maps starting from January indicated that the maximum rainfall amounts appear in the southwestern part of the country. For instance, from May to October, maximum rainfall appears continuously in the north and northeastern parts, whereas during November-April period, the maximum amounts are along the southwestern corner at the intersection of the Mediterranean Sea with the Aegean Sea. The reason for the maximum rainfall occurrences in the southwestern corner during the winter months is due to the moist air from the Mediterranean sea.
2) There are two pronounced seasons in Turkey, wet and dry periods depending on the region. During summer seasons north and northwestern regions are dry.
3) The overall monthly average rainfall amount patterns, in general, show that the eastern parts of the Black Sea region get the excessive rainfall amounts.
4) Starting from May onwards, the rainfall amounts less than 100 mm remain in the south regions of Turkey. This contour line (100 mm) shifts to the outmost northern part in September.

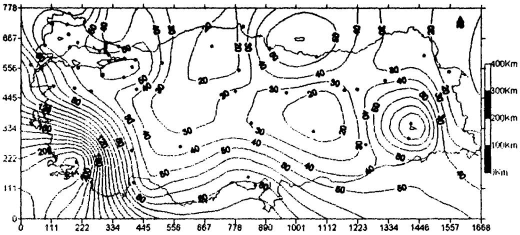
Conclusions

Spatial rainfall pattern identification in the case of sparse distribution of stations is possible through the nearest number of influencing station based spatial correlation function (SCF) rather that the distance based SCF concept. It has been emphasized in this paper that the distance based SCFs do not yield acceptable prediction errors, especially at moderate and large distances. However, in the case of a dense station network availability, both SCF approaches yield almost the same results.

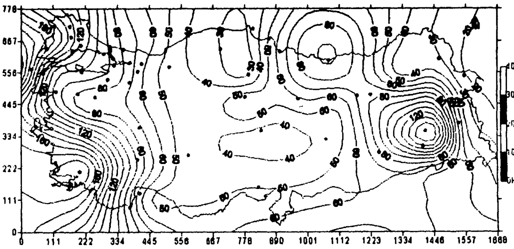
The optimum interpolation procedure as applied in this paper shows that the quality of the analysis is not very sensitive to changes in the maximum number of influencing stations which is found as 4 stations for Turkish monthly rainfalls. For station numbers over 4 the maximum mean square expected error (low accuracy estimation) appears in summer months because the precipitation is a more discontinuous field variable due to convective rainfall occurrences. Additionally, the minimum square error (high accuracy estimation) appears in winter months because the precipitation in this season is areally extensive and more continuous due to frontal rainfall occurrences.



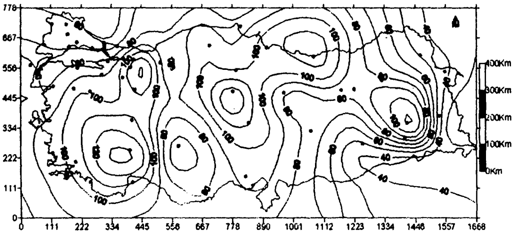
January



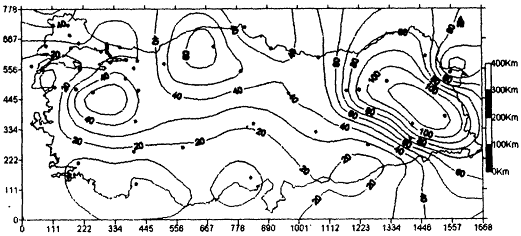
February



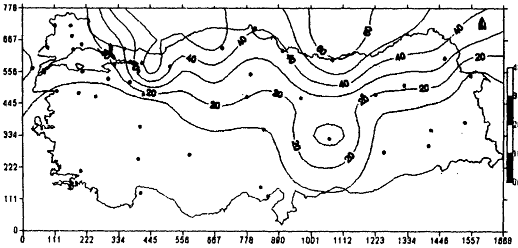
March



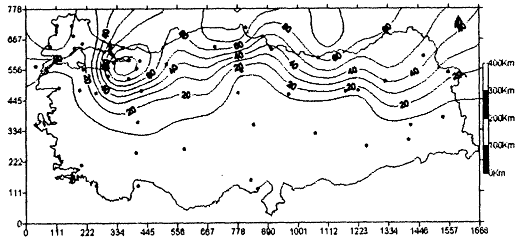
April



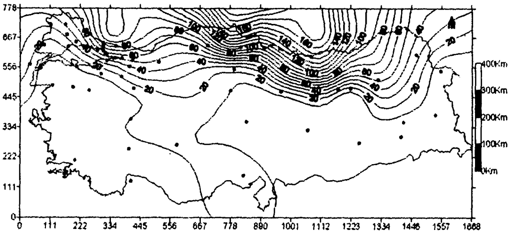
May



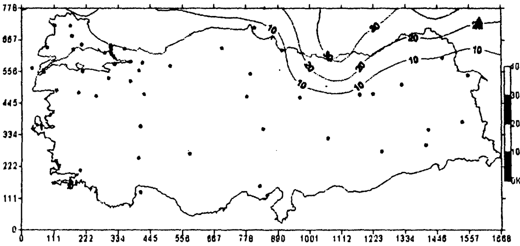
June



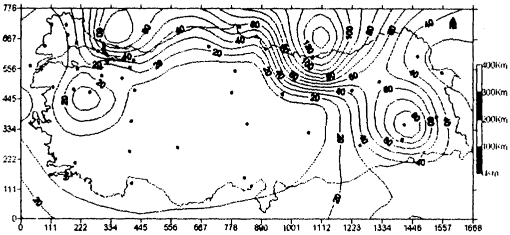
July



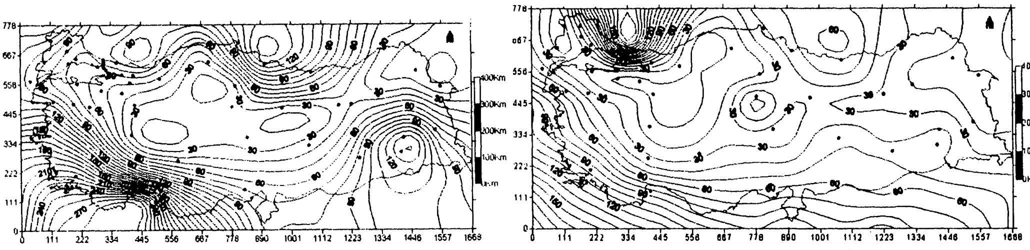
August



September



October



November

December

Fig. 8. Total rainfall contour map in 1984.

References

Barnes, S. L. (1964) A technique for maximizing the details in numerical weather map analysis, *J. of Applied Met., Vol. 3 (4)*, pp. 396-409.

Buell, C. E. (1958) The correlation between wind and height on a isobaric surface II: summer data, *J. of Applied Met., Vol. 15 (12)*, pp. 502-512.

Cressmam, G. Å. (1959) An operational objective analysis system, *Mon.Weath.Rev., Vol. 87(10)*, pp. 367-374.

Dixon, R. (1969) Orthogonal polynomial as a basis for objective analysis, Meteorological Office, Scientific Paper No. 30, Her Majesty's Stationery Office, London, 20 pp.

Dixon, R. (1976) An objective analysis using orthogonal polynomials, in Proceedings of the JOC Study Group Conference on four-dimensional Data Assimilation Paris, GARP Program on Numerical Experimentation, Report No. 11, pp. 73-85.

Flattery, T. W. (1970) Spectral models for global analysis and forecasting, Proceedings of the Sixth AWS Technical Exchange Conference, U.s. Naval Academy, 21 September, Air Weather Service technical Report 242, pp. 42-54.

Fritsch, J. M. (1971) Objective analysis of a two-dimensional data field by the cubic spline technique, *Mon.Weath.Rev., Vol. 99(5)*, pp. 1122-1143.

Gandin, L. S. (1963) *Objective Analysis of Meteorological Fields*. Leningrad, Gidromet; Translated from Russian, Jerusalem, Israel Program for Scientific Translations, 1965, 242 pp.

Gandin, L. S. (1965) Objective Analysis, Lectures on Numerical Short-range Weather Prediction, WMO, Regional Training Seminar. pp. 633-677.

Habib, Z. (1993) Objective analysis with point cumulative point semivariogram in meteorology. Unpublished M.SC Thesis, Istanbul Tech.Univ., Met. Dept., 87 pp.

Habib, S. (1999) Optimum interpolaltion modeling supported by the cumulative semivariogram for spatial-temporal meteorological variables. Unpublished Ph.D. Thesis, Istanbul Tech. Univ. 156 pp.

Hevesi, J. S., Istok, J. D., and Flint, A. I. (1992) Precipitation estimation in mountainous terrain using multivariate geostatistics, Part I: structural analysis . *J.Applied Met., Vol. 31*, pp. 661-676.

Isaaks, E. H., and Srivastava, R. M. (1989) *An Introduction to Applied Geostatistics*, Oxford, Univ. Press, Oxford, 561 pp.

Kadioglu, M., Erdun, H., Öztürk, N., and Sen, Z. (1999) On the precipitation climatology of Turkey by harmonic analysis, *Int. J. Climatol., Vol. 19,* pp. 1717-1728.

Kadioglu, M. (1997) Trends in Turkish Precipitation data. Int. Conf. Water Problems in the Mediterranean Countries, Nicosia, Northern Cyprus.

Kalman, R. E. (1960) A new approach to linear filtering and prediction problem, *J. Basic Eng., Vol. 82, series D (1),* pp. 35-45.

Ledimet, F. and Talagrand, O. (1986) Variational algorithms for analysis and assimilation of meteorological observation: Theoretical aspects, *Tellus, Vol. 38A (2),* pp. 97-110.

Lorenc, A. C. (1981) A global three-dimensional multivariate statistical interpolation scheme, *Mon.Weath.Rev., Vol. 109,* pp. 701-721.

Panofsky, H. A. (1949) Objective weather map analysis, *J..Met., Vol. 6(6),* pp. 386-392.

Sasaki, Y. (1958) An objective analysis based on variational method, *J.Met.Soc.Japan, vol. 36(3),* 77-88.

Schlatter, T. W., and Branstator, G.W. (1987) Experiments with a three-dimensional statistical objective analysis scheme using FGGE data, *Mon.Weath.Rev., Vol. 115,* pp. 272-296.

Schlatter, T. W. (1988) Past and present trends in the objective analysis of meteorological data for nowcasting and numerical forecasting, Eight Conference on Numerical Weather Prediction, American Meteorological Society, pp. J9-J25.

Tabois, G. W., and Salas, J. D. (1985) A comparative analysis of techniques for spatial interpolation of precipitation, *Wat.Resour.Bull., Vol. 21,* pp. 265-380.

Şen, Z. (1980) Adaptive Fourier analysis of periodic-stochastic hydrologic series, *J. Hydrol.,Vol. 46,* 239-249.

Şen, Z. (1983) Predictive hydrologic models in hydrology, *Nordic Hydrol., Vol. 14,* pp. 19-32.

Şen, Z. (1997) Objective analysis of cumulative semivariogram technique and its application in Turkey, *J. Applied Met., Vol. 36,* pp. 1712-1720.

Şen, Z., and Habib, Z. (1998) Point cumulative semivariogram of areal precipitation in mountainous regions, *J. Hydrol, Vol. 105,* pp. 81-91.

Received: 21 February, 2000
Revised: 3 August, 2000
Accepted: 10 October, 2000

Address:
Zekai Şen,
Technical University of Istanbul,
Civ.Eng.Faculty, Hydraulics Div.,
Maslak 80626 Istanbul,
Turkey.
Email: zsen@itu.edu.tr

Zeyad Habib,
University of Al-Fateh,
Faculty of Science, Dept. of Met.,
P.O.Box 13488 Tripoli,
Libya.