

OBJECTIVE ANALYSIS OF RAINFALL DATA FROM DIGITAL RADAR AND RAIN GAGE MEASUREMENTS

Witold F. Krajewski and Kenneth C. Crawford¹

ABSTRACT: Two devices are used to perform measurements of rainfall: digital radar and rain gages. Each of them gives different types of information. Digital radar data represent a spatial pattern of a storm but the point estimates may be biased through, for example, the reflectivity-rainfall transfer relationship. A standard rain gage optimally sited generally gives a very good estimate of rainfall at a point, but the measurement often is not representative over a larger area. A multivariate regression model has been developed to merge those two kinds of data into an optimal (in a statistical sense) estimate of rainfall. Such optimal estimates can then be used as input to hydrological models. Plans for operational implementation of the regression model is presented as well as results of example data analysis from the National Weather Service Radar at Oklahoma City. Computer requirements are discussed and some aspects of potential future improvements are mentioned. These include the possibility of accounting for the autocorrelation of errors and the inclusion of satellite data.

(KEY TERMS: radar; precipitation; rain gage; objective analysis.)

INTRODUCTION

The Hydrologic Research Laboratory of the National Weather Service (NWS) has many ongoing activities. One research effort is the Hydrologic Rainfall Analysis Project (HRAP), that addresses development of objective techniques for preprocessing, quality controlling, and operationally merging rainfall data from different sources such as multi-radars, rain gages, and satellites (Hydrologic Research Laboratory, 1980; Greene, *et al.*, 1979).

This research deals with the objective analysis of rainfall data from digital radars and standard rain gage stations. It will be shown later that satellite data also can be incorporated after it becomes available "on-line" in required form (digitized).

Our goal is the development of a relatively simple, from a computational viewpoint, method which can be implemented into the operational environment and produce final rainfall analyses to several end users, in particular NWS River Forecast Centers (Figure 1).

ON SITE PROCESSING

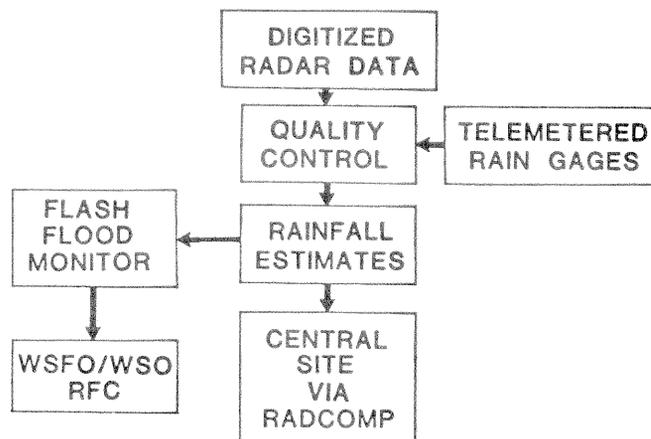


Figure 1. Block Diagram Illustrating Steps Involved in Rainfall Data Processing.

Our portion of producing a uniform grid point rainfall analysis from multiple data sources is illustrated by Figure 2. The grid has been especially designed for HRAP purposes and is fully described by Greene and Hudlow (1982). The rainfall estimates are desired at the grid node locations as a convenience in estimating areal rainfall distributions and other uses as well.

One of the first attempts to solve the problem of merging gage and radar data was made by Brandes (1975) but many consider the best possible results to be provided by optimum interpolation first introduced by Eddy (1963) and Gandin (1963). This technique has been gradually modified, especially for surface rainfall analysis using radar and rain gage data, as examined by Brady (1976), Crawford (1977), Hembree (1980), and others. The characteristics of this methodology and its superiority qualities are given in detail

¹ Respectively, NOAA/NRC Postdoctoral Associate, Hydrologic Research Laboratory, Office of Hydrology National Weather Service, 8060 13th Street, Room 530, Silver Spring, Maryland 20910 (permanent affiliation with the Systems Research Institute of Polish Academy of Sciences); and Deputy Meteorologist in Charge, National Weather Service, New Orleans Area Weather Service Forecast Office, 1120 Old Spanish Trail, Slidell, Louisiana 70458.

by Crawford (1978 and 1979). Using this methodology, rainfall estimates for each grid point are computed from a regression equation of the form:

$$Y_i = X_{li}^G b_{li}^G + \dots + X_{ni}^G b_{ni}^G + X_{li}^R b_{li}^R + \dots + X_{mi}^R b_{mi}^R$$

where:

Y_i is final estimate of rainfall at i^{th} grid point;

$X_{li}^G, \dots, X_{ni}^G$ are n observations of rainfall from gages used for i^{th} grid analysis;

$X_{li}^R, \dots, X_{mi}^R$ are m observations of rainfall from radar bins used for i^{th} grid analysis; and

$b_{li}^G, \dots, b_{ni}^G$ and $b_{li}^R, \dots, b_{mi}^R$ are respective regression coefficients.

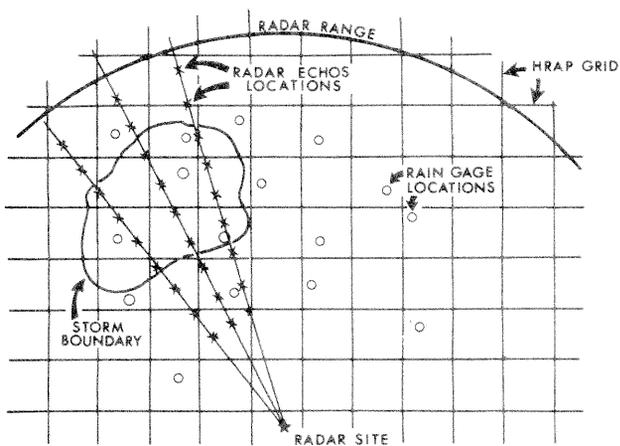


Figure 2. Schematic Illustration of Data Fields Used in Analysis.

If the b coefficients are chosen properly, then the final analysis provides optimal merging of the two data sets with pattern "accuracy" given by radar and point accuracy enhanced by gages.

The algorithm showing the basic concept of the method is described in the following section.

THE METHOD DESCRIPTION

The method applied is a multivariable linear regression,

$$V = Z\beta + \epsilon \quad (1)$$

where:

V is a predictand matrix;

$Z = (Z_G, Z_R)$ is a matrix of observations consisting of two submatrices containing the observations from gage stations and radar bins, respectively; and

$\beta = (\beta_G, \beta_R)$ is a matrix of respective regression coefficients; and ϵ is a matrix of population errors (i.e., it contains among other features a measure of our model inaccuracies).

The solution for b , the minimum variance estimators of β of Equation (1) is:

$$b = (Z^t Z)^{-1} Z^t V \quad (2)$$

where $Z^t Z$ is a matrix of covariance between the predictors and $Z^t V$ is a matrix of covariance between the predictand and the predictors. Usually in actual computer regression analysis, the covariance matrix is replaced by a correlation matrix in order to avoid the round-off errors.

The modification in our approach resulted from the fact that the observation matrix V is not known; therefore, the product $Z^t V$ in Equation (2) is not known either. Using the structural similarity of $Z^t Z$ and $Z^t V$ (let us recall that V is also the rainfall estimate but at the location where data were not being collected) and the fact that $Z^t Z$ is a correlation between two variables separated in space, it is possible to estimate $Z^t V$ given the spatial distribution of correlation $Z^t Z$. In other words, each element of $Z^t Z$ is a value of a function of distance and direction only.

For a bivariate regression, Equation (1) takes the form

$$V = Z_G \beta_G + Z_R \beta_R + \epsilon \quad (3)$$

and therefore $Z^t Z$ matrix may be presented as

$$Z^t Z = \begin{vmatrix} Z_G^t Z_G & Z_G^t Z_R \\ Z_R^t Z_G & Z_R^t Z_R \end{vmatrix} \quad (4)$$

which is a symmetric matrix since $(Z_R^t Z_G)^t = Z_G^t Z_R$. Also,

$$Z^t V = \begin{vmatrix} Z_G^t V \\ Z_R^t V \end{vmatrix} \quad (5)$$

Thus, we see that to perform a regression, we have to model three terms: $Z_G^t Z_G$, $Z_G^t Z_R$, and $Z_R^t Z_R$. The first

two terms would be used to estimate $Z_G^t V$ and $Z_R^t V$ (since V should be treated as a gage rather than radar bin). $Z_G^t Z_G$ and $Z_R^t Z_R$ will be called auto-correlation matrices and $Z_G^t Z_R$ will be called cross-correlation matrix. Once the models are available, the matrices $Z^t Z$ and $Z^t V$ can be constructed, b coefficients determined, and the analysis performed.

However, two important questions remain to be answered: how do we model these correlations and how many variables should be taken into regression? To answer the first question, let us explain our construction of raw correlation matrices. Correlations are calculated using each set of paired observations. The atmosphere's anisotropy is incorporated through discrete but variable lag separation distances in both the X and Y directions. The result is that not only station separation distances, but also directional separations are considered. Figure 3 shows an example of three points. More detail is provided in Crawford (1977).

Thus, for the case of using n stations of the same kind, the derived raw auto-correlation matrix is based on a total of $n(n-1)$ station pairs. Of course, constructing the cross-correlation matrix for n_1 stations of the first kind and n_2 of the second, we consider $n_1 \cdot n_2$ pairs. Our next step is to approximate the raw correlation matrix with a two-dimensional function satisfying certain important conditions. More detailed discussion of this matter can be found in Crawford (1977).

In our current study the following function, similar to that used by Brady (1976), Crawford (1977), and then Hembree (1980), is implemented:

$$f(x, y) = \rho \exp \left[- \frac{K}{2(1-\alpha^2)} \right], \quad (6)$$

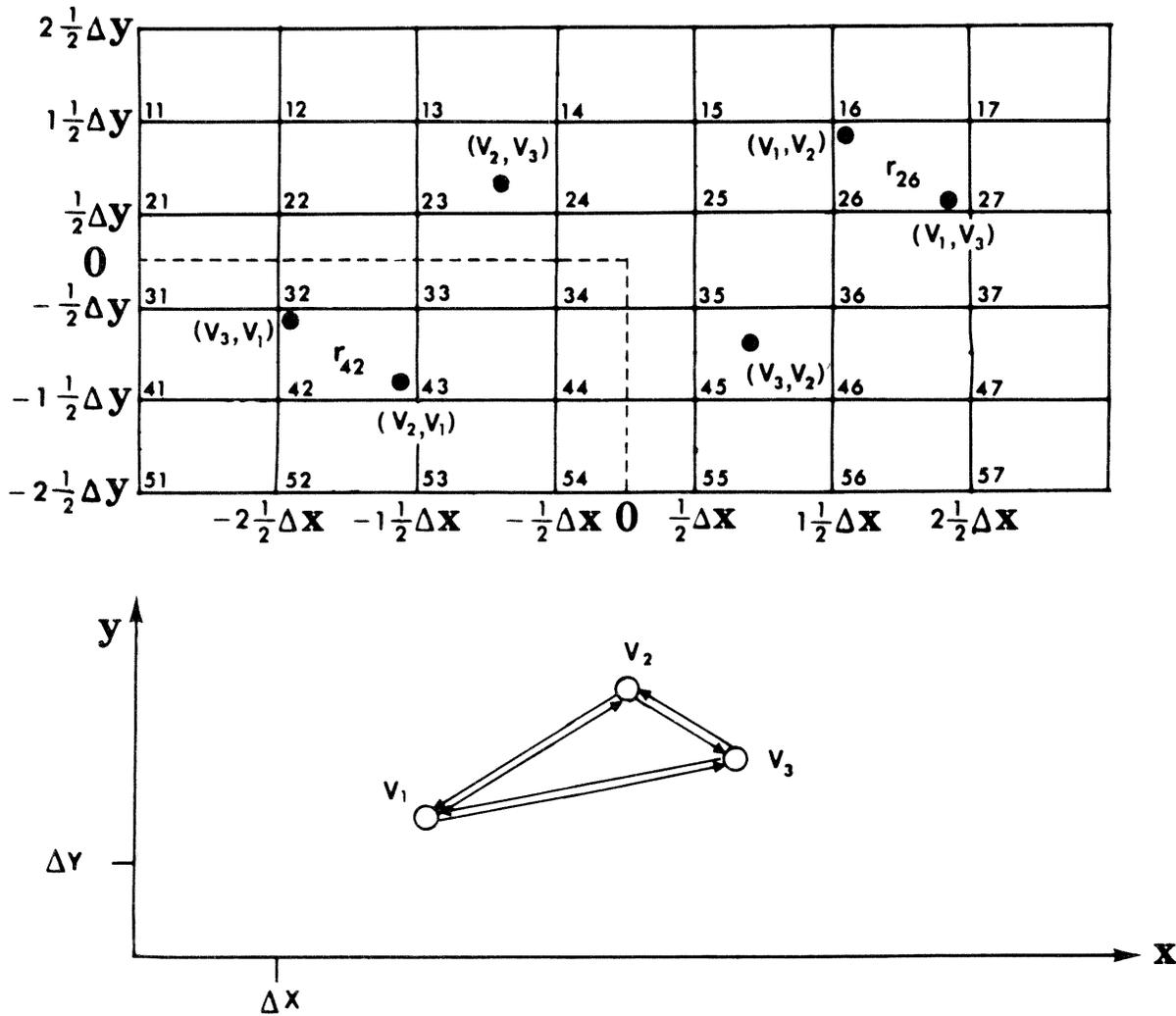


Figure 3. Construction of a Raw Correlation Matrix.

where

$$K = \frac{x^2}{\sigma_x^2} - 2\alpha \frac{xy}{\sigma_x \sigma_y} + \frac{y^2}{\sigma_y^2}, \text{ and} \quad (7)$$

α is an ellipticity parameter in the X-Y plane;

ρ is the lag-zero correlation coefficient; and

σ_x, σ_y are decorrelation distances in X and Y.

Allowing the point of maximum correlation to occur at other than lag zero in X-Y plane, the expression for K takes the form:

$$K = \frac{(x-x_0)^2}{\sigma_x^2} - 2\alpha \frac{(x-x_0)(y-y_0)}{\sigma_x \sigma_y} + \frac{(y-y_0)^2}{\sigma_y^2} \quad (8)$$

where x_0, y_0 are now the coordinates of the maximum.

Approximation of the raw covariance matrix using Equation (6) can be obtained by solving the following nonlinear optimization problem:

$$\text{minimize } Q = \sum_{i=1}^M N_i [r_i - f_i(x, y, x_0, y_0, \alpha, \sigma_x, \sigma_y, \rho)]^2, \quad (9)$$

where

M is the total number of new correlation matrix values over which the function is fit;

N_i is the number of station pairs used to estimate r_i (the raw correlation coefficient);

f_i is a value of the approximation function; and

$x_0, y_0, \alpha, \sigma_x, \sigma_y$, and ρ are parameters to be estimated.

The nonlinear optimization algorithm used was based on the Levenberg-Marquardt method (Brown, *et al.*, 1972).

The second question: "How many variables should be incorporated into regression" can be answered with the help of a step-wise regression algorithm (Draper and Smith, 1966). The set of predictors examined consists of all points (gages and radar bins) located within the "radius of influence" which is fully defined by ellipticity parameter α and decorrelation distances σ_x and σ_y measured in X and Y directions from a grid point under investigation. After selecting possible predictors to be used in Equation (3), the previously derived correlation functions are used to calculate Equation (2). Stepwise regression is performed to then choose the "best" set of predictors.

Finally, the regression coefficients so calculated are used to "filter" the "best" set of predictors to estimate the

predictand value at each location. This completes the main part of the analysis. All that remains is an analysis of residual errors and plotting final analysis for an end user.

Some of the computational aspects will be discussed in the following section.

SOME ASPECTS OF THE METHOD APPLICATION

The above described method is being implemented by NWS using data for the Oklahoma region. Data from several radars, including the Oklahoma City radar, and rain gage reports used by the Tulsa River Forecast Center are transmitted to the NWS IBM 360/195 computer in Suitland, Maryland. These digitized data have been previously checked for large errors and thus are ready to be processed. The radar data are full resolution data sets (2 degree azimuthal by 1 nautical mile radial data bins). This gives $180 \times 115 = 20,700$ possible bins from a single radar. About 700 rain gage reports are available across the region covered by the Oklahoma City radar umbrella. Our grid consists of 101×101 nodes with its center located near the radar site. The number of grid points means that for a widespread storm we have to estimate rainfall at 10,201 points. Obviously efficient management of computer resources is vital to our timely analysis. For example, in order to minimize storage, special procedures for rainfall data management have been developed. These include subroutines for bin coordinates computation. To minimize CPU time for each complete analysis, the so-called convex hull (Bentley, *et al.*, 1982) is constructed for each storm, limiting computations only to the region where rainfall actually occurred. Nevertheless, the CPU time consumed for a solution to the total grid network can be considerable (an important factor for using only two-dimensional correlations). Incorporation of vertical correlation and time correlation would: (1) increase the data volume significantly, and (2) increase the computational time for the first phase of our analysis (i.e., construction of the correlation function). The second phase, step-wise regression, and the final analysis would not be affected very much since there is always a possibility of limiting the number of variables in the step-wise regression to be checked.

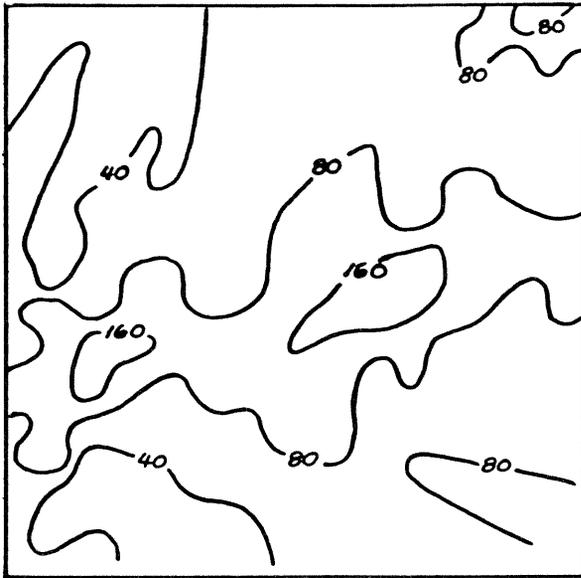
The inclusion of satellite data into the analysis would expand the covariance matrix to:

$$Z^t Z = \begin{vmatrix} Z_G^t Z_G, Z_G^t Z_R, Z_G^t Z_S \\ Z_R^t Z_G, Z_R^t Z_R, Z_R^t Z_S \\ Z_S^t Z_G, Z_S^t Z_R, Z_S^t Z_S \end{vmatrix}$$

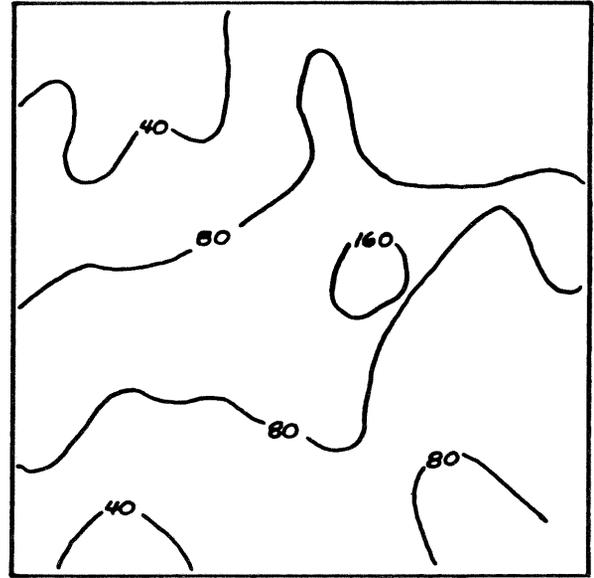
where Z_S is satellite observations matrix. This would necessitate the calculation of three additional correlation matrices.

Figure 4 gives an example for the data obtained from the numerical experiment described by Greene, *et al.* (1980). An advantage for the use of these data is that the true field (original field) is known so the results can be compared against it.

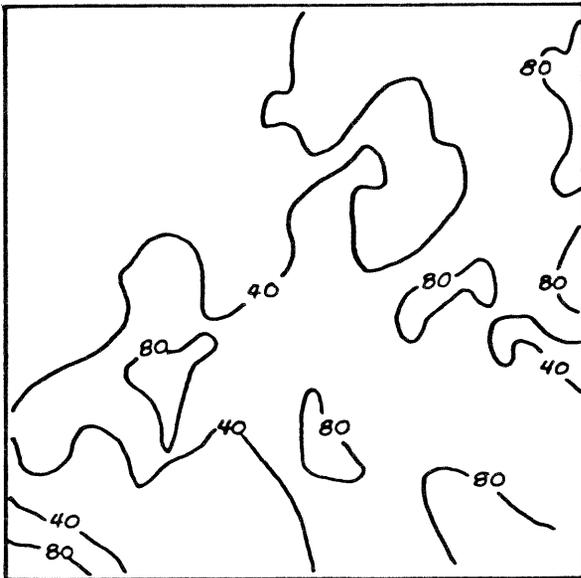
A final point that should be mentioned is that some further investigation of the spatial auto-correlation of errors is needed. Such investigation is planned and will be done in the near future along with other tests of the method.



Original field



Simulated gage field
(disturbed original field)



Simulated radar field
(disturbed original field)



Rainfall analysis based
on simulated fields

Figure 4. The Results of the Numerical Experiment. The simulated fields were generated out of the original field and then merged using the method described.

ACKNOWLEDGMENTS

This research has been done partly under a National Oceanic and Atmospheric Administration/National Research Council Associateship at the Hydrologic Research Laboratory of the National Weather Service, Silver Spring, Maryland.

LITERATURE CITED

- Bentley, J. L., M. G. Faust, and F. P. Preparata, 1982. Approximation Algorithms for Convex Hulls. *Communications of the ACM* 25(1):64-68.
- Brady, P. J., 1976. *Multivariate Experimental Design in Meteorology*. Dept. of Meteorology, University of Oklahoma, Norman, Oklahoma, 121 pp.
- Brandes, E. A., 1975. Optimizing Rainfall Estimates With the Aid of Radar. *Journal of Applied Meteorology* 14:1339-1345.
- Brown, K. M. and J. E. Dennis, 1972. Derivative Free Analogues of the Levenberg-Marquardt and Gauss Algorithms for Nonlinear Least Squares Approximations. *Numerische Mathematik* 18:289-297.
- Crawford, K. C., 1977. The Design of a Multivariate Mesoscale Field Experiment. Dept. of Meteorology, University of Oklahoma, Norman, Oklahoma, pg. 157.
- Crawford, K. C., 1978. On the Bivariate Objective Analysis of Surface Rainfall Using Optimum Interpolation. *Proceedings of the 18th Radar Meteorology Conference*. American Meteorological Society, Boston, Massachusetts, pp. 336-341.
- Crawford, K. C., 1979. Considerations for the Design of a Hydrologic Data Network Using Multivariate Sensors. *Water Resources Research* 15(6):1752-1761.
- Draper, N. and H. Smith, 1966. *Applied Regression Analysis*. John Wiley & Sons, New York, New York.
- Eddy, A., 1963. Kinematic Divergence and Large Scale Energy Conservation. Dept. of Meteorology, McGill University, Montreal, Canada.
- Gandin, L. S., 1963. *Objective Analysis of Meteorological Fields*. Leningrad, USSR.
- Greene, D. R. and M. D. Hudlow, 1982. *Hydrologic Grid Mapping Procedures*.
- Greene, D. R., M. D. Hudlow, and R. K. Farnsworth, 1979. A Multiple Sensor Rainfall Analysis System. *Preprints, Third Conference on Hydrometeorology*, Bogota, Columbia, pp. 44-53.
- Greene, D. R., M. D. Hudlow, and E. R. Johnson, 1980. A Test of Some Objective Analysis Procedures for Merging Radar and Rain-Gage Data. *Preprints, 19th Conference on Radar Meteorology*, Miami, Florida. American Meteorological Society, Boston, Massachusetts, pp. 470-478.
- Hembree, L. A., 1980. An Investigation of Satellite-Radar Correlation Functions. Report by Amos Eddy, Inc., under Bureau of Reclamation Contract No. 9-07-85-V0024, pg. 192.
- Hydrologic Research Laboratory, 1980. *Hydrologic Rainfall Analysis Project (HRAP) Activities*, NWS, Hydrologic Research Laboratory, Silver Spring, Maryland, 7 pp.