January 1988

Stochastic Interpolation of Precipitation Data From Multiple Sensors

D. J. Seo
A. Azimi-Zonooz
D. S. Bowles
W. F. Krajewski
C. J. Duffy
K. P. Georgakakos

See next page for additional authors

Follow this and additional works at: http://digitalcommons.usu.edu/water_rep

Part of the Civil and Environmental Engineering Commons, and the Water Resource Management Commons

Recommended Citation
http://digitalcommons.usu.edu/water_rep/109

This Report is brought to you for free and open access by the Utah Water Research Laboratory at DigitalCommons@USU. It has been accepted for inclusion in Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.
Authors
Final Report

on

NSF Grant No. ECE-8419189

STOCHASTIC INTERPOLATION OF PRECIPITATION DATA
FROM MULTIPLE SENSORS

by

D. J. Seo, A. Azimi-Zonooz, D. S. Bowles
W. F. Krajewski, C. J. Duffy, K. P Georgakakos,
and T. H. Lee

to

NATIONAL SCIENCE FOUNDATION
Washington, D.C.

Utah Water Research Laboratory
Utah State University
Logan, Utah 84322-8200

February 1988
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>3.1</td>
</tr>
<tr>
<td>3.2</td>
</tr>
<tr>
<td>3.3</td>
</tr>
<tr>
<td>3.4</td>
</tr>
<tr>
<td>3.5</td>
</tr>
<tr>
<td>3.6</td>
</tr>
<tr>
<td>3.7</td>
</tr>
<tr>
<td>3.8</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>5.1</td>
</tr>
<tr>
<td>5.2</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>APPENDICES</td>
</tr>
</tbody>
</table>

**APPENDIX A:** CUMULATIVE LIST OF PROJECT PUBLICATIONS

**APPENDIX B:** RAINFALL ESTIMATION BY LINEAR CO-KRIGING OF DATA FROM MULTIPLE SENSORS (Preliminary Draft)

**APPENDIX C:** OPTIMAL RAINFALL ESTIMATION BY DISJUNCTIVE CO-KRIGING OF RADAR-RAINFALL AND RAINGAGE DATA (Preliminary Draft)

**APPENDIX D:** HERMITE APPROXIMATION OF ANAMORPHOSIS FUNCTION

**APPENDIX E:** ESTIMATION OF MEAN PRECIPITATION FIELDS USING OPERATIONALLY AVAILABLE HYDROMETEOROLOGICAL DATA AND A TWO-DIMENSIONAL PRECIPITATION MODEL

**APPENDIX F:** CO-KRIGING OF RADAR-RAINFALL AND RAINGAGE DATA
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Software system for this project</td>
<td>4</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Summary of Computer Programs used on the Project</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1 INTRODUCTION

This report summarizes the work conducted under Grant No. ECE-8419189, "Stochastic Interpolation of Precipitation Data from Multiple Sensors," which was awarded to Utah State University in September, 1985, and completed February 29, 1988. It also covers work under a supplemental award made in February, 1986.

The final report is organized into four sections. The following section presents the objective of the research and a brief problem statement. Section 3 contains a summary of second-year work including the project team, work plan, work completed, and publications. In Section 4, project conclusions are summarized. A summary of on-going future work is given in Section 5, together with our plans for publication of research results from this project. Copies of preliminary draft manuscripts and completed technical reports which have been prepared as a result of second-year activities are contained in the Appendices. A cumulative summary of project publications is presented in Appendix A.

2 RESEARCH OBJECTIVE AND PROBLEM STATEMENT

The overall objective of this research is to investigate the feasibility of using stochastic interpolation (co-kriging) techniques for "merging" rainfall data from three different types of sensors: raingages, radars, and satellites.

Rainfall estimates are the single most important input to hydrologic models which are used to forecast river flooding or as the basis for operating water resource systems during periods of high flows. Estimates of mean areal rainfall can be made from direct measurements of rainfall obtained from raingages, or from indirect measurements obtained from radars or satellites. However, each sensor measures a different physical variable, at a
different height above the ground, over different temporal and spatial averaging scales, and with errors which differ in statistical structure and origin. The goal of "merging" is to exploit the complimentarities in the characteristics of rainfall estimates from all three types of sensors, by combining the three types of estimates to obtain a merged estimate of the rainfall field.

A more detailed presentation of the problem statement is presented in a paper which was prepared for the Chapman Conference on Modeling of Rainfall Fields held in Caracas, Venezuela in March 1986 (Bowles et al 1986). A copy of that paper was appended to the First-Year Progress Report. It also contains a description of the previous approaches for utilizing multiple sensors and their limitations, the proposed stochastic interpolation approaches, and the overall evaluation methodology which was implemented on this project.

3 SUMMARY OF SECOND-YEAR WORK

3.1 Project Team

The individuals who were involved in this project and their project roles are listed below:

Dr. David S. Bowles - Principal Investigator
Dr. Witold F. Krajewski - Co-Principal Investigator
Dr. Christopher J. Duffy - Co-Principal Investigator
Mr. Dong Jun Seo - Ph.D. Graduate Student
Mr. Ali Azimi-Zonooz - Ph.D. Graduate Student

In addition, through a subcontract with the University of Iowa funded by an amendment to the original award, the following individuals were also involved in this project:

Dr. Konstantine P. Georgakakos - Principal Investigator for Subcontract
Mr. Tim H. Lee - Ph.D. Graduate Student
3.2 Second-Year Work Plan

In October 1986, approximately one year after commencement of the project, the project team was assembled for an in-depth one week review of the first year's progress and to develop a detailed work plan for the second year of the project. The principal activities which were scheduled for the second year are listed below:

1) Complete development and testing of the disjunctive co-kriging algorithm.

2) Evaluate alternative approaches to detrending, distributions, gage field covariance estimation and anisotropy for rainfall fields and select appropriate approaches.

3) Combine software into a linked system of computer programs capable of data analysis and plotting, detrending, simulation, kriging, and analysis and plotting of results (see Figure 3.1 and Table 3.1).

4) Complete sensitivity analyses and verification of physically-based two-dimensional precipitation trend model, including determination of a functional relationship between ground-level precipitation rate and cloud condensed liquid water equivalent and interfacing of the trend model with the co-kriging software.

5) Perform evaluation of the use of universal and disjunctive co-kriging for merging rainfall measurements from multiple sensors and compare their performance with other rainfall estimation methods.

6) Explore the potential for using a static filter approach to rainfall estimation as an alternative approach to merging by co-kriging of measurements from multiple sensors.

A brief statement on each of these activities is provided in the following subsections.
Figure 3.1. Software system for this project.
<table>
<thead>
<tr>
<th>Program Name</th>
<th>Brief Description</th>
<th>Source of Program</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kriging Algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMPLE</td>
<td>Performs ordinary kriging and ordinary co-kriging in 2-D</td>
<td>Developed with common routines from ACOKRIP (Kafritsas and Bras, 1984)</td>
</tr>
<tr>
<td>UNIVERS</td>
<td>Performs universal kriging and universal co-kriging in 2-D</td>
<td>ACOKRIP is modified to perform automatic structure identification</td>
</tr>
<tr>
<td>DJUNC1</td>
<td>Performs disjunctive kriging and disjunctive co-kriging in 2-D</td>
<td>Disjunctive kriging based on Yates et al. (1985) and Puente and Bras (1982). Disjunctive co-kriging program is developed.</td>
</tr>
<tr>
<td><strong>Rainfall Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAIN6</td>
<td>Generates artificial space-time rainfall</td>
<td>RAIN6 (Valdes, 1985)</td>
</tr>
<tr>
<td>TWODPM</td>
<td>Two-dimensional physically-based rainfall model</td>
<td>TWODPM (Geogakakos and Lee, 1986)</td>
</tr>
<tr>
<td><strong>Rainfall Measurement Field Simulation Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAGESGEN</td>
<td>Generates point raingage measurements</td>
<td>GAGESGEN (Krajewski, 1985)</td>
</tr>
<tr>
<td>RADARGEN</td>
<td>Generates radar rainfall</td>
<td>RADARGEN (Krajewski, 1985)</td>
</tr>
<tr>
<td><strong>Evaluation Algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERRSTAT</td>
<td>Computes me, rmse, smse, MAP errors, error histogram, and residual power spectrum</td>
<td>Developed with 2-D FFT from Press et al. (1986)</td>
</tr>
<tr>
<td><strong>Data Analysis Algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2DSTAT</td>
<td>Computes directional correlogram, variogram, and power spectrum in 2-D</td>
<td>Developed with FFT from Press et al. (1986)</td>
</tr>
<tr>
<td>LN2(3), PT2(3)</td>
<td>Computes MM and ML estimates for 2(3) parameter lognormal and gamma distributions, respectively</td>
<td>Kite (1977), with chi-square and K-S tests added</td>
</tr>
</tbody>
</table>
3.3 Disjunctive Co-kriging Algorithm

Disjunctive block-kriging and co-kriging programs were written in FORTRAN-77 and tested on a VAX-8650 computer. The computations for obtaining a disjunctive (co-) kriging estimate begin by sorting the radar and raingage measurements from smallest to largest and obtaining the empirical cumulative frequency distribution for each data set. An estimate of cumulative probability for each data point is then obtained and inverted to obtain the corresponding univariate and bivariate distributed standard Gaussian random variables. This transformation function is denoted by $\phi(y)$.

The next step is to approximate the function $\phi(y)$ using a limited expansion of Hermite polynomials. In the testing process, sample mean and variance were computed and the appropriate number of terms in the expansion was calculated so that the mean and variance obtained by using the expansion is close enough to the mean and variance computed from the data. It was found that 10 terms is usually sufficient to model the sample data. A visual inspection of the $\phi(y)$ fit to the original sample data was made to further test validity of the fit. Two different algorithms for approximation of this transform were tested (see Appendix D for details). The results of extensive numerical experiments showed that piecewise linear approximation of the function, $\phi(y)$, is more efficient in terms of CPU time and avoids numerical problems that can be encountered with the numerical integration approach.

Next is to determine the sample correlation function from the semivariograms of transformed radar and raingage observations, used in the disjunctive co-kriging equations to obtain a spatial rainfall estimate (see Appendix C for a detailed presentation of Equations). Disjunctive and ordinary co-kriging programs utilized in this study were tested and verified against results published in Yates (1986) by obtaining the same data set used in that paper.
3.4 Resolution of Issues Discussed in First-Year Progress Report

Several issues were presented in the First-Year Progress Report as issues that we were addressing at the end of the first year. Below is a summary of the outcome of work on each of these issues.

a) Detrending

Rainfall fields often contain a large scale trend that may not be described by a simple function, particularly in convective type situations or under orographic effects. Analysis of Fourier series fit and polynomial function fit indicates that for the GATE rainfall fields, presence of a trend cannot be clearly identified by any of the fitting attempts. Universal (co-) kriging, which assumes the trend to be of a polynomial type, does not show any improvement over ordinary (co-) kriging. Two alternatives are suggested. One is to estimate the trend a priori from the past measurements. This approach, however, requires a substantially large number of data. The other approach is to use a physically-based rainfall model, and this is currently being investigated (see Sections 3.5 and 5.1).

b) Distribution

Disjunctive kriging, not requiring any distributional assumption, is computationally expensive. If the rainfall measurements are known to follow a specific distribution, both the computational requirement and the errors associated with the variable transformation will be substantially reduced. Analysis of the GATE and Oklahoma data shows that they follow neither gamma nor lognormal distributions. A theoretical study conducted as part of this research project (Seo, 1988), indicated that the probabilistic nature of the factors shaping the raindrop size distribution, such as the raindrop number concentration and the mean diameter of raindrops, determine the distributional character of rainfall depth.
c) **Covariance estimation for a sparse raingage field**

The number of raingage measurements available from an operational raingage network is too small to obtain a reliable estimate of the covariance structure. The effect of unreliable gage measurement field covariance structure is found to be significant in that the measures for the performance of block-kriging have high variance. To obtain a more reliable covariance estimate, it is suggested that Bayesian updating be used, making use of both the past and the currently available measurements.

d) **Anisotropy**

When the (cross-) covariance is anisotropic, due to advection for example, kriging must be able to handle the anisotropy either directly or by scaling and/or rotating. Due to the small number of gage measurements, it is almost impossible to identify anisotropy from the gage measurement field. The radar rainfall field is found to be helpful in identifying the direction of anisotropy, but not as helpful in obtaining the degree of anisotropy, since radar rainfall data often have large measurement errors.

e) **Co-kriging via sensitivity analysis vs. co-kriging with acknowledged gage measurement error**

Relatively small raingage measurements are also often in error due to, for example, wind, topography and exposure. Since kriging cannot account for the measurement error, it will give poor estimates of the ground truth if the measurement error is high. Co-kriging via the sensitivity analysis (Krajewski, 1987) cannot be performed in practice. When the gage measurement error has zero mean, or, if undercatch due to wind can be estimated, the procedure used in this research can be performed in practice.
f) Static Filter

In kriging, explicit incorporation of the effects of the sensor measurement errors is not possible. Application of static filtering, which explicitly accounts for the measurement error, was found (Seo, 1988) to be restricted for the rainfall estimation problem because 1) the sensor measurement error structures are not fully known, and 2) the known measurement error structures, particularly for the radar rainfall, are generally nonlinear.

3.5 Physically-Based Two-Dimensional Precipitation Trend Model

Previous work in modeling mesoscale rainfall fields was reviewed. A two-dimensional precipitation estimation model was formulated based on the principles of conservation of mass of liquid water equivalent and of heat conservation. Advection of liquid water equivalent is accomplished by the middle tropospheric wind velocity (assumed to be the storm velocity). Spatial interpolation of the spatially- and temporally-sparse wind observations is done based on objective interpolation techniques.

The model formulation explicitly accounts for condensation of vapor, advection and sub-cloud evaporation of liquid water equivalent. The two free model parameters that determine updraft-velocity strength and particle-size distribution were estimated based on contours of various performance criteria in the parameter space. The contours were generated from available real-time meteorological and rainfall hourly observations of convective storms in Oklahoma. The issue of grid size determination was approached from a practical, CPU-time viewpoint and with consideration of the precipitation estimation accuracy.

The model is suitable for use in detrending precipitation fields observed by various types of sensors for the purpose of merging observed-precipitation-fields. Also, because of its state-space mathematical form, it is suitable for use in real-time precipitation forecasting.
3.6 Evaluation of Universal and Disjunctive Co-kriging

The evaluation of universal and disjunctive co-kriging for spatial rainfall estimation was the main focus of this research. The algorithms, experimental designs, results and conclusions for universal and disjunctive co-kriging are summarized in the draft papers in Appendices B and C, respectively.

3.7 Static Filter Approach to Rainfall Estimation

The Static Filter approach was not a part of the original scope of work for this project. However, some investigation of this approach was performed under the NSF project and is continuing under UWRL funding (see Section 5.1).

To investigate the possibility of using static filters for rainfall estimation problem, the basic measurement error structure of the sensors, i.e., radar and raingages, were identified (Seo, 1988). Raingage measurement error is location-dependent, due to such factors as wind, topography, and exposure. Radar measurement error structure was found to be highly nonlinear. Both measurement error structures are only known qualitatively. It is concluded that, to be able to implement static filtering for rainfall estimation, sensor measurement error structures must be further studied to allow their mathematical formulation. As for the nonlinearity, no simple answer exists. Once the problem can be fully formulated, possibilities such as variable transformation and second-order approximation should be studied.

3.8 Publications

The following papers or presentations have been prepared during the second year of this grant. A complete list of project publications is contained in Appendix A.


4 SUMMARY OF PROJECT CONCLUSIONS

All stated objectives for the proposed research have been met. A summary of the conclusions of our work is presented below. The journal articles which are in preparation (see Section 5.2) will contain a more detailed presentation of these conclusions and their basis. Copies of these articles will be forwarded to NSF as they become available.
1) Utilization of radar rainfall data by co-kriging does improve rainfall estimation over gage-only estimation or the Brandes method. The improvement is consistent under various measurement error characteristics of radar rainfall and varying raingage density.

2) Performance of co-kriging is insensitive to varying radar measurement error characteristics. In particular, co-kriging is very effective in bias removal from the radar rainfall field.

3) Universal co-kriging, as implemented in this project, does not seem to warrant its routine application for the following reasons: i) structure identification is very difficult and computationally very expensive, and ii) the validity of the intrinsic hypothesis is questionable.

4) Performance of linear co-kriging is found to deteriorate substantially due to i) uncertain covariance structures and ii) raingage sampling error. If reliable estimates of the covariance structures could be obtained, a significant improvement can be expected.

5) Disjunctive co-kriging improves rainfall estimation consistently over ordinary and universal co-kriging.

6) For approximation of the anamorphosis function the analytical method is much more efficient in terms of CPU time and avoids certain instabilities of numerical integration approach.

5 ON-GOING FURTHER WORK

5.1 On-going Activities - UWRL Funding

Research work proposed for the NSF funded project has been completed and journal articles are in preparation to describe our work and findings. Additional research activities
which have been stimulated by work under this project are continuing under funding from the Utah Water Research Laboratory, Utah State University. These are as follows:

1) Evaluation of the use of the Physically-Based Two-Dimensional Precipitation Trend Model for rainfall estimation by co-kriging.

2) Further development of the Bayesian approach to rainfall estimation.

3) Evaluation of alternative approaches to estimation of raingage semi-variograms and raingage-radar cross variograms.

The following discussion amplifies the work being conducted under activity 3, listed above.

The ability to obtain higher resolution estimates of spatial variability in the rainfall fields is one of the primary areas of interest in this study. Rainfall fields at spatial scales of our interest are highly variable in space and time, and this variability tends to introduce a great deal of uncertainty into the estimation process. Spatial variability of rainfall fields is generally described by the first two moments, the mean and covariance from a single time period. We are investigating the use of multiple time periods.

The procedure that will be explained will make use of Phase I and Phase II GATE data provided by National Weather Service. Multiple time periods of hourly rainfall fields will be used to obtain semi-variograms and coefficients for the anamorphosis function expansion in a Bayesian Framework. The parametric model for the mean will be provided by the Hermite coefficients, and for covariance by the updated semivariograms based on past observations. The parameters for the models will be constructed for various levels of raingage and radar-rainfall densities and will further be used in the areal rainfall estimation using disjunctive co-kriging. The data will be divided into various classes of density and,
depending on the variability of data and a trial-and-error procedure, the number of classes will be determined and used to obtain parameters of the model for each class. The spatial rainfall estimates obtained by this approach will be compared with estimates obtained from single time period data and evaluated using the ground-truth rainfall fields as was done before.

5.2 Publication Plans

In addition to the publications and presentations listed in Section 3.8 of this report, it is planned to publish the results of this research in a series of refereed journal articles. A tentative list of these papers, with the names of the first authors only shown at this time, is presented below, together with other papers or presentations which are planned:


4) Rainfall Estimation by Linear Co-kriging of Data from Multiple Sensors, Part II: Results and Conclusions. To be submitted to Water Resources Research. (In preparation) (D. J. Seo),

6) A Bayesian approach to rainfall estimation using data from multiple sensors. To be submitted to Water Resources Research. (In preparation) (D. J. Seo)

Also, the following individuals will submit Ph.D. dissertations based on research which was supported under this grant:

Dong Jun Seo - Utah State University
Ali Azimi-Zonooz - Utah State University
Tim H. Lee (Partial support for this project) - University of Iowa

These dissertations will include user's manuals for all software developed or modified under this grant.

6. REFERENCES


APPENDICES
APPENDIX A

CUMULATIVE LIST OF PROJECT PUBLICATIONS
Cumulative List of Project Publications


3) Stochastic-dynamic short-term prediction of space-time rainfall, by K. P. Georgakakos. Presented at the Chapman Conference on Modeling of Rainfall Fields. Caracas, Venezuela, March. 1986. (This paper was prepared with support from this project and from the Hydrology Research Laboratory, National Weather Service.) (See Appendix C of First-Year Progress Report).


APPENDIX B

RAINFALL ESTIMATION BY LINEAR CO-KRIGING OF DATA FROM MULTIPLE SENSORS
RAINFALL ESTIMATION BY LINEAR CO-KRIGING OF DATA FROM MULTIPLE SENSORS

by

Dong-Jun Seo
Utah Water Research Laboratory
Utah State University
Logan, Utah 84322-8200

February, 1988
RAINFALL ESTIMATION BY LINEAR CO-KRIGING OF DATA FROM MULTIPLE SENSORS

Seo, D. J., W. F. Krajewski, and D. S. Bowles

THE MULTI-SENSOR RAINFALL ESTIMATION PROBLEM

To present the problem of rainfall estimation using raingage measurements and radar rainfall data in a more general context, and to justify the choice of kriging, the following potentially useful approaches are briefly described. Throughout this paper, the term "multi-sensor" is used to refer to raingages and radar, thus excluding satellite, another potentially useful sensor in rainfall estimation.

The problem can be defined as follows: We want to obtain an estimate (point or spatially averaged) of the ground-truth rainfall at an arbitrary location using the rainfall measurements from a network of gages and the rainfall data from a radar. In addition, measurements of some meteorological variables may be available. In this work, we concern ourselves only with the conventional weather radar, and thus only the reflectivity factor is available for rainfall estimation.

Physically-Based Statistical Approach 1

One of the major factors affecting the accuracy of radar rainfall is the uncertainty in the Z-R relationship. Radar measures average returned power, from which the reflectivity factor can be obtained via a radar equation (Battan, 1973). Throughout this work, we make the following distinctions: 1) we equate radar measurement with the reflectivity
factor, not the average returned power, and 2) radar rainfall is referred to as the rainfall amount obtained via a Z-R relationship from the radar measurement of reflectivity.

Ground-level rainfall can be accurately obtained if: 1) raindrop size distribution at the ground-level is known, and 2) if the fall velocities of the raindrops at the ground-level is known. Under some simplifying assumptions (exponential distribution and constant raindrop number concentration), raindrop size distribution at the cloudbase may be inferred from the reflectivity factor.

Raindrop size distribution at the cloudbase, however, differs from that at the ground level since the raindrops undergo various physical processes such as coalescence, collision, breakup, evaporation, and advection (Pruppacher and Klett, 1978), before they reach the ground.

If these physical processes could be modeled adequately, requiring only the readily available meteorological variables as the model's input, we could obtain the ground level rainfall estimates from the radar measurements of reflectivity. These estimates, obtained from the physically-based "process model," may then be combined with the raingage measurements to give the final rainfall estimates.

Unfortunately, the physical processes involved are not so well-known as to allow rigorous modeling of them. Also, the current radar technology is limited to measuring average returned power within ±3 db, which introduces uncertainty of a magnitude comparable with the uncertainty due to the Z-R relationship (Stout et al., 1968).

Due to these reasons, it seems that realization of the type of physically based approach described above is a number of years away, and, for the time being, we have to resort to purely statistical approaches.
Physically-Based Statistical Approach 2

Success of the above approach will rely heavily on how accurately radar can measure the reflectivity factor. If the radar measurements have large error variance, little information can be extracted on the spatial variability. If a physically-based rainfall model can provide the larger scale spatial variability of the rainfall, its output may be used to describe the mean of the rainfall field, thus reducing the burden on statistical spatial prediction. An application of this approach will be given later in this work.

Purely Statistical Approaches

i) Uncoupling of Ground-Truth and Measurement Error

Though raingages give more accurate measurements of rainfall than radar, their measurements are seldom free of measurement error. When strong wind is present, undercatch by raingages has been reported to exceed 20 percent on a relative scale (Larson and Peck, 1974). Including other factors such as exposure and topography, even the measurements from a highly dense network of raingages may not represent the ground-truth rainfall field.

Ideally, we want to obtain an estimate of the ground truth, which requires, at least, the sensors' error structures to be known. Though undercatch by raingages due to wind may be modeled from site to site, the error structure for the radar rainfall is known only qualitatively, as will be described in detail later.

ii) Bayesian Updating of Prior Information

Due to the highly variable nature of the rainfall process, both in space and time, it is difficult to accurately estimate statistics such as mean, variance, and correlation scale,
particularly on a small time scale such as hourly. It will require, at least, systematic archiving of data.

If some information is available at the time of spatial prediction, the Bayesian estimation approach can be used (see, e.g., Kitanidis, 1986). One important application of this approach will be on the estimation of the gage measurement field covariance. Due to sparse nature of the raingage network, the gage measurement field covariance, estimated from the measurements at hand alone, will almost always suffer from a large estimation error.

In view of the limitations associated with the above approaches, co-kriging is considered to be a very attractive method in that it requires only the currently available measurements.
PROBLEM FORMULATION

Suppose that our raingage network lies in space $\Omega$, and our radar umbrella includes this space. The raingage network provides measurements of point rainfall accumulated over the period of interest. The radar provides spatially averaged rainfall accumulated over the same time period on a grid of squares, or, "bins." Typically, bin size ranges from 4x4 km to 5x5 km, and the radius of the radar umbrella is about 200 km. The sensors' data configuration is shown in Figure 1.

Given the two sets of rainfall data, the problem is to obtain an estimate $\hat{Z}_g(U_0)$ of $Z_g(U_0)$ defined as

$$Z_g(U_0) = \frac{1}{A} \int Z_{pg}(U) \, dU$$

where $A$ is the size of a radar bin, $Z_{pg}(U)$ is the point gage rainfall at location $U$, and $Z_g(U_0)$ is the gage rainfall averaged over $A$ centered at an arbitrary location $U_0$ in $\Omega$.

The choice of $\hat{Z}_g(U_0)$ over $\hat{Z}_{pg}(U_0)$, the estimate of spatially averaged gage rainfall over the estimate of point gage rainfall, stems from the fact that, in many hydrological applications, point rainfall estimates are seldom of interest.

Though direct co-kriging of the point raingage measurements and the (spatially averaged) radar rainfall data is possible to obtain estimates of spatially averaged gage rainfall, it makes structure identification and its interpretation more difficult. In this work, we adopt the following two-step approach: 1) block-krige the point gage measurements, rendering the measurement scale of the gage measurements compatible with that of the radar rainfall data, and 2) co-krige the block-kriged gage measurements and radar rainfall data as if each block-kriged gage measurement or radar rainfall is sampled at the center of the block or bin.
LINEAR CO-KRIGING

It is not our intention to give an exhaustive presentation on linear kriging, as it can be found elsewhere (see, for example, Journel and Huijbregts, 1978, Kafritsas and Bras, 1984). In this section, we limit ourselves to a brief description of linear co-kriging.

The linear co-kriging estimator has the following form.

\[ \hat{Z}_g(U_0) = \sum_{i=1}^{N_g} \lambda_{gi} Z_g(U_i) + \sum_{j=1}^{N_r} \lambda_{rj} Z_r(U_j) \]  

(2)

where \( Z_g(U_i) \) is the spatially averaged gage measurement centered at \( U_i \), \( \lambda_{gi} \)'s and \( \lambda_{rj} \)'s are weighting coefficients to be determined, \( N_g \) is the number of spatially averaged gage measurements, \( Z_r(U_j) \) is the radar rainfall at the bin centered at \( U_j \), \( N_r \) is the number of radar rainfall data, and \( Z_g(U_0) \) is the estimated gage rainfall averaged over \( A \) centered at an arbitrary location \( U_0 \) in \( \Omega \).

When the mean and the covariance of both fields are perfectly known, the weighting coefficients that give the unbiased, minimum error variance estimate can be found by minimizing

\[ E[(\hat{Z}_g(U_0) - Z_g(U_0))^2] \]

(3)

The solution for the above problem can be easily obtained by the simple application of the Gauss-Markoff theorem (Liebelt, 1968), and constitute probably the most important building block in estimation theory.
where $Q_{or}$ is the $(1\times N_r)$ cross-covariance vector of the unknown gage measurement and the sampled rainfall data $Q_{og}$ is the $(1\times N_g)$ covariance vector of the unknown gage measurement and the sampled gage measurement $Q_{rr}$ is the $(N_r \times N_r)$ covariance matrix of the sampled radar rainfall data $Q_{rg}$ is the $(N_r \times N_g)$ cross-covariance matrix of the sampled radar rainfall data and the gage measurements $Q_{gg}$ is the $(N_g \times N_g)$ covariance matrix of the sampled gage measurements

When the measurements are normally distributed, $^{^\wedge}Z_{go}$ is also the conditional expectation $E[Z_{go}|Z_g, Z_r]$ (Schweppe, 1973). In many applications, however, the statistical quantities in the above expressions are not perfectly known. Of particular interest is the case when no a priori information is available about the mean, for which the above optimal linear estimator reduces to the co-kriging estimator (Kitanidis, 1986).

**Ordinary Co-Kriging**

When both the (spatially averaged) raingage measurement field and the radar rainfall field are second-order homogeneous, ordinary or simple co-kriging can be performed. If both the mean of the gage measurement field and the mean of the radar rainfall field are
unknown (but constant), only ordinary co-kriging is of interest, and the minimization of Eq. (3) is made subject to the following constraints to force unbiasedness.

\[ \sum_{i=1}^{N_g} Z_g(U_i) = 1 \text{ and } \sum_{j=1}^{N_r} Z_r(U_j) = 0 \]  \hspace{1cm} (6)

The role of the above constraints are easily seen from

\[
E[Z_g(U)] = \sum_{i=1}^{N_g} \lambda_{g_i} E[Z_g(U_i)] + \sum_{j=1}^{N_r} \lambda_{r_j} E[Z_r(U_j)]
\]

\[= m_g \sum_{i=1}^{N_g} \lambda_{g_i} + m_r \sum_{j=1}^{N_r} \lambda_{r_j}\]  \hspace{1cm} (7)

\[= m_g\]

where \(m_g\) is the unknown, but constant mean of the gage measurement field, and \(m_r\) is the unknown, but constant mean of the radar rainfall field.

In other words, whatever the mean of the radar rainfall field may be, and whatever the mean of the gage measurement field may be, the estimate is unbiased. This property is very attractive for the rainfall estimation problem in that radar rainfall data often has unknown bias in the mean even when they provide a good spatial description of the ground-truth field.
Universal Co-Kriging

When the mean of each field can be described by one of the following polynomial functions, but with unknown coefficients, universal co-kriging may be performed.

\[ m_s(u,v) = a_0 \]  \hspace{1cm} \text{order 0}  \\
\[ m_s(u,v) = a_0 + a_1 u + a_2 v \]  \hspace{1cm} \text{order 1}  \\
\[ m_s(u,v) = a_0 + a_1 u + a_2 v + a_3 u^2 + a_4 v^2 + a_5 u v \]  \hspace{1cm} \text{order 2}  \\

where \( s \) denotes the measurement field of a sensor, either radar or raingages, \( m_s(u,v) \) is the mean of the measurement field at \((u,v)\); and \( a_0, a_1, a_3, \text{ and } a_5 \) are coefficients

When the above is the case, a generalized increment can be defined as follows.

\[ G_s = \sum_{i=1}^{N} \lambda_i Z_s(U_i) \]  \hspace{1cm} (9)

where \( Z_s(U_i) \) is measurement from a sensor sampled at \( U_i \), and \( \lambda_i \)'s are coefficients

is a generalized increment of order \( k \), if it filters out polynomials or order \( k \), i.e.,

\[ E[G_s] = \sum_{i=1}^{N} \lambda_i E[Z_s(U_i)] = 0 \]  \hspace{1cm} (10)

where the mean of \( Z_s(\ ) \) is of polynomial of order \( k \)
The generalized covariance $K_{SS}(\ )$ is defined as satisfying

$$\text{Var}[G_s] = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j K_{SS}(U_i, U_j)$$

(11)

If Eqs. (10) and (11) hold, the function $Z_s(\ )$ is called an intrinsic random function of order $k$. The generalized cross-covariance $K_{gr}(\ )$ between the gage measurement field and the radar rainfall field is defined as satisfying

$$\text{E}[G_g G_r] = \sum_{i=1}^{N} \sum_{j=1}^{M} \lambda_{gi} \lambda_{rj} K_{gr}(U_i, U_j)$$

(12)

As was the case in ordinary co-kriging, the minimization of Eq. (3) requires constraints to force unbiasedness.

The co-kriging estimator will be unbiased, i.e.,

$$\text{E}[\hat{Z}_g(U_o) - Z_g(U_o)] = 0, \quad \text{if both}$$

$$\sum_{i=1}^{N_g} \lambda_{gi} Z_g(U_i) - \hat{Z}_g(U_o) \quad \text{and} \quad \sum_{j=1}^{N_r} \lambda_{rj} Z_r(U_j)$$

(13)

are generalized increments of order $k$, which translates into the following constraints.
\begin{align*}
\sum_{i=1}^{N} \lambda_{gi} z_{g}(U_{i}) &= 1, & \sum_{j=1}^{N} \lambda_{rj} z_{r}(U_{j}) &= 0 & \text{if } k=0, 1, \text{ or } 2 \\
\sum_{i=1}^{N} \lambda_{gi} z_{g}(U_{i}) &= u_{o}, & \sum_{j=1}^{N} \lambda_{rj} z_{r}(U_{j}) &= 0 & \text{if } k=1 \text{ or } 2 \\
\sum_{i=1}^{N} \lambda_{gi} z_{g}(U_{i}) &= v_{o}, & \sum_{j=1}^{N} \lambda_{rj} z_{r}(U_{j}) &= 0 & \text{if } k=1 \text{ or } 2 \\
\sum_{i=1}^{N} \lambda_{gi} z_{g}(U_{i}) &= u_{o} v_{o}, & \sum_{j=1}^{N} \lambda_{rj} z_{r}(U_{j}) &= 0 & \text{if } k=2 \\
\sum_{i=1}^{N} \lambda_{gi} z_{g}(U_{i}) &= u_{o}^{2}, & \sum_{j=1}^{N} \lambda_{rj} z_{r}(U_{j}) &= 0 & \text{if } k=2 \\
\sum_{i=1}^{N} \lambda_{gi} z_{g}(U_{i}) &= v_{o}^{2}, & \sum_{j=1}^{N} \lambda_{rj} z_{r}(U_{j}) &= 0 & \text{if } k=2 \\
\sum_{i=1}^{N} \lambda_{gi} z_{g}(U_{i}) &= u_{o} v_{o}, & \sum_{j=1}^{N} \lambda_{rj} z_{r}(U_{j}) &= 0 & \text{if } k=2
\end{align*}

(14)
DESIGN OF EXPERIMENT 1

Since the estimation procedure described in the previous sections give estimates of the spatially averaged raingage rainfall, the most natural way of evaluating its performance would be to compare the estimated fields with the spatially averaged raingage measurement fields obtained from a very dense raingage network. Unfortunately, such a raingage network with the corresponding radar rainfall data, is not available, and thus we resort to a simulation experiment.

In this first simulation experiment, we assume that the GATE radar rainfall data represent the spatially-averaged ground-truth fields. The GATE radar rainfall data, taken during the GARP Atlantic Tropical Experiment in 1974, are considered to be of high quality in that: 1) anomalous propagations (AP) and outliers are removed (Krajewski, 1987), and 2) the Z-R relationship is established from dropsize measurements made on ships and aircraft (Austin and Geotis, 1978).

Given the assumed spatially averaged ground-truth field, or the "original" field, the radar rainfall field and the point raingage measurement field are artificially generated as follows.

Generation of Radar Rainfall

To generate the radar rainfall field, the model by Krajewski and Georgakakos (1985) is used. As noted earlier, the error structure of radar rainfall is not fully known. Radar measures average returned power. When the radar hardware is accurately calibrated, the average returned power can be converted to reflectivity factor via a radar equation. Expressing the measurement error in average returned power in decibels, we have (Battan, 1973).
\[ p = 10 \log_{10}\left(\frac{P_1}{P_0}\right) \]  \hspace{1cm} (15)

where \( P_1 \) and \( P_0 \) are power levels in watts, and \( p \) is the difference in power levels in dB

Making use of the linearity between the average returned power and the reflectivity factor, we have the following equation for the basic radar measurement error structure.

\[ Z_1 \sim Z_0 \ 10^{(p/10)} \]  \hspace{1cm} (16)

where \( Z_0 \) is the true reflectivity factor, and \( Z_1 \) is the measured reflectivity factor

Replacing \( Z \) with \( A R^b \), we have

\[ R_1 \sim R_o \ 10^{(p/10b)} \]  \hspace{1cm} (17)

where \( R_o \) is the true rainfall, and \( R_1 \) is the estimated rainfall

The above expression is only approximate in nature in that uncertainty in the \( Z-R \) relationship is not represented. Observations indeed show that the following qualitative features are present (Krajewski, 1987):

1) Errors are higher in high-rainfall gradient areas
2) Errors are higher in high-rainfall intensity areas
3) Errors are correlated in space
The radar rainfall generator incorporates the above characteristics and has the following structure:

\[ R(U) = O(U) 10^{c(U)} S(U) \]  

where \( R(U) \) is the generated radar rainfall at location \( U \), \( O(U) \) is the assumed spatially averaged ground-truth at the same location, \( c(U) \) is the random component of the noise at the same location, and \( S(U) \) is the deterministic component of the noise at the same location.

The random noise field \( c \) is assumed to be homogenous and isotropic, and is generated using the Turning Bands method (Mantoglou and Wilson, 1982). The deterministic noise field \( S \) simulates the first two qualitative features of radar rainfall noted above, and has the following form (Greene et al, 1980).

\[ S(U) = \frac{\overline{\nabla O(U)} \cdot O_{\text{max}}(U) + O(U) \cdot \overline{\nabla O(U)}_{\text{max}}}{2 \overline{\nabla O(U)}_{\text{max}} \cdot O_{\text{max}}(U)} \]  

where \( \overline{\nabla O(U)} \) is the average absolute value of the gradient in four directions around \( U \), \( \overline{\nabla O(U)}_{\text{max}} \) is the maximum absolute gradient, and \( O_{\text{max}} \) is the maximum value.

There are three parameters that control the statistical properties of the generated radar rainfall field. The first parameter specifies the ratio \( E[R]/E[O] \), or bias of the radar rainfall field in the mean. The second parameter, the variance of the logarithmic ratio of \( R/O \),
Var[log_{10}(R/O)] controls the amount of noise added in the radar rainfall field. The third parameter specifies the correlation distance of the random noise field, $e$.

**Generation of Raingage Measurements**

To generate the raingage measurements from the assumed spatially averaged ground-truth field, the model by Krajewski (1987) is used. As noted earlier, raingage measurements can have large catch deficiencies due to wind. In this work, however, it is assumed that the point raingage measurements are not biased against the spatially averaged ground truth rainfall, i.e.,

$$E[Z_{pg}(U_i)] = X(U_i)$$ for all $i$ \hspace{5em} (20)

where

- $U_i$ is the $i$th gage location,
- $X(\cdot)$ is the spatially averaged ground truth rainfall over that location,
- $Z_{pg}(\cdot)$ is the point raingage measurement at that location.

The above assumption makes direct comparison between the estimated field and the spatially-averaged ground-truth field possible, without introducing the meteorological variable, wind velocity. With this assumption, the generation of raingage measurements then amounts to converting a spatially averaged ground-truth to a point rainfall. Once a specified number of raingage locations are randomly generated, then, for each gage location, the relationship between the variances of the point process and the spatial average process presented by Rodriguez-Iturbe and Mejia (1974) is used to obtain the point variance at that gage location.
\[ \sigma_p^2 = \sigma_A^2 \int_0^d r(v) f(v) \, dv \]  

(21)

where \( \sigma_p^2 \) is the point process variance,
\( \sigma_A^2 \) is the spatial average process variance,
\( r(\cdot) \) is the correlation function for the point process,
\( f(\cdot) \) is the probability density function of the distance between two randomly chosen points in the square averaging area, and
\( d \) is the diagonal of the square.

The correlation function \( r(\cdot) \) is assumed to be exponential type, and its correlation scale is obtained by solving

\[ r_A(L_1) = \int_{A_1} \int_{A_2} r(U_1) \, r(U_2) \, dU_1 \, dU_2 \]  

(22)

where \( A_1 \) and \( A_2 \) are the adjacent averaging areas, and
\( r_A(L_1) \) is the lag-1 correlation coefficient for the spatial average process.

The lag-1 correlation coefficient, as well as the spatial average process variance, is computed from the assumed ground truth rainfall values surrounding the raingage location.

Once the point process variance is obtained at a gage location, the gage measurement \( Z_{pg}(U_i) \) is generated from

\[ Z_{pg}(U_i) \sim LN(X(U_i), \sigma_p^2) \]  

(23)

where \( LN(\cdot) \) denotes lognormal distribution with the given mean and variance,
\( X(\cdot) \) is the assumed spatially averaged ground truth over the area in which the gage is located, and
\( \sigma_p^2 \) is the point process variance.
The lognormality assumption implies that the error due to point sampling is lognormally distributed with mean zero. It, however, does not imply that the generated point gage measurements, $Z_g(U_j)$'s, will be distributed lognormally, since each measurement is generated with locally different mean and variance.

**Description of Simulation Experiment 1**

Given a GATE radar rainfall field, which acts as the spatially averaged ground-truth field, a total of twenty-four combinations of radar rainfall field and point raingage measurement field are generated to evaluate the estimators under various radar rainfall error characteristics and raingage densities.

Due to excessive computational requirements, a compromise has to be made between a larger number of combinations and a reduction in the amount of computation. The parameters chosen for the generators are as follow.

For the generation of the radar rainfall fields:
1) $\text{sigmar} = 0.005, 0.02$, representing the low and high degrees of corruption of the original field, respectively
2) $\text{cordis} = 8 \text{ km}, 16 \text{ km}$, representing the shorter and longer correlation distances in the random noise field, $e$, respectively
3) $\text{bias} = 1, 2$, representing no bias and 100 percent bias in the mean of the radar rainfall field, respectively

For the generation of the raingage measurement fields, three gage densities, 32, 160, and 286 gages over a 200x200 km area, are selected. The first gage density represents approximately the raingage density over the continental U.S.A. (one gage per 1000~2000 km$^2$, Wilson and Brandes, 1979). The second gage density represents approximately the
gage density above which the radar-gage estimates are no longer better than the gage-only estimates for Illinois convective storms (Hildebrand et al., 1979).

For each GATE radar rainfall field selected, a single simulation run then constitutes the following steps.

1) Generate a radar rainfall field (using one of the eight parameter combinations)
2) Generate a gage measurement field, each time with randomly varying raingage network configuration (one of three raingage densities)
3) Perform estimation
4) Compute estimation error statistics
5) Go to step 2, and repeat until convergence in estimation error statistics is achieved

The GATE radar rainfall data selected for this work are eight hourly rainfall fields from Julian day 245 of phase 2 of the GATE experiment. From hour 1 through hour 24, the data covers all of the stages, i.e., developing, mature, and dissipating, of a tropical convective storm. Again, to reduce the amount of computation, only eight representative hourly fields, hour 3, hour 6, hour 9, hour 12, hour 15, hour 18, hour 21, and hour 24 are selected. It is noted that the choice of hourly, convective rainfall fields provides a much more stringent test on the estimators, as opposed to using daily or stratiform rainfall fields.
ESTIMATORS AND THEIR COMPUTATIONAL ALGORITHMS

Ordinary Block-Kriging of Point Gage Measurements

The point raingage measurements are first block-kriged over the whole domain $\Omega$, with
the blocksize being 4x4 km, to render the spatial scale of the gage measurements
compatible with that of the radar rainfall. The resulting field will also represent the gage
measurement-only estimation.

The point gage measurement field semi-variogram is estimated using the following
non-parametric estimator (Journel and Huijbregts, 1978).

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^{N} \left( Z_{pg}(U_i + h) - Z_{pg}(U_i) \right)^2$$
(24)

where $\gamma(h)$ is the semi-variogram at lag distance $h$, assumed isotropic,
$Z_{pg}(U_i)$ is the point gage measurement at location $U_i$, and
$N$ is the total number of pairs with lag distance $h$.

The sample semi-variogram is then fitted with spherical, exponential, and gaussian
variogram models following the weighted least squares criterion given below.

$$\text{Minimize} \sum_{i=1}^{L} N_i \left( \gamma_i - f_i (c_0, c, r) \right)^2$$
(25)

where $N_i$ is the number of pairs at lag $i$,
$\gamma_i$ is the sample semi-variogram at lag $i$, and
$f_i (c_0, c, r)$ is the value of the variogram model at lag $i$, a function of $c_0$ (nugget effect), $c$ (sill), and $r$ (range or correlation length)
Among the three fitted variogram models, the one with the smallest sum of residuals squared is then chosen for the block-kriging.

**Ordinary Co-Kriging of the Block-Kriged Gage Field and the Radar Rainfall Field (Ordinary Co-Kriging 1)**

Before estimating the semi-variogram for the spatially averaged gage measurements, only those blocks that contain at least a single raingage are retained. The reason for using only the gage-containing blocks is that we want to view the gage-containing block estimates as the pseudo-measurements for the spatially averaged point raingage measurements with negligible (pseudo-) measurement error variance.

As noted earlier, kriging assumes perfectly known semi-variogram, where as the semi-variogram estimated from the sparse gage measurements will certainly suffer from a large estimation error. Block estimates for those blocks that are distant from the gage locations will be associated with larger estimation variances (kriging variances), and thus are more likely to be in error due not only to the error in the semi-variogram estimation but also to the larger kriging variances. The estimation (kriging) variance associated with the gage-containing block estimates, however, will be minimal, and thus the pseudo-measurements are not burdened further with other errors than the semi-variogram estimation error.

The semi-variograms for the spatially averaged gage measurements and the radar rainfall are estimated in the same way the point gage measurement field semi-variogram is estimated. The cross-variogram between the two fields is estimated from the following non-parametric estimator (Journel and Huijbregts, 1978).

\[
\gamma_{gr}(hl) = \frac{1}{2N} \sum_{i=1}^{N} \left( (Z_g(U_i + h) - Z_g(U_i)) (Z_r(U_i + h) - Z_r(U_i)) \right)
\]  

(26)
where \( \gamma_{gr}(hl) \) is the isotropic cross-variogram,
\( Z_g(U_i) \) is the block-kriged gage measurement at \( U_i \),
\( Z_r(U_i) \) is the radar rainfall at \( U_i \), and
\( N \) is the number of pairs with lag distance \( h \).

Once the three variograms are obtained and fitted by spherical, exponential, and gaussian models, the model which gives the smallest sum of the squared residuals is selected.

**Ordinary Co-Kriging of Ground-Truth Field and Radar Rainfall Field (Ordinary Co-Kriging 2)**

As noted earlier, co-kriging assumes perfectly known variograms, and thus will not account for the uncertainty associated with the estimated variograms. This is particularly important for the rainfall estimation problem since both the semi-variogram for the block-ckriged gage measurement field and the cross-variogram between the two fields are estimated from a very small number of data.

In co-kriging 2, we test the full potential of rainfall estimation using co-kriging by minimizing the uncertainty associated with the variograms and by eliminating the point sampling error associated with the gage measurements.

In place of the spatially averaged gage measurement field, the ground truth field is used as follows. The semi-variogram for the ground truth field and the cross-variogram between the ground truth field and the radar rainfall field are estimated using all the data in the ground truth field. Once the aforementioned variograms are estimated, only the gage-containing ground truth blocks are retained, as was in co-kriging 1, before co-kriging between the "sparse" ground truth field and the radar rainfall field is performed.
The differences between the co-kriging 1 and the co-kriging 2 can be summarized as follows.

1) Whereas, in co-kriging 1, the semi-variogram for the block-kriged gage measurement field and the cross-variogram between the block-kriged gage measurement field and the radar rainfall field are obtained from a very small number of data (at most in the order of hundreds, as the number of gage-containing blocks is bounded by the number of point gage measurements), the variograms for co-kriging 2 are estimated from a much larger number of data (in the order of thousands, as the number of ground truth data is the same as the number of radar rainfall data).

2) The ground truth rainfall has already the same spatial measurement scale as the radar rainfall data, and thus the block-kriging of the point gage measurement field is not necessary. Subsequently, the errors associated with the variogram estimation from the sparse gage measurements and the error due to point sampling of raingages are eliminated.

The estimation steps using universal kriging are exactly the same as the ordinary kriging case, except that now universal block-kriging and universal co-kriging replaces ordinary block-kriging and ordinary co-kriging, respectively.

In this work, the two-dimensional co-kriging program by Kafritsas and Bras (1984) is used. Here, we give only a brief description of the structure identification procedure adapted for the rainfall estimation problem. For detail, the reader is requested to refer to the aforementioned reference.

**Universal Block-Kriging of Point Gage Measurement Field**

The structure identification consists of two steps, order identification and generalized covariance estimation. First, the order of the point gage measurement field is identified as
follows. For all three orders, all the known point gage measurements are kriged assuming
\( K(lhl) = -h \). for each gage measurement at \( U_i \), the order that gives the smallest kriging error
\( \hat{\Delta} Z_{pg}(U_i) - Z_{pg}(U_j) \), is given the grade 1, and the two remaining orders, the grades 2 and
3. The grade of each order is then accumulated over all the points kriged, and the order
with smallest average grade is chosen as the best order.

Once the order is chosen, the generalized covariance is estimated as follows. First,
generalized increments of the form

\[
G_k = \sum_{i=1}^{N_g} \lambda_i Z(U_i) - Z(U_k)
\]  

(27)

where \( G_k \) is the kth generalized increment,
\( Z(U_i) \)'s are the point gage measurements surrounding \( Z(U_k) \),
\( \lambda_i \)'s are the weighting coefficients obtained by kriging for the
point \( U_k \) under the identified order,
\( N_g \) is the number of surrounding point gage measurements used, and
\( Z(U_k) \) is the kth known point gage measurement

are constructed for all the known point gage measurements. The variance of the
generalized increment is defined as satisfying

\[
\text{Var}[G_k] = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j K(U_i - U_j)
\]  

(28)

To estimate the generalized covariance \( K() \) that satisfies most closely for all the generalized
increments constructed, the following least squares regression is solved iteratively.
To start the iteration, \( K(lhl) = -lhl \) is initially assumed. The iteration is continued until a convergence is achieved, or stopped when no convergence is reached within a specified number of iterations. Computational experience shows that three iterations are enough if the generalized covariance converges at all.

There are 3, 7, and 11 possible forms for the generalized covariance function for order 1, 2, and 3, respectively (Table 1). Once all the valid, i.e., convergent and conditionally positive semi-definite, sets of coefficients for the generalized covariance are obtained, the one which gives the following ratio closest to 1 is selected.

\[
Q = \sum_{k=1}^{N} \left\{ G_k^2 - \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j K(lU_i - U_j) \right\}^2
\] (29)

To start the iteration, \( K(lhl) = -lhl \) is initially assumed. The iteration is continued until a convergence is achieved, or stopped when no convergence is reached within a specified number of iterations. Computational experience shows that three iterations are enough if the generalized covariance converges at all.

There are 3, 7, and 11 possible forms for the generalized covariance function for order 1, 2, and 3, respectively (Table 1). Once all the valid, i.e., convergent and conditionally positive semi-definite, sets of coefficients for the generalized covariance are obtained, the one which gives the following ratio closest to 1 is selected.

\[
r = \frac{\sum_{k=1}^{N_i} G_k^2}{\sum_{k=1}^{N_i} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j K(lU_i - U_j)}
\] (30)

Universal Co-Kriging 1

Once the point gage measurement field is block-kriged, the order of the spatially averaged gage measurement field and the order of the radar rainfall field are identified in the same way described above. In case the two orders are not the same, the higher order is selected as the common order. The generalized covariance of each field is estimated in the same way as described above. The generalized cross-covariance is obtained in a similar
manner. First, two sets of generalized increments are constructed from the two sets of data.

\[ G_{gk} = \sum_{i=1}^{N} \lambda_{gi} Z_g(U_i) \]
and
\[ G_{rk} = \sum_{j=1}^{M} \lambda_{rij} Z_r(U_j) \] (31)

The generalized cross-covariance is defined as satisfying

\[ E[G_{gk}G_{rk}] = \sum_{i=1}^{N} \sum_{j=1}^{M} \lambda_{gi} \lambda_{rij} K_{gr} (|U_i - U_j|) \] (32)

The generalized cross-covariance that satisfies the above most closely is obtained by minimizing the following iteratively.

\[ Q = \sum_{k=1}^{N} \left( G_{gk}G_{rk} - \sum_{i=1}^{N} \sum_{j=1}^{M} \lambda_{gi} \lambda_{rij} K_{gr} (|U_i - U_j|) \right)^2 \] (33)

Once the generalized covariances and the generalized cross-covariance are estimated, all the valid generalized covariance matrices are used to compute
\[
 r = \frac{\sum \sum G_{s_1} G_{s_2}}{\sum \sum \sum \sum \sum \sum \lambda_{s_1i} \lambda_{s_2j} K_{s_1s_2} (U_i - U_j)} 
\]  

(34)

where $S_1$ and $S_2$ denote the two sensors

The generalized covariance matrix that yields the value of $r$ closest to 1 is then selected for co-kriging 1.

The number of generalized increments that can be constructed is equal to the number of data available. The more number of generalized increments is constructed, the more reliable the generalized covariance estimate will be. Most radar rainfall fields contain data points in the order of thousands, and, in an iterative estimation scheme such as the one used in this work, constructing many number of generalized increments for each coefficient combination at each iteration can be computationally very expensive.

Starks and Fang (1982) have shown that, when the weighted least squared regression is used, the coefficient $a_1$ of $K(hh) = a_1 \cdot h$ approximately follows chi-square distribution with the degree of freedom equal to the number of generalized increments constructed. The confidence intervals of $a_1$, for various numbers of generalized increments constructed are shown as fractions of $a_1$, the estimate of $a_1$. 
<table>
<thead>
<tr>
<th>N</th>
<th>lower bound</th>
<th>upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.775</td>
<td>1.356</td>
</tr>
<tr>
<td>200</td>
<td>0.831</td>
<td>1.233</td>
</tr>
<tr>
<td>300</td>
<td>0.859</td>
<td>1.184</td>
</tr>
<tr>
<td>400</td>
<td>0.876</td>
<td>1.184</td>
</tr>
<tr>
<td>500</td>
<td>0.888</td>
<td>1.138</td>
</tr>
<tr>
<td>600</td>
<td>0.896</td>
<td>1.125</td>
</tr>
<tr>
<td>700</td>
<td>0.904</td>
<td>1.114</td>
</tr>
<tr>
<td>800</td>
<td>0.909</td>
<td>1.106</td>
</tr>
<tr>
<td>900</td>
<td>0.914</td>
<td>1.100</td>
</tr>
<tr>
<td>1000</td>
<td>0.918</td>
<td>1.094</td>
</tr>
</tbody>
</table>

In this work, a compromise is made between the number of generalized increments constructed and the amount of computation. In any case, however, a minimum of 700 generalized increments are constructed whenever the number of data exceeds the minimum.
DESIGN OF EXPERIMENT 2

The first experiment is a limited test of the estimation procedure in that the assumed ground truth fields are of oceanic convective type only. In this second simulation experiment, we test the estimation procedure further by using artificially generated rainfall fields of various characteristics as the ground truth fields. To generate the assumed ground truth fields, the space-time rainfall model by Waymire et al. (1984).

In this work, the sets of model parameters chosen by Valdes et al. (1985) are used to generate hourly rainfall fields of three different storm types (referred to as climates in the reference). The storm types are characterized by high intensities and large numbers of cells (Type 1), less high intensities and small numbers of cells (Type 2), less high intensities and intermediate numbers of cells (Type 3).

The theoretical work by Waymire et al. (1984) and the numerical simulation experiment by Valdes et al. (1985) deal mainly with the long-term, stationary aspects of the model. Our time scale of interest, however, is much smaller, i.e., hourly. Sivapalan and Wood (1987), using a conceptual description of the storm system similar to that of Waymire et al. (1984), provide the following non-stationary mean and covariance.

\[ [x(t, U)] = \langle v \rangle \rho L 2\pi D^2 \left( \frac{\beta}{\alpha - \beta} \right) i_o (e^{-\beta t} - e^{-\alpha t}) \]  \hspace{1cm} (35)
COV\[X(t_1, U_1), X(t_2, U_2)\]

\[= \langle v(n - 1) \rangle \rho_L i_o^2 (2\pi D_e^2)^2 (\frac{\beta}{\alpha - \beta})^2 \]

\[\cdot \eta[u_{11} + U_{b1}(t_2 - t_1); 2D^2 + 2\sigma_1^2] (u_{21})\]

\[\cdot \eta[u_{12} + U_{b2}(t_2 - t_1); 2D^2 + 2\sigma_2^2] (u_{22})\]

\[\cdot (e^{-\beta t_1} - e^{-\alpha t_1}) (e^{-\beta t_2} - e^{-\alpha t_2})\]

\[+ \langle v \rangle \rho_L i_o^2 (2\pi D_e^2)^2 (\frac{\beta}{2\alpha - \beta})\]

\[\cdot \eta[u_{11} + U_{b1}(t_2 - t_1); 2D^2] (u_{21})\]

\[\cdot \eta[u_{12} + U_{b2}(t_2 - t_1); 2D^2] (u_{22})\]

\[\cdot e^{-\alpha(t_2 - t_1)} (e^{-\beta t_1} - e^{-2\alpha t_1})\]

(36)

where

\[\eta[a, b^2] (x) = \frac{1}{\frac{1}{2} e^{\frac{(x - s)^2}{2b^2}}} \]

\[U_1 = (u_{11}, u_{12}),\]

\[U_2 = (u_{21}, u_{22}),\]

\[\langle \rangle \] is the expectation operator
\( v \) is the number of cells per cluster potential,
\( \rho_L \) is the mean density of cluster potentials,
\( i_0 \) is the rainfall intensity at cell center at the time of birth
\( D \) is a measure of a cell's spatial extent,
\( \beta \) is the cellular birth rate,
\( \alpha \) is a measure of cell duration,
\( U_{b1} \) and \( U_{b2} \) are the storm velocity along x and y directions, respectively, and
\( \sigma_1 \) and \( \sigma_2 \) are the cell location parameters within a cluster potential region along x and y directions, respectively.

The above expressions are for the instantaneous point rainfall field. We can obtain the expressions for the hourly, spatially averaged rainfall field by integrating Eqs. 35 and 36. Numerical results, however, show that spatial averaging does not significantly affect the covariance structure. In this work, we assume that hourly point rainfall field represent the hourly spatially averaged rainfall field. The mean and the covariance for the hourly rainfall field, accumulated over the \( i \)th hour, then, are given by

\[
\begin{align*}
\mathbb{E}[i_i] &= \langle v \rangle \cdot \rho_L (2\pi D^2) i_o \beta - \beta \\
\text{COV}_i[U_1, U_2] &= \int_\tau^\tau \int \text{Eq. (36)} \, ds \, dt
\end{align*}
\]
The rainfall fields generated by the model are homogeneous in space. For this reason, we omit the estimation procedure using universal kriging. Otherwise the design of this experiment is essentially the same as the first experiment. In this section, we only describe the steps that are different from those of Experiment 1.

Description of Simulation Experiment 2

First, 286 randomly scattered raingage locations are generated over the 200 x 200 km area. The first 32 raingage locations represent the sparse network. The first 32 raingage locations plus the next 128 raingage locations represent the dense network. The three raingage networks are fixed throughout the simulation experiment. A single simulation run then constitutes the following steps.

1) generate the assumed ground truth field (one of three storm types)
2) select a raingage network (one of three)
3) generate the radar rainfall field (one of eight parameter combinations)
4) perform the estimation
5) compute the error statistics
6) go to step 1 until convergence in error statistics is achieved

In Simulation 1, the gage network configuration is changed for each radar rainfall field generated since we have only one realization of the ground truth field for each hourly stage of the storm. In Simulation 2, however, multiple number of realizations is available from the rainfall model, and thus the gage configurations are left unchanged.

One distinct advantage of having a model as the one used is that we have a priori knowledge of the second-order statistics of the ground truth fields.
In ordinary co-kriging 2, we may use the theoretical semi-variogram for the ground truth field instead of the fitted semi-variogram. The particular case of interest is using the theoretical correlation function and the sample variance. Numerical experiments, however, show that the covariance matrix thus constructed is seldom semi-positive definite when the cross-variogram and the radar rainfall field semi-variogram are fitted with spherical, exponential, or gaussian model. Due to complex, and thus non-parsimonious nature of the theoretical covariance functional, no further attempts are made to fit the other two variograms, with the theoretical covariance functional to help induce positive definiteness.
DESIGN OF EXPERIMENT 3

The success of the estimation procedures used in this work hinges mostly upon whether a strong enough spatial predictability is present in the rainfall field, the spatial predictability being expressed as the correlation scale. There are situations, however, such as earlier stages of a convective rainfall, where rainfall fields simply do not possess strong enough spatial correlation for the estimation procedures to be effective. In other situations, such as orography-enhanced rainfall, rainfall fields may not be assumed to be homogeneous, or satisfying the intrinsic hypothesis. In this experiment, we make use of a physically-based rainfall estimation model to obtain the mean field of the rainfall measurement fields. The idea is that, if the mean field captures the larger scale spatial variability of the rainfall, the burden on the part of the residual prediction will be reduced.

In this work, the two dimensional precipitation model by Georgakakos (1987) is used. It is an extended version of the station precipitation model by Georgakakos and Bras (1984), and details of the model description can be found in Georgakakos (1987).
SUMMARY OF RESULTS AND CONCLUSIONS

Results

i) Global Statistics

This result represents average performance of each estimator over all the parameter combinations for all the stages of storm development.

All the kriging estimators give smaller mean error than the radar or the Brandes method (Fig. 3). Throughout the figures, the following abbreviations are used:

RA = radar-only estimation
BR = Brandes method
OB = ordinary block-kriging (gage-only estimation)
O1 = ordinary co-kriging 1
O2 = ordinary co-kriging 2
DB = disjunctive block-kriging (gage-only estimation)
D1 = disjunctive co-kriging 1
D2 = disjunctive co-kriging 2
UB = universal block-kriging (gage-only estimation)
U1 = universal co-kriging 1
U2 = universal co-kriging 2

Among the kriging estimators, co-kriging 2 estimators give smaller mean error, and co-kriging 1 estimators give larger mean error (Fig. 2).

All the kriging estimators give smaller root mean square error than the radar or the Brandes method (Fig. 2).

Disjunctive block-kriging gives smaller root mean square error than ordinary or universal block-kriging (Fig. 2).

Co-kriging 1 estimators give improvements in root mean square error over block-kriging estimators (gage-only estimation). The margin of improvement is larger for disjunctive co-kriging and smaller for universal co-kriging (Fig. 2). Among
co-kriging 1 estimators, disjunctive co-kriging gives the smallest root mean square error, followed by ordinary co-kriging and universal co-kriging (Fig. 2).

Co-kriging 2 estimators give much smaller root mean square error than co-kriging 1 estimators (and, subsequently, block-kriging estimators). The margin of improvement is larger for ordinary and disjunctive co-kriging estimators and smaller for universal co-kriging estimators (Fig. 2). Among co-kriging 2 estimators, disjunctive co-kriging gives the smallest root mean square error, followed by ordinary and universal co-kriging (Fig. 2).

All the kriging estimators tend to overestimate estimation variance. Among them, disjunctive co-kriging 1 and universal co-kriging 2 give better standardized mean square error (Fig. 2).

As for the MAP error, the kriging estimators exhibit performance characteristics very similar to that of the root mean square error case. At smaller averaging areas, however, kriging estimators sometimes do no better than radar. In Fig. 3, cumulative error in total rainfall depth is shown over different areas.

Also, for the mean error over various ranges of rainfall depth, the kriging estimators exhibit performance characteristics very similar to that of the root mean square error case. At ranges of higher rainfall depth, however, radar and Brandes method do occasionally better than the kriging estimators (Fig. 4).

Error histograms for all the ordinary and universal kriging estimators show well-defined peak at zero. Disjunctive kriging estimators give flatter peaks that are often negatively skewed. This characteristic is more pronounced for disjunctive block-kriging, less pronounced for disjunctive co-kriging 1, and further less pronounced for disjunctive co-kriging 2, which locates mode at zero (Fig. 5).
Residual power spectra show that all the kriging estimators are effective in capturing smaller scale fluctuations. Radar and Brandes method often leave out smaller scale fluctuations (Fig. 6). Comparison among the kriging estimators shows that, even after eliminating the effect of bias in the mean (of the kriging error), co-kriging gives only a small improvement in residual variance reduction over block-kriging. The margin of this improvement is larger for disjunctive co-kriging and smaller for ordinary and universal co-kriging.

ii) Effect of sigmar (radar field noise parameter)

In general, sigmar has no significant effect on mean error for all the co-kriging estimators. When its effect is noticeable, co-kriging 1 estimators tend to give higher mean error when sigmar is high. Both radar and Brandes method are more sensitive to sigmar, and its effect depends on the development stage of the storm (Fig. 7).

In general, sigmar has no significant effect on root mean square error for all the co-kriging estimators. When its effect is noticeable, co-kriging 2 estimators tend to be more adversely affected by high sigmar than co-kriging 1 estimators. Both radar and Brandes method are more sensitive to sigmar, and its effect depends on the development stage of the storm (Fig. 7).

In general, all the kriging estimators tend to further overestimate the estimation variance when sigmar is high (Fig. 7).

iii) Effect of bias (bias of radar field in the mean)

In general, bias has no significant effect on mean error and root mean square error for all the co-kriging estimators (Fig. 8). When its effect is noticeable, universal co-kriging 2 tends to be more adversely affected by high bias. Brandes method is shown to be very sensitive to bias (Fig. 8). When there is no bias, radar alone gives smaller root mean
square error than all the block-kriging and co-kriging estimators. Only ordinary co-kriging 2 and disjunctive co-kriging 2 give smaller root mean square error than radar for all the values of bias (Fig. 8).

All the co-kriging estimators tend to overestimate estimation variance further when the bias is high (Fig. 8). This tendency is more pronounced for ordinary co-kriging 1, disjunctive co-kriging 1, and universal co-kriging 2.

iv) Effect of cordis (correlation length of random noise)

Cordis has no significant effect on all the co-kriging estimators (Fig. 9). In some cases, longer cordis tends to lower root mean square error. Both radar and Brandes method has more significant dependence on cordis (Fig. 9).

Cordis shows no consistent effect on standardized mean square error. In some cases, co-kriging 1 estimators give better standardized mean square error when cordis is longer (Fig. 9).

v) Effect of gage density

When the gage density is low, substantially higher mean error is shown, particularly for co-kriging 1 estimators (Fig. 10). Error histograms show that, when the gage density is low, no clear peak is defined (Fig. 11).

For the root mean square error, the performance of the estimators is characterized by the following (Fig. 10):

1) Brandes method is not consistent in that a larger number of gages does not necessarily result in reduction of root mean square error.
2) All the kriging estimators are consistent in that a larger number of gages results in reduction of root mean square error.

3) Marginal improvement by co-kriging 1 estimators over block-kriging estimators is most pronounced when the gage density is low. This is more pronounced for disjunctive kriging and less pronounced for universal kriging.

4) In general, for ordinary and universal co-kriging estimators, marginal improvement by co-kriging 2 over co-kriging 1 is larger than that by co-kriging 1 over block kriging for all the gage densities. For disjunctive kriging, the reverse is observed.

5) In general, increase in the number of gages from low to medium results in more reduction in root mean square error than the reduction due to increase from medium to high.

In general, increase in gage density results in more accurate standardized mean square error for all the kriging estimators, in particular, co-kriging 1 estimators and universal co-kriging 2 estimator.

Conclusions

Use of radar rainfall via co-kriging does improve rainfall estimation over gage-only estimation or Brandes method. The improvement is consistent under various measurement error characteristics of radar rainfall and varying raingage density.

Universal co-kriging, as implemented in this work, does not seem to warrant its routine application for the following reasons: 1) structure identification is very difficult and, thus, computationally expensive, and 2) the assumption of the intrinsic hypothesis may not be valid.

In general, disjunctive kriging is superior to ordinary or universal kriging.
Performance of linear co-kriging is found to deteriorate substantially due to 1) uncertain covariance structures and 2) raingage sampling error. If reliable estimates of the covariance structures can be obtained, a significant improvement is expected.
Fig. 1. Schematic representation of the radar (squares) and rain gage (dots) data in the domain $\Omega$.

(from Krajewski, 1987)
<table>
<thead>
<tr>
<th>$k = 0$</th>
<th>$k = 1$</th>
<th>$k = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K(h) = c\delta(h)$</td>
<td>$K(h) = c\delta(h)$</td>
<td>$K(h) = c\delta(h)$</td>
</tr>
<tr>
<td>$K(h) = a_1</td>
<td>h</td>
<td>$</td>
</tr>
<tr>
<td>$K(h) = c\delta(h) + a_1</td>
<td>h</td>
<td>$</td>
</tr>
<tr>
<td>$K(h) = c\delta(h) + a_1</td>
<td>h</td>
<td>$</td>
</tr>
<tr>
<td>$K(h) = a_1</td>
<td>h</td>
<td>+ a_3</td>
</tr>
<tr>
<td>$K(h) = c\delta(h) + a_1</td>
<td>h</td>
<td>+ a_3</td>
</tr>
<tr>
<td>$K(h) = a_1</td>
<td>h</td>
<td>+ a_5</td>
</tr>
</tbody>
</table>

Table 1

Generalized Covariance Models and Corresponding Orders
(from Kafritsas and Bras, 1984)
ME, RMSE, AND SMSE
GATE DAY 245 ENDING HOUR 9

Figure 2
MAP ERROR
GATE DAY 245 ENDING HOUR 3

LEGEND
- BRANDES
- ORD BLK
- ORD CO1
- ORD CO2
- RADAR

ERROR IN M/HR

MUTLIPLE OF STANDARD DEVIATION

Figure 4
ERROR HISTOGRAM
GATE DAY 245 ENDING HOUR 12
BIAS 1.0

LEGEND
--- = BRANDES
----- = ORD BLK
------ = ORD CO1
----- = ORD CO2

LEGEND
--- = BRANDES
----- = DIS BLK
------ = DIS CO1
----- = DIS CO2

LEGEND
--- = BRANDES
----- = UNI BLK
------ = UNI CO1
----- = UNI CO2

Figure 5
RESIDUAL POWER SPECTRUM
GATE DAY 245 ENDING HOUR 21
SIGMAR 0.020

LEGEND
--- BRANDES
----- DIS BLK
--- DIS CO1
--- DIS CO2

LEGEND
--- BRANDES
----- DIS BLK
--- DIS CO1
--- DIS CO2

LEGEND
--- BRANDES
----- DIS BLK
--- DIS CO1
--- DIS CO2

LEGEND
--- BRANDES
----- DIS BLK
--- DIS CO1
--- DIS CO2

Figure 6
ME, RMSE
GATE DAY 245 ENDING HOUR 9

LEGEND
○ - SIGMAR-0.005
● - SIGMAR-0.02

ME IN MMAR

LEGEND
○ - SIGMAR-0.005
● - SIGMAR-0.02

Rmse In MmAr

SMSE
GATE DAY 245 ENDING HOUR 12

LEGEND
○ - SIGMAR-0.005
● - SIGMAR-0.02

Figure 7
ME, RMSE
GATE DAY 245 ENDING HOUR 6

Legend
- BIAS-1.
- BIAS-2.

Figure 8
ME, RMSE, AND SMSE
GATE DAY 245 ENDING HOUR 12

Figure 9
ME, RMSE, AND SMSE
GATE DAY 245 ENDING HOUR 3

ME IN MM/HR

Legend:
- NO-32
- NO-160
- NO-286

RISSE IN MM/HR

Legend:
- NO-32
- NO-160
- NO-286

SMSE

Legend:
- NO-32
- NO-160
- NO-286

Figure 10
ERROR HISTOGRAM
GATE DAY 245 ENDING HOUR 9
NG 32

LEGEND
--- = BRANDES
--- = ORD BLK
--- = ORD CO1
--- = ORD CO2

LEGEND
--- = BRANDES
--- = DIS BLK
--- = DIS CO1
--- = DIS CO2

LEGEND
--- = BRANDES
--- = UNI BLK
--- = UNI CO1
--- = UNI CO2

Figure 11
REFERENCES


APPENDIX C

RAINFALL ESTIMATION BY NON-LINEAR CO-KRIGING OF DATA FROM MULTIPLE SENSORS
RAINFALL ESTIMATION BY NON-LINEAR CO-KRIGING OF
DATA FROM MULTIPLE SENSORS

by

Ali Azimi-Zonooz
Utah Water Research Laboratory
Utah State University
Logan, Utah 84322-8200

February, 1988
Optimal Rainfall Estimation by Disjunctive Co-kriging of Radar-Rainfall and Raingage data

Ali Azimi-Zonooz, W. F. Krajewski, and D. S. Bowles

The feasibility of geostatistical estimation technique, disjunctive co-kriging, for optimal merging of rainfall data from raingage and radar observations is investigated in this study by use of controlled numerical experiments. Synthetic radar and raingage data are generated with their inherent error characteristics, using high quality radar data considered to be the ground-truth field. Mean areal precipitation estimates are obtained based on the minimum variance, unbiased property of kriging techniques under the second order homogeneity assumption of rainfall fields. The evaluation of estimated rainfall fields is done based on refinement of spatial predictability from what is provided from each sensor alone. The true rainfall fields consist of high quality hourly radar data from the international Global Atlantic Tropical Experiment (GATE), and numerically simulated rainfall fields with various climatic characteristics. Attention is mainly given to removal of noise and bias that are synthetically introduced to radar measurements. The influence of raingage network density on estimated rainfall fields is also examined.

INTRODUCTION

For operational forecasting of river flows in real time, the accurate estimation of the spatial distribution of precipitation rate over river basins is of paramount importance. High spatial variability of precipitation at river basin scale and sparse network of raingage data are known to be the major cause of uncertainty in forecasting streamflows. The accurate estimation of spatial distribution of rainfall from only raingage data for operational purposes requires a very dense network of automated instruments, which entails large installation and operational costs. In recent years, ground based radar systems have been implemented to provide continuous measurements of rainfall amounts in space and time over river basins. It is generally recognized that proper prediction of river basin response to rainfall are highly dependent on the accuracy of determining storm system locations within watershed boundaries. Thus, the ability to obtain higher resolution estimates of spatial variability in the rainfall fields becomes important in the case
of identification of locally intense storms which could lead to floods and particularly to flash floods. Since weather-radar could detect the precipitation patterns over a large area at short periods of time, it could provide the continuous sampling gaps which is not provided from raingage data. However, experience with the implementation of weather-radar for operational hydrologic applications have illustrated that radar measurements of rainfall are corrupted by various random and systematic errors (Zawadzki, 1984) which could lead to errors of as much as 200% (Wilson, 1970). These errors create large discrepancies between radar measurements of rainfall rate and the true rainfall rate at the ground surface. Therefore it is believed that due to high accuracy of raingage data, the combination of observations from both sensors will improve the areal estimates of rainfall. An important consideration in this research is selection of a suitable estimation technique that produces sufficiently accurate mean areal precipitation estimates for input to hydrologic models, which will make operational forecasting of river flows and flash floods possible. By assessing errors in the estimation procedures the adequacy of these techniques to achieve this goal will be evaluated.

PROBLEM FORMULATION

We begin the problem of estimating the spatial distribution of rainfall by assuming that the rainfall field \( Z(u) \) constitutes a second order homogeneous and isotropic random process in 2-D space and that rainfall observations include: (1) point raingage measurements of rainfall scattered over a certain region, and (2) radar measurements of rainfall which consist of areally averaged observations over a regular grid. Both measurements are accumulated over time intervals \( \Delta T \) which could range from hourly to daily periods. Our goal is to further improve the areal estimates of rainfall which is traditionally provided from raingage data alone, by including radar in the estimation, in the hope of providing more accurate estimates. Consequently the problem is formulated as finding a function \( f(G(u_i), R(v_j)), i=1, \ldots, n_g; j=1, \ldots, n_r, \) which is an unbiased minimum variance estimator of \( Z(u_0) \), where \( G \) and \( R \) are used to identify gages and radar respectively and, \( n_g \) and \( n_r \) are the
number of observations used in the estimation. $Z(u_0)$ is theoretically obtained from integrating the rainfall rate $Z(u)$ over the area $A$ and averaging over $A$

$$Z(u_0) = \frac{1}{A} \int_A Z(u) du$$  \hspace{1cm} (1)$$

where $A$ is used to denote the unit area where both rainfall estimates or radar observations are obtained. The best approximation of $Z(u_0)$ by the measurable function $f(G_i, R_j)$ is the conditional expectation of $Z(u_0)$ given the raingage and radar observations (Matheron, 1976).

$$Z^*(u_0) = f(G(u_i), R(v_j)) = E[Z(u_0) | G(u_i), R(v_j)], \quad i=1, \ldots, n_g; j=1, \ldots, n_r$$  \hspace{1cm} (2)$$

Computation of the above conditional expectation requires knowledge of the joint probability density function of $(n_g + n_r + 1)$ variables $Z(u_0)$, $G(u_i)$, and $R(v_j)$, which is a difficult task to obtain in reality. This is due to the unavailability of sufficient number of rainfall field realizations to construct the joint distribution of the rainfall observations or the high computational costs of obtaining such an estimate. However, if the assumption is made that the rainfall process is multivariate Gaussian, which is seldom the case, then the above conditional expectation becomes a linear operator, which requires nothing more than the knowledge of the covariance function which could be estimated from observations. It is only in this case that the assumption of second order homogeneity of the rainfall process leads to the full definition of the multivariate density function. Therefore, in the framework of rainfall estimation from raingage and radar observations, the disjunctive co-kriging estimator (DCK) is proposed, which requires no prior assumptions about the distribution of rainfall fields or knowledge of the covariance function. However, in DCK estimator the assumption is made that the rainfall process can be obtained as a transformation of a second order homogeneous field which has a bivariate standard Gaussian distribution.

In the ordinary co-kriging of raingage and radar-rainfall data, proposed by Krajewski (1987), the estimator $Z^*(u_0)$ was obtained as a linear combination of $G(u_i)'s$ and $R(u_j)'s$ given by ordinary co-kriging
(OCK) estimator

\[ \hat{z}_{OCK}(u_0) = \sum_{i=1}^{n} \lambda_i G(u_i) + \sum_{j=1}^{n} \gamma_j R(v_j) \]  

in which the coefficients \( \lambda_i \)'s and \( \gamma_j \)'s are obtained as the LaGrange multipliers for a constrained optimization problem. The DCK estimator is obtained by forming the estimator

\[ \hat{z}_{DCK}(u_0) = \sum_{i=1}^{n} f_i[g(u_i)] + \sum_{j=1}^{n} h_j[r(v_j)] \]

where \( f_i \)'s and \( h_j \)'s are a sequence of nonlinear functions, which makes DCK a nonlinear estimator and is more general than the OCK estimator. As explained before the DCK method makes use of the transformed variables which are assumed to be uni and bivariate normally distributed. These transforms are defined as

\[ G(u_i) = \Phi_g[g(u_i)] = \sum_{k=0}^{K} C_k H_k[g(u_i)] \]  

\[ R(v_j) = \Phi_r[r(v_j)] = \sum_{k=0}^{K} D_k H_k[r(v_j)] \]

where \( H_k \) is a Hermite polynomial of order \( k \), and \( g(u_i) \) and \( r(v_j) \) are the standard normal random variables which are obtained from transforms

\[ g(u_i) = \Phi_g^{-1}[G(u_i)] \]  

\[ r(v_j) = \Phi_r^{-1}[R(v_j)] \]

A Hermite polynomial of order \( k \) may be evaluated by the recursive relationship

\[ H_{k+1}(y) = yH_k(y) - kH_{k-1}(y) \]

where \( H_0 = 1 \) and \( H_1 = y \). Making use of the orthogonality of Hermite polynomials with Gaussian density the Coefficients \( C_k \) can be determined by

\[ C_k = \frac{1}{k!} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \Phi(y) H_k(y) e^{-y^2/2} dy \]

The coefficients \( C_k \) can be obtained in different ways. The most widely
used method in geostatistics is by numerical integration using the
Gauss-Hermite quadrature (Abramowitz and Stegun, 1970)
\[
C_k = \frac{1}{J!\sqrt{2\pi}} \sum_{j=1}^{J} \Phi(Y_j) W_j H_k(Y_j) e^{Y_j^2/2} \tag{11}
\]
where \( J \) is the total number of terms used for abscissas \( Y_j \) and weight
factors \( W_j \). The above numerical integration is employed if \( \Phi_g \) and \( \Phi_r \)
are nonlinearly approximated. However, for piecewise linear
interpolation an analytical solution for Hermite integration is possible
as given by Puente and Bras (1982). The following approximation of the
anamorphosis function was proposed
\[
\Phi(y) = \begin{cases} 
  z_1 & y \leq y_1 \\
  a_k y + b_k & y_k \leq y \leq y_{k+1} \quad k = 1, \ldots, n-1 \\
  z_n & y \geq y_n 
\end{cases} \tag{12}
\]
where \( z_1, z_2, \ldots, z_n \) are ordered data observations and \( y_1, y_2, \ldots, y_n \)
are corresponding values of standard Gaussian distribution function.
From (12) it follows that
\[
a_k = \frac{z_{k+1} - z_k}{y_{k+1} - y_k} \tag{13}
\]
\[
b_k = \frac{z_k y_{k+1} - z_{k+1} y_k}{y_{k+1} - y_k} \tag{14}
\]
Incorporating (12), (13), and (14) into (10) and integrating results in
an analytic solution. The final expressions are
\[
C_0 = z_1 G(y_1) + \sum_{i=1}^{c_0} \left\{ b_i G(y) \left| \begin{array}{c} Y_{i+1} \\ Y_i \end{array} \right. \right. - a_i G(y) \left| \begin{array}{c} Y_{i+1} \\ Y_i \end{array} \right. \right. + z_n [1-G(y_n)] \tag{15}
\]
\[
C_1 = - z_1 g(y_1) + \sum_{i=1}^{c_1} \left\{ -b_i g(y) \left| \begin{array}{c} Y_{i+1} \\ Y_i \end{array} \right. \right. + a_i [G(y) - yg(y)] \left| \begin{array}{c} Y_{i+1} \\ Y_i \end{array} \right. \right. + z_n g(y_n) \tag{16}
\]
and for $k \geq 2$

$$C_k = \left[-z_1 H_{k-1}(y_1) g(y_1) + z_n H_{k-1}(y_n) g(y_n) + \sum_{i=1}^{n-1} \left\{ g(y) H_{k-1}(y) \left[ Y_{i+1}^{1} \right] + a_1 \left[ y g(y) H_{k-1}(y) + g(y) H_{k-2}(y) \left[ Y_{i+1}^{1} \right] \right] \right\} \right] (k!)^{-1}$$

where $g(y)$ is standard normal density

$$g(y) = (2\pi)^{-1/2} e^{-1/2y^2}$$

and

$$G(y) = (2\pi)^{-1/2} \int_{-\infty}^{y} g(y) \, dy$$

Expanding the DCK estimator (Eq. 4) in a series of Hermite polynomials gives

$$Z_{\text{DCK}}^*(u_o) = \sum_{k=0}^{\infty} \sum_{i=1}^{n} f_{ik} H_k[g(u_1)] + \sum_{j=0}^{K} \sum_{j=1}^{\infty} h_{jk} H_k[x(v_j)]$$

where $f_{ik}$ and $h_{jk}$ are coefficients of Hermite expansion. The problem is to determine the weights $f_{ik}$ and $h_{jk}$ such that the following holds

1. The estimator $Z_{\text{DCK}}^*(u_o)$ is a minimum variance estimator

$$\text{Var}[Z_{\text{DCK}}^*(u_o) - Z(u_o)] = \mathbb{E}[\left\{ Z_{\text{DCK}}^*(u_o) - Z(u_o) \right\}^2] \rightarrow \min$$

2. The estimator $Z_{\text{DCK}}^*(u_o)$ is an unbiased estimator of $Z(u_o)$

$$\mathbb{E}[Z_{\text{DCK}}^*(u_o) - Z(u_o)] = 0$$

Following the same approaches given by Journel and Huijbregts (1978) and Yates (1986) we obtain the following system of equations

$$\mathbb{E}[\Phi_g[g(u_o)] | g(u_a)] = \sum_{j=1}^{n} \mathbb{E}[f_1[g(u_1)] | g(u_a)] + \sum_{j=1}^{n} \mathbb{E}[h_j[x(v_j)] | g(u_a)]$$

$$\mathbb{E}[\Phi_g[g(u_o)] | r(v_\beta)] = \sum_{j=1}^{n} \mathbb{E}[f_1[g(u_1)] | r(v_\beta)] + \sum_{j=1}^{n} \mathbb{E}[h_j[x(v_j)] | r(v_\beta)]$$

where $\alpha = 1, \ldots, n_g$ and $\beta = 1, \ldots, n_r$. Expanding the bivariate standard
Gaussian densities in terms of Hermite polynomials gives (Matheron, 1976)

\[ f(x, y) = \sum_{k=0}^{\infty} (\rho_{xy})^k H_k(x) H_k(y) \frac{g(x) g(y)}{k!} \]  

(25)

Then the conditional expectation of function \( \Phi(x) \) in (23) and (24) could be written as

\[ \mathbb{E}[\Phi(x) | y] = \sum_{k=0}^{\infty} (\rho_{xy})^k C_k H_k(y) \]  

(26)

Incorporating (26), (24), and (23) into (20) leads to the DCK estimate

\[ Z_{\text{DCK}}(u_0) = \sum_{k=0}^{K} C_k H_k^*[g(u_0)] \]  

(27)

where \( H_k^*[g(u_0)] \) is given by

\[ H_k^*[g(u_0)] = \sum_{i=1}^{n_g} a_{ik} H_k[g(u_i)] + \sum_{j=1}^{n_r} b_{jk} H_k[r(v_j)] \]  

(28)

where \( f_{ik} \) and \( h_{jk} \) in equation (20) are written as \( C_k a_{ik} \) and \( C_k b_{jk} \). The \( a_{ik} \) 's and \( b_{jk} \) 's are determined from

\[ \sum_{i=1}^{n_g} a_{ik} \rho_{gai}^k + \sum_{j=1}^{n_r} b_{jk} \rho_{r\beta j}^k = \rho_{g\alpha}^k \quad \alpha = 1, \ldots, n_g \]  

(29)

\[ \sum_{i=1}^{n_g} b_{ik} \rho_{r\beta i}^k + \sum_{j=1}^{n_r} a_{jk} \rho_{r\beta j}^k = \rho_{r\beta \beta}^k \quad \beta = 1, \ldots, n_r \]  

(30)

where \( \rho_{gai} \) and \( \rho_{r\beta j} \) are correlation functions of the gage and radar fields, \( \rho_{g\beta j} \) and \( \rho_{r\beta i} \) are cross correlation functions between radar and gage fields, \( \rho_{g\alpha} \) is the correlation function of point to be estimated and the gage field, and \( \rho_{r\beta \beta} \) is the cross correlation function between point to be estimated on the gage field and the radar field.
APPROXIMATION OF THE ANAMORPHOSIS FUNCTION

A fully controlled numerical experiment was designed and carried out to compare the two approaches for approximation of anamorphosis function. The approaches are the numerical solution given by Equation 11, and the analytic solution given by Equations (15-17). Samples of various sizes (50, 100, 200, 400, 800) were drawn from two-dimensional random, homogeneous, and isotropic fields with exponential covariance function. The correlation length of the fields varied from 5.0 to 40.0 units and the sampling domain was 100.0 by 100.0 units. The data were point observations at randomly generated locations. The locations were kept constant for all the realizations of the fields. The smaller samples were always included in the larger samples, so that we could simulate the expansion of our observational network.

The fields were generated using the Turning Band Method (TBM), described by Mantoglou and Wilson (1982). Ten realizations of the fields were used for each set of parameters. All the fields were N(0,1) (normal with zero mean and unit variance). The observations were generated for various levels of measurement error: 0.0, 10.0, and 50.0 percent. The measurement error is expressed as having a standard deviation equal to the specified percentage of the field value at a given point. That way the higher the absolute field values are the higher are the observational error. However, since the mean of the error term is always zero, the generated noise does not introduce a bias into the samples.

Normally distributed samples were generated and transformed to obtain lognormally distributed data and performance of the two approaches were compared. Results show that for large samples the fit of the Hermite approximation is poor in the upper tail of distribution for the numerical solution method (see Figure 1a, for example). This behavior of numerical solution approach is attributed to the initial step in the algorithm, where fitting a polynomial regression to the sample to obtain points necessary for the numerical integration scheme is made. The fit is dominated by a large number of points in the middle region of the distribution. The problem could be alleviated by careful selection of the polynomial degree or a weighted least squares fit, such a procedure, however, would increase computational costs, and make the algorithm even
Figure 1. Examples of anamorphosis function approximation. R is measurement error, B is correlation length of simulated random field, and N is the sample density.
more complicated. Another problem, common to both algorithms is that under certain sampling conditions there is no unique solution to inverse problem, Equations (7) or (8). An example of such situation is shown in Figure 1b. The approach that was taken was to search for the roots of (7) or (8) in the interval

\[ y_i - 2u \leq y \leq y_i + 2u \]

where \( y_i \) corresponds to \( z_i \) for which the solution is needed and \( u \) is the \( \max \{ y_i - y_{i-1} \mid i=1, \ldots, n \} \). If two or more solutions were present in the specified range then the minimum solution was selected in the upper tail and maximum in the lower tail. The investigated sampling conditions, such as correlation distance and measurement error, did not seem to affect the performance of the two algorithms.

EXPERIMENTAL DESIGN

A controlled numerical experiment was carried out to analyse the accuracy of ordinary block kriging, disjunctive block kriging, ordinary co-kriging, and disjunctive co-kriging for rainfall estimation. Precipitation fields having three different climatic characteristics were generated as simulated original fields by using the space-time rainfall model developed by Waymire et al. (1984). The three simulated climates (examples of which are shown in Figures 2), range from frequent storms, with high intensities and large number of cells (climate 1) to climates with less frequent storms and smaller intensities (climate 2). Details of the characteristics for each climate are given in Valdes et al. (1985). The experiment was carried out by repeating the analysis for 10 realizations from each climate type, while keeping the same noise parameters and raingage configuration unchanged. The results were then averaged across the ensemble of realizations for examining the effects of noise parameters on the estimated fields. The radar and gage generators developed and described by Krajewski (1987) were used to simulate radar and gage fields with their prespecified sampling characteristics. The radar noise parameters that chosen were bias = 1 and 2, noise variance = .005 and .02, and the noise correlation distance = 8.0 and 16.0 km. The raingage densities were set to 32, 160,
Figure 2. Examples of simulated hourly rainfall fields for various climates.
and 286, and an error of 0% was specified for gage measurements in all the runs. Climate 1, with maximum number of cluster potential centers and highest cell intensity (Figure 2a) provides a very rigorous evaluation of estimation techniques for testing the noise parameters which have the highest influence on the outcome of the estimated rainfall fields.

RESULTS OF NUMERICAL EXPERIMENTS

Comparison among various estimators are made in terms of their effectiveness to filter out measurement noise and bias from radar field, and their ability to estimate spatial variability for various climates. Since we assume a known true rainfall field and sampled raingage and radar-rainfall fields with known noise parameters, measures such as how well the merged field describes the spatial variability in the true rainfall field after removing noise and bias introduced by measurement error is possible. The first check on the performance of estimators is made to verify whether or not the methods effectively remove bias from the radar field. Figures (3-5) show the average power spectrum across all realizations for bias = 1, and 2 cases and various climates. As is evident from these results low frequency components of the radar-residual field are effectively attenuated by co-kriging techniques. Disjunctive co-kriging 2 provides the best performance to filter out the superimposed bias, where the residual spectrum is remarkably flat and is not significantly different from a white noise signal. Disjunctive co-kriging 1 provides the next best performance which is very close to the ordinary co-kriging 2 case. These plots also demonstrate that inclusion of the radar field data in the ordinary co-kriging 1 case does not lead to significant improvements over just the ordinary block kriging. Brandes method provides the worst performance for removal of bias from the radar field. The radar bias and noise removal capability of estimators are also examined in terms of mean error (ME) and root mean square error (RMSE) between the estimated and true fields. These results are plotted and shown in Figure 6. Disjunctive co-kriging 2 again shows the best performance for removal of noise and bias from the radar field. In general co-kriging methods are effective for removal of noise, where
Figure 3. Plots of average spectral density of residual fields for bias = 1, and 2 cases (climate 1).
Figure 4. Plots of average spectral density of residual field for bias = 1, and 2 cases (climate 2).
Figure 5. Plots of average spectral density of residual fields for bias = 1, and 2 cases (climate 3).
Figure 6. Illustration of the effects of bias and noise variance on the estimated rainfall fields using ME and RMSE for various climates.
markers showing ME and SMSE for 2 different noise levels are superimposed. Brandes method, which does not account for sampling characteristics of the radar field, shows the worst performance for noise removal.

The mean areal absolute deviations between estimated and true rainfall fields are examined for various rainfall intensities by using multiples of standard deviation of true rainfall field. These results are presented in Figure 7. Disjunctive co-kriging 2 provides the best estimates of spatial variability for all climate types. Kriging techniques consistently underestimate the true rainfall field, while the Brandes method provides a consistent overestimation for various intensity levels. Comparison of ordinary co-kriging 1 and block kriging shows that incorporation of radar field in the estimation does not lead to significant improvements of estimated spatial variability. In contrast for disjunctive co-kriging 1 inclusion of radar field in the estimation results in improved estimates of spatial variability.

Figure 8 summarizes the effect of radar noise correlation distance and raingage density on the estimated field. The two noise correlation distances that were examined show a similar performance on the outcome of the estimated fields, and higher raingage densities lead to improved areal rainfall estimates.

Table 1 is a summary of ME and RMSE for various spatial rainfall estimators that were examined in this study. These results were obtained from averaging these two error statistics across entire realizations of the rainfall fields. As was evident from previous results it appears that ordinary co-kriging 1 does not offer any improvements over just block kriging the gage field. Disjunctive co-kriging 1 gives a 33% improvement in RMSE over ordinary co-kriging 1 for climate 1, about 40% improvement for climate 2, and 21% improvement for climate 3. Disjunctive co-kriging 2 again has the best performance for spatial rainfall estimation.

CONCLUSIONS

The main purpose of this research has been to compare the performance
of various (co)-kriging methods for rainfall estimation under a radar umbrella. The results of numerical experiments indicate that spatial rainfall estimation by merging radar-rainfall and raingage data is more accurately performed using disjunctive co-kriging technique. One important conclusion that could be drawn from this work is that rainfall fields are generally described by nonlinear functions that can not be adequately estimated using linear estimators. The theoretical advantage of disjunctive co-kriging, as presented in this paper, is to estimate bivariate distributions of rainfall fields so that the conditional expectation of two observations at a time can be computed, using these distributions. The disjunctive co-kriging procedure is specifically aimed at deriving the probability distribution of rainfall fields, which in turn, preserves the spatial properties of rainfall fields more accurately. The main conclusions of the study could be summarized as follows:

1) For approximation of the anamorphosis function the analytical method is much more efficient in terms of CPU time and avoids certain instabilities of the numerical integration approach.

2) The disjunctive co-kriging estimator gives more accurate mean areal precipitation estimates than ordinary co-kriging, as is evident from root mean square error and mean error statistics obtained by averaging entire realizations of the estimated rainfall fields for each climate type. Disjunctive and ordinary co-kriging provide a substantial increase in accuracy over the Brandes method.

3) Examination of the residual power spectrum for all kriging estimators indicates that the best performance for the most uncorrelated noise is provided by disjunctive co-kriging 2, followed by disjunctive co-kriging 1, ordinary co-kriging 2, and disjunctive block kriging, ordinary co-kriging 1, and ordinary block kriging showing similar performance. Brandes method is found to be inadequate for an effective removal of noise and bias from the radar field.
4) Inclusion of radar-rainfall data in the estimation by co-kriging provides an improvement in the accuracy of estimated rainfall fields over only block-kriging in case of disjunctive co-kriging, but does not lead to significant improvements in case of ordinary co-kriging.

5) Disjunctive co-kriging gives better results when variograms estimated from transformed Gaussian data are used, over estimates provided from variograms of the radar and raingage fields.

6) Disjunctive co-kriging generally provides better estimates of spatial variability as is evident from error statistics that were obtained for various levels of rainfall intensities.

7) The effect of correlation distance of the radar error field was found to be insignificant on the estimated rainfall fields.

8) Spatial rainfall estimates obtained by disjunctive block kriging do not offer any improvements over estimates from ordinary block kriging.

RECOMMENDATIONS FOR FUTURE WORK
(1) The noise parameters that were used in this study were the measurement error, correlation distance, and bias of the radar field. The effect of measurement error in the gage observations was not studied and is recommended as a possibility for future study.

(2) Since the proposed methods of co-kriging for rainfall estimation are geared toward implementation of these techniques in an operational environment. therefore we like to propose that these methods be compared based on observed streamflows and streamflows predicted using merged rainfall estimates by input to a rainfall-runoff model.

(3) Since the requirements of forecasting services are different depending on required forecast lead times, size of watershed, rainfall intensities in space and time due to climatic differences, etc., so we like to recommend that estimators that are suitable for various
conditions to be used.

(4) As an alternative for non-linear estimation that was performed in this study, for future study a better approach is the estimates obtained from conditional expectation from joint distribution of rainfall measurements obtained from multiple realizations of the rainfall fields.

(5) Variogram estimation in present study was performed under the ergodicity assumption from single realization data. An alternative method of variogram estimation is from multiple measurements of the rainfall fields for various rainfall intensities, and updating in time as additional measurements become available. This approach is proposed as a possible direction for future research which will be performed as a continuation of present work.

(6) By use of disjunctive co-kriging one could obtain an estimate of a probability distribution of rainfall intensities at a certain unsampled location over a river basin where measurements can not be obtained. Through the use of this estimated probability distribution, probability of rainfall events exceeding a certain tolerance level (i.e., rainfall intensities that could lead to flash floods) could be computed.

In real time forecasting of floods the capability to forecast probability that a certain critical flood stage will be exceeded at various times in the future, or the probability of flood occurrence is crucial to quantifying the uncertainties due to high precipitation intensities. Since occurrence of floods from the time storm begins are controlled in large part by watershed hydrogeology, we like to propose as a direction for future research, development of a model that translates the probability distribution of rainfall intensities, obtained form nonlinear property of disjunctive co-kriging estimator into probability of flood occurrence from catchment area. This model could be utilized as one component of a flood prediction system.


Figure 7. Plots of mean areal precipitation error statistics for various levels of rainfall intensities using multiples of standard deviation of the true rainfall fields, a) comparison among different estimations, and b) comparison between Brandes method and radar field, for various climates.
Figure 8. Illustration of the effects of radar noise correlation distance and rain gage network density on the estimated fields using ME and RMSE for various climates.
Table 1. Comparison of ME and SMSE for various estimators and climates.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19.58</td>
<td>79.91</td>
<td>3.75</td>
<td>1.86</td>
<td>1.91</td>
<td>1.60</td>
<td>1.07</td>
<td>0.91</td>
</tr>
<tr>
<td>2</td>
<td>4.22</td>
<td>17.79</td>
<td>1.91</td>
<td>1.76</td>
<td>1.56</td>
<td>0.97</td>
<td>1.64</td>
<td>0.83</td>
</tr>
<tr>
<td>3</td>
<td>8.36</td>
<td>34.86</td>
<td>2.68</td>
<td>0.79</td>
<td>1.13</td>
<td>0.34</td>
<td>1.21</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>9.81</td>
<td>13.15</td>
<td>13.19</td>
<td>13.09</td>
<td>9.81</td>
<td>13.15</td>
<td>10.31</td>
<td>5.70</td>
</tr>
</tbody>
</table>
APPENDIX D

HERMITE APPROXIMATION OF ANAMORPHOSIS FUNCTION
HERMITE APPROXIMATION OF ANAMORPHOSIS FUNCTION

by

Witold F. Krajewski
Hydrologic Research Laboratory
National Weather Service
Silver Spring, Maryland

and

Ali Azimi-Zonooz
Utah Water Research Laboratory
Utah State University
Logan, Utah

May 1987
INTRODUCTION

Transformation of non-Gaussian statistical data into Gaussian samples is an important problem often met in geophysical data analysis. Due to many very useful properties of Gaussian variables, it is often desirable to find such transformation or to use some transformation that makes highly skew data to "look" more Gaussian. In general, it is very difficult to find (and/or prove) the analytical form of such transformation and, therefore, empirical transformation based on the data samples have to be used. In some applications, such as disjunctive kriging (Matheron, 1976; Randu, 1980; Puente and Bras, 1982; and Yates et al., 1985) a numerical approximation of such empirical transformation is required. In the estimation framework of disjunctive kriging, it is convenient to use the Hermite polynomial approximation. In this study, we investigated two algorithms that perform such approximation. The first one is proposed by Yates et al. (1985) and the second one by Puente and Bras (1982). We compared the two algorithms in terms of accuracy and efficiency (CPU time) under a wide range of sampling conditions.

HERMITE APPROXIMATION OF ANAMORPHISIS

Let us assume that the original variable \( Z \) is non-Gaussian and that there exists a transformation \( \phi \) such that

\[
Z = \phi^{-1}(Y) \quad Y = \Phi^{-1}(Z)
\]

where \( Y \) is a standard normal variable. In the following, \( Z = <z_1, z_2, \ldots, z_N> \) and \( Y = <y_1, y_2, \ldots, y_N> \) are realizations of variables \( Z \) and \( Y \), respectively.

The function \( \phi(y) \) can be approximated by a linear combination of Hermite polynomials

\[
\phi(y) = \sum_{k=0}^{\infty} C_k H_k(y)
\]  \hspace{1cm} (1)

where \( C_k \) are coefficients and \( H_k \) is the Hermite polynomial of order \( k \) and can be defined here as

\[
H_k(y) = (-1)^k e^{\frac{y^2}{2}} \frac{d^k}{dy^k} e^{-\frac{y^2}{2}}
\]  \hspace{1cm} (2)

Higher order polynomials are easily calculated using the following recurrence formula:

\[
H_{k+1}(y) = yH_k(y) - kH_{k-1}(y)
\]  \hspace{1cm} (3)

The choice of Hermite approximation of the anamorphosis function results from the domain of \( \phi(y) \) which covers the range \((-\infty, \infty)\) and the fact that the Hermite polynomials are orthogonal with respect to the standard normal distribution function. It can be shown that:
The formula for calculations of the $C_k$ coefficients results from the orthogonality of the Hermite approximation:

$$
C_k = (k!)^{-1} (2\pi)^{1/2} \int_{-\infty}^{\infty} \phi(y) H_k(y) e^{-y^2/2} \, dy
$$

The coefficients $C_k$ can be computed from (5) in several different ways. Probably the most natural one is by numerical integration using the Gauss-Hermite quadrature. According to that method, expression (5) is approximated by a finite sum.

$$
C_k = (k!)^{-1} (2\pi)^{1/2} \sum_{j=1}^{J} \phi(y_j) H_k(y_j) \cdot w_j e^{-y_j^2/2}
$$

The optimal values of $y_j$ and the corresponding weights $w_j e^{-y_j^2}$ for different values of $J$ can be found in Abramowitz and Stegun (1970). The choice of $J$ affects the accuracy of integration, but in general $J=10^2$ provides sufficiently good results.

An important problem that needs to be addressed is the computation of $\phi(y_j)$ for the values of $y_j$ given by the Gauss-Hermite integration scheme. Since the scheme may require (and in general it does) values of $\phi(y_j)$ at $y_j$ different.

Figure 1. Piece-wise linear approximation of the anamorphosis function.
than those corresponding to the data sample values, some kind of interpolation of the empirical distribution is needed. Figure 1 demonstrates the problem in cases where the empirical distribution is approximated with a piece-wise linear function.

Two basic simple approaches are possible. The first one is to fit a polynomial of high-order to the empirical distribution. This was the approach implemented by Yates et al. (1986). The second is to employ a simple linear approximation as presented on Figure 1. Moreover, in the case of single, linear approximation of empirical distribution, one can construct a corresponding linear approximation of the anamorphosis function and solve Eq. (5) analytically. This was done by Puente and Bras (1982). The following approximation of the anamorphosis was proposed:

\[
\phi(y) = \begin{cases} 
Z_1 & y \leq y_1 \\
a_k y + b_k & y_k \leq y \leq y_{k+1} \quad k = 1, \ldots, N-1 \\
Z_N & y \geq y_N
\end{cases}
\] (7)

where \(z_1, z_2, \ldots, z_N\) are ordered data observations and \(y_1, y_2, \ldots, y_N\) are corresponding values of standard Gaussian distribution.

From (7) it follows that:

\[
a_k = \frac{z_{k+1} - z_k}{y_{k+1} - y_k}
\] (8)

and

\[
b_k = \frac{z_k y_{k+1} - z_{k+1} y_k}{y_{k+1} - y_k}
\] (9)

Substituting (8) and (9) into (5) yields

\[
C_k = (k!)^{-1} (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} (a_i y + b_{i+1}) H_k(y) e^{-\frac{1}{2}y^2} dy 
\] (10)

for \(y_i \leq y \leq y_{i+1}\)

\[i = 1, \ldots, N-1\]

\[
C_k = (k!)^{-1} (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} z_{i} H_k(y) e^{-\frac{1}{2}y^2} dy
\] (11)

for \(y \leq y_{i+1}\)

\[
\text{and}
\]

\[
C_k = (k!)^{-1} (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} z_{i} H_k(y) e^{-\frac{1}{2}y^2} dy
\] (12)

for \(y \geq y_{i+1}\)
Let us denote

\[ g(u) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}u^2} \]  

and

\[ G(u) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} g(x)dx \]

Eq. (13) is the standard Gaussian density function and Eq. (14) is the standard Gaussian distribution function.

Then we have

\[ C_k = (k!)^{-1} \left\{ \int z_1 \cdot h_k(y) g(y) dy + \sum_{i=1}^{N-1} \int (a_i y + b_i) \cdot h_k(y) \cdot g(y) dy \right. \\
\left. + \int z_N \cdot h_k(y) g(y) dy \right\} \]

To evaluate the Eq. (15) we use the following relationships which result directly from (2):

\[ \int h_i(y) \cdot g(y) dy = -h_{i-1}(y) \cdot g(y) \quad \text{for } i \geq 1 \]  

and

\[ \int y \cdot h_i(y) \cdot g(y) dy = -y \cdot h_{i-1}(y) \cdot g(y) - h_{i-2}(y) \cdot g(y) \quad \text{for } i \geq 2 \]

The final expressions are:

\[ C_0 = z_1 \cdot G(y_1) + \sum_{i=1}^{N-1} \left[ \frac{y_{i+1}}{y_i} \right] - a_i g(y) \left\{ y_{i+1} \right\} + z_N \cdot [1 - G(y_N)] \]

\[ C_1 = -z_1 g(y_1) + \sum_{i=1}^{N-1} -b_i g(y) \left\{ y_{i+1} \right\} + a_i [G(y) - y \cdot g(y)] \left\{ y_{i+1} \right\} + z_N g(y_N) \]
The two approaches (Yates, et al, 1986, Eq. (6) and Puente and Bras, 1982, Eqs. (18)-(20)) were compared via a numerical simulation experiment. The Eqs. (18)-(20) are slightly different from the corresponding equations given in Puente and Bras, 1982, due to a different definition of Hermite polynomials used.

**NUMERICAL EXPERIMENT**

A fully controlled experiment has been designed and carried out to compare the two approaches. The samples of various size (50, 100, 200, 400, 800) were drawn from two-dimensional random, stationary, and isotropic fields with exponential covariance functions. The correlation length of the fields varied from 5.0 to 40.0 units and the sampling domain was 100.0 by 100.0 units. The data were point observations at randomly generated locations. The locations were then kept constant for all the realizations of the fields. Also, the smaller samples are always included in the larger samples. This way we can simulate the "expansion" of our observational network.

The fields were generated using the Turning Bands Method (TBM), very well described by Mantoglou and Wilson (1982). Ten realizations of the fields were used for each set of parameters. All the fields were N(0,1) (normal with zero mean and unit variance). The observations were generated for various levels of measurement error: 0.0, 10.0, and 50 percent. The measurement error is expressed as having a standard deviation equal to the specified percentage of the field value at a given point. That way the higher the absolute field values are the higher are observational error. However, since the mean of the error term is always zero, the generated noise does not introduce a bias into the samples.

The distribution of fields generated with the TBM is Gaussian, therefore, to test the anamorphism approximation, we have to transform the Gaussian data into some other distribution and then identify the inverse transformation via a Hermite approximation. We decided to use the convenient lognormal distribution. Thus, the N(0,1) fields are first transformed by exponential transformation and then sampled by a white noise adding mechanism:

\[
    z_i = \exp(y_i) + N(0, \sigma_i^2) \quad \text{for } i = 1, \ldots, N \tag{21}
\]
where

\[ r_i = R \cdot \exp(y_i) \quad \text{and} \quad R = \langle 0.0, 0.1, 0.5 \rangle \]

Since the transformation is \( Y = \ln(Z) \), the performance of both algorithms was compared in terms of mean square error with respect to the true anomorphosis:

\[
\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left[ \phi(y_i^*) - \sum_{k=0}^{K} C_k H_k(y_i) \right]^2
\]

(22)

and in terms of best fit to the data sample

\[
\text{ZSE} = \frac{1}{N} \sum_{i=1}^{N} \left[ Z_i - \sum_{k=0}^{K} C_k H_k(y_i) \right]^2
\]

(23)

For the lognormal data, the Hermite coefficients of higher order tend quickly to zero (see Table 1) therefore, the maximum order of approximation, \( K \), was taken as 9. Another criterion used was the consumption of the CPU time. Table 2 gives CPU time for various sampling size on VAX 11/780 computer.

Table 1. Example of Hermite coefficient values for two different sampling densities \((R=0.0, B=5.0)\)

<table>
<thead>
<tr>
<th>( k )</th>
<th>( N = 50 )</th>
<th>( N = 200 )</th>
<th>( N = 800 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3.3585</td>
<td>3.5218</td>
<td>3.8330</td>
</tr>
<tr>
<td>1</td>
<td>3.0810</td>
<td>3.5417</td>
<td>4.0483</td>
</tr>
<tr>
<td>2</td>
<td>1.1537</td>
<td>1.5775</td>
<td>1.9023</td>
</tr>
<tr>
<td>3</td>
<td>0.1429</td>
<td>0.3187</td>
<td>0.4121</td>
</tr>
<tr>
<td>4</td>
<td>-0.0336</td>
<td>-0.0203</td>
<td>0.0361</td>
</tr>
<tr>
<td>5</td>
<td>-0.0222</td>
<td>-0.0306</td>
<td>-0.0561</td>
</tr>
<tr>
<td>6</td>
<td>-0.0065</td>
<td>-0.0070</td>
<td>-0.0145</td>
</tr>
<tr>
<td>7</td>
<td>-0.0001</td>
<td>0.0004</td>
<td>0.0014</td>
</tr>
<tr>
<td>8</td>
<td>0.0007</td>
<td>0.0005</td>
<td>0.0017</td>
</tr>
<tr>
<td>9</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
Table 2. CPU time (in seconds) on VAX 11/780 computer

<table>
<thead>
<tr>
<th>N</th>
<th>A1</th>
<th>A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.08</td>
<td>0.20</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
<td>0.35</td>
</tr>
<tr>
<td>200</td>
<td>0.18</td>
<td>0.63</td>
</tr>
<tr>
<td>400</td>
<td>0.36</td>
<td>1.24</td>
</tr>
<tr>
<td>800</td>
<td>0.79</td>
<td>2.44</td>
</tr>
</tbody>
</table>

RESULTS

The results of the experiment proved a very similar performance of the two methods. However, it identified one very significant deficiency of the Yates algorithm. In the estimation framework of disjunctive kriging the Hermite approximation of the anamorphosis functions is used to solve the problem

\[ y = \phi(z) \]

given \( z \). It turns out that for certain sampling conditions, especially for large sample cases, the fit of the Hermite approximation is poor in the upper tail of the distribution (see, for example, Figure 2). In such a case the solution of (23) would be burdened with high error (on Figure 2 the solution is outside of the shown region, probably around 5.0, while the solution of Puente and Bras algorithm is around 3.0). This behavior of the Yates algorithm is attributed to the initial step

\[ R = 0.0 \quad B = 10.0 \quad N = 800 \]

Figure 2. Hermite approximation of anamorphosis function.
in the algorithm i.e. fitting a polynomial regression to the sample to obtain points necessary for the numerical integration scheme. The fit is dominated by a large "mass" of the point in the middle region of the distribution. The problem could be alleviated by careful selection-of-the-polynomial-degree or a weighted-least square-fit, such procedure, however, would further complicate the algorithm.

\[ R = 10.0 \quad B = 10.0 \quad N = 50 \]

Figure 3. Hermite approximation of anomorphosis function.

Another problem, common for both algorithms is that under certain conditions there is no unique solution to (23). An example of such situation is shown on Figure 3. As an "ad hoc" procedure, we propose to search for the roots of (23) in the interval

\[ y_1 - 2u \leq y_1 \leq y_1 + 2u \]

where \( y_1 \) corresponds to \( z_1 \) for which the solution is needed and \( u \) is the max \( \{ y_1 - y_{i-1} : i = 2,3,\ldots \} \). If two or more solutions are present in the specified range then the minimum of them should be accepted. The investigated sampling conditions, such as correlation distance and measurement error, did not
seem to affect the performance of the algorithms. The full results of the study are presented in the Appendix, the first row of each segment (see Appendix) contains the mean values of 10 realizations for given statistics and second one contains the corresponding variances. The statistics are (from left to right):

1. the mean of the sample
2. the variance of the sample
3. the mean from the algorithm #1 (Puente and Bras)
4. the mean from the algorithm #2 (Yates et al)
5. the variance from the algorithm #1
6. the variance from the algorithm #2
7. MSE from algorithm #1 (Eq.21)
8. MSE from algorithm #2
9. ZSE from algorithm #1 (Eq.22)
10. ZSE from algorithm #2

As the conclusion of this short study we recommend that Puente and Bras algorithm is used. It is much more efficient in terms of CPU time (see Table 2) and avoids certain instability of the Yates algorithm. However, it should be pointed out that under most sampling conditions the performance of both algorithms is very similar.

ACKNOWLEDGEMENTS:

This research was supported by the National Science Foundation under grant ECE-8419189 and by the Hydrologic Research Laboratory of the National Weather Service, NOAA. This support is greatly acknowledged.

REFERENCES:


### APPENDIX

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 5.0</th>
<th>( N ) = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.100</td>
<td>7.696</td>
<td>2.090</td>
</tr>
<tr>
<td>0.353</td>
<td>17.818</td>
<td>0.349</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 5.0</th>
<th>( N ) = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.936</td>
<td>6.386</td>
<td>1.932</td>
</tr>
<tr>
<td>0.464</td>
<td>16.460</td>
<td>0.464</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 5.0</th>
<th>( N ) = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.887</td>
<td>6.116</td>
<td>1.885</td>
</tr>
<tr>
<td>0.469</td>
<td>25.054</td>
<td>0.468</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 5.0</th>
<th>( N ) = 400</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.907</td>
<td>6.582</td>
<td>1.906</td>
</tr>
<tr>
<td>0.542</td>
<td>48.208</td>
<td>0.542</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 5.0</th>
<th>( N ) = 800</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.949</td>
<td>7.605</td>
<td>1.949</td>
</tr>
<tr>
<td>0.435</td>
<td>41.161</td>
<td>0.435</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 10.0</th>
<th>( N ) = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.656</td>
<td>5.266</td>
<td>1.652</td>
</tr>
<tr>
<td>0.354</td>
<td>23.537</td>
<td>0.353</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 10.0</th>
<th>( N ) = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.596</td>
<td>4.967</td>
<td>1.594</td>
</tr>
<tr>
<td>0.370</td>
<td>29.147</td>
<td>0.368</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 10.0</th>
<th>( N ) = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.609</td>
<td>4.916</td>
<td>1.608</td>
</tr>
<tr>
<td>0.354</td>
<td>23.537</td>
<td>0.353</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 10.0</th>
<th>( N ) = 400</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.626</td>
<td>4.805</td>
<td>1.626</td>
</tr>
<tr>
<td>0.370</td>
<td>19.759</td>
<td>0.369</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 10.0</th>
<th>( N ) = 800</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.597</td>
<td>4.809</td>
<td>1.597</td>
</tr>
<tr>
<td>0.361</td>
<td>18.945</td>
<td>0.361</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 20.0</th>
<th>( N ) = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.035</td>
<td>3.381</td>
<td>2.031</td>
</tr>
<tr>
<td>0.493</td>
<td>7.639</td>
<td>0.493</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 20.0</th>
<th>( N ) = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.055</td>
<td>3.415</td>
<td>2.053</td>
</tr>
<tr>
<td>0.560</td>
<td>12.003</td>
<td>0.560</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 20.0</th>
<th>( N ) = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.030</td>
<td>3.432</td>
<td>2.029</td>
</tr>
<tr>
<td>0.603</td>
<td>13.974</td>
<td>0.604</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 20.0</th>
<th>( N ) = 400</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.043</td>
<td>3.767</td>
<td>2.042</td>
</tr>
<tr>
<td>0.661</td>
<td>16.941</td>
<td>0.661</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R ) = 0.0</th>
<th>( B ) = 20.0</th>
<th>( N ) = 800</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.029</td>
<td>3.733</td>
<td>2.028</td>
</tr>
<tr>
<td>0.678</td>
<td>17.465</td>
<td>0.677</td>
</tr>
<tr>
<td>R = 0.0</td>
<td>B = 40.0</td>
<td>N = 50</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>1.128</td>
<td>1.074</td>
<td>1.126</td>
</tr>
<tr>
<td>1.126</td>
<td>1.074</td>
<td>1.126</td>
</tr>
<tr>
<td>1.050</td>
<td>2.272</td>
<td>0.507</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 0.0</th>
<th>B = 40.0</th>
<th>N = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.096</td>
<td>0.844</td>
<td>1.094</td>
</tr>
<tr>
<td>0.439</td>
<td>1.426</td>
<td>0.435</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 0.0</th>
<th>B = 40.0</th>
<th>N = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.075</td>
<td>0.752</td>
<td>1.073</td>
</tr>
<tr>
<td>0.397</td>
<td>0.969</td>
<td>0.395</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 0.0</th>
<th>B = 40.0</th>
<th>N = 400</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.081</td>
<td>0.832</td>
<td>1.080</td>
</tr>
<tr>
<td>0.434</td>
<td>1.373</td>
<td>0.433</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 10.0</th>
<th>B = 5.0</th>
<th>N = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.880</td>
<td>6.185</td>
<td>1.877</td>
</tr>
<tr>
<td>0.482</td>
<td>27.884</td>
<td>0.481</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 10.0</th>
<th>B = 5.0</th>
<th>N = 400</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.906</td>
<td>6.785</td>
<td>1.905</td>
</tr>
<tr>
<td>0.557</td>
<td>54.046</td>
<td>0.558</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 10.0</th>
<th>B = 5.0</th>
<th>N = 600</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.948</td>
<td>7.687</td>
<td>1.947</td>
</tr>
<tr>
<td>0.442</td>
<td>45.327</td>
<td>0.442</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 10.0</th>
<th>B = 10.0</th>
<th>N = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.662</td>
<td>5.487</td>
<td>1.657</td>
</tr>
<tr>
<td>0.441</td>
<td>40.730</td>
<td>0.435</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 10.0</th>
<th>B = 10.0</th>
<th>N = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.605</td>
<td>5.093</td>
<td>1.602</td>
</tr>
<tr>
<td>0.354</td>
<td>24.051</td>
<td>0.353</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 10.0</th>
<th>B = 10.0</th>
<th>N = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.612</td>
<td>4.977</td>
<td>1.611</td>
</tr>
<tr>
<td>0.358</td>
<td>18.245</td>
<td>0.357</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 10.0</th>
<th>B = 10.0</th>
<th>N = 400</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.625</td>
<td>4.775</td>
<td>1.624</td>
</tr>
<tr>
<td>0.367</td>
<td>19.102</td>
<td>0.367</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R = 10.0</th>
<th>B = 10.0</th>
<th>N = 800</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.594</td>
<td>4.764</td>
<td>1.593</td>
</tr>
<tr>
<td>0.358</td>
<td>18.245</td>
<td>0.357</td>
</tr>
<tr>
<td>R</td>
<td>10.0</td>
<td>20.0</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>2.040</td>
<td>3.443</td>
<td>2.036</td>
</tr>
<tr>
<td>0.504</td>
<td>8.338</td>
<td>0.502</td>
</tr>
<tr>
<td>2.051</td>
<td>3.413</td>
<td>2.049</td>
</tr>
<tr>
<td>0.557</td>
<td>12.596</td>
<td>0.556</td>
</tr>
<tr>
<td>2.030</td>
<td>3.482</td>
<td>2.030</td>
</tr>
<tr>
<td>0.603</td>
<td>13.585</td>
<td>0.603</td>
</tr>
<tr>
<td>2.044</td>
<td>3.871</td>
<td>2.044</td>
</tr>
<tr>
<td>0.664</td>
<td>17.155</td>
<td>0.664</td>
</tr>
<tr>
<td>2.031</td>
<td>3.847</td>
<td>2.030</td>
</tr>
<tr>
<td>0.678</td>
<td>18.002</td>
<td>0.678</td>
</tr>
<tr>
<td>1.124</td>
<td>1.044</td>
<td>1.123</td>
</tr>
<tr>
<td>0.513</td>
<td>2.714</td>
<td>0.511</td>
</tr>
<tr>
<td>1.120</td>
<td>0.974</td>
<td>1.120</td>
</tr>
<tr>
<td>0.502</td>
<td>2.169</td>
<td>0.501</td>
</tr>
<tr>
<td>1.093</td>
<td>0.840</td>
<td>1.092</td>
</tr>
<tr>
<td>0.435</td>
<td>1.377</td>
<td>0.434</td>
</tr>
<tr>
<td>1.073</td>
<td>0.754</td>
<td>1.073</td>
</tr>
<tr>
<td>0.393</td>
<td>0.950</td>
<td>0.393</td>
</tr>
<tr>
<td>1.080</td>
<td>0.844</td>
<td>1.080</td>
</tr>
<tr>
<td>0.428</td>
<td>1.389</td>
<td>0.428</td>
</tr>
<tr>
<td>2.026</td>
<td>10.365</td>
<td>2.013</td>
</tr>
<tr>
<td>0.453118725</td>
<td>0.446</td>
<td>0.458</td>
</tr>
<tr>
<td>1.855</td>
<td>8.172</td>
<td>1.849</td>
</tr>
<tr>
<td>0.596</td>
<td>60.138</td>
<td>0.592</td>
</tr>
<tr>
<td>1.852</td>
<td>8.590</td>
<td>1.848</td>
</tr>
<tr>
<td>0.537</td>
<td>66.497</td>
<td>0.537</td>
</tr>
<tr>
<td>1.904</td>
<td>9.751</td>
<td>1.902</td>
</tr>
<tr>
<td>0.624120761</td>
<td>0.624</td>
<td>0.625117.745117.196</td>
</tr>
<tr>
<td>1.941</td>
<td>10.313</td>
<td>1.940</td>
</tr>
<tr>
<td>0.472</td>
<td>84.690</td>
<td>0.472</td>
</tr>
<tr>
<td>R</td>
<td>B</td>
<td>N</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>-50.0</td>
<td>-10.0</td>
<td>50</td>
</tr>
<tr>
<td>0.511</td>
<td>75.414</td>
<td>0.503</td>
</tr>
<tr>
<td>-50.0</td>
<td>-10.0</td>
<td>100</td>
</tr>
<tr>
<td>0.419</td>
<td>51.963</td>
<td>0.416</td>
</tr>
<tr>
<td>-50.0</td>
<td>-10.0</td>
<td>400</td>
</tr>
<tr>
<td>0.362</td>
<td>34.737</td>
<td>0.362</td>
</tr>
<tr>
<td>-50.0</td>
<td>-10.0</td>
<td>800</td>
</tr>
<tr>
<td>0.359</td>
<td>29.310</td>
<td>0.359</td>
</tr>
<tr>
<td>-50.0</td>
<td>-20.0</td>
<td>50</td>
</tr>
<tr>
<td>0.576</td>
<td>22.073</td>
<td>0.572</td>
</tr>
<tr>
<td>-50.0</td>
<td>-20.0</td>
<td>100</td>
</tr>
<tr>
<td>0.554</td>
<td>30.146</td>
<td>0.551</td>
</tr>
<tr>
<td>-50.0</td>
<td>-20.0</td>
<td>200</td>
</tr>
<tr>
<td>0.623</td>
<td>31.289</td>
<td>0.622</td>
</tr>
<tr>
<td>-50.0</td>
<td>-20.0</td>
<td>400</td>
</tr>
<tr>
<td>0.576</td>
<td>22.073</td>
<td>0.572</td>
</tr>
<tr>
<td>-50.0</td>
<td>-20.0</td>
<td>800</td>
</tr>
<tr>
<td>0.576</td>
<td>22.073</td>
<td>0.572</td>
</tr>
<tr>
<td>-50.0</td>
<td>-40.0</td>
<td>50</td>
</tr>
<tr>
<td>0.683</td>
<td>38.422</td>
<td>0.682</td>
</tr>
<tr>
<td>-50.0</td>
<td>-40.0</td>
<td>100</td>
</tr>
<tr>
<td>0.683</td>
<td>38.422</td>
<td>0.682</td>
</tr>
<tr>
<td>-50.0</td>
<td>-40.0</td>
<td>200</td>
</tr>
<tr>
<td>0.683</td>
<td>38.422</td>
<td>0.682</td>
</tr>
<tr>
<td>-50.0</td>
<td>-40.0</td>
<td>400</td>
</tr>
<tr>
<td>0.683</td>
<td>38.422</td>
<td>0.682</td>
</tr>
<tr>
<td>-50.0</td>
<td>-40.0</td>
<td>800</td>
</tr>
<tr>
<td>0.683</td>
<td>38.422</td>
<td>0.682</td>
</tr>
</tbody>
</table>
APPENDIX E

ESTIMATION OF MEAN PRECIPITATION FIELDS USING OPERATIONALLY AVAILABLE HYDROMETEOROLOGICAL DATA AND A TWO-DIMENSIONAL PRECIPITATION MODEL
ESTIMATION OF MEAN PRECIPITATION FIELDS USING
OPERATIONALLY AVAILABLE HYDROMETEOROLOGICAL DATA
AND A TWO-DIMENSIONAL PRECIPITATION MODEL

by

K.P. Georgakakos and T.H. Lee

Sponsored by

Utah State University, Logan Utah
Subcontract No. 86-079 to NSF Contract No. ECE-8419189
and
United States Geological Survey, Reston, Virginia
Grant Award No. 14-08-0001-G1297

IIHR Report No. 315
Iowa Institute of Hydraulic Research
The University of Iowa
Iowa City, Iowa 52242-1585
July 1987
Executive Summary

Previous work in modeling mesoscale rainfall fields is reviewed. A two-dimensional precipitation estimation model is formulated based on the principles of conservation of mass of liquid water equivalent, and of heat conservation. Advection of liquid water equivalent is accomplished by the middle tropospheric wind velocity (assumed to be the storm velocity). Spatial interpolation of the spatially- and temporally-sparse wind observations was done based on objective interpolation techniques. The model formulation explicitly accounts for: condensation of vapor, advection and sub-cloud evaporation of liquid water equivalent. The two free model parameters that determine updraft-velocity strength and particle-size distribution were estimated based on contours of various performance criteria in the parameter space. The contours were generated from real-time available meteorological and rainfall hourly observations of convective storms in Oklahoma. The issue of grid size determination is discussed from a practical, CPU-time viewpoint and with consideration of the precipitation estimation accuracy. The model is suitable for use in detrending precipitation fields observed by various types of sensors for the purpose of merging observed-precipitation-fields. Also, because of its state-space mathematical form, it is suitable for use in real-time precipitation forecasting.
ACKNOWLEDGEMENTS

This work was supported by the Utah State University Subcontract No. 86-079 (to NSF Contract No. ECE-8419189) and by the United States Geological Survey Grant No. 14-08-0001-G1297. Substantial computer resources support was provided by the WEEG Computer Center of the University of Iowa.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
</tbody>
</table>

1. INTRODUCTION ................................................................. 1
   1.1. Statement of Problem .............................................. 1
   1.2. Objective and Scope ............................................... 2

2. LITERATURE REVIEW ............................................................. 3
   2.1. Introduction .......................................................... 3
   2.2. Statistical Precipitation Models ............................... 4
   2.3. Dynamical Precipitation Models ............................... 7
   2.4. Statistical-Dynamical and Stochastic-Dynamic Precipitation Prediction Models ............................... 10

3. A SPATIALLY-LUMPED PRECIPITATION MODEL ............................ 14
   3.1. Formulation Using Surface Meteorological Data ................. 14
   3.2. Parameter Estimation Studies .................................. 20
   3.3. Input Spatial Interpolation ..................................... 23

4. A TWO-DIMENSIONAL PRECIPITATION MODEL ................................ 28
   4.1. Introduction ......................................................... 28
   4.2. Formulation .......................................................... 28
   4.3. Spatial Interpolation of Horizontal Wind ..................... 38
   4.4. Initial and Boundary Conditions ................................ 41

5. PARAMETER ESTIMATION OF A TWO-DIMENSIONAL PRECIPITATION MODEL .... 44
   5.1. Introduction ......................................................... 44
   5.2. Parameter Estimation ............................................... 44
       5.2.1. Measures of Performance .................................... 44
       5.2.2. Data and Basin Characteristics ............................ 47
       5.2.3. Results ....................................................... 50
   5.3. Sensitivity Analysis .............................................. 64

6. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS ............................ 66

7. REFERENCES ................................................................. 72
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Interpolation Error and Observation Statistics</td>
<td>27</td>
</tr>
<tr>
<td>5.1 Storm Dates and Durations</td>
<td>48</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.</td>
<td>Schematic representation of a one-dimensional dynamic precipitation model.</td>
<td>15</td>
</tr>
</tbody>
</table>
| 3.2.         | (a) Region in parameter space of optimal performance for the stratiform group.  
(b) Region in parameter space of optimal performance for the convective group. | 22   |
| 3.3.         | Forecasts (dashed line) vs. observations (solid line) for the verification storm group (Tulsa, OK). One-hour time steps. | 24   |
| 3.4.         | Time trace of the estimated hourly averaged mass liquid water content for the convective storm group used in calibration. | 25   |
| 4.1.         | Elementary volume of moist air with influxes $Q_{x1}$, $Q_{y1}$, $Q_{z1}$ and outfluxes $Q_{x2}$, $Q_{y2}$, $Q_{z2}$ of condensed liquid water equivalent in the positive directions of the axis system $(x,y,z)$. | 30   |
| 4.2.         | Integration volume and associated in- and out-flows of liquid water equivalent. The positive direction of the axes is shown. | 33   |
| 4.3.         | Quantities for the computation of a boundary inflow along the positive $x$ direction for an Integration Volume of horizontal dimensions $D_x$ and $D_y$. | 35   |
Figure 4.4. Horizontal discretization grid for the two-dimensional precipitation model..........................37

Figure 4.5. Schematic for the spatial interpolation procedure applied to observations of the middle tropospheric horizontal wind speed.................................40

Figure 5.1. Schematic illustration of threat score (TS) and bias........46

Figure 5.2. Model domain spanning 120.5km x 189km. The location of meteorological surface-data and upper-air-data stations and raingage stations is shown........49

Figure 5.3. Contours of the time-averaged value of the areal-mean estimation error in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. Advection is based on the 500mbar wind. Grid size: 37.8x24.1 km$^2$..............................52

Figure 5.4. Contours of the time-standard-deviation of the areal-mean estimation error in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. Advection is based on the 500mbar wind. Grid size: 37.8x24.1 km$^2$..............................53

Figure 5.5. Contours of the time-averaged value of the Threat Score in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. Advection is based on the 500mbar wind. Grid size: 37.8x24.2 km$^2$............54

Figure 5.6. Contours of the time-averaged value of the Bias Score in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. Advection is based on the 500mbar wind. Grid size: 37.8x24.1 km$^2$............55
Figure 5.7. Contours of the time-averaged value of the areal-mean estimation error in the space of parameters $\epsilon_1$ and $\epsilon_4$.* Advection is based on the 700mbar wind. Solid line corresponds to the numerical scheme of Equation (4.16). Dashed line corresponds to the upstream numerical scheme of Equation (5.6). Grid size: 37.8x24.1 km$^2$.

Figure 5.8. Contours of the time-standard-deviation of the areal-mean estimation error in the space of parameters $\epsilon_1$ and $\epsilon_4$. Advection is based on the 700mbar wind. Solid line corresponds to the numerical scheme of Equation (4.16). Dashed line corresponds to the upstream numerical scheme of Equation (5.6). Grid size: 37.8x24.1 km$^2$.

Figure 5.9. Contours of the time-averaged value of the Threat Score in the space of parameters $\epsilon_1$ and $\epsilon_4$. Advection is based on the 700mbar wind. Solid line corresponds to the numerical scheme of Equation (4.16). Dashed line corresponds to the upstream numerical scheme of Equation (5.6). Grid size: 37.8x24.1 km$^2$.

Figure 5.10. Contours of the time-averaged value of the Bias Score in the space of parameters $\epsilon_1$ and $\epsilon_4$. Advection is based on the 700mbar wind. Solid line corresponds to the numerical scheme of Equation (4.16). Dashed line corresponds to the upstream numerical scheme of Equation (5.6). Grid size: 37.8x24.1 km$^2$.

Figure 5.11. Contours of the time-averaged value of the areal-mean estimation error in the space of parameters $\epsilon_1$ and $\epsilon_4$. No advection is accounted for. Solid lines stand for a grid size of 37.8x24.1 km$^2$. Dashed lines stand for a grid size of 18.9x12.05 km$^2$. 

vii
Figure 5.12. Contours of the time-standard-deviation of the areal-mean estimation error in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. No advection is accounted for. Solid lines stand for a grid size of 37.8x24.1 km$^2$. Dashed lines stand for a grid size of 18.9x12.05 km$^2$.

Figure 5.13. Contours of the time-averaged value of the Threat Score in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. No advection is accounted for. Solid lines stand for a grid size of 37.8x24.1 km$^2$. Dashed lines stand for a grid size of 18.9x12.05 km$^2$.

Figure 5.14. Contours of the time-averaged value of the Bias Score in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. No advection is accounted for. Solid lines stand for a grid size of 37.8x24.1 km$^2$. Dashed lines stand for a grid size of 18.9x12.05 km$^2$.

Figure 5.15. Time-averaged value of areal-mean, estimation error as a function of grid-square area. No advection was included. $\varepsilon_1 = 6 \times 10^{-3}$, $\varepsilon_4 = 6 \times 10^{-5}$ m.

Figure 5.16. Time-standard-deviation of areal-mean estimation error as a function of grid-square area. No advection was included. $\varepsilon_1 = 6 \times 10^{-3}$, $\varepsilon_4 = 6 \times 10^{-5}$ m.

Figure 5.17. Threat Score as a function of grid-square area. No advection was included. $\varepsilon_1 = 6 \times 10^{-3}$, $\varepsilon_4 = 6 \times 10^{-5}$ m.

Figure 5.18. Bias Score as a function of grid-square area. No advection was included. $\varepsilon_1 = 6 \times 10^{-3}$, $\varepsilon_4 = 6 \times 10^{-5}$ m.

Figure 5.19. Standard deviation of point residual errors as a function of grid-square area. No advection was included. $\varepsilon_1 = 6 \times 10^{-3}$, $\varepsilon_4 = 6 \times 10^{-5}$ m.
1. INTRODUCTION

1.1 Statement of Problem

Accurate measurement of rainfall is important to disciplines, such as hydrology, climatology and weather forecasting. For the estimation and prediction of streamflows in large river basins, areal-mean values of rainfall may be adequate. However, for small basins and for convective storms, where both the temporal and spatial distribution of rainfall is important (Hamlin, 1983), measurement of rainfall with a high degree of spatial and temporal detail is desirable. Measurement of rainfall in space using raingages can be impractical due to installation and maintenance costs of a large number of raingages. Thus, there has been considerable interest recently in utilizing weather radar in the determination of the spatial distribution of rainfall.

Radar overcomes the primary difficulties associated with rain gages, i.e., the radar measurements are over an area, available in real time and measured from a single location. The main disadvantage of the radar sensor is the lack of a unique relationship between the energy return to the radar and the precipitation reaching the ground. Various scientists have found that the relatively large radar-raingage measurement differences average 30% at best and reach 200% at times (Wilson, 1968).

Past work has shown that the combination of data from raingages and radar should provide improved rainfall measurements. The simplest and most commonly used method consists of merging the two data sets so that the radar is used for areal rainfall measurement and the raingage are used to remove the mean storm bias of the radar measurements (Wilson and Brandes, 1979). It is expected that the available data merging methods can be improved if the cokriging method is used to merge the detrended raingage and radar data. This is mainly because the kriging method makes optimal use of available information (Journel and Huijbregts, 1978). Use of this modern estimation technique requires reliable estimates of the mean precipitation field for detrending the observed fields.
1.2 Objective and Scope

The objective of this study was to develop a physically-based, two-dimensional precipitation model. The model should utilize only operationally available meteorological data from stations within the area of interest. The model output should be the accumulated (over prespecified time intervals and spatial domains) precipitation at various regions within the area of interest. The model could be used for the purposes of: 1) determination of the mean of rainfall field under a radar umbrella for merging rainfall data from various sensors, and 2) real-time precipitation prediction.

The second section reviews past efforts in the area of precipitation modeling. Section 3 gives a brief review of the one-dimensional station precipitation model developed by Georgakakos and Bras (1984). Section 4 extends the one-dimensional model to two dimensions in order to capture the spatial distribution of precipitation. Section 5 presents the application of the two-dimensional model to the Oklahoma–Tulsa area for model parameter estimation and for the examination of various performance statistics for mean rainfall field estimation. Conclusions are presented in Section 6.
2. LITERATURE REVIEW

2.1. Introduction

Historically, two main approaches have been followed to model precipitation. One utilizes statistical techniques which are used to characterize the time and space rainfall pattern reaching the ground and is called, hereafter, the statistical method. The other approach models the cloud rainfall dynamics and physics aloft and, thus, it is called the dynamical method. The development of statistical methods stemmed from recognition of the difficulties in physical modeling owing to the diversity of climatic processes in different geographic regions and to the lack of detailed understanding of the rainfall producing mechanism. As these models are developed on the basis of regional historical data, the statistical models are applicable to a storm by storm analysis. The difficulty with the method that uses dynamical models is that it utilizes various weather parameters as input, and the available observations of those parameters have a poor resolution in space and time. Also, because physical/dynamical processes interact at different scales, it seems impossible, with today's knowledge and computer technology, to formulate a single model that includes all the processes at all scales. In recent years, a third approach incorporating the dynamical models with statistical inference techniques showed promising results. The principal reasons for the existence of this stochastic-dynamic approach (that brings together statistics, probability theory and dynamical models) to precipitation prediction are:

1. Inclusion of precipitation physics and dynamics is important for accurate real-time rainfall forecasting.

2. Knowledge of the statistical spatial and temporal structure of precipitation can be useful for predicting precipitation behavior in the future.

3. Initial conditions and parameter input to physical models are characterized by inherent uncertainties. Also, simplification of mathematical models is usually required in order to arrive at a manageable computational algorithm. A satisfactory prediction
model should be able to incorporate the uncertainties and predict not only the expected precipitation but also the error expected in the precipitation forecasts.

(4) In the absence of perfect forecasts, users usually require estimates of the likelihood of occurrence of the relevant events in order to make optimal decisions in uncertain situations.

A more detailed discussion of the need and benefit of this method is presented by the Committee on Precipitation, AGU Hydrology Section in EOS Transactions, 1984.

2.2. Statistical Precipitation Models

For the pure statistical precipitation prediction models (or 'classical approach' as by Klein, 1969), future precipitation is estimated based on past observable data. This situation can be expressed as:

$$\hat{Y}_t = f(a, X_o)$$

(2.1)

Where $\hat{Y}_t$ is the predictand at some time 't' in the future, $X_o$ is the predictor vector which consists of current and past observations, and $a$ is the vector of model parameters. The functional form of $f$ and the parameters of the model are to be determined from past data. Based on the differences in the definition of $X_o$, statistical precipitation prediction models can be separated into various types. The first type of models is classified under the broad title 'time series models'. Past precipitation data are used as predictors. Some most commonly used time series forecasting models are: Box-Jenkins models, Bayesian forecasting models, and exponential smoothing models. Miller and Leslie (1984) applied a second-order Markov chain model with a time-of-day dependent transition, using 3-hourly observations, to forecast probabilities of rain up to 12 hours in advance. Johnson and Bras (1980) utilized the Bayesian linear filtering method (Kalman filtering) to predict short-term (of the order of 1 hr or less) quantitative rainfall rates. The model was nonstationary and predicted rainfall rates at multiple points and at multiple forecast-lead times, thus, providing some degree of spatial and temporal detail.
In recent years significant effort has been spent on the characterization of the statistical pattern of rainfall in both time and space, and it has resulted in the class of 'stochastic precipitation models'. Excellent reviews of these models are reported by Waymire and Gupta (1981); Stern and Coe (1984); Guttorp (1986); and Georgakakos and Kavvas (1987). Not all the stochastic precipitation models can be applied for rainfall prediction. As it is stated by the AGU Committee on Precipitation (1984): 'The only classes of space-time models for which satisfactory inference procedures are available are those that can be specified by (simple) spatial and temporal autocovariance functions.'

Other statistical precipitation prediction models utilize precipitation-relevant meteorological parameters, e.g., dew point temperature, ceiling height, as predictors for rainfall occurrence or amount. These models are known to the statistician as 'causal models'. They differ from dynamical methods for the reason that they do not involve precipitation-producing physics; that is, they are empirical. Petterssen (1956) introduced a statistical technique for quantitative precipitation prediction. The technique consists of relating the occurrence and amount of rainfall to a number of meteorological variables which are independent (or nearly so) of one another, and which are considered to provide prognostic information concerning precipitation. The chosen variables are then grouped in pairs and used as coordinates of a scatter diagram on which the rainfall is plotted. The analysis is carried out by constructing contours of rainfall amount on each chart. Recently, Miller (1981) used this technique for the prediction of a large number of meteorological variables one hour in advance.

Another group of statistical precipitation models involves estimation of movement parameters and extrapolation of the current rain-cloud pattern observed by remote sensors for a few hours into the future. This simple forecasting method is useful in rainfall forecasting at the 0-6 hr. scale with a low computation cost. Many applications of this approach have appeared in the literature in recent years. The radar rainfall forecast procedure of Bellon and Austin (1984) 'consists of making digital Cartesian maps at a height of 3 km and computing the velocity of the precipitation pattern on two maps by locating the direction of maximum cross-correlation coefficient. The forecast is then obtained by translating the entire pattern in the computed direction.'
Their experimental results for the City of Montreal showed a mean absolute deviation that varies from 50 to 60% for 0.5 to 3 hr forecasts. A similar mapping technique with satellite-radar combined data was used by Bellon, Lovejoy and Austin (1980) for 0-6 hr rainfall forecasts. Linear extrapolation of radar data for 0-6 hr into the future was used by Browning et al (1982) for the forecasting of frontal rain.

Scofield and Spayd (1984) produced real time, short-term (up to 3 hours) predictions of precipitation from extratropical cyclones using satellite, radar and conventional data. Their method was based on the analysis of satellite visible and InfraRed (IR) imagery of storm clouds, on previous standardized patterns of known meteorological characteristics and precipitation. The authors reported verification of the technique, used operationally, for the period from September 1982 to April 1983. Six-hourly accumulations of three-hourly precipitation forecasts were verified. The authors found that about 45 percent of the rainfall forecasts were within 35 percent of the reported values. Several other techniques and applications along this line of research can be found in Nowcasting edited by Browning (1982).

The fundamental procedure in simple statistical extrapolation precipitation methods is the linear extrapolation of rainfall patterns. Linear proportionality of rainfall intensity and area distribution with time is assumed. Intuitively, one may suspect that the model can be made more sophisticated by including the storm development statistics. However, as shown in the work by Tsonis and Austin (1981), marked improvement is not expected. The major cause of poor forecasts is not due to errors from the linear displacement simplification but due to the inaccuracies in remotely-sensed data caused by unpredictable rainfall/moisture configurations (Bellon and Austin, 1978).

In addition to the deficiency of not incorporating the precipitation producing physics, there are three major problems associated with statistical models. First, the precipitation time-correlation coefficient dies out very quickly as the prediction lead-time increases, which makes prediction with a lead-time more than one hour very difficult (Johnson and Bras, 1980). Second, these models do not differentiate between precipitation types (e.g. convective vs. stratiform), which provide significant information about the rainfall duration and intensity. Third, they do not include the upwind cloud-rainfall
information. However, statistical models are not totally unattractive. They appeal to scientists for their computational efficiency, which is very important for nowcasting (0-6hr). And the time structure of rainfall, which contains statistical information of past behavior of rainfall, may be useful in developing statistical-dynamical prediction models.

2.3. Dynamical Precipitation Prediction Models

This group of models recognizes precipitation forecasting as an initial value problem whereby the governing equations are integrated forward from fully determined initial values of the meteorological fields at some initial time. This group of methods is usually referred to as 'conventional dynamical methods' and is particularly attractive for short- to medium-range prediction.

Dynamical precipitation models can be classified in terms of the scale length (time and space) or the dynamics and cloud physics that produce rainfall. As stated previously, this work is concentrated on mesoscale, short-term precipitation estimation and forecasting, and thus, only representative precipitation forecasting models that fit into these time and length scales are reviewed. A more comprehensive review can be found in Frank (1983). Also Georgakakos and Hudlow (1984) review the current hydrometeorological forecasting programs of the U.S. National Weather Service, with classifications according to time and space scale. In terms of dynamics, there are two basic types of precipitation: stratiform and convective (Houze, 1981). Stratiform and convective storms are differentiated based on whether the atmosphere is convectively stable or unstable. In the case that the atmosphere is convectively stable throughout a grid volume, only the stratiform clouds will develop. Phase changes of water can be represented using cloud physics formulations. The precipitation particles initiate at or near the top of the cloud system, and tend to form aggregates as they move downward, particularly as they approach to within 1 km of the melting level. Stratiform-type precipitation is easy to predict because it is long lived and the melting layer is easily observed by weather radar. A fairly reliable criterion for determining whether or not precipitation will occur in stratiform cloud systems is the degree of saturation (refer to the review by Anthes, 1983). Examples of applications are the Limited area Fine Mesh (LFM) model (Gerrity, 1977; Newell and Deaven, 1981), ANMRC (Australian Numerical
Meteorology Research Center, McGregor et al., 1978), Sundqvist (1978), and Colton (1976).

Convective precipitation prediction modeling involves one of the following parameterization methodologies and is classified accordingly (Pielke, 1981):

(1) convective adjustment,
(2) use of one-dimensional cloud models,
(3) use of a cumulus field model (3D) or set of equivalent observations, and
(4) explicit representation of moist thermodynamics.

The first two classes arise for the reason that the larger scale models cannot resolve the transport of convective heat, moisture and momentum at the cumulus cloud scale. Parameterizations are developed to relate these processes to the variables predicted. The convective adjustment method 'assumes that there exists a critical state of the large-scale thermodynamical field. When the field tends to become unstable, it is adjusted, under some constraints, to a new stable or mild state' (Kurihara, 1973). The advantages to this type of parameterization are its simplicity and reliability for representing the gross stabilizing effects of convection on the environment. However, this bulk-adjustment scheme contributes little to the understanding of how convection and large-scale processes interact (Anthes, 1977). Moreover, using this method the regions of potential instability are removed too rapidly (Pielke, 1981). Examples of applications of this method for convective precipitation prediction modeling can be found in the formulation of the Limited area Fine Mesh model (LFM), models of the Japanese Meteorological Agency and of the British Meteorological Office (Anthes, 1983; Krishnamurti et al., 1980; and Colton 1976).

Kuo's parameterization scheme (Kuo, 1965, 1974; Anthes, 1977), one-dimensional entrainment models (Weinstein and Davis, 1968; Simpson and Wiggert, 1971), Arakawa and Schubert's (1974) parameterization scheme belong to the second class. In using one-dimensional models, vertical thermodynamics and conservation of heat, moisture and momentum are simulated on an area-averaged basis. The cloud properties are computed using a cumulus cloud model. The
one-dimensional models are superior to the convective adjustment models in the sense that they can represent the feedback of cumulus scales to the larger scale. The shortcomings of these models are that they require arbitrary inputs and assumptions, such as cloud radius and a quasi-equilibrium with the large-scale environment (Pielke, 1981). Also, as shown by Cotton and Tripoli (1978), they cannot simultaneously and accurately predict both cloud top height and profile of moisture. Examples of applications of these methods for precipitation forecast modeling are the Movable Fine Mesh (MFM) model, ANMRC, Navy-NORAPS, PSU model (Anthes, 1983; Gruber, 1973; and Krishnamurti et al., 1980).

Three-dimensional (3-D) models are important for: 1) capturing the 3-D characteristics of the wind field since subtle changes in the environmental winds can produce significant differences in relative storm motion (Miller, 1978; Thorpe and Miller, 1978); and 2) lending a degree of credibility to model prediction and to the use of model predicted data in interpreting sparse observations (Cotton and Tripoli, 1978). These techniques utilize 3-D cumulus field model simulations or sets of observations in order to determine the temporal and spatial response of cumulus clouds to a particular set of meso-scale dependent variables, and their interactions. Recent formulations utilizing this kind of technique have been presented by Cotton and Tripoli (1978), Clark (1979), Chen and Orville (1980), and Nickerson et al. (1986), among others. Example of applications of this technique for rainfall prediction are: Clark's (1979) 3-D numerical simulation of a severe storm; Hsie et al.'s (1981) extension model by using the concept of eddy transport and their illustration of the importance of wind shear on precipitation; Ross and Orlanski's (1982) 3-D numerical simulation of frontal rain; Nickerson et al.'s (1986) 2-D simulation of orographically forced clouds, rain and air flow.

Rosenthal (1978) reported that while tropical cyclogenesis can be well represented with the traditional cloud-and-large scale coupled cumulus parameterization scheme, some tropical squall lines (Pielke, 1981) cannot. He claimed that, in this kind of situation, explicit representation of moist thermodynamics is necessary. Also, explicit representation has to be utilized for small grid-size simulations. Anthes (1983) stated that: 'Parameterizing cumulus convective effects as a function of the resolvable scale becomes ques-
tionable as the grid size becomes smaller than 100 km.' Thus, 'For high-resolution models, therefore, it may be preferable to abandon the concept of parameterization in favor of explicit treatments of condensation and evaporation through the introduction of prediction equations for cloud and precipitation water.' Because, 'For finer mesh, the separation between the resolvable and convective scales becomes small and the model may begin to simulate the same cloud it is also trying to parameterize.' Successful applications can be seen in studies by Rosenthal (1978) and Ross and Orlanski (1982).

However, because dynamical processes interact on different scales, it seems that it will be impossible within the foreseeable future to formulate a single model that includes all of the processes operating at these scales. Also, because the operation of dynamical models requires high quality meteorological data inputs and a large amount of computation, it is difficult to use them for real-time, short-term prediction of flash-flood inducing rainfall.

2.4. Statistical-Dynamical and Stochastic-Dynamic Precipitation Prediction Models

Meteorological physics is attractive for its ability to explain the behavior of meteorological processes. Statistical models produce forecasts based on past information of rainfall and on extrapolation techniques. The purpose for developing the class of statistical-dynamical models is to combine physically-based meteorological knowledge and past observations and experience in an efficient way in real-time in order to construct a better precipitation prediction model. Different strategies for combining past information and predictions from large-scale simulation models give rise to different classes of models. The three most commonly applied statistical-dynamical models are reviewed below: perfect prog, model output statistics and physical Bayesian filtering (Kalman filtering) models.

Perfect prog method: Observed historical data are used to derive statistical relationships between a desired weather element and concurrent values of relevant circulation parameters. The relationship can be represented as:

\[ Y_0 = f(a, X_0) \]  
(2.2)
The next step is to predict the vector $\hat{X}_t$ from a dynamical prognostic model given $X_0$, and then input $\hat{X}_t$ to the statistical relation in order to yield weather forecasts $\hat{Y}_t$. That is, in mathematical notation:

$$\hat{Y}_t = f(a, \hat{X}_t)$$

(2.3)

The perfect prog technique was used extensively in the 1960's and early 1970's to produce categorical forecasts. Klein (1971) employed this approach for precipitation forecasts at 108 cities in the United States on an experimental basis. He developed a multiple regression equation as the 'statistical relationship'. The regression equation features three predictors: initial 850-mb height, initial 850-700 mb mean dewpoint spread, and previous 12-hr precipitation at the network of surface stations. The predictors were determined by applying the stepwise screening regression technique. The 850-mb heights are obtained from the National Meteorological Center (NMC) primitive equation model (Shuman and Hovermale, 1968), and 850- and 700-mb dew-point spreads are obtained from the laminated moisture version of the primitive equation model by Stackpole and Bedient (1970).

The perfect prog approach assumes the model output, $\hat{X}_t$, is 'perfect' (hence its name). Thus this technique uses dynamical models but ignores the uncertainty in the prediction generated by these models. As the lead time increases, the uncertainty becomes greater, thereby limiting the range of applicability of this procedure. Since the statistical relationship is based entirely on observations, a large data set is usually needed to ensure a stable relationship (i.e., parameters not sensitive to additional data). The availability of observations also may allow useful stratifications of the data; that is, different relations can be developed for different months of the year, hours of the day, etc.

A similar approach is presented by Gleeson (1970, 1975). The initial state of the $m$ parameters is regarded as the vector of coordinates of the $m$ dimensional phase space. Each parameter is assumed random, and is characterized by a probability density function. Dynamical equations provide standard deterministic predictions, while a general continuity equation transforms initial probability distributions into final distributions which, in turn, yield probability forecasts. This continuity equation is resolvable into
component equations of "probability diffusion" for all coordinates of the phase space. This method is classified as a Perfect Prog method because the technique that determines the time evolution of the parameter probability density functions via diffusion equations is deterministic.

Model output statistics method: The statistical-dynamical technique that has received the greatest attention in the last decade is the model output statistical (MOS) technique (Glahn and Lowry, 1972; Klein and Glahn, 1974; and Lowry and Glahn, 1976). In the MOS approach, the prediction equation based on past data relates $\hat{Y}_t$ to $\hat{X}_t$, with the latter taken from a large-scale dynamical model. MOS allows uncertainty in the prediction generated by a dynamical model, but it requires a substantial number of forecasts from the model to derive the prediction model, therefore, dynamical model outputs and MOS forecasts must be archived. Moreover, since the prediction equations developed in MOS are associated with particular dynamical models, these equations may need to be redeveloped if the models undergo major changes. Fortunately, the evidence available to date indicates that the decrease in skill is minimal when a MOS equation is used in conjunction with a model other than the one from which it was developed (Glahn, 1985).

Accuracy of MOS is discussed in Lowry and Glahn (1976) and Murphy and Winkler (1984). Examples of recent applications of MOS for quantitative precipitation forecasting can be found in Bermowitz (1975), which uses output from the LFM model; Bermowitz and Zurndorfer (1979), which uses output from Primitive Equation (PE) (Shuman and Hovermale, 1968) and Trajectory (TJ) (Reap, 1972) models; Arritt and Frank (1985) and Tapp, Woodcock and Mills (1986), which uses data from the Australian Region Primitive Equation (ARPE) (McGregor et al., 1978). Leith (1978, 1980) proposed an alternative approach to statistical-dynamical prediction based on the use of Monte Carlo methods. This approach, which involves repeated application of the first moment equation (starting from different initial conditions), substantially reduces the computational burden but still provides an estimate of the uncertainty inherent in the forecasts.

Physical Bayesian filtering methods: Recently, Georgakakos and Bras (1984a, 1984b) utilized state-space modeling techniques to formulate a one-dimensional physically based model for precipitation forecasting. Model input
consists of ground level station temperature, pressure, and dew-point temperature observations. Key physical parameters in the formulation are the pressure at the cloud top, the height-averaged updraft velocity, and the inverse of the average diameter of the hydrometeors at cloud base. The model predicts the spatially averaged precipitation at the ground surface. A Kalman filter (Bras and Iturbe, 1985), suitable for use with the precipitation model was formulated. The filter updated the model state in real time from observations of surface precipitation. Tests of the formulated stochastic-dynamic precipitation model with operationally available hourly data revealed consistently good predictions of precipitation at a precipitation observation station when the input variables are from the same station. Efficiencies (proportion of variance in observed station precipitation accounted for by the model) ranged from 0.28 to 0.38.

For several reasons this model is appealing for the formulation of a high quality precipitation forecasting model:

1. It is statistically-physically based, and it has the ability to incorporate sophisticated cloud-rainfall models and observation from remote sensors for improved forecasts.

2. Uncertainty in the input variables, initial conditions, model structure, and output forecasts are explicitly accounted for.

3. Through the use of an 'Extended Kalman Filter' the model compares the model forecasts with observations and make necessary corrections in the model states (by which error accumulation is avoided.)

4. It produces computationally efficient quantitative precipitation forecasts, and it is believed that through careful formulation it should be suitable for operating on a mini or microcomputer, which is important to local precipitation and flash-flood forecast systems.

5. It can be coupled with hydrological catchment models (Georgakakos, 1986a-b) with improved performance in flow predictions compared to decoupled hydrological models.
3. A SPATIALLY-LUMPED PRECIPITATION MODEL

3.1. Formulation Using Surface Meteorological Data

Georgakakos and Bras (1984a,b) formulated a spatially-lumped precipitation model in state-space form. Based on the surface pressure, temperature and dew-point temperature, their model gives as an output the precipitation rate. The model state is the mass of the condensed liquid water equivalent in the area characterized by the input temperature and pressure indices. The model formulation is based on pseudo-adiabatic ascent of the air-masses and on simplified cloud microphysics with exponential particle-size distribution and linear dependence of the particle terminal fall-velocity on the particle diameter. Evaporation of the falling particles, for unsaturated sub-cloud layer is explicitly taken into account by the model. Predictions of snowfall or rainfall are distinguished based on the surface air-temperature.

Figure 3.1 presents a sketch of the physical mechanisms that are modeled. The upper part of the figure is a plan-view of the moving (velocity denoted by \( u \)) storm clouds, while the lower part is a cross-section through them. The shaded regions correspond to a cloud-column characterized by the input variables: air-temperature, \( T_0 \); air-pressure, \( p_0 \); and dew-point temperature, \( T_d \); at the ground level. The model simulates the dynamics in this column. Air rises pseudo-adiabatically in the clouds with updraft velocity \( v \) (possibly height-varying), producing an input rate of condensed water equivalent \( I \). The input mass of condensed water is distributed to various droplet diameters according to an exponential particle size distribution, \( n(D) \), with parameter \( c \) representing the inverse of the cloud particle average diameter. Due to the action of the updraft at the cloud top, a portion of the water mass leaves the column with a rate \( Q_t \). The larger droplets fall through the cloud bottom with a rate \( Q_b \). The precipitation rate \( P \) at the ground level is computed from \( Q_b \) by subtraction of the mass evaporated due to possible unsaturated conditions below the cloud base. The model dynamics equation consists of a statement of the conservation of the condensed water equivalent mass \( X \) within the cloud column. Heat-adiabatic ascent is used to determine the cloud-base (level \( Z_b \)) pressure, \( p_s \), and temperature, \( T_s \). Pseudo-adiabatic ascent and the terminal pressure \( p_t \) at the cloud-top (level \( Z_t \)) are used to determine the terminal temperature \( T_t \) and, subsequently, the water vapor condensed per unit mass of moist air.
Figure 3.1. Schematic representation of a one-dimensional dynamic precipitation model. (Adopted from Georgakakos and Bras, 1984a).
The physical quantities $v$, $c$ and $p_t$ are parameterized using the input variables $p_0$, $T_o$ and $T_d$ in an effort to obtain a storm and location invariant structure. The updraft velocity $v$ is parameterized to represent convective ascent of the air masses. The terminal pressure $p_t$ is based on an empirical relationship between $p_t$ and $v$ that quantifies the fact that the stronger the updraft, the more rigorous the storm cloud development is and, consequently, the lower $p_t$ is. The cloud particle average diameter $1/c$ is given as a power function of the updraft, so that the stronger the updraft the larger the cloud particle diameters will be.

As a first step toward model verification, Georgakakos and Bras (1984b) considered uniform vertical profiles of the updraft velocity and the cloud particle average diameter. In addition, the cloud particle average diameter was held constant, independent of the updraft velocity. The free model parameters in this case are:

1) the ratio $\text{EPS1}$ of the updraft velocity to the square root of the potential thermal energy per unit mass of the ascending air at the height of average updraft velocity, and

2) the time- and storm-constant cloud particle average diameter denoted by $\text{EPS4}$ (equal to $1/c$).

Deterministic simulation runs were used to obtain contour maps of various least squares performance criteria in the parameter space. Values for the free parameters were selected from these maps.

Mathematical formulation of the precipitation model consists of the following scalar equations.

a) The Dynamics Equations:

\[
\frac{dX(t)}{dt} = f(t) - h(t) X(t); \quad t_k \leq t \leq t_{k+1}; \quad k = 0, 1, \ldots \quad (3.1)
\]

b) The Prediction Equation:

\[
Y(t_k) = \Delta t \phi(t_k) X(t_k); \quad k = 1, 2, \ldots \quad (3.2)
\]
where \( X(t) \) is the mass of condensed liquid water equivalent (CLWE) in the cloud column at time \( t \), \( f(t) \) is the input rate of CLWE in the cloud column due to condensation (=I), \( h(t) X(t) \) is the mass flux of CLWE that leaves the cloud bottom \( (=O_t + O_b) \), \( \Delta t \) \( (=t_k-t_{k-1}, k=1,2,\ldots) \) is the duration of the interval between precipitation observation times, \( \phi(t_k) X(t_k) \) is the mass flux of CLWE that reaches the ground as precipitation, and \( Y(t_k) \) is the predicted rainfall volume at the observation site during the interval of duration \( \Delta t \).

The input rate \( f(t) \) was approximated at each time by:

\[
f = \Delta w \rho_m v \, dA
\]  

(3.3)

where \( \Delta w \) is the mass of liquid water resulting from the condensation during the pseudo-adiabatic ascent of a unit mass of moist air, \( \rho_m \) is the vertically averaged (from cloud top to cloud bottom) density of moist air, \( v \) is the average updraft velocity in the cloud column, and \( dA \) is the unit area measure.

The condensed mass \( \Delta w \) can be computed from pseudo-adiabatic charts given ground-surface air temperature, pressure and dew-point temperature. The ideal gas law gives expressions for \( \rho_m \) given cloud top and bottom pressure and temperature. The updraft velocity \( v \) was parameterized by the convective expression:

\[
v = EPSI \, \sqrt{\frac{c_p}{p} (T_m - T_s')}
\]

(3.4)

where \( EPSI \) is a model parameter, \( c_p \) is the specific heat of dry air under constant pressure, \( T_m \) is the cloud temperature at a certain pressure level \( p' \), and \( T_s' \) is the corresponding ambient air temperature determined from surface data using heat-adiabatic ascent of air from the ground surface. The level \( p' \) is defined by

\[
p' = p_s - \frac{1}{4} (p_s - p_t)
\]

(3.5)

where \( p_s \) is the cloud bottom pressure computed from heat-adiabatic ascent and \( p_t \) is the cloud top pressure.
The cloud top pressure is given as a function of the updraft velocity according to:

\[
\frac{p_t - p_L}{p_o - p_L} = \frac{1}{1 + (v/v_o)}
\]

(3.6)

where

\[ p_L = 200 \text{ mbar} \]
\[ p_o = 500 \text{ mbar} \]
\[ v_o = 1 \text{ m/sec} \]

Equation 3.6 ensures that \( p_t \) will vary from \( p_o \) to \( p_L \), being inversely proportional to the updraft velocity.

The rate function \( h(t) \) at each time \( t \) is given as:

\[
h = \frac{v}{Z_c} \left[ 2 \frac{1 + \frac{3}{4} N_v + \frac{2}{4} + \frac{3}{24} + \frac{N_v}{4} - 1}{e} \right]
\]

(3.7)

and the rate function \( \phi(t) \) is given at each time \( t \) as:

\[
\phi = \frac{v}{Z_c} \left[ \xi \left( \frac{N_D}{N_v} \right) \left( 1 - \frac{N_v}{4} \right) \left( 1 + \frac{N_D}{2} \right) + \frac{N_D^2}{8} \right. \\
\left. + \left( 1 - \xi \left( \frac{N_D}{N_v} \right) \right) \left( 1 + \frac{3}{4} N_v + \frac{1}{4} N_v^2 + \frac{1}{24} N_v^3 - \frac{1}{24} N_D^3 \right) \right]
\]

(3.8)

The dimensionless numbers \( N_v \) and \( N_D \) are defined as:

\[
N_v = \frac{v}{aEPS4}
\]

(3.9)

\[ N_D = \frac{D}{EPS4} \]

(3.10)

with \( EPS4 \) a model parameter denoting the average hydrometeor diameter in the cloud column, \((aEPS4)\) the average terminal velocity of hydrometeors in the
cloud column and D_c a measure of the subcloud evaporation computed from air temperature, pressure and dew-point temperature near the ground level (see Georgakakos and Bras, 1984a-b, for details). Z_c denotes cloud height in meters. The velocity v_p is defined by:

\[ v_p = 4 a \text{ EPS4} \quad (3.11) \]

Testing of the formulated one-dimensional, dynamic, precipitation model complemented with a state estimator was done (Georgakakos and Bras, 1984b) simulating real-time forecast conditions, utilizing hourly point data from several storms of both the convective and the stratiform type. The main conclusions were:

1) The stochastic-dynamical precipitation model gives consistently good predictions of the hourly precipitation rates at a precipitation observation station when the input variables T_o, P_o and T_d are obtained from the same station. Efficiencies range from 0.28 to 0.38 (proportion of the variance in observed station precipitation accounted for by the model).

2) The forecast lead time of good predictions (efficiency greater than 0.10) varied from storm to storm. Nevertheless, it was never found less than 2 hours, with values of 0.20 persisting for 6 hours for some storms. These values of efficiency were obtained by using observed input data. Therefore, the forecast error of input prediction is not included in the aforementioned values.

3) The persistence coefficient, comparing the model performance to the performance of the scheme that predicts the previous precipitation observation, ranged from 0.07 to 0.50 for hourly forecast lead time indicating better (significantly better in some cases) performance for the stochastic-dynamic model.

4) The hourly model predictions compared favorably to the predictions of regression models calibrated on the same data that they were to
predict. The regression coefficients of the regression models varied considerably from storm to storm, thus, rendering these models inappropriate for the real-time precipitation forecasting. The parameters of the stochastic-dynamic model remained constant throughout the tests.

5) Due to the exclusive use of surface meteorological variables as input, the model predictions are poor in cases when thermal inversions are present in a moist lower atmosphere.

3.2. Parameter Estimation Studies

Georgakakos (1984) examined in detail the parameter estimation issues of the one-dimensional, stochastic-dynamic, precipitation model described previously. In his work, both the two free dynamic-model parameters and the state-estimator model-error statistics were estimated from hourly station precipitation data.

The methodology of parameter estimation consisted of: 1) the construction of contour maps of various performance indices in the space of the two free model parameters, 2) the use of a stochastic approximations algorithm that determined the minimum variance estimate of the model error variance parameter Q for each point considered in the parameter space, and 3) the use of equilibrium conditions for the dynamic precipitation model in the determination of the initial value of Q (to be used as the starting estimate for the stochastic approximations algorithm) as a function of the dynamic-model parameters.

Three performance criteria were used in an effort to examine various aspects of the model performance:

1) Errors in the total mass of each storm-group were represented by the absolute proportional mean error (APME). This criterion is the absolute value of the ratio of the 1-step predicted residual mean to the mean of corresponding observations for the period under study. A value of zero represents optimal performance with respect to this criterion.

2) The standard least-squares criterion was represented by the proportional standard error (PSE). It is the ratio of the 1-step predicted resi-
duals standard deviation to the standard deviation of the corresponding observations. It gives the proportion of the observations standard deviation unexplained by the model. A value of zero corresponds to perfect performance with respect to PSE.

3) Maximum likelihood estimation was represented by the average value of the log-likelihood (ALL) over the period of interest. The greater the value of this criterion the better the model performance is. Optimization with respect to ALL gives the parameter values with the highest probability of having generated the observed sequence, under the assumption that the model structure is the true one.

The calibration data consisted of hourly storm data from the meteorological station at Logan Airport, Boston, Massachusetts. The storms were divided into two groups:

1) A convective group consisting of a line storm and a tropical storm with a total of 110 wet hours.

2) A stratiform group consisting of low-pressure frontal storm with a total of 125 wet hours.

The space of the two free dynamic-model parameters was divided by grids and the value of each performance criterion was computed for each grid point via simulation with the model error variance parameter Q of the state estimator recursively estimated by the stochastic approximations algorithm.

Because of the anticipated discretization and computational inaccuracies, regions of best performance rather than a single optimal performance location were identified for each of the performance indices and for each of the storm groups. The regions that correspond to the upper 5% of the performance values for PSE and ALL and to values of APME less than 0.1 are presented in Figure 3.2. Figure 3.2a corresponds to the stratiform storm group, while Figure 3.2b corresponds to the convective storm group. The thick solid line in Figure 3.2 delineates the region of optimal performance with respect to PSE. The thin solid line defines the region of optimal APME and the dotted line defines the region of optimal ALL. The shaded area on both Figures 3.2a and 3.2b signifies the overlap of optimal regions for all performance criteria and both storm groups. The considerable overlap in the parameter space suggests robust
Figure 3.2. (a) Region in parameter space of optimal performance for the stratiform group.
(b) Region in parameter space of optimal performance for the convective group.
model structure and parameterization with respect to storm types and performance criteria.

Contour plots of the final Q estimates and of the coefficient of variation of Q during the adaptation period showed: 1) that the final estimates depend on the dynamic model parameters with a characteristically flat estimate surface at the optimal parameter region, and 2) that, near and at the optimal parameter region, low coefficients of variation of Q were observed suggesting fast convergence to final estimates.

Verification of the estimated stochastic-dynamic model parameters using data from convective storms in Tulsa, Oklahoma, resulted in performance indices that were close to the optimal values obtained for the Boston storm data. This supports the conclusion that the model structure and parameter values are reasonably robust to changes in climatic and topographic regime. Figure 3.3 presents an example of real-time hourly rainfall predictions for the verification storm group.

Physically realistic values of the dynamic model physical variables were obtained when the calibrated stochastic-dynamic model was used to forecast hourly precipitation rates in real time. Figure 3.4 presents, as an example, the time trace of the estimated hourly-averaged mass liquid water content in grams/m^3 for the convective storm group used in calibration.

3.3. Input Spatial Interpolation

The studies presented in Section 3.2 used meteorological and precipitation data from a single station to feed the stochastic-dynamic model that in turn produced precipitation rate forecasts at the location of the station. It is very usual, however, to need forecasts of precipitation rate in areas where no meteorological stations exist. Given an average horizontal separation distance of about 100 miles between first-order meteorological stations in the United States, it follows that many watersheds are devoid of such stations. It was necessary, therefore, to develop an interpolation procedure for the precipitation model meteorological input. Georgakakos (1986a) examined the spatial interpolation of surface air temperature T_o, pressure p_o, and dewpoint temperature T_d in terrain of varying altitude.
Figure 3.3. Forecasts (dashed line) vs. observations (solid line) for the verification storm group (Tulsa, OK). One-hour time steps. (Adopted from Georgakakos, 1984).
Figure 3.4. Time trace of the estimated hourly averaged mass liquid water content for the convective storm group used in calibration.
The input $u$ is decomposed into two parts $u_t(z)$ and $u_a$, corresponding to topography and atmospheric disturbance respectively, according to

$$u = u_t(z) + u_a$$

(3.12)

where $u$ denotes any of $T_o$, $p_o$, $T_d$ and $z$ denotes altitude.

For the determination of the topographic component of input an air parcel is followed as it is forced by the topographic relief to ascend from the lowest point in the area under consideration. The thermal properties of the parcel are determined from its initial properties at the lowest observation point, and from the assumption of heat-adiabatic ascent in unsaturated environment or pseudo-adiabatic ascent in saturated environment. Solution of the simultaneous algebraic equations that relate altitude with pressure and temperature (hypsometric), and equivalent potential temperature with mixing ratio and latent heat release, gives the topographic component $u_t(z)$ of the input for all the altitudes of interest. Once the topographic component of the input has been obtained, it is subtracted from the observations at the meteorological observation stations in the area. The residual atmospheric disturbance component $u_a$ is interpolated at the location of interest using a distance-weighted average of the residuals at all nearby stations. Use of (3.12) for that location gives the total value of the input $u$.

Georgakakos (1986a) verified the aforementioned methodology with six-hourly temperature, pressure and dewpoint data from Tulsa, Oklahoma, and Lewistown, Montana areas. Data from Wichita (Kansas), Springfield (Missouri) and Oklahoma City (Oklahoma) were utilized to determine the meteorological variables at Tulsa. Data from Billings (Montana) and Great Falls (Montana) were utilized for Lewistown. The nearest station in the case of Tulsa, Oklahoma, was 105 km away from the point of interest. In the case of Lewistown the point of interest was 145 km away from the nearest station. The difference in the two test cases lies in the different topographic and climatic regimes. The Tulsa case is characterized by a flat topographic regime at low elevations with an average air temperature of 15°C. The Montana case is characterized by pronounced topographic relief with cold temperatures averaging 5°C.
A total of 5 years of data were used for the Tulsa tests and a total of 2 years of data were used for Lewistown. Table 3.1 presents the interpolation error standard deviation together with the observations mean and standard deviation corresponding to temperature, pressure and dewpoint temperature for both test sites.

Table 3.1
INTERPOLATION ERROR AND OBSERVATION STATISTICS
(Temperature: °K; Pressure: mbars)

<table>
<thead>
<tr>
<th></th>
<th>ERROR ST. DEVIATION</th>
<th>OBSERVATIONS MEAN</th>
<th>OBSERVATIONS STANDARD DEVIATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) TULSA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>1.1</td>
<td>287.9</td>
<td>10.6</td>
</tr>
<tr>
<td>Pressure</td>
<td>0.8</td>
<td>992.8</td>
<td>6.9</td>
</tr>
<tr>
<td>Dew Point Temperature</td>
<td>1.5</td>
<td>281.6</td>
<td>10.4</td>
</tr>
<tr>
<td>2) LEWISTOWN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>2.1</td>
<td>278.4</td>
<td>12.0</td>
</tr>
<tr>
<td>Pressure</td>
<td>1.0</td>
<td>871.8</td>
<td>6.3</td>
</tr>
<tr>
<td>Dew Point Temperature</td>
<td>1.9</td>
<td>270.4</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Inspection of the statistics in Table 3.1 reveals that the interpolation methodology had a good overall performance.

Once an interpolation methodology for the input meteorological variables was developed, then the stochastic-dynamic precipitation model of the previous sections could be used to predict the precipitation rate distribution both in time and space given spatially sparse input data.
4. A TWO-DIMENSIONAL PRECIPITATION MODEL

4.1. Introduction

In order to provide information on the spatial distribution of rainfall, a two-dimensional precipitation model was formulated. The model uses temperature, pressure, dew-point temperature and wind direction and magnitude as its input and produces estimates of space-time precipitation at ground level as its output. The model is suitable for detrending precipitation fields observed by various sensors (remote and on-site) for the purpose of merging the observed precipitation fields. It is designed for use in real-time precipitation estimation with precipitation accumulation periods as short as one-hour long. It uses all available surface meteorological observations (excluding rainfall observations) in the area of interest. Upper-air meteorological observations are used (when and where available) to determine the middle tropospheric winds that it is assumed define the storm velocity. In formulating the model, parsimonious formulations were sought.

4.2. Formulation

Pielke (1984) provides the full set of partial differential equations that govern the time and space evolution of the air density, wind velocity components, potential temperature, and water substance in various phases at the mesoscale. The equations are expressions of the conservation of dry air mass, momentum in the three spatial directions, heat, and water substance mass. Given: 1) initial fields of the thermodynamic variables and the wind field, and 2) boundary conditions of the area of interest, it is possible, in principle, to integrate the governing equations and determine the three-dimensional distribution of water at each time. From this distribution and based on microphysical parameterizations (as for example in Georgakakos and Bras, 1984a,b) one can estimate the precipitation field at ground level in the area of interest.

Such an approach would be justified in cases where: 1) adequate observations of meteorological variables are available within the area of interest to define the initial and boundary conditions, and 2) adequate time and computer resources are available for timely integration of the governing equations.
Given an average distance of about 100 miles for meteorological stations (where hourly surface meteorological data are available in real time), the 12-hour interval between upper-air meteorological observations, and the short time increments between observations considered in this work (down to 1 hour), such a general approach was abandoned. Instead, it was considered adequate to use only the governing equations corresponding to the conservation of water mass law together with the heat conservation law (to determine mass of condensate) and substitute the conservation of momentum of air with a spatial interpolation procedure of the observed wind vectors. Initial and boundary conditions were obtained by objective interpolation procedures from available data.

Derivation of the partial differential equation that represents the water mass conservation law can be based on the elementary volume concept depicted in Figure 4.1. Denote by $Q_{j1}$ (in Kg/sec) the influx of liquid water equivalent in the positive direction of any one of the axes $x$, $y$ and $z$, and by $Q_{j2}$ the corresponding outflux in the same direction. If the local liquid water equivalent content of air is represented by $q$ (in Kg/m$^3$) and the local velocity of the liquid water equivalent by $u$, then for the $x$ axis

$$Q_{x1} = (qu)_{x1} \Delta y \Delta z$$ (4.1)

and

$$Q_{x2} = (qu)_{x2} \Delta y \Delta z$$ (4.2)

where subscript $(x,1)$ denotes conditions upstream, while subscript $(x,2)$ denotes conditions downstream on the $x$ axis. Assuming that $\Delta x$ is small (the volume considered is elementary), $(qu)_{x2}$ is given by the first two terms of a Taylor series expansion about $(qu)_{x1}$, as follows:

$$(qu)_{x2} = (qu)_{x1} + \frac{\partial (qu)}{\partial x} \Delta x$$ (4.3)

Thus, the net influx into the elementary volume along the $x$ axis is

$$Q_{x1} - Q_{x2} = - \frac{\partial (qu)}{\partial x} \Delta x \Delta y \Delta z$$ (4.4)

Similarly, the net influx in the $y$ and $z$ directions is
Figure 4.1. Elementary volume of moist air with influxes $Q_{x1}$, $Q_{y1}$, $Q_{z1}$ and outfluxes $Q_{x2}$, $Q_{y2}$, $Q_{z2}$ of condensed liquid water equivalent in the positive directions of the axis system $(x,y,z)$. Also shown is the net-source of liquid water equivalent $S_w$. 
respectively, where \( v \) and \( w \) denote the local velocities of the liquid water equivalent in the \( y \) and \( z \) directions respectively.

Because of possible condensation, evaporation, deposition, and sublimation there could exist sources and sinks of liquid water equivalent within the elementary volume. Given the small dimensions of the elementary volume, these sources and sinks are represented by a net-source term \( S_w \) (in Kg/sec/m\(^3\)) per unit volume. Accumulation of liquid water equivalent mass \( M \) is expected in the elementary volume with the rate of accumulation \( (\Delta M/\Delta t) \) (in Kg/sec) given by

\[
\frac{\Delta M}{\Delta t} = - \left( \frac{\partial q_u}{\partial x} \right) x_1 + \left( \frac{\partial q_v}{\partial y} \right) y_1 + \left( \frac{\partial q_w}{\partial z} \right) z_1 \Delta x \Delta y \Delta z + S_w \Delta x \Delta y \Delta z \tag{4.7}
\]

with

\[
\frac{\Delta M}{\Delta t} = \Delta x \Delta y \Delta z \quad \frac{\Delta q}{\Delta t} \tag{4.8}
\]

It has been assumed that the volume remains constant in time and that the anelastic conservation of mass law (usually utilized in mesoscale numerical models, Pielke (1984)) holds. Substitution of Equation (4.8) into Equation (4.7), elimination of \( (\Delta x \Delta y \Delta z) \) from both sides of the resulting equation, and taking the limits \( \Delta t \to 0, \Delta x \to 0, \Delta y \to 0, \Delta z \to 0 \), results in the following form of the conservation of liquid water equivalent mass law:

\[
\frac{\partial q}{\partial t} = - \frac{\partial q_u}{\partial x} - \frac{\partial q_v}{\partial y} - \frac{\partial q_w}{\partial z} + S_w \tag{4.9}
\]

The main assumption in the derivation of (4.9) is that no change of density of the air occurs within the elementary volume. Given the velocity field, \( u, v, w \), initial and boundary conditions on \( q \), and expressions for the term \( S_w \), the previous equation can be integrated to determine the field \( q \) at future times. Then, based on the predictions of the \( q \) field, accumulated precipitation can be estimated. Such an approach, although considerably simpler than the initially described one (based on Pielke's (1984) equations),
would still require considerable computer time to carry out if a three dimen-
sional grid was to be established over the area of interest. It is also not
appealing conceptually given the sparse nature of the meteorological observa-
tions in time and space.

A simpler model results if the terms of Equation (4.9) are integrated in
space, with the integration along the vertical z axis extending from the cloud
base to the cloud top. In symbolic form:

\[
\int_V \frac{\partial q}{\partial t} \, dV = \int_V (- \frac{\partial q u}{\partial x} - \frac{\partial q v}{\partial y} - \frac{\partial q w}{\partial z}) \, dV + \int_S q \, dV
\]  

(4.10)

where \( V \) represents the volume of integration (depicted in Figure 4.2).

Using Leibnitz's rule (Hildebrand, pg. 365, 1976) and assuming that the
boundaries of the integration volume \((X_1, X_2)(Y_1, Y_2), (Z_1, Z_2)\) are time invar-
ant, one obtains:

\[
\int_V \frac{\partial q}{\partial t} \, dV = \frac{d}{dt} X(t)
\]  

(4.11)

where

\[
X(t) = \int_V q \, dV
\]  

(4.12)

with \( X(t) \) representing the mass of liquid water equivalent in the integration
volume. The assumption of time invariant boundaries is weakest for the verti-
cal z direction. However, for short time intervals, the spatial-average
altitude of the top and bottom of the cloud can be assumed to be slowly vary-
ing.

The first term on the right-hand side of equation (4.10) can be written
as

\[
\int_V - \frac{\partial q u}{\partial x} \, dV = U_1(t) - U_2(t)
\]  

(4.13)

where

\[
U_1(t) = \int_{Z_1}^{Z_2} \int_{Y_1}^{Y_2} (qu)_X \, dy dz
\]  

(4.14)

with \((qu)_X\) representing the influx of liquid water at the boundary of the
integration volume defined by \( x = X_1 \). \( U_2(t) \) is similarly determined for the
boundary defined by \( x = X_2 \).
Figure 4.2. Integration volume and associated in- and out-flows of liquid water equivalent. The positive direction of the axes is shown.
The second and third terms in the right-hand side of Equation (4.10) can be expressed in an analogous to Equation (4.13) manner (see also Figure 4.2). The symbols \( V_1(t) \) and \( V_2(t) \) were used for the fluxes through the boundaries defined by \( y = Y_1 \) and \( y = Y_2 \), respectively. For the boundaries defined by \( z = Z_1 \) and \( z = Z_2 \), \( W_1(t) \) and \( W_2(t) \) represented the fluxes of liquid water, respectively.

The volume-integrated net-source term (last term) in Equation (4.10) was represented by \( I(t) \). Since the integration volume contains only cloudy air (assumed saturated with respect to water vapor) the total net-source term \( I(t) \) represents condensation and deposition only.

Using the previous definitions in Equation (4.10) results in:

\[
\frac{dX(t)}{dt} = U_1(t) - U_2(t) + V_1(t) - V_2(t) + W_1(t) - W_2(t) + I(t) \quad (4.15)
\]

and the task now is to determine expressions for the terms on the right-hand side of (4.15) in terms of the observable meteorological variables.

The condensation-deposition term \( I(t) \) can be modeled as in Georgakakos and Bras (1984a,b) and Section 3 of this document. The term \( W_2(t) \) signifies the liquid water equivalent in the form of minute droplets lost locally through the top of the clouds due to the action of the updrafts (e.g., anvil formation). The term \( W_1(t) \) signifies the influx of liquid water equivalent through the bottom of the cloud. Since liquid water equivalent only passes through the cloud bottom as precipitation, this term is negative and represents precipitation at cloud base.

It is noted at this point that in the absence of lateral fluxes (U and V terms in Equation (4.15)) Equation (4.15) describes a spatially-lumped precipitation model of the type presented in Section 3 of this document. Thus, the \( W \) terms were modeled as in Georgakakos and Bras (1984a,b).

The lateral-flux terms were all defined based on the same principle. Using the nomenclature of Figure 4.3, the term \( U_1(t) \) was defined as follows:

\[
U_1(t) = \overline{u(t)} \frac{X(t) + X'(t)}{2 \text{Dx}} \quad (4.16)
\]
BOUNDARY OF MASSES \( X(t) \) AND \( X'(t) \)

**NEIGHBORING INTEGRATION VOLUME**

**INTEGRATION VOLUME**

**AVERAGE WIND VELOCITY AT MIDDLE TROPOSPHERIC LEVEL THROUGH BOUNDARY**

**Figure 4.3.** Quantities for the computation of a boundary inflow along the positive x direction for an Integration Volume of horizontal dimensions \( D_x \) and \( D_y \).
where \( \bar{u}(t) \) represents the spatially-averaged velocity of liquid water equivalent at the boundary of interest; \( X(t) \) and \( X'(t) \) represent the mass of liquid water equivalent in the two neighboring integration volumes; and \( \Delta x \) represents the length of the integration volume along the \( x \) axis. Analogous representations were obtained for the rest of the lateral-flux terms in Equation (4.15). The wind velocity in the middle troposphere (500 mb to 700 mb) was used as the storm (and, consequently, the liquid water equivalent) velocity.

Referring to the horizontal discretization of the area of interest into integration volumes depicted in Figure 4.4, and using the previously discussed representation of lateral-flux terms, Equation (4.15) for the \((i,j)\) th integration volume becomes:

\[
\frac{dX_{i,j}(t)}{dt} = \bar{u}_{j-1,j}(t) \frac{X_{i-1,j}(t) + X_{i,j}(t)}{2 \Delta x} - \bar{u}_{j,j+1}(t) \frac{X_{i,j}(t) + X_{i,j+1}(t)}{2 \Delta x} + \bar{v}_{i-1,i}(t) \frac{X_{i-1,j}(t) + X_{i,j}(t)}{2 \Delta y} - \bar{v}_{i,i+1}(t) \frac{X_{i,j}(t) + X_{i,j+1}(t)}{2 \Delta y} - O_{b,i,j}(t) - O_{c,i,j}(t) + I_{i,j}(t)
\]

The terms \(-O_b\) and \(-O_c\) replaced \( W_1 \) and \( W_2 \) in Equation (4.15), respectively, and \( \bar{v} \) represents the storm velocity at the perpendicular to the \( y \) axis boundaries of the integration volumes. Given initial conditions, integration of the system of Equations (4.17) for all \((i,j)\) in the area of interest gives the time evolution of all \( X_{i,j} \).

Assuming that the time evolution of the state has been computed for \( N \) consecutive intervals of time \( t \), the accumulated precipitation at ground level \( P_{i,j} \), generated by the \((i,j)\) th integration volume is given by

\[
P_{i,j} = \Delta t \sum_{k=1}^{N} \phi_{i,j}(t_k) X_{i,j}(t_k)
\]

where \( \phi_{i,j} \) has been defined in Section 3 (Eq. (3.8)), and subscripts \( i \) and \( j \) signify that the precipitation generated by the \((i,j)\) th integration volume reaches the ground within the area of intersection of the vertical projection of the \((i,j)\) th integration volume and the ground surface. Equation (4.18) is a direct consequence of Equation (3.2).
Figure 4.4. Horizontal discretization grid for the two-dimensional precipitation model.
The assumption of precipitation falling directly below the integration volume where it was generated is not a very restrictive one for the case of rainfall that we are interested in. It could be a problem in case of snowfall. For example, consider the horizontal wind speed $c$ near the ground. In order to have significant drift of precipitation, $c$ should be at least equal to the value $c_0$ given by

$$c_0 = \frac{v_T}{Z_b} \quad (4.19)$$

where $v_T$ represents the terminal velocity of raindrops, and $Z_b$ represents the spatially averaged distance of the cloud base from the ground. For the rather extreme values: $Z_b = 2$ km, $v_T = 18$ km/hr, and $Dx = 5$ km, the limiting velocity $c_0$ becomes equal to 45 km/hr (about 28 miles/hr). Usually, the values of $Z_b$ much smaller than 2 km are expected during significant rainfall. Also, as it will be seen later in Section 5, constraints on the computer time required to integrate the model equations for an area the size of the area of the radar umbrella impose values of $Dx$ greater than 10 km. Thus, no significant drifts are expected below cloud base during significant rainfall.

4.3. Spatial Interpolation of Horizontal Wind

In previous sections (e.g. Section 3.3) the issue of the spatial interpolation of surface temperature, pressure, and dew-point temperature was addressed for the spatially-lumped precipitation prediction model. The two-dimensional precipitation model requires values of these meteorological variables for each integration volume. These values are needed for the computation of the terms $O_b$, $O_t$, and $I$ in Equation (4.17). Thus, the interpolation procedure of Section 3.3 was followed to interpolate the observations of surface temperature, pressure and dew-point temperature to the center of all the integration volumes in the area of interest.

In addition to the above mentioned meteorological variables, the two-dimensional precipitation model requires input values of the transport velocities $\bar{u}$ and $\bar{v}$ at the boundaries between neighboring integration volumes (see Figure 4.3). Since the conservation of momentum law is very expensive to integrate and requires high quality and quantity data, spatial interpolation was utilized instead.
There is evidence (e.g., Wallace and Hobbs, 1977), that supports the assumption that storms move with the middle tropospheric wind velocities. This assumption was made in the development of the two-dimensional model and, thus, it was necessary to obtain values of the middle tropospheric wind velocities at the boundaries between neighboring integration volumes. Assuming that upper-air wind data exists for the area of interest and for the time intervals of interest (down to one hour), a linear interpolation procedure for each one of the wind components (of the type used for the atmospheric disturbance component $u_a$ in Section 3.3) could be used. However, upper-air data is available only once every 12 hours with a spacing of stations that is very large compared to the scale length of the integration volume (order of 10 km). Both a temporal and spatial interpolation is necessary, and, under the data availability conditions, considerable smoothing of the wind field is expected.

The interpolation procedure is presented below on a step-by-step basis. The nomenclature of Figure 4.5 is used. It is assumed that there is only one upper-air data station within the area of interest, which is a reasonable assumption for areas of the order of $10^4$ km$^2$.

Step 1. For the location of the station $s_1$ in Figure 4.5 with upper-air and surface data, the magnitude ratio $\frac{c_1}{c_1}$ and the angle $\phi = \text{Angle} (c_1, c_1)$ is computed at the times when upper-air data are recorded (every 12 hours). The vectors $c_1$ and $c_1$ signify the recorded horizontal wind vectors at the surface and at the middle-tropospheric level, respectively.

Step 2. For the times when upper-air data are not available at station $s_1$, and for all times at the location of the other stations, the middle level tropospheric wind vectors $c_k$ are determined from the following magnitudes and directional angles:

$$c_k = c_k \frac{c_1}{c_1}$$

and

$$\text{Angle} (c_k, c_k) = \phi$$
Figure 4.5. Schematic for the spatial interpolation procedure applied to observations of the middle tropospheric horizontal wind speed.
Step 3. Based on the computed $c_k$ and $\phi_k$ at the locations of the surface-data stations, the horizontal components $\bar{u}_k$, $\bar{v}_k$ of the middle tropospheric wind for all the $N_s$ surface-data stations (within the area of interest) can be determined. Then, the middle tropospheric wind components at the boundaries between neighboring integration volumes are computed from:

\[
\bar{u}_{x,y} = \sum_{k=1}^{N_S} \frac{\bar{u}_k}{d_k} \quad (4.22)
\]

and

\[
\bar{v}_{x,y} = \sum_{k=1}^{N_S} \frac{\bar{v}_k}{d_k} \quad (4.23)
\]

with $\bar{u}_{x,y}$ and $\bar{v}_{x,y}$ representing the middle-tropospheric wind components along the $x$ and $y$ axes, respectively, at the location of a boundary surface between neighboring integration volumes. The variable $d_k$ signifies the distance of the $k$th station from the center of the boundary surface.

The above presented interpolation methodology preserves the ratio of magnitudes and directional phase angle between the recorded surface and upper-air wind for all locations and all times in the area of interest for which no observations of upper-air wind exist.

4.4. Initial and Boundary Conditions.

Integration of the governing differential equation (4.17) for all the integration volumes of interest (see Figure 4.4) requires initial conditions for the state variables $X_{i,j}$. Also, in case the middle-tropospheric wind at the boundaries of the area of interest (for $i=1$, $i=n_y$, $j=1$, $j=n_x$, in Figure
4.4) has a direction toward the area of interest, then, the boundary values of $X$ are also needed.

Two methods of initial condition determination were investigated. The first starts from some time before the beginning of the period of interest when no precipitation is observed within the area of interest, and assuming that $X_{i,j} = 0$, as an initial condition, integrates (Equation 4.17) forward in time up to the beginning of the period of interest. The resultant $X_{i,j}$ field is taken as the initial field for the period of interest. Experience with the first method with data from Oklahoma has indicated that, for hourly observations, a "warm up" period of at least 5 hours is required.

The second method of initial condition determination was based on the widely-used experimental, Marshall-Palmer drop-size distribution (Marshall and Palmer, 1948):

$$N_D = N_0 e^{-AD}$$

(4.24)

where $N_D dD$ represents the number of drops of diameter between $D$ and $D+dD$ in a unit volume of space, $A$ is a parameter depending on rainfall rate, and $N_0$ is the value of $N_D$ for $D = 0$. Marshall and Palmer (1948) give:

$$N_0 = 0.08 \text{ cm}^{-4}$$

(4.25)

and

$$A = 41 R^{-0.21} \text{ cm}^{-1}$$

(4.26)

with $R$ being the rainfall rate in mm/hr.

Consider the expression:

$$X = Z_c \int_0^\infty \rho_w \frac{\pi}{6} D^3 N_D dD$$

(4.27)

where $X$ is the mass of liquid water equivalent within an integration volume, $Z_c$ is the cloud height, $\rho_w$ is the density of liquid water ($= 1,000 \text{ kg/m}^3$), and where it has been assumed that the cloud base coincides with the ground surface which is a reasonable assumption for periods of significant precipitation. Equation 4.27 relates the model state variable $X$ with the precipitation
rate $R$, in terms of $N_D$. Integration of the right-hand side of Equation (4.27) yields:

$$X = \frac{Z_{c^*} \rho w \pi N_o}{(41)^4} R^{0.84}$$  \hspace{1cm} (4.28)$$

where $Z_{c^*}$ is in m, $R$ is in mm/hr and $X$ is in kg/m².

The second method of initial condition determination consists of determining $X$ from Equation (4.28) for the locations where observations of rainfall are available, and, then, interpolating in space to determine values of $X$ for all the integration volumes of interest. The first method is applied for some time before the beginning of the time period of interest. Then the model equations (4.17) are integrated forward to obtain the $X$ field at the beginning of the time period of interest. Experience with the second method indicated that a "warm up" period of 2 hours is adequate for the Oklahoma area and for hourly observations.

The results of application of the two methods were very similar. This is explained by the sparse nature of observations of rainfall within the area of interest and the considerable smoothing of the $X$ field of the second method due to linear interpolation.

The boundary conditions were determined based on the second method of initial condition determination. A row (column) of boundary integration volumes was created along the horizontal (vertical) boundaries of the area of interest. Then, the second method was used to obtain $X$ values at the boundary integration volumes using only observed precipitation at stations outside the boundaries of the area of interest. Then, the enlarged state vector equations were integrated forward in time. The decision to use precipitation data from stations outside the boundaries of the area of interest stems from the foreseen use of the model with kriging procedures that utilize observed precipitation within the area of interest. Potential use of the observed precipitation data twice has thus been avoided.
5. PARAMETER ESTIMATION FOR THE TWO-DIMENSIONAL PRECIPITATION MODEL

5.1. Introduction

This section proposes a procedure for the identification of the two free model parameters, $\varepsilon_1$ and $\varepsilon_4$. The quantity $\varepsilon_1$ is analogous to the ratio of kinetic to thermal energy per unit mass of ascending air. Parameter $\varepsilon_4$ represents the mean diameter of raindrop sizes (Georgakakos and Bras, 1984b, and Section 3.2). By definition, $\varepsilon_1$ is dimensionless and $\varepsilon_4$ has dimensions of length (in meters).

Optimal values of the two parameters were determined by examining contours of various performance indices. Section 5.2 presents the performance indices and the procedures used to determine the optimal parameter values.

Sensitivity analysis is presented in Section 5.3. The changes in the precipitation estimates of the model were determined resulting from changes in: 1) the size of the cross-sectional area of the integration volume, and 2) the use of upstream ($X'(t)$) instead of average ($\frac{X'(t) + X(t)}{2}$) conditions for the determination of the inflow rate of liquid water equivalent into the integration volume (see Equation (4.16)).

5.2. Parameter Estimation

5.2.1. Measures of Performance

Unbiasedness and high-accuracy are the two most desirable properties of the predictions of a model. Since one of the main purposes of this study was to provide a mean precipitation-field for detrending observed precipitation fields, unbiasedness is a required property of the model predictions. Therefore, the "optimal" parameter values are traced on the zero-bias contour line in the parameter space.

Accuracy is an absolute measure of performance as provided by scoring rules. One of the direct measures of accuracy would be the standard deviation of the point prediction error. Also, the time series of the standard deviation of the areal-mean estimation error was utilized since the mean value over the precipitation field is of prime interest in this study. The performance of the model predictions in space is measured by the threat score and bias (Charba and Klein, 1980).
The time-averaged value \( \mu_T \) of the areal-mean estimation error, and the time-standard-deviation \( \sigma_T \) of the areal-mean estimation error were computed by

\[
\mu_T = \frac{1}{N_T} \sum_{t=1}^{N_T} \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (P_{i,j}(t) - \hat{P}_{i,j}(t))
\]  

(5.1)

and

\[
\sigma_T = \frac{1}{N_T} \sum_{t=1}^{N_T} \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (P_{i,j}(t) - \hat{P}_{i,j}(t))^2 - \mu_T^2
\]  

(5.2)

where \( P_{i,j}(t) \) is the observed precipitation in the \((i,j)\)th grid square and for time step \( t \), and \( \hat{P}_{i,j}(t) \) is the corresponding model-estimated precipitation. \( N_T \) represents the total number of time steps, and \( n_x \) and \( n_y \) represent the number of grid squares in the \( x \) and \( y \) directions respectively (see Figure 4.4). \( \mu_T \) and \( \sigma_T \) were selected as parameter estimation performance indices due to their relevance to possible model uses as a detrending scheme for observed rainfall fields.

The threat score and bias have been developed for categorical forecasts. Their definitions are illustrated in Figure 5.1. The bias will be referred to as "bias score" to avoid confusion with the "bias in the mean". The threat score, as it was originally defined in Charba and Klein (1980), is not suitable for parameter estimation purposes. It can be shown that the threat score contour lines in the parameter space are open lines (in contrast to closed curves) ranging from zero, when model always predicts no rain at small \( \epsilon_1 \) values, to the highest value, when the model always predicts rain at high \( \epsilon_1 \) values. Similar behavior was also observed for the bias score. The threat score and bias score are then redefined in order to provide meaningful measures.

The surface rainfall observations over all air columns over the wet hours are ordered in terms of magnitude and the first, the second and the third quantiles are found, which divide the rainfall magnitude into four categories instead of the two categories defined in Charba and Klein (1980). (The original score definition involved only two categories: rain and no rain.) The threat score then can be defined by:

\[
TS = \sum_{i=1}^{4} C_i (P_i + O_i - C_i)
\]  

(5.3)
(Adopted from Charba and Klein, 1980).

Figure 5.1 Schematic illustration of threat score (TS) and bias (B)

\[ TS = \frac{C}{(F + O - C)} \]
\[ B = \frac{F}{O} \]

\( C \) - correctly forecast area
\( F \) - forecast area
\( O \) - observed area

\( TS = 0 \), when \( C = 0 \) (Incorrect forecasts)
\( TS = 1 \), when \( F = 0 = C \) (Perfect forecasts)
\( B = 1 \), when \( F = 0 \) (Perfect bias)
where \( C_i, F_i \) and \( O_i \) are the correctly forecast, forecast and observed area of rainfall in the \( i \)th category, respectively.

The bias score is redefined as:

\[
B = \frac{4}{\sum_{i=1}^{4} b_i}
\]  
(5.4)

with \( b_i \) defined by

\[
b_i = \begin{cases} 
\frac{O_i}{F_i} & \text{if } O_i < F_i \\
\frac{F_i}{O_i} & \text{if } F_i < O_i 
\end{cases}
\]  
(5.5)

Equations (5.4) and (5.5) differ from the definition Equation for \( B \) in Figure 5.1 in two respects: 1) the definition of more than one category, and 2) \( b_i \), for each category, is always less than or equal to one under the new definition. The second difference is necessary to avoid average bias scores that appear better than each of the bias scores averaged. For example, under the definition of Figure 5.1, the average of bias scores \( B_1 = 0.5 \) and \( B_2 = 2 \) would be equal to 1.25, which is a better score than either one of \( B_1, B_2 \).

The new definitions of the threat score and bias score measure not only the accuracy of spatial distribution as before, but also the accuracy of the precipitation magnitude. The best values of threat score and bias are now: one for perfect forecasts, and zero for worst forecasts. Because there are more restrictions under the new definitions, we expect smaller threat scores and bias scores values. Selection of these scores for parameter estimation was motivated by the possibility of model use in rainfall prediction.

5.2.2. Data and Basin Characteristics

Data from the Oklahoma City area was used for the study of the performance of the proposed two-dimensional precipitation model and for the parameter estimation studies.

Surface meteorological data corresponding to the first order National Weather Service stations: Oklahoma City and Tulsa, and to the station: Vance Air Force Base at Enid, Oklahoma, were obtained from the National Climatic Data Center in Asheville, North Carolina. Also, precipitation data from all the hourly stations in the state of Oklahoma were acquired. The data span the period January 1, 1985 through December 31, 1985, except for those
corresponding to station at Enid, Oklahoma, which were available for only the period July 1, 1985 through September 30, 1985. Oklahoma City was the only station in Oklahoma with upper-air data for the aforementioned period. These upper air data were also obtained.

The model domain span 189 km x 120.5 km defined by 35° 20'N, 36° 25'N, 95° 50'W and 97° 55'W. The model domain and the locations of the meteorological stations and raingages are presented in Figure 5.2.

There are 41 raingages in the area that were used to estimate of rainfall intensity which in turn was used to estimate the initial and boundary conditions. Data from the rain-gages was also used to produce the "observed precipitation field" utilized in the computation of the performance scores. Unfortunately one of the most valuable raingages, located close to the center of the model domain at Lincoln, Oklahoma (35° 42'N, 96° 53') with gage number 1684, has missing data for the period from June 1, 1985 through October 31, 1985. The raingage with missing data was treated as if it did not exist. This makes the total number of raingages used equal to 40.

A total of 99 wet hours of data (spanning six storm events) during the summer months were used for model parameter estimation. An hour is considered to be a "wet hour" if at least 5 percent of the raingages within the model domain show measurable precipitation. Table 5.1 lists the storm dates utilized in this work.

<table>
<thead>
<tr>
<th>STORM NO.</th>
<th>STARTING DATE (1985)</th>
<th>ENDING DATE (1985)</th>
<th>DURATION (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>07-25-00:00</td>
<td>07-25-20:00</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>08-14-00:00</td>
<td>08-14-22:00</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>08-19-03:00</td>
<td>08-19-12:00</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>09-13-05:00</td>
<td>09-13-21:00</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>09-25-00:00</td>
<td>09-25-09:00</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>09-29-00:00</td>
<td>09-29-17:00</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TOTAL HOURS</td>
<td>99</td>
</tr>
</tbody>
</table>
Figure 5.2. Model Domain spanning 120.5km x 189km. The location of meteorological surface-data and upper-air-data stations and raingage stations is shown.
In order to have a measure of the performance of the precipitation model, we need a methodology to transfer rainfall data from point observations at raingages to areal-mean estimates for air columns. The model domain was discretized into 60 x 60 grid squares of equal area. To avoid confusion with grid squares that represent integration volume we will call the observation-grid-squares, elements. The deviation of rainfall at any point in an element from the rainfall at the center of the element is small, due to the small size of the grid. Thus, the interpolated rainfall at the center of an element was taken as the mean rainfall rate over that element. The National Weather Service (NWS) interpolation technique (e.g., Larson, 1975, Larson and Peck, 1974) for point precipitation observations was used. Centered at the point where interpolated rainfall is needed, the method divides the two-dimensional space into four quadrants. The nearest raingage to the point of interest, in each quadrant, is used in the computation of the point rainfall estimate. The point rainfall estimate at the origin is computed as the weighted sum of the rainfall recorded by the four raingages. The weights are equal to the inverse of the squared distance of each of the four raingages from the origin. The areal-mean rainfall rate at ground level for an integration volume was then estimated by averaging the rainfall rates of the elements contained within the integration volume.

5.2.3. Results

This section discusses the results of parameter estimation for the two free model parameters $\varepsilon_1$ and $\varepsilon_4$. Three cases corresponding to three different advection hypotheses were examined: advection using the 500 mbar wind field, advection with the 700 mbar wind field and no advection. For each case the parameter space of the two free model parameters was discretized in grid squares and the value of each performance criterion was computed at each grid point. The values of parameter $\varepsilon_1$ ranged from $1.2 \times 10^{-3}$ to $7.2 \times 10^{-3}$, while the values of parameter $\varepsilon_4$ ranged from $2 \times 10^{-5}$m to $6 \times 10^{-5}$m. The ranges of values for $\varepsilon_1$ and $\varepsilon_4$ were established after some experience was gained as to the shape of the surface of the performance criteria. The model domain was discretized into five grid lengths in each direction. This discretization results in a 37.8 x 24.1 km² area for each spatial grid square. The rather coarse spatial discretization was imposed by the excessive CPU-time require-
ments of the parameter estimation procedure. Section 5.3 presents results of sensitivity analysis regarding spatial grid-square size.

Contour plots of the time-averaged value of the areal-mean estimation error $\mu_T$, the time standard-deviation of the areal-mean estimation error $\sigma_T$, the time-averaged value of the Threat Score $TS$ and of the Bias Score $B$, in the space of $\varepsilon_1$ and $\varepsilon_4$, are presented in Figures 5.3 through 5.6 for the case of advection with a 500mbar wind. Based on Figures 5.3, 5.5 and 5.6 the optimal region of the parameter space appears to be near the point $(\varepsilon_1 = 6 \times 10^{-3}, \varepsilon_4 = 6 \times 10^{-5})$ with performance indices: $\mu_T = 0$, $\sigma_T = 1.6$mm/hr, $TS = 0.07$ and $B = 0.26$.

Figures 5.7 through 5.10 present contour maps for $\mu_T, \sigma_T, TS$ and $B$ for the case of advection with the 700mbar wind (solid lines). The optimal parameter values for this case are the same as for the case of 500mbar wind with similar values of the performance indices.

The no-advection case is presented in Figures 5.11 through 5.14 (solid lines). Even though the contours of performance criteria differ from the previous two "advection" cases, the optimal parameter values are very close to the ones previously determined.

The parameter estimation results indicate that positive Threat Scores and Bias Scores are obtained with the two dimensional precipitation model while a value of $\sigma_T$ equal to 1.6mm/hr was obtained by the calibrated model. The main feature of the contour maps of $TS$ and $B$ is a ridge in parameter space defined by $\varepsilon_4 = 6.1 \times 10^{-5}$ m. Restrictions on CPU-time prevented exploration of the parameter space much beyond a value of $7.2 \times 10^{-3}$ for $\varepsilon_1$. The presented contour maps, however, indicate that the performance criteria are much more sensitive to the cloud-averaged droplet diameter $\varepsilon_4$ than they are to the updraft velocity parameter $\varepsilon_1$. Also, the performance criteria $TS$ and $B$ indicate that sensitivity to $\varepsilon_4$ is much greater for values less than the optimal value of $6 \times 10^{-5}$m than it is for values greater than $6 \times 10^{-5}$m.

The results also indicate that no-advection is as good as advection with either 500mbar or 700mbar wind. This is attributed to the significant smoothing of the model input data due to the very sparse available observations of meteorological input in the model domain.
Figure 5.3 Contours of the time-averaged value of the areal-mean estimation error in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. Advection is based on the 500mbar wind. Grid size: 37.8x24.1 km$^2$. 
Figure 5.4 Contours of the time-standard-deviation of the areal-mean estimation error in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. Advection is based on the 500mbar wind. Grid size: 37.8x24.1 km$^2$. 
Figure 5.5  Contours of the time-averaged value of the Threat Score in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. Advection is based on the 500mbar wind. Grid size: 37.8x24.1 km$^2$. 
Figure 5.6 Contours of the time-averaged value of the Bias Score in the space of parameters $e_1$ and $e_4$. Advection is based on the 500mbar wind. Grid size: 37.8x24.1 km$^2$. 
Figure 5.7. Contours of the time-averaged value of the areal-mean estimation error in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. Advection is based on the 700mbar wind. Solid line corresponds to the numerical scheme of Equation (4.16). Dashed line corresponds to the upstream numerical scheme of Equation (5.6). Grid size: 37.8x24.1 km$^2$. 
Figure 5.8. Contours of the time-standard-deviation of the areal-mean estimation error in the space of parameters $c_1$ and $c_4$. Advection is based on the 700mbar wind. Solid line corresponds to the numerical scheme of Equation (4.16). Dashed line corresponds to the upstream numerical scheme of Equation (5.6). Grid size: 37.8x24.1 km$^2$. 
Figure 5.9. Contours of the time-averaged value of the Threat Score in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. Advection is based on the 700mbar wind. Solid line corresponds to the numerical scheme of Equation (4.16). Dashed line corresponds to the upstream numerical scheme of Equation (5.6). Grid size: 37.8x24.1 km².
Figure 5.10. Contours of the time-averaged value of the Bias Score in the space of parameters $\epsilon_1$ and $\epsilon_4$. Advection is based on the 700mbar wind. Solid line corresponds to the numerical scheme of Equation (4.16). Dashed line corresponds to the upstream numerical scheme of Equation (5.6). Grid size: 37.8x24.1 km$^2$. 
Figure 5.11. Contours of the time-averaged value of the areal-mean estimation error in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. No advection is accounted for. Solid lines stand for a grid size of 37.8x24.1 km$^2$. Dashed lines stand for a grid size of 18.9x12.05 km$^2$. 
Figure 5.12. Contours of the time-standard-deviation of the areal-mean estimation error in the space of parameters $e_1$ and $e_4$. No advection is accounted for. Solid lines stand for a grid size of 37.8x24.1 km$^2$. Dashed lines stand for a grid size of 18.9x12.05 km$^2$. 
Figure 5.13. Contours of the time-averaged value of the Threat Score in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. No advection is accounted for. Solid lines stand for a grid size of 37.8x24.1 km$^2$. Dashed lines stand for a grid size of 18.9x12.05 km$^2$. 
Figure 5.14 Contours of the time-averaged value of the Bias Score in the space of parameters $\varepsilon_1$ and $\varepsilon_4$. No advection is accounted for. Solid lines stand for a grid size of $37.8 \times 24.1$ km$^2$. Dashed lines stand for a grid size of $18.9 \times 12.05$ km$^2$. 
5.3 Sensitivity Analysis

Sensitivity of model performance to the method of computation of the inflow of liquid water equivalent into an integration volume (as for example in Equation (4.16)) is examined next.

As an alternative to Equation (4.16), the boundary inflow was given by:

\[ U_1(t) = \overline{u}(t) \frac{X'(t)}{Dx} \] (5.6)

for the positive x-axis direction of Figure 4.3. This scheme differs from the "average" scheme used in Equation (4.16) in that it only uses the upstream integration-volume mass (in this case \( X'(t) \)) to compute inflow. It will be called the "upstream" scheme in the following.

The contours of all performance indices near the region of optimal performance in the parameter space were recomputed for the 700mbar wind case and they are presented in dashed lines in Figures 5.7 through 5.10. Even though the shape and location of the contour lines has slightly changed compared to the solid-line contours, the optimal values of the performance indices remained the same. The optimal value of \( \varepsilon_4 \) for the upstream scheme remained 6x10^{-5}, while the optimal value of \( \varepsilon_1 \) became equal to about 6.5 x 10^{-3}. Thus, for all practical purposes and given the available data, the model performance is not significantly sensitive to the scheme used for the determination of the boundary inflows into integration volumes.

Next, sensitivity of the parameter estimation results with respect to the size of the grid square of spatial discretization was examined. Since the parameter estimation results with and without advection were very similar, only the case of no advection, which was inexpensive in CPU time, was studied.

Figures 5.11 through 5.14 present the contours of all performance indices near the optimal region of parameter space (as defined by the parameter estimation runs) for a spatial discretization with a 18.9x12.05 km² grid-square (dashed lines). The results are very similar to the results corresponding to a 37.8x24.1 km² grid size showing little sensitivity of the parameter estimation results to spatial discretization.
Figures 5.15 through 5.18 present the dependence of the performance criteria on the grid-square area. For these runs no advection was included and the parameter values were fixed at their optimal values ($e_1 = 6 \times 10^{-3}$, $e_4 = 6 \times 10^{-5}$). Small sensitivity of model performance to changes in grid-square area is observed.

Increase of the standard deviation of the point residual errors as the grid-square area decreases is observed in Figure 5.19. The standard deviation $\sigma$ was defined in this case by:

$$
\sigma^2 = \frac{\sum_{t=1}^{N_T} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (P_{i,j}(t) - \hat{P}_{i,j}(t))^2}{N_T n_x n_y} - \left( \frac{\sum_{t=1}^{N_T} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \hat{P}_{i,j}(t)}{N_T n_x n_y} \right)^2
$$

From a value of about 2 mm/hr for an area of $10^4$ km$^2$ one arrives at a value of about 2.9 mm/hr for an area of $10^2$ km$^2$. This figure points to the dependence of the point residual error statistics on the scale of discretization. It also indicates that small-scale features of the observed data are not well estimated by the model. This last conclusion should be expected for areas with spatially sparse meteorological data. It is noted that the standard deviation of the observed point hourly precipitation is equal to 2.39 mm/hr.
6. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Previous work in modeling mesoscale rainfall fields was reviewed. A two-dimensional precipitation estimation model was formulated based on the principles of conservation of mass of liquid water equivalent, and of heat conservation. Advection of liquid water equivalent was accomplished by the middle tropospheric wind velocity which was assumed to be the storm velocity. Spatial interpolation of the spatially- and temporally-sparse wind observations was performed based on objective interpolation techniques. The model formulation explicitly accounts for: condensation of vapor, advection and sub-cloud evaporation of liquid water equivalent. The two free model parameters that determine updraft-velocity strength and particle-size distribution were estimated based on contours of various performance criteria in the parameter space. The contours were generated from real-time available meteorological and rainfall hourly observations of convective storms in Oklahoma.

Results of parameter estimation and sensitivity analysis showed that the two-dimensional precipitation model performance is very robust to changes in grid size of spatial discretization and to whether or not advection is included. This is attributed to the fact that available meteorological data is very sparse in space within the model domain. It is expected that, in data-rich areas, the model performance will become much more sensitive to grid size and advection computation. However, use of the model as a detrending scheme for real-time observed precipitation fields will imply in most cases data situations similar to the one examined in this work; and, for those cases, robust performance with $\mu_T = 0$ is expected.

In terms of the model structure, inclusion of orographic enhancement is recommended for the future. In terms of applications, use of the model with hydrometeorological data from other regions is recommended. Finally, the model mathematical form is suited for use with a state estimator for the short-term prediction of space-time rainfall. Such a research direction should be also pursued.
Figure 5.15 Time-averaged value of areal-mean, estimation error as a function of grid-square area. No advection was included. $\epsilon_1 = 6 \times 10^{-3}$, $\epsilon_4 = 6 \times 10^{-5}$ m.
Figure 5.16 Time-standard-deviation of areal-mean estimation error as a function of grid-square area. No advection was included.

$\varepsilon_4 = 6 \times 10^{-3}, \varepsilon_4 = 6 \times 10^{-5}$ m.
Figure 5.17 Threat Score as a function of grid-square area. No advection was included. $\epsilon_1 = 6 \times 10^{-3}, \epsilon_4 = 6 \times 10^{-5} \text{m}$. 
Figure 5.18 Bias Score as a function of grid-square area. No advection was included. $e_1 = 6 \times 10^{-3}$, $e_4 = 6 \times 10^{-5}$ m.
Figure 5.19 Standard deviation of point residual errors as a function of grid-square area. No advection was included.

\[ \varepsilon_1 = 6 \times 10^{-3}, \varepsilon_4 = 6 \times 10^{-5} \text{ m}. \]
7. REFERENCES

AGU Hydrology Section, Committee on Precipitation, 1984: A new interdisciplinary focus on precipitation research. EOS, Vol 65, No. 23.


Stackpole, J. B., and H. A. Bedient, 1970: PEP computation of dew point temperature at mandatory levels. NMC office Note 36, 2pp.


APPENDIX F

CO-KRIGING OF RADAR-RAINFALL AND RAIN GAGE DATA
Cokriging Radar-Rainfall and Rain Gage Data

WITOLD F. KRAJEWSKI

Hydrologic Research Laboratory, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, Maryland

An ordinary cokriging procedure has been developed to optimally merge rainfall data from radars and standard rain gages. The radar-rainfall data are given in digitized form. The covariance matrices required to perform cokriging are computed from single realization data, using the ergodicity assumption. Since the ground truth and the error structure of the radar data are unknown, parameterization of the covariance between radar data and the true rainfall is required. The sensitivity of the procedure to that parameterization is analyzed within a controlled simulation experiment. The experiment is based on a hypothesized error structure for the rainfall measurements. The effect of measurement noise and network density is examined. The usefulness of the procedure to remove the bias in radar is tested. Daily data are used.

INTRODUCTION

Recent progress in quantitative hydrology brings out in strong relief the need for accurate real-time analysis of precipitation, probably the single most important hydro-meteorological input to streamflow prediction models. Because of its large variability in space and time, precipitation is difficult to measure accurately with a network of rain gages. For real-time hydrologic applications of rain gage data, automated gages should be used. Large numbers of automated rain gages are both expensive and difficult to maintain and operate, even with today's sophisticated communication networks. An alternative device which is potentially useful for precipitation measurement is meteorological radar [Kessler and Wilk, 1968; Hudlow, 1973; Austin and Austin, 1974; Anderl et al., 1976].

Land-based weather radar provides capability to measure precipitation continuously in time and space, typically within a radius of up to 200 km. Radar measurement of precipitation is indirect, and raw reflectivity data have to be converted into rainfall units, using a "Z-R relationship" [Battan, 1973]. In order to estimate the coefficients of a Z-R relationship, rain gage data are used. Many radar systems are equipped with digital processors which allow them not only to convert the raw data into rainfall, but also to integrate them into a desired time and space scale.

Unfortunately, radar data, as well as data from other remote sensors, are characteristically in error because of equipment and meteorological variabilities. Austin [1964], Harrold et al. [1973], and Wilson and Brandes [1979], among others, discuss the various causes of these errors. The errors exhibit both systematic and random behavior and quite often can exceed 100% on a relative scale. It is impossible to eliminate these errors directly by using rain gage data to calibrate the radar, because of the generally low density of rain gage networks and the different sampling characteristics of the two sensors. Rain gages measure point precipitation on the ground level, while radar-based precipitation represents a volumetric (or areal) average above the ground at a level depending on the distance from the radar site.

In this paper an optimal estimation approach to the problem of measuring precipitation using both radar and rain gage rainfall data will be described. This represents a philosophy similar to that of Eddy [1979] and Crawford [1979], who studied the problem of radar and rain gage data merging in a multivariate analysis framework. Here a well-known geostatistical interpolation technique called kriging is examined. The use of kriging for merging radar and rain gage data was also studied by Lebel [1986] and Creutin and Debiets [1986]. In the study reported here, a numerical simulation experiment has been designed and carried out for the purpose of testing this technique.

The study was part of the design and implementation of a precipitation-processing system being developed for hydrologic use. The system is designed to be used with the NEXRAD (Next Generation Weather Radar) radar systems and will also include satellite data. The usage of satellite data is not addressed in this paper. Ultimately, the system will work in real time, providing hourly rainfall data for input into hydrologic models. Operational constraints dictated the choice of an ordinary cokriging algorithm instead of more sophisticated methods of universal cokriging [Myers, 1982] or disjunctive cokriging [Yates, 1986]. While these latter methods are perhaps more accurate and are theoretically justified for rainfall estimation, their computational requirements currently prohibit real-time applications. For more details on the future precipitation-processing systems of the National Weather Service, refer to papers by Hudlow et al. [1983, 1984] and Ahnert et al. [1983]. In the following sections a multisensor rainfall estimation problem will be formulated and a methodology to solve this problem described. Also, a test experiment design will be discussed, along with the results.

FORMULATION OF THE PROBLEM

Let us assume that our precipitation measurements network covering space \( \Omega \) consists of two sensors: a weather radar and a set of \( N \) rain gages. Let us further assume that the radar is equipped with a digital processor, which produces accumulated rainfall estimates on a rectangular grid over time period \( \Delta T \) and space \( \Omega \). Similarly, rain gage data represent point measurements for the same time period \( \Delta T \). Both data sets are schematically depicted in Figure 1.

The motivation to use both data sets to estimate rainfall stems from the error characteristics of the two sensors. Rain gage data is typically considered to provide good point accuracy, but it offers little information on the spatial distribution of rain storms, especially in convective type situations. Radar, on the other hand, is capable of accurately delineating rainfall boundaries but, because of various meteorological, equipment, and methodological factors, its estimates of rainfall are burdened with errors that are very often quite significant. Thus it is
The problem of rainfall estimation, using the two sensors, can then be formulated as follows: Find the best estimate $V^*(u_o)$ of $V(u_o)$, defined as

$$V(u_o) = \frac{1}{|A|} \int_A Z(u_o) \, du$$

where $|A| = |A_k|$.

Thus $V^*(u_o)$ is an estimate of the mean areal precipitation on the ground level over the same area as sampled by radar.

**Model Description**

As a solution to the problem formulated above, a linear model is proposed:

$$V^*(u_o) = \sum_{i=1}^{N_G} \lambda_{o,i} G_i(u_o) + \sum_{j=1}^{N_R} \lambda_{R,j} R_j(u_o)$$

where $N_G \leq N$ is the number of gages in the local vicinity of location $u_o$, $N_R$ is the number of radar bins surrounding the location $u_o$, and $\lambda_{o,i}$ and $\lambda_{R,j}$ are corresponding coefficients (weights) that need to be estimated.

It is assumed in the model that the rainfall field $Z(u)$ is second-order stationary and ergodic over the space $\Omega$. It is also assumed that rain gage observation errors are random with zero mean, and variance $\sigma_o^2$ and uncorrelated in space. The radar observation errors are random with mean $m_{rx}$ and spatial covariance $\text{Cov}_{RG}(u)$. Both assumptions have bases in various experiments with real world data; however, the spatial error structure of radar is, to the best of the author's knowledge, unknown at this point.

The weights $\lambda_{o,i}$ and $\lambda_{R,j}$ can be obtained minimizing the estimation variance:

$$\sigma_v^2 = E[(V - V^*)^2] = \frac{1}{|A|^2} \int_A \int_A \text{Cov}_v(u - v) \, du \, dv$$

$$= \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} \lambda_{o,i} \lambda_{R,j} \text{Cov}_{RG}(u_o - u_j)$$

$$+ \frac{1}{|A|^2} \int_A \int_A \text{Cov}_R(u - u_j) \, du \, dv$$

$$+ \sum_{i=1}^{N_G} \lambda_{o,i}^2 \text{Cov}_{GG}(u_o - u_j)$$

$$+ \sum_{j=1}^{N_R} \lambda_{R,j}^2 \text{Cov}_{RR}(u_o - u_j)$$

$$+ 2 \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} \lambda_{o,i} \lambda_{R,j} \text{Cov}_G(u_i - u_j) \text{Cov}_R(u_o - u_j)$$

where $E[\ ]$ is the expectation operator, $u_o$ is the middle point in block of area $A$ at the location of interest, and $u$ and $v$ are other points within the same block $A$. The covariance of the true area average process is denoted as $\text{Cov}_v(\cdot)$; $\text{Cov}_{RG}(\cdot)$ is an unknown covariance between the integrated process $Z$ and the rain gage observations; $\text{Cov}_{RG}(\cdot)$ is an unknown covariance between the integrated process $Z$ and the radar observations; $\text{Cov}_{GG}(\cdot)$ is the covariance of the rain gage data; $\text{Cov}_{GG}(\cdot)$ is the cross covariance between gage and radar data; and $\text{Cov}_{RR}(\cdot)$ is the covariance of the radar data.
For the estimate \( V^* \) to be unbiased, it has to satisfy
\[
E\{V^*\} = E\{V\}
\]
(6)

Under our assumptions about the stationarity and error structure, the following conditions apply:
\[
\sum_{i=1}^{n_0} \lambda_{g1} = 1
\]
(7)
\[
\sum_{j=1}^{n_R} \lambda_{Rj} = 0
\]
(8)

It should be noted, however, that if \( m_{g0} = 0 \), i.e., the radar-rainfall field is unbiased, then (7) and (8) reduce to
\[
\sum_{i=1}^{n_0} \lambda_{g1} + \sum_{j=1}^{n_R} \lambda_{Rj} = 1
\]
(9)

The problem of \( \sigma^2 \) minimization under unbiased conditions can be solved using the Lagrange multiplier technique. Minimization of the Lagrangian function leads to a set of simultaneous linear equations that can be written in matrix form as:
\[
\begin{bmatrix}
  \text{Cov}_{RR} & \text{Cov}_{g0} & 1 & 0 & \lambda_{g1} \\
  \text{Cov}_{gR} & \text{Cov}_{RR} & 0 & 1 & \lambda_{R1} \\
  1 & 0 & 0 & 0 & \mu_k \\
  0 & 1 & 0 & 0 & \mu_G
\end{bmatrix}
\begin{bmatrix}
  \text{Cov}_{g0} \\
  \text{Cov}_{RR} \\
  \text{Cov}_{gR} \\
  \text{Cov}_{RR}
\end{bmatrix} = \begin{bmatrix}
  0 \\
  0 \\
  \mu_k \\
  \mu_G
\end{bmatrix}
\]
(10)

where
\[
\text{Cov}_{RR} = \begin{bmatrix}
  \text{Cov}_{RR}(u_{g1}, u_{g1}) & \cdots & \text{Cov}_{RR}(u_{RR}, u_{RR}) \\
  \vdots & \ddots & \vdots \\
  \text{Cov}_{RR}(u_{RR}, u_{RR}) & \cdots & \text{Cov}_{RR}(u_{RR}, u_{RR})
\end{bmatrix}
\]
\[
\text{Cov}_{gR} = \begin{bmatrix}
  \text{Cov}_{gR}(u_{g1}, u_{g1}) & \cdots & \text{Cov}_{gR}(u_{g1}, u_{gRR}) \\
  \vdots & \ddots & \vdots \\
  \text{Cov}_{gR}(u_{gRR}, u_{g1}) & \cdots & \text{Cov}_{gR}(u_{gRR}, u_{gRR})
\end{bmatrix}
\]
\[
\text{Cov}_{gG} = \begin{bmatrix}
  \text{Cov}_{gG}(u_{g1}, u_{g1}) & \cdots & \text{Cov}_{gG}(u_{gG}, u_{g1}) \\
  \vdots & \ddots & \vdots \\
  \text{Cov}_{gG}(u_{g1}, u_{gRR}) & \cdots & \text{Cov}_{gG}(u_{gG}, u_{gRR})
\end{bmatrix}
\]
\[
\text{Cov}_{gR} = [\text{Cov}_{RR}]^T
\]
\[
\text{Cov}_{gG} = \langle \int_{\Omega} \text{Cov}_{gG}(u, u_{g1}) du, \cdots, \int_{\Omega} \text{Cov}_{gG}(u, u_{gRR}) du \rangle^T
\]
\[
\lambda_{g1} = \langle \lambda_{g1}, \cdots, \lambda_{gRR} \rangle^T
\]
\[
\lambda_{R1} = \langle \lambda_{R1}, \cdots, \lambda_{RR} \rangle^T
\]
\[
\lambda_{g1} = \langle \lambda_{g1}, \cdots, \lambda_{gRR} \rangle^T
\]

and \( \mu_k \) and \( \mu_G \) are scalar Lagrangian multipliers. The superscript \( T \) denotes matrix or vector transpose operator.

The system (equation (10)) yields a unique solution if the covariance matrix is positive definite. Because of the irregular pattern of the network of rain gages the system needs to be solved for each location, and it is necessary to approximate the matrices \( \text{Cov}_{gG} \) and \( \text{Cov}_{gR} \) with a function model. Although the relative configuration of radar data is not changed from location to location and the number of data values used in the \( \text{Cov}_{gR} \) computation is typically large, the matrix \( \text{Cov}_{gR} \) also needs to be approximated to ensure positive definiteness.

To avoid solving the system (Equation (10)) at each location in \( \Omega \), a modified algorithm is proposed. First, the rain gage data are interpolated onto the same grid blocks as those for which the radar data are given. This is done using block kriging estimation [see Journel and Huijbregts, 1978]. Then the above described algorithm is used to cokrige the two fields: the radar data field and the field obtained by block kriging the rain gage data. Now the relative geometry in either field and between the two fields is constant throughout \( \Omega \) (except on the edges), the system (10) needs to be solved only once, and the same weights can be applied at each location in the field. The matrices \( \text{Cov}_{gG}, \text{Cov}_{gR}, \) and \( \text{Cov}_{RR} \) are approximated by exponential models.

The computational algorithm can be summarized as follows:

1. Estimate the rain gage data covariance function, using the exponential isotropic model. This is done by least squares fit.
2. Block krig the rain gage data to estimate
\[
V_g^*(i, j) = \frac{1}{|A_i|} \int Z(u_i) du_{ij} \quad u_{ij} \in \Omega
\]
3. Estimate \( \text{Cov}_{gG}, \text{Cov}_{gR}, \) and \( \text{Cov}_{RR} \) from the radar and the new kriged rain gage fields. Now the elements of \( \text{Cov}_{gG} \) represent estimates of the cross-covariance between two areal average observation fields. The estimation of \( \text{Cov}_{gR} \) and \( \text{Cov}_{gG} \) will be explained later.
4. Model \( \text{Cov}_{gG}, \text{Cov}_{gR}, \) and \( \text{Cov}_{RR} \) using exponential isotropic models. This is done by least squares fit, also.
5. Compute \( \lambda_{g1} \) and \( \lambda_{R1} \).
6. Cokrige the two fields.

The choice of exponential model was made for two main reasons: first, the behavior of the model near the origin, which seems appropriate for rainfall estimation [Rodriguez-Iturbe and Mejia, 1974b], and second, its computational efficiency. Computational efficiency was also the main reason for using least squares as the method of covariance estimation. The proposed algorithm is much faster than using direct rain gage observations and varying network configuration from location to location, and at the same time it results in minimal degradation of accuracy.

A few additional remarks are in order. First is the problem of a changing pattern along the edges of the domain \( \Omega \). In the work presented here, no special accommodation has been made for this problem. The same weights are applied along the edges as in the middle of \( \Omega \), but the bins located outside of \( \Omega \) are treated as missing values. Such a procedure results in local bias, but it did not affect the results, since they were based only on the points located inside \( \Omega \), separated by a few bins from the edges. The second important problem is that of the small sample size of the rain gage data. In a real-world application of the method presented here, where there is not enough data to compute reliable covariance, computed covariances can be substituted with climatological covariances computed from historical data. However, for reasons explained in the next section, this problem is not of concern in this paper.

To solve the system (10), one needs to estimate the vectors \( \text{Cov}_{gR} \) and \( \text{Cov}_{gG} \). Elements of these vectors are covariances between radar and rain gage data, respectively, and the true precipitation \( V \). Since \( V \) is unknown, \( \text{Cov}_{gR} \) and \( \text{Cov}_{gG} \) have to be approximated. The approximation used has the form:
\[
\text{Cov}_{gR} = \beta_R \text{Cov}_{gG} \quad \beta_R \in (0, 1)
\]
(11)
KRAJEWSKI: RADAR RAINFALL AND RAIN GAGE DATA

Rain gage field

Radar field

Fig. 2. Schematic representation of the configuration of the data included in the cokriging system from each field.

\[ \text{Cov}_{Y_R} = \beta_R \cdot \text{Cov}_{R_G} \quad \beta_R \in (0, 1) \] (12)

where the elements of \( \text{Cov}_{R_G} \) have the meaning described in point 3 of the algorithm. The values of \( \beta_R \) and \( \beta_G \) are unknown scalars and, in general, are extremely difficult (if not impossible) to estimate from the data. However, since they represent a measure of the accuracy of radar and gage measurement in relation to \( V \), they can be estimated based on experience or specially designed experiments. The sensitivity analysis of the previously presented method with respect to these two parameters will follow, and some recommended values will be given.

It should be pointed out that \( \text{Cov}_{R_R} \) and \( \text{Cov}_{G_G} \) can be expressed in terms of data error characteristics. However, because of the generally unknown statistical error structure of radar data, it was decided to investigate the previously described approach first.

Before the weights \( \lambda_R \) and \( \lambda_G \) are computed, one needs to decide how many data points should be used for estimation at each location. The configuration of the data used in the present study was constant and is schematically presented in Figure 2. The relatively small number of data points (five from each field) was dictated again by the computational efficiency.

The presented algorithm accounts for the sampling differences of the two sensors. It also uses the spatial cross-correlation function to account for differences between the error fields involved.

Once the coefficients \( \lambda_R \) and \( \lambda_G \) are determined, one can calculate the estimation variance as

\[ \sigma^2 = \text{Cov}_{R_G}(u_0, u_0) - \mu_R - \sum_{i=1}^{N_R} \lambda_R \cdot \text{Cov}_{Y_R}(u_0, u_i) \]

\[ - \sum_{i=1}^{N_G} \lambda_G \cdot \text{Cov}_{R_G}(u_0, u_i) \] (13)

In this expression \( \text{Cov}_{R_G} \) was substituted for \( \text{Cov}_{Y_R} \).

TEST OF THE METHOD

Probably the most natural way to test any method of estimating mean areal precipitation would be to compare the results with data from a very dense network which would constitute a "ground truth." In the present case, however, such an approach seems to be infeasible. First, there are not many (if any) rain gage networks with high enough density and large enough coverage. Second, it may be impossible to find corresponding radar data, since the systematic archiving of the high-resolution RADAP II data by the National Weather Service (NWS) started in 1985. Also, a significant data management effort is required to handle enough data to give the experiment statistically meaningful results.

In order to avoid these problems, a numerical experiment has been designed following the ideas of Greene et al. [1980]. In the experiment the sampling and measurement properties of radar and rain gages are mathematically simulated by generating radar and rain gage data from an original rainfall field which constitutes the "ground truth." Such an experiment has many advantages: (1) full control of the experiment with minimum effort; (2) knowledge of the true field; (3) control of the measurement errors; (4) control of the measurement network configuration (sampling scheme, network density); (5) feasibility of performing sensitivity analysis with respect to measurement errors and network density; (6) statistically valid conclusions.

The experimental system which is schematically depicted in Figure 3 consists of four elements. These are described in the following sections.

Original Field

The original field could be generated by a space-time rainfall model, such as developed by Waymire et al. [1984], or it could be a high-quality radar field. The latter was selected, and the original fields are the radar-rainfall fields from the GARP Atlantic Tropical Experiment (GATE) conducted in 1974. The GATE data represent convective systems and there-
fore provide a stringent test for the present estimation procedure. The data are described by Hudlow and Patterson [1979]. The GATE data underwent substantial analysis prior to their release. They are considered to be of high quality in that anomalous propagation and other outliers are removed, and all the major features of real storm events are preserved. The GATE radar data defines the $\Omega$ region as being a circle inscribed in a $400 \times 400$ km square. Data points are given on a rectangular $4 \times 4$ km grid.

**Radar Generator**

The radar generator used in this study is described by Krajewski and Georgakakos [1985]. The radar field is generated as

$$ R(i, j) = \theta(i, j)10^{\mu_{R}+\sigma_{R}^{2}\rho_{R}(i, j)} $$

where $R(i, j)$ is the radar field at the location $(i, j)$; $\theta(i, j)$ is the original field at the same location; $\sigma_{R}^{2}$ is the random component of the noise field and is a stationary, isotropic random field, with the mean $\mu_{R}$, variance $\sigma_{R}^{2}$, and spatial correlation function $\rho_{R}(i, j)$; and $S(i, j)$ is the deterministic component of the noise field.

The deterministic component $S(i, j)$ accounts for the measurement behavior of the radar as a sensor. It usually exhibits higher errors in high rainfall intensity and high gradient areas. The form of the function $S(i, j)$ was borrowed from Greene et al. [1980]:

$$ S(i, j) = 0.5\left\{\langle\nabla \theta(i, j)\rangle_4 \right\}^{-1} + \Theta(i, j)\theta_{\text{m}}(i, j)^{-1} $$

(15)

where $\langle\nabla \theta(i, j)\rangle_4$ is the average absolute value of the gradient computed in four directions around the point $(i, j)$ in the original field; $\nabla \theta(i, j)$ is the maximum absolute gradient in the original field; and $\theta_{\text{m}}(i, j)$ is the maximum value in the original field. The parameters of the random component $\epsilon$ are, in general, unknown but can be estimated for the purpose of the generation by setting requirements on the resultant radar field $R$. In the present study these requirements were (1) The mean of the radar field is required to be $M_{\text{R}}^{\ast}$. Thus

$$ M_{\text{R}}^{\ast} = \frac{1}{|\Omega|} \int_{\Omega} E(R) \ d\Omega $$

(16)

(2) The logarithmic variance of the ratio $R/\theta$ is required to be $L_{\text{R}}^{\ast}$:

$$ L_{\text{R}}^{\ast} = \frac{1}{|\Omega|} \int_{\Omega} E\left[\log_{10}^{2}(R/\theta)\right] \ d\Omega - \frac{1}{|\Omega|} \int_{\Omega} E\left[\log_{10}(R/\theta)\right] \ d\Omega $$

(17)

Solving the system of (16) and (17), one can obtain $\mu_{R}$ and $\sigma_{R}^{2}$. The correlation function $\rho_{R}(i, j)$ of the $\epsilon$ is assumed to be isotropic and exponential. Its parameter was specified directly in the present experiment and was a subject of the sensitivity analysis. Once the parameters of $\epsilon$ are specified, $\epsilon$ is generated using, for example, the Turning Bands method [Montogoliou and Wilson, 1982]. After $\epsilon$ is generated and the values of the function $S$ are computed for each location in $\Omega$, the radar field can be generated using (14).

**Rain Gage Data Generator**

Rain gage data are generated in two steps. First, the locations of the gages are selected based on a uniform random distribution in $\Omega$. Second, a rainfall value is assigned to each location, based on the original field values. The original field of choice (GATE radar data) represents areal averages, but we want to generate the corresponding point process values. To do that, the relationship between the variances of the point
process and the areal average process given by Rodriguez-Iturbe and Mejia [1974a] was used:

\[ \sigma_p^2 = \sigma_A^2 \left[ \int_0^d r(v)G(v) \, dv \right]^{-1} \]  

(18)

where \( \sigma_p^2 \) and \( \sigma_A^2 \) are the point and areal average process variances, respectively; \( r(\cdot) \) is the correlation function of the point process; \( G(\cdot) \) is the probability density function for random distribution between two points in the averaging area (a square in the present case); and \( d \) is the largest distance in that area. The distribution \( G(\cdot) \) for a rectangular area is given by Gosh [1951].

It was assumed that the rainfall data are lognormally distributed; consequently, the gage values \( G_k(i, j) \) are generated as

\[ G_k(i, j) = LN\{ \theta(i, j), \sigma_p^2 \} + LN\{0, 0.0001m^2\theta^2(i, j)\} \]  

(19)

\( k = 1, \cdots, N \)

where \( m \) is measurement error expressed as a percentage of the mean, and \( LN\{a, b\} \) denotes lognormal distribution, with the mean \( a \) and variance \( b \).

The correlation function \( r(\cdot) \) was assumed to be exponential and was estimated for the local vicinity of \( G_k(i, j) \). The correlation parameter \( h \) was obtained from the equation

\[ r_A(\tau_k) = \int_A r(u, h) \, du \]  

(20)

where \( r_A(\tau_k) \) is the lag one correlation of the areal average process, as computed from the data.

**Performance Criteria**

The evaluation of the cokriging algorithm to merge radar and rain gage rainfall data is based on the comparison of the cokriged field and the original field. The comparison can be done by visual inspection of the resultant maps or by a set of statistics. The latter approach is more appropriate if the objective of the experiment is sensitivity analysis. The statistics selected for comparison include the mean and variance of the fields, the correlation coefficient with the original field, the estimation variance (both computed and estimated), the mean square error for averaging areas ranging from 16 km\(^2\) to 1000 km\(^2\), and the maximum mean square error in the field for the same areas. Inspection and proper interpretation of all of these statistics allows us to evaluate the proposed methodology in a fair way. For the sake of clarity, the expressions used to compute the mean square error and estimation variance are given:

**Mean square error**

\[ \frac{1}{N_A} \sum_{i=1}^N (\theta_i - M)^2 \]  

(21)

**Estimation variance**

\[ \frac{1}{N_A} \sum (\theta_i - M)^2 - \left[ \frac{1}{N_A} \sum (\theta_i - M) \right]^2 \]  

(22)

where \( N_A \) is the number of points used in comparison and \( \theta_i \) and \( M \) are the original and merged field values, respectively.

It should be realized that the general validity of such a numerical experiment is limited by the validity of the error models of radar and rain gage data. There have been several
studies done on the error characteristics of rain gage data (see, for instance, Larson and Peck [1974] and Sevruk [1982]). In the light of these studies the proposed model seems to be adequate. The situation is quite different as far as radar data are concerned. There were many studies done in the past on the comparison of radar and rain gage data, but the question of what the statistical structure of radar rainfall errors is in space remains unanswered. Such a question, however, is critical in the design of an experiment like the one described here. The radar error model presented has certain qualitative features identified by previous studies:

1. The model delineates rainfall patterns correctly, i.e., anomalous propagation (AP) is not modeled in clear air. It was assumed that in the operational environment the radar data would undergo quality control steps that would, perhaps with the aid of satellite data, eliminate AP.
2. Errors are higher in high-rainfall gradient areas.
3. Errors are higher in high-rainfall intensity areas.
4. Errors are correlated in space.

Among the effects that are not modeled are (1) Range effect due to attenuation of electromagnetic wave (it was assumed that for the radars with parameters corresponding to those planned for the NEXRAD, this effect is negligible [Hudlow et al., 1984]); (2) Beam-filling effect (however, the effect of this problem also will be significantly reduced using NEXRAD equipment and processing); and (3) Complete evaporation of rain before hitting the ground (thus application of such an error model in some situations may be inappropriate). Summarizing, one could say that if the error models are valid, then the results of this study are valid also.

RESULTS OF THE SENSITIVITY ANALYSIS

The sensitivity analysis of the merging method described previously was performed using daily data from the GATE experiment. It is very expensive in terms of computer central processing unit (CPU) time to perform a truly comprehensive experiment. It was estimated that such an experiment would take over 5 years of CPU time on the PRIME 750 computer. Therefore a limited experiment was performed instead. The optimal combination of the parameters $\beta_0$ and $\beta_k$ was sought as a function of various noise parameters and rain gage densities. The noise parameters selected here for investigation were the bias of the radar field, the variance $L^*_R$ (see equation (17)), and the correlation distance $h_k$ in the $e$ field. The effect of measurement error in the gage observations was not studied. An error of 10% was specified for all the runs.

Since the framework of the methodology described here is the estimation of random fields, the proper way of conducting this experiment would be to repeat the analysis for a number of realizations (at least 30) for each field, keeping the same noise parameters, and then to average the results across the ensemble of realizations. Such a procedure, however, is very costly and prohibits even a limited experiment. To evaluate the variability of the results across realizations, a few (five) realizations were used for a selected set of radar noise parameters and a network of 50 gages. The data for GATE day 245 (September 2, 1974) were used. The radar noise parameters were (1) no bias in the radar field, (2) high noise ($L^*_R = 0.03$, which is in the range given by Greene et al. [1980] and was also found to generate outliers [Krajewski and Georgakakos, 1985]), and (3) correlation distances in the $e$ field of 8 and 20

$Fig. 6$. Contours of the mean square error. Parameters are $L^*_R = 0.03, \rho_g = 8 \text{ km}, \text{ no bias, and } 50 \text{ gages.}$
Two basic situations were investigated: the unbiased radar field case and the biased radar field case. It is important to distinguish between these two cases, since in the operational environment of NWS, where the previously described method is to be implemented [Ahnert et al., 1983; Hudlow et al., 1983], there will be a bias removal procedure, based on the Kalman filter concept [Ahnert et al., 1986]. The procedure will attempt to remove overall bias often present in the radar data, based on limited number of rain gage observations. If such a procedure does not precede the merging step, or does not work properly because of a lack of adequate information, the bias has to be removed by the cokriging algorithm.

First, let us consider the no-bias case. Two levels of $L_R^*$ were considered: $L_R^* = 0.03$, a rather high noise, and $L_R^* = 0.01$, a medium-to-low noise. The correlation distance $\rho_s$ of the field was $8, 20,$ and $40$ km. The number of rain gages ranged from 50 to 200. The 50-gage case corresponds to approximately 1 gage per 2500 km$^2$ and the 200-gage case to 1 gage per 600 km$^2$. In order to study the selected statistics in the $\beta_R, \beta_G$ parameter space, 100 runs were made for each noise parameter combination. The statistics are correlation coefficient with the original field, the variance of residuals, and the mean square error for various averaging areas. Figures (4-6) are examples of the plots of the correlation coefficient, the variance of residuals, and the mean square error surfaces, respectively. The shaded area corresponds to a region where radar alone did better than the merging procedure. For case presented here, rain gage analysis based on kriging 50 gages was worse than the analysis based on radar data. The cross-hatched areas correspond to those combinations of $\beta_R$ and $\beta_G$ which result in a worse performance than the gage data analysis. The shape of the surfaces is very regular with a flat optimum vicinity, which means that very precise location of the optimum combination of $\beta_G$ and $\beta_R$ is not needed. Also, note that the optima for all three criteria have approximately the same location. This is probably due to the quadratic character of all the selected statistics. The method seems to be more sensitive to proper specification of $\beta_G$ than $\beta_R$. The location of optimum moves to higher values of $\beta_R$ with noise $L_R^*$ decreased and moves to higher values of $\beta_G$ with increased density of the rain gage network. This is obviously an expected behavior. Figure 7 shows the mean square error plot as a function of the number of rain gages for $L_R^* = 0.03$ and $\rho_s = 20$ km.

In the case of the biased radar field one would like to distinguish between overestimation and underestimation of the rainfall field by radar. An underestimated radar-rainfall field was generated by requiring the mean of the radar field to be half of the original field mean. It was found that in such a case, for all the combinations of other parameters specified, the model was practically insensitive to the choice of $\beta_R$ and $\beta_G$ (in terms of selected criteria). The bias was effectively removed for any number of rain gages in the range 50–200, but the improvement offered by the merging procedure over rain...

\[ \text{LEGEND} \]
\[ \square - \text{RADAR} \]
\[ \triangledown - \text{gages} \]
\[ \bullet - \text{merged} \]

Fig. 7. Mean square errors as a function of rain gage network density. Parameters are $L_R^* = 0.03$, $\rho_s = 20$ km, no bias, $\beta_R = 0.45$, and $\beta_G = 0.35$. It was found that the variability of the location of the optimal set of $\beta_R$ and $\beta_G$ was negligible compared to the effects of other noise parameters, and therefore only one realization was used for all other runs.
gage data only was marginal (≤5%). Whether this is a consistent result remains to be seen until more data fields are investigated.

The second situation, with bias present, was the case of the overestimated field, i.e., the radar field mean was generated as 150% of the mean of the original field. Again, the bias was effectively removed even by 50 gages (error in the mean was less than 5%). This time, however, there was a well-defined optimum in the $\beta_w, \beta_o$ parameter space corresponding to approximately $\beta_w = 0.2$ and $\beta_o = 0.3$. The model is more sensitive to the $\beta_w$ parameter than to $\beta_o$. It is interesting that although even 50 gages merged with radar improved the mean square error (see Figure 8) and the estimation variance, 100 gages were required to produce a merged field with a correlation coefficient statistic equal to that of the radar field, and 150 gages were required to produce a merged field with a slightly improved correlation coefficient. The effect of the correlation distance of the radar error field was found to be minimal in the investigated range of 8–20 km. Less-correlated noise leads to some (≤3%) degradation of performance statistics and slight increase of $\beta_w$ values.

Two general results were evident. First, the maximum square error in the radar field (which was in all cases the more noisy field), was always reduced to a level limited by the accuracy of the rain gage data. Second, estimates for larger areas showed reduced average error characteristics. This reduction of errors was again limited by the accuracy of the rain gage field.

It should also be pointed out that although the noise ($L*_{\beta}$) was set to the same value of 0.03 for both bias and no-bias cases, the "amount" of random noise was really higher in the no-bias case (in which all the noise expressed in terms of error characteristics can be attributed to $L*_{\beta}$), while in the bias cases (especially in the case of overestimation) the bias is responsible for high mean square values, but the correlation of the radar and original fields is still good. Thus the improvement of the mean square error is the most that should be expected from the merging procedure.

**Conclusions**

A method for merging radar-rainfall and rain gage data was presented. The method was tested via a numerical experiment, with error fields of both sensors being modeled. The results presented represent a preliminary testing phase and are limited to daily rainfall data. Ongoing testing of the described technique, prior to its operational implementation by the National Weather Service, proceeds along two paths. One is based on the methodology described here and will be followed by similar analyses for more daily fields and also for 6-hourly and 1-hourly data fields. The second (semioperational test) is based on real-time data from the Oklahoma City radar and Tulsa River Forecast Center rain gage data. The comparison is based on hydrographs resulting from mean areal precipitation estimated from rain gages only (as is currently being done in the operational environment) and via the merging procedure for selected basins. For more details on this ap-
proach and some preliminary results, see Krajewski and Ahnert [1986]. As the results presented in this paper suggest, the best configuration of a precipitation-processing system, using data from radar and a network of rain gauges, includes a separate bias removal procedure, so that an unbiased radar-rainfall field enters the merging step. Then, if the bias is effectively removed and the noise in the radar field is low, the merging will not substantially alter the rainfall field. If, however, the noise is high, then a substantial improvement can be expected.

As far as the specification of the values of the parameters $\beta_R$ and $\beta_D$ is concerned, the preliminary results show that the robust region is somewhere in the range of 0.2–0.4 for both parameters. The robustness of these results needs to be further investigated. Also, for data collected at other than daily intervals (hourly, 6-hourly, etc.), these results may not be valid. It is expected that the performance of the presented method, relative to rain gage data analysis, should be better for hourly data.

The described approach, since geared toward an operational environment with its computational time constraints, presents some compromises between mathematical and physical appropriateness and practical efficiency. These are manifested by using ordinary kriging versus the way of the previously mentioned methods of universal and disjunctive kriging, parameter estimation, and also, now from the physical point of view, a lack of accounting for temporal correlation of the rainfall process. The practical consequences of these compromises are being investigated and will be reported.

Acknowledgments. The study reported here began while the author was a National Research Council and National Oceanic and Atmospheric Administration Research Associate. The sponsorship of NRC and NOAA is gratefully acknowledged. The author would also like to thank Michael Hudlow for guidance, encouragement, and many helpful discussions, and Dr. Konstantine Georgakakos for many invaluable comments and suggestions, and friendship. The editorial and graphical work of Ruth Ripkin, Lianne Iseley, and my wife, Diana Thrift, are very much appreciated.

REFERENCES


W. F. Krajewski, Hydrologic Research Laboratory, National Weather Service, National Oceanic and Atmospheric Administration, 8060 13th Street, Room 330, Silver Spring, MD 20910.

(Received July 10, 1986; revised March 31, 1987; accepted April 1, 1987.)