WATER RESOURCES RESEARCH, VOL. 26, NO. 5, PAGES 915–924, MAY 1990

Stochastic Interpolation of Rainfall Data From Rain Gages and Radar Using Cokriging
2. Results

DONG-JUN SEO[1]

Utah Water Research Laboratory, Utah State University, Logan

WITOLD F. KRAJEWSKI

Institute of Hydraulic Research, University of Iowa, Iowa City

ALI AZIMI-ZONOOZ AND DAVID S. BOWLES

Utah Water Research Laboratory, Utah State University, Logan

Various estimation procedures using ordinary, universal, and disjunctive cokriging are evaluated in merging rain gage measurements and radar rainfall data. The estimation procedures and the simulation experiments were described in part 1 (Seo et al., this issue) of this two-part work. In this part, the experiments are described in detail. An objective comparison scheme, devised to compare a large number of estimators, is also described. The results are presented focusing upon (1) the potential of radar-gage estimation using cokriging over radar-only estimation and gage-only estimation under widely varying conditions of gage network density and the error characteristics of radar rainfall, (2) the potential for using universal or disjunctive cokriging over ordinary cokriging, (3) how the uncertain second-order statistics affect the estimators, due to lack of rain gage measurements, and (4) how the statistical characteristics of ground truth rainfall affect the estimators.

INTRODUCTION

This paper is part 2 of the two-part series. In the first part [*Seo et al.*, 1990], the experimental design for testing several algorithms of optimally merging radar and rain gage data is described. The testing is based on two numerical experiments. Both experiments use computer-generated fields of radar and rain gage observations of an original rainfall field. In experiment I the original field is a radar rainfall field from the GATE experiment. In experiment II the original field is a realization of a space-time rainfall model.

The organization of this paper is as follows: details of the simulation used in experiment I, the procedure used in comparing the estimators, results from experiment I, details of the simulation used in experiment II, results from experiment II, and conclusions and future research recommendations. To follow the material presented in this paper, it is essential to first read part 1 [*Seo et al.*, 1990].

EXPERIMENT I SIMULATION

The GATE radar rainfall data selected were eight hourly fields from day 245, phase 2 of the GATE experiment [*Hudlow and Patterson*, 1979]. The data covered the developing, maturing, and dissipating stages of a tropical convective storm. Eight hours were selected (hours 3, 6, 12, 15, 18, 21, and 24) from the 24-hour period. For each hour, a 200 × 200 km area showing active rainfall was chosen from the radar umbrella of radius of 200 km. The reader is referred to *Seo* [1988] for the rainfall maps and the empirical semivariograms. Temporal variation of the sample mean, variance, and skewness coefficient of the chosen rainfall fields are shown in Figure 1.

For each hourly rainfall field, assumed as the ground truth rainfall field, a total of 24 combinations of the radar rainfall field and the point gage rainfall field were generated in a single simulation run. The values of the parameters used for the radar rainfall generator were (1) variance of the logarithmic ratio of the radar-rainfall to the ground-truth rainfall (denoted "SIGMAR") (values of 0.005 and 0.02 were used representing small and large measurement error in the radar rainfall field, respectively), (2) bias in the mean of the radar rainfall field (denoted "BIAS") (values of 1 and 2 were used representing no bias and an overestimation bias of 100 percent, respectively), (3) correlation distance of the random noise in the radar rainfall field (denoted "CORDIS") (values of 8 and 16 km were used representing, relatively speaking, short and long correlation distances, respectively). These values of the parameters result in eight combinations. In addition, three rain gage network densities of 32, 160, and 286 gages over the 200 × 200 km area were used to generate the point gage rainfall fields. Gages were assumed to be randomly scattered. This is not an unrealistic assumption in that towns (which are likely to have gages) are known to be randomly scattered [*Karr*, 1986]. The number of rain gages is denoted "NG." The first density approximately corresponds to the average rain gage network density over the continental United States (one gage per 1000–2000 km^2 [*Wilson and*

[1]Now at Hydrologic Research Laboratory, National Weather Service, Silver Spring, Maryland.

Paper number 89WR02992.
0043-1397/90/89WR-02992$05.00

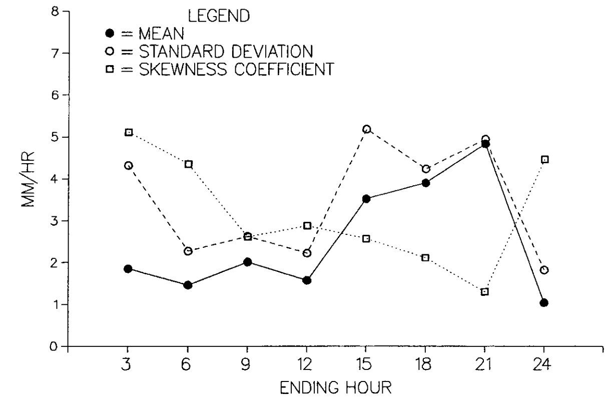

Fig. 1. Temporal variation of sample mean, variance, and skewness coefficient of the ground truth rainfall.

TABLE 1. Example of Pairwise Comparison Among the Estimators in Experiment II

	BR	RA	OB	O1	O2	DB	D1	D2
BR	...	Y						
RA		...						
OB		Y	...					
O1		T		...				
O2	Y	Y			...			
DB		Y				...		
D1	Y	Y				Y	...	
D2	Y	Y	Y	Y		Y	Y	...

For storm type I under the percent mean error criterion under the global pooling. "Y" denotes that estimates from the estimator in the column were better than the estimates from the estimator in the row. For example, the BR estimates were better than the RA estimates only, and the O2 estimates were better than the BR estimates and the RA estimates only.

Brandes, 1979]) and is referred to as the low gage density. The second is approximately the gage density above which radar-gage estimates are found to be no better than gage-only estimates for Illinois convective storms [*Hildebrand et al.*, 1979]; this is referred to as the medium gage density. The third is referred to as the high gage density.

For each assumed ground truth rainfall field, a single simulation run involved the following steps:

1. Generate a radar rainfall field (using one of the eight combinations of parameter values).
2. Generate a point gage rainfall field (using one of the three rain gage densities).
3. Perform rainfall estimation.
4. Compute estimation error statistics.
5. Go to step 2 until all three gage densities are used.
6. Go to step 1 until all eight combinations are used.

Following this procedure, there were 24 combinations of the radar rainfall fields and the point gage rainfall fields in a single simulation run. For each combination, both gage-only and radar-gage estimation were performed.

Ideally, we would have repeated the above steps to obtain enough simulation runs to achieve a clear convergence in the estimation error statistics. Due to the heavy computational requirements, however, a compromise had to be made. To check the convergence, a total of 20 simulation runs were made using the assumed ground truth field of hour 3. It has the largest coefficient of variation, and the estimation error statistics had the slowest rate of convergence. After 10 simulation runs, the estimation error statistics were always observed to be converging. Therefore the number of simulation runs was set at 10 for the rest of the experiment. For further details on the convergence characteristics, the reader is referred to *Seo* [1988].

If the estimation in step 3 had always been successful, the total number of estimated rainfall fields from 10 simulation runs for each estimator would have been 240 (or 480 for hour 3 from 20 simulation runs), excluding the radar-only estimation. However, the estimation was not always successful because (1) the variogram(s) could not be fitted due to lack of raingage data, (2) the generalized (cross) covariance(s) did not converge and thus could not be estimated, (3) the semipositive definiteness conditions were not met, or (4) the conditionally semipositive definiteness conditions were not met. For these reasons, the total number of estimated rainfall fields varied from one estimator to another and from one assumed ground truth field to another. The minimum number of estimated rainfall fields was set at 55. Whenever the number of estimated rainfall fields fell below this minimum, the corresponding estimator was excluded from the comparison. The average numbers of estimated rainfall fields obtained from a single estimator were 274, 131, 145, 103, 173, 187, 163, and 141 for hours 3, 6, 9, 12, 15, 18, 21, and 24, respectively.

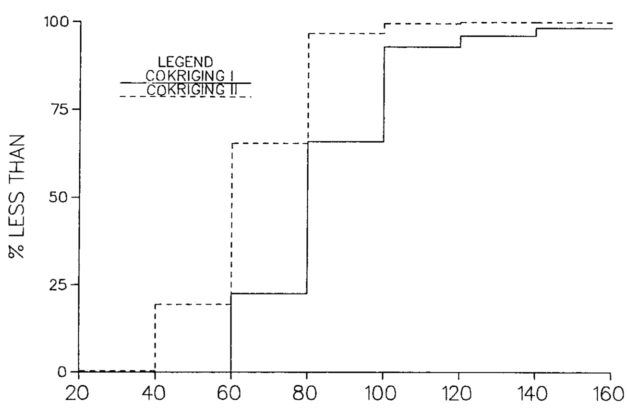

Fig. 2. An example showing that the estimates from ordinary cokriging II are better than the estimates from ordinary cokriging I under the PRMSE criterion (arbitrarily scaled).

COMPARISON OF THE ESTIMATORS

The following abbreviations will be used throughout to designate the estimators: BR, the Brandes' method; RA, the radar-only estimation; OB, DB, UB, the gage-only estimation using ordinary, disjunctive, and universal block kriging, respectively; O1, D1, U1, the radar-gage estimation using ordinary, disjunctive, and universal cokriging I, respectively; O2, D2, U2, the radar-gage estimation using ordinary, disjunctive, and universal cokriging II, respectively;

To compare the estimators, we used the following types of pooling of the estimation error statistics:

1. For global pooling, for each hour of the eight hourly ground truth fields, the estimation error statistics were pooled over all the simulation runs made using all 24 combinations. Thus the comparisons made under global

TABLE 2. Mathematical Expressions of the Performance Measures

Performance Measure	Expression	Definitions
Percent mean error	$\mathrm{PME} = \frac{100}{m_x} \frac{1}{n} \sum_{i=1}^{n} (\hat{x}_i - x_i)$	where m_x is the mean of the ground truth rainfall, n is the total number of data in the ground truth rainfall field, $\hat{x}_i$ is the rainfall estimate at location i, and x_i is the ground truth rainfall at location i.
Percent root mean square error	$\mathrm{PRMSE} = \frac{100}{\sigma_x} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\hat{x}_i - x_i)^2 \right\}^{1/2}$	where σ_x is the standard deviation of the ground truth rainfall.
Standardized mean square error	$\mathrm{SMSE} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{x}_i - x_i)^2}{\sigma_i^2}$	where σ_i^2 is the kriging or the cokriging variance at location i.
Maximum error of overestimation	$\mathrm{MAXE}_0 = \max_{i=1,\cdots,n} \{\hat{x}_i - x_i\}$	
Maximum error of underestimation	$\mathrm{MAXE}_u = \min_{i=1,\cdots,n} \{\hat{x}_i - x_i\}$	
Relative percent mean error (at level j)	$\mathrm{RPME} = 100 \frac{\sum_{i=1}^{n_j} (\hat{x}_{ji} - x_{ji})}{\sum_{i=1}^{n_j} x_{ji}} \ (\geq -100)$	where x_{ji} is the ground-truth rainfall at location i satisfying $(j-1)\sigma_x < x_{ji} \leq j\sigma_x$, $\hat{x}_{ji}$ is the coinciding estimated rainfall, and n_j is the number of ground truth rainfall data satisfying the above set of inequalities.

pooling reflect the overall relative performance among the estimators. Whenever a result obtained from global pooling is given, it will be indicated by the term "global(ly)."

2. For marginal pooling, for each of the four parameters used in generating the radar-rainfall field and the point gage rainfall field, for each value of the parameter, the estimation error statistics were pooled over all the simulation runs made, only when the estimation error statistics were obtained using the radar rainfall field and the point gage rainfall field generated from that particular value of the particular parameter. Therefore there were two pools each for SIGMAR, BIAS, and CORDIS, and three pools for the gage density. In this way, the effect of each parameter could be evaluated. Whenever a result obtained from marginal pooling is given, it will be indicated by the term "marginal(ly)."

Once the estimation error statistics were pooled for all the estimators for each performance measure, for each hour, the estimators were ranked under each pooling type, as follows:

1. The empirical cumulative distribution of the estimation error statistics were constructed for all estimators. If the empirical cumulative distribution for estimator A lay above that for estimator B, the estimates from estimator A were considered to be better than the estimates from estimator B. An example is shown in Figure 2 involving ordinary cokriging I and ordinary cokriging II under the percent root mean square error (PRMSE) criterion. This approach has a definite advantage over using, for example, only the mean of the estimation error statistics as it also accounts for the variability of the estimation error statistics.

2. The step 1 results of the pairwise comparisons among the estimates were tabulated as, for example, in Table 1 for experiment II.

3. Estimator A is considered to have performed better than estimator B, if, in step 2, the number of estimators, against which the estimates from estimator A were better, is larger than the number of estimators against which the estimates from estimator B were better. For example, the final ranking obtained from Table 1 may be written in the following form of groups of equal performance (from the best to the worst group): D2-D1-O2-BR, OB, O1, DB-RA. The rationale behind the above criterion is that every estimator has a certain merit of its own. No disagreement between step 1 and step 3 can arise: if estimator A is ranked higher than estimator B in step 3, the estimates from estimator A would never be worse than the estimates from estimator B in step 1.

A similar ranking procedure was used for percent mean error (PME), standardized mean square error (SMSE), maximum errors (MAXE), and relative percent mean error (RPME). It differs only in that the absolute values of the estimation error statistics (absolute values were taken after the log transform for SMSE) were used to construct the empirical cumulative distribution. By using the absolute values, no preference was given to either overestimation or underestimation. Table 2 shows the mathematical expressions of the performance measures.

The performance measures were computed for only the rain area in the ground truth field. Thus they measure average performance given that it is actually raining (i.e., when the rain area is perfectly known). This procedure was selected for two reasons: the radar rainfall generator and the rain gage rainfall generator were already assumed to be perfect sensors in identifying the presence or absence of rain [*Seo et al.*, 1990], and the amount of computation was substantially reduced since estimation was performed for only those points or blocks inside the rain area. It is important to note that this procedure is more favorable to gage-only estimation than to radar-gage estimation. In real world situations, it is generally unknown as to whether or not it is actually raining at all points in the area of interest. Since rain gage data generally contain much less information on rain area than radar data, the assumption of perfectly known rain area should benefit gage-only estimation more than radar-gage estimation.

EXPERIMENT I RESULTS

The results were compiled focusing on the following issues:

1. Are the radar-gage estimates better than both the radar-only estimates and the gage-only estimates? Particular attention was given to the following situations: For the low gage density (i.e., with the average gage density over the continental United States), are the radar-gage estimates better? For the case of no bias in the mean of the radar rainfall field (i.e., when the radar-only estimates are likely to be of good quality), are the radar-gage estimates better than the radar-only estimates?
2. Does any particular cokriging estimator perform better than the others?
3. How much improvement do cokriging II estimators provide over cokriging I estimators?
4. How do the characteristics of the ground truth rainfall affect the estimators?

In comparing a pair of estimators, an ideal situation would be when one estimator performed better (i.e., was ranked higher) than the other at all stages of the storm development. In such a case, a clear conclusion can be drawn. In many cases, however, one estimator performed better than the other at one hour but did worse at another hour. For this reason, the following definitions were made:

1. Estimator A is "superior" to estimator B if A performed better than B more times than A did worse than B.
2. Estimator A is "clearly superior" to estimator B, if A performed never worse than B, and if A did better than B at least once.
3. Estimator A is "tied" with estimator B if A performed better than B as many times as A did worse than B, or if A and B did equally well at all times.

The distinction between "superiority" and "clear superiority" is an important one. Only "clear superiority" implies that an estimator never performed worse than another at all stages of storm development.

In the next subsections, we present the findings for each performance measure. Most of the figures are given in the form of a modified box-whisker plot. The upper and lower ends of the whisker represent 95 and 5% levels, respectively. The upper and the lower sides of the box represent 75 and 25% levels, respectively. The median is shown by the horizontal bar inside of the box. The arithmetic mean is shown as the horizontal bar which is wider than the box. It is not possible to present all the relevant figures in a journal article. The reader is reminded that most of the figures in this paper are only examples and therefore do not support all of our conclusions. More detailed results are given by *Seo* [1988] and *Azimi-Zonooz* [1989].

Comparisons involving UB, U1, and U2 are given separately since their performance depends on the presence or absence of a nonconstant trend. For brevity, only OB was used to represent the gage-only estimation. BR is not included in the general comparison as its poor performance was evident. Thus the estimators included in the general comparison below were RA, OB, O1, O2, D1, and D2 only. Since O2 and D2 made use of near perfectly known variograms and error-free spatially averaged gage rainfall data (for D2, also assumed to be constant, perfectly known mean), it was expected that O2 performs better than OB and O1, and that D2 performs better than OB, O1, O2, and D1. According

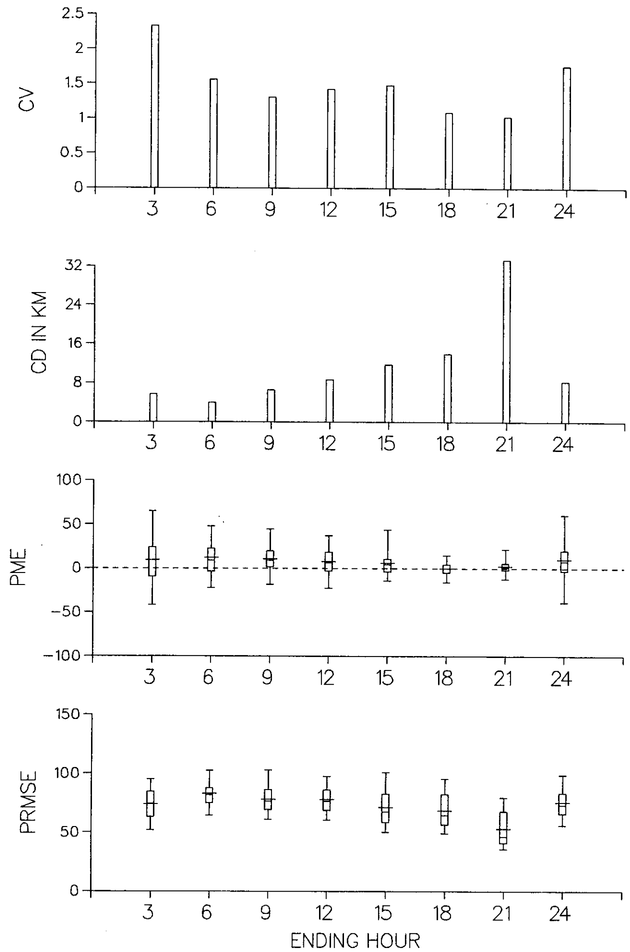

Fig. 3. From the top to the bottom, coefficient of variation of the ground truth rainfall, correlation distance of the ground truth rainfall, PME of the O2 estimates, and PRMSE of the O2 estimates.

to this reasoning, certain comparisons are not mentioned if the relative performance among the estimators were as expected.

Percent Mean Error

The PME measures how well the unbiasedness property of the kriging and the cokriging estimators was realized. Visual inspection showed that the unbiasedness of the kriging and cokriging estimates was most noticeably affected by the coefficient of variation of the ground truth rainfall. The smaller the coefficient of variation, the smaller the variance of the PME. An example involving O2 is shown in Figure 3. The pattern was similar for the other kriging and cokriging estimators.

Globally, the O1 estimates were as unbiased as the OB estimates, because O1 and OB were tied in rank. The relative unbiasedness of the O1 estimates, however, showed a strong dependence on the error characteristics of the radar rainfall field: O1 was superior to OB only when there was no bias or when CORDIS was long, whereas O1 was inferior to OB when SIGMAR was high, BIAS was high, or CORDIS was short. When the gage density was low or medium, O1 and OB were tied in rank. When the gage density was high,

TABLE 3. Percentages of the Times UB, U1, and U2 Identified the Order of the Trend to be Greater Than Zero

	Hour							
	3	6	9	12	15	18	21	24
UB	5	6	12	12	7	10	13	10
U1	10	11	25	37	97	55	96	100*
U2	0	0	10	14	100*	44	100*	†

For example, at hour 3, UB identified the order of the trend in the point gage rainfall field to be nonconstant 5% of the total number of estimation completed.

*The number of estimated rainfall fields obtained was insufficient for the estimator to be included in the comparison.

†Not a single estimated rainfall field was obtained, and thus the estimator was excluded from the comparison.

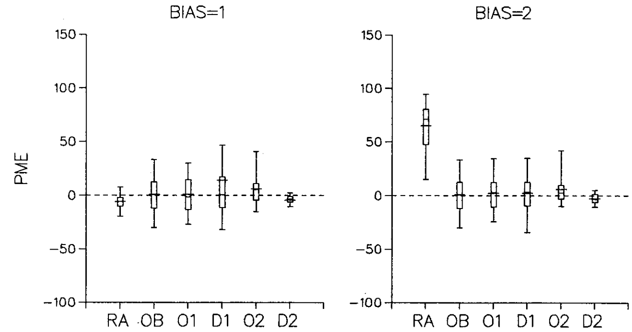

Fig. 5. An example showing the relative unbiasedness of the estimates at different values of bias (hour 15).

however, the unbiasedness was better achieved by the gage-only estimation, as O1 was clearly inferior to OB.

Globally, D1 estimates were more biased than O1 estimates, as D1 was inferior to O1. This was not totally unexpected. Disjunctive kriging or cokriging, unlike its ordinary or universal counterpart, makes use of the sample mean, which is assumed perfectly known. The sample mean used in DB or D1, however, is likely to be in error due to lack of data. The D1 estimates are thus likely to be biased since D1 takes the error-prone sample mean as its estimate whenever the point, for which the estimate is desired, lies outside the zone of influence of surrounding data. Marginally, D1 was superior to O1 only when the gage density was low.

Globally, D2 was superior to O2. At hour 21, the D2 estimates were more biased than the O2 estimates. It was unusual in that the D2 estimates were fully expected to be the least biased since D2 made use of the perfectly known (assuming that it is constant) mean of the ground truth rainfall field. Under the PME criterion, it is not expected to achieve the level of performance of D2 in reality. The structure identification used in universal cokriging indicated the presence of a nonconstant trend at hour 21 (see Table 3). It was suspected that the unusual biasedness of the D2 estimates at hour 21 might be due to the nonhomogeneity in the mean. A possible numerical problem could not, however, be ruled out. Marginally, O2 was comparable to D2 only when the gage density was high, as they were tied in rank.

Globally, O1 and D1 were clearly superior and superior to RA, respectively. When there was no bias, however, only the D2 estimates were as unbiased as the radar estimates, as D2 and RA were tied in rank. When the gage density was low, both O1 and D1 were inferior to RA, but O2 and D2 were superior and clearly superior to RA, respectively. When the gage density was medium or high, both O1 and D1 were clearly superior to RA. This indicates that, for example, for the ordinary cokriging estimates to be less biased than the radar estimates, either of the following two may be required: (1) with the average gage network density over the continental United States, at least the covariances and the cross covariance must be known near perfectly; and (2) the gage network density must be higher than the average gage density over the continental United States, when the covariances and the cross covariance are estimated, with large uncertainty, only from the presently available data.

Figures 4 and 5 show examples of the relative unbiasedness among the estimates at various gage densities and at different values of BIAS, respectively.

The dependence of the unbiasedness of the kriging and the cokriging estimates on the characteristics of the ground-truth rainfall suggests that an estimator may be more useful at one stage of storm development than at others. For example, when the gage density was low, the O1 estimates were less biased than the radar estimates only at hours 15, 18, and 21. These three hours constitute the mature stage of the storm. The corresponding ground truth rainfall fields are characterized by a lower coefficient of variation and a longer correlation distance. When the gage density was low, the O2 estimates were still more biased than the radar estimates at

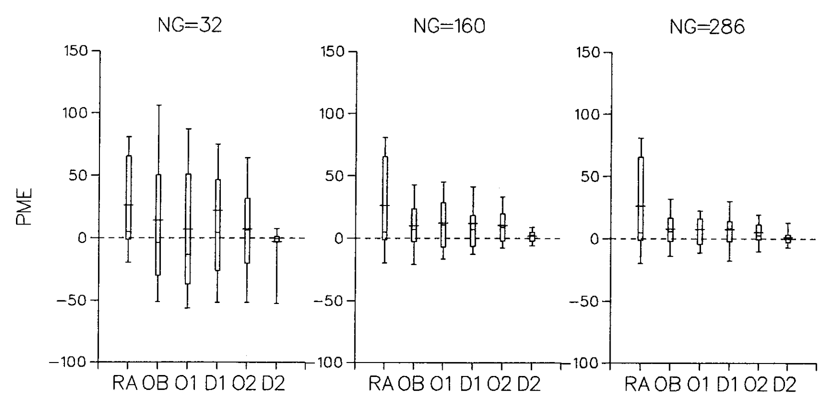

Fig. 4. An example showing the relative unbiasedness of the estimates at various gage network densities (hour 12).

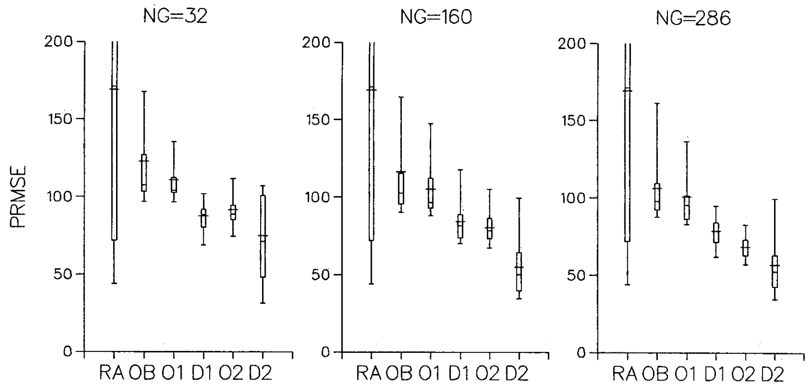

Fig. 6. An example showing the relative performance among the estimators under the PRMSE criterion at various gage network densities (hour 9).

hours 3, 6, and 24. The first two hours constitute the early developing stage of the storm, and hour 24 corresponds to the dissipating stage. The three corresponding ground truth rainfall fields are characterized by a higher coefficient of variation and a shorter correlation distance.

Percent Root Mean Square Error

The PRMSE is a measure of performance in the minimum-error-variance sense. The smaller the PRMSE, the more accurate, on the average, the estimate for a 4 × 4 km block and thus the more accurate the mean areal precipitation (MAP) estimate over an arbitrary area. Visual inspection showed that the PRMSE of the kriging and cokriging estimators was affected most noticeably by the correlation distance of the ground truth rainfall. The longer the correlation distance, the smaller the PRMSE (see Figures 3 and 8).

Better performances of O1 over OB, and D1 over O1 were clearly seen. Globally, O1 was clearly superior to OB. Throughout the marginal comparison, O1 was never inferior to OB. Globally, as well as throughout the marginal comparison, O1 was clearly superior to RA, and thus we focused on the comparison among the cokriging estimators.

Globally, D1 was clearly superior to O1. Throughout the marginal comparison, D1 was never inferior to O1. Globally, D1 was clearly inferior to O2. Marginally, D1 was superior to O2 only when the gage density was low. D1 performed particularly well when the gage density was low. Recall that when the gage density was low, D1 was also superior to O1 under the PME criterion. Under the PRMSE criterion, D1 was superior not only to O1 but also to O2 when the gage density was low.

Globally, D2 was superior to O2. At hours 3 and 21, D2 performed worse than O2. Recall that, at hour 21, D2 also performed poorer than O2 under the PME criterion. The relative performance of D2 showed a strong dependence on the error characteristics of the radar rainfall field: D2 was inferior to O2 when SIGMAR was high, BIAS was high, or CORDIS was long. When the gage density was low or medium, D2 was superior to O2. When the gage density was high, however, D2 was inferior to O2. When the gage density is high enough, D2 is seen to offer little advantage over O2. Recall that when the gage density was high, D2 was only tied in rank with O2, even under the PME criterion.

Figures 6 and 7 show examples of the relative performance among the estimators at various gage densities and at different values of BIAS, respectively. Figure 8 shows the mean PRMSEs of the estimates at all hours under the global pooling. The improvement by O1 over OB is small. D1 improves over both OB and O1 substantially and reaches almost the performance level of O2. Unusually poor performance by D2 at hour 21 seems to be due to a numerical problem rather than nonhomogeneity.

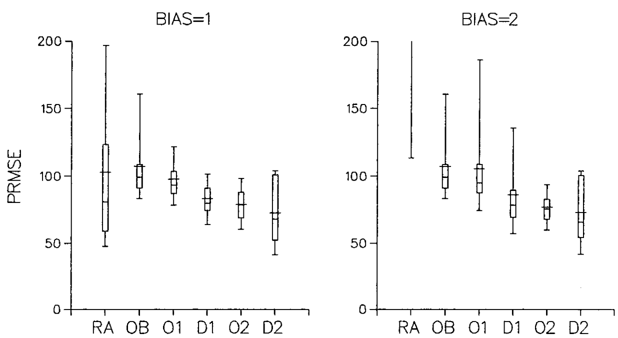

Fig. 7. An example showing the relative performance among the estimators under the PRMSE criterion at different values of bias (hour 12).

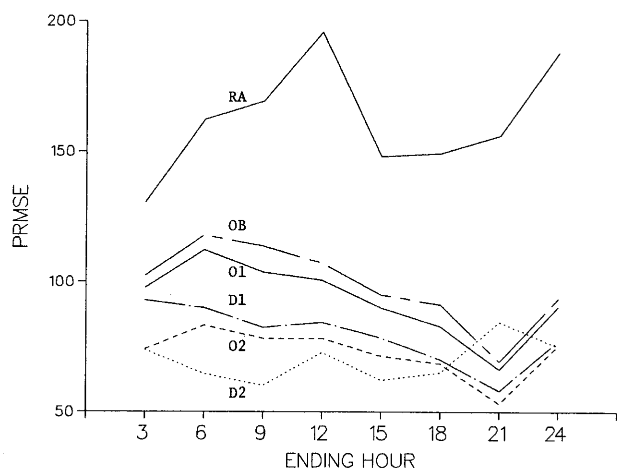

Fig. 8. Mean PRMSEs of the estimates at all hours under the global pooling.

Standardized Mean Square Error

The SMSE measures accuracy of the kriging or cokriging variances. A perfect estimator will yield an SMSE of 1. The SMSE also indicates how closely the assumptions required for kriging or cokriging are met. In this subsection, only the results from the global pooling are given.

A rather uniform pattern of relative performance was found over the 8 hours among all the kriging and the cokriging estimators, including UB, U1, and U2. It may be summarized by the following ranking (from the best group to the worst group): O2-D2-UB-OB-DB-U1-U2-O1-D1. An example is shown in Figure 9. Only O2 and D2 were seen to have met the assumptions required for kriging or cokriging. At hour 21, unusually poor SMSEs were observed for D2. Both O1 and D1 tended to grossly underestimate the estimation variances more than OB and DB. This suggests that the uncertainties associated with both the semivariogram of the spatially averaged gage rainfall field, and the cross variogram deteriorated further for O1 and D1 than the uncertainty associated with the semivariogram of the point gage rainfall field alone did for OB and DB.

Maximum Errors

To measure the extreme behavior of the estimators, the maximum errors were examined. In this subsection, only the results from the global pooling are given. As for the maximum error of overestimation, the following ranking summarizes the relative performance among the estimators (from the best group to the worst group): D2-D1-O2-O1-OB-RA. Similarly, for the maximum error of underestimation: RA-O2, D2-O1-D1-OB. Better performances by the cokriging II estimators over the cokriging I estimators and the cokriging I estimators over OB are readily seen. Figure 10 gives an example of the relative performance among estimators.

Relative Percent Mean Error

For flood forecasting, one of the most important aspects of rainfall estimation is the ability to locate areas of intense rainfall and predict high rainfall depths. To measure this ability and to examine the characteristics of the estimators, we decomposed the rainfall occurring area into six subareas of equal rainfall depth levels. Level 1 denotes the rainfall depth greater than zero and less than one standard deviation of the ground truth rainfall, level 2 denotes the rainfall depth

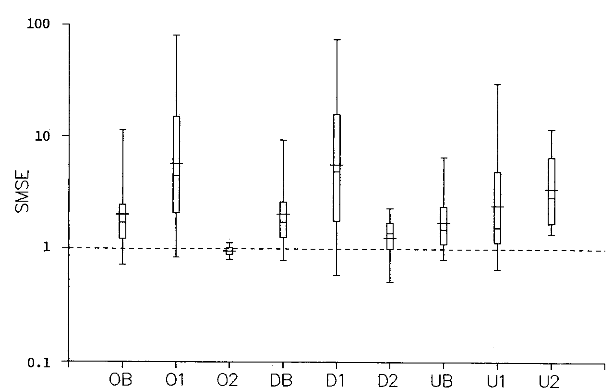

Fig. 9. An example showing the SMSEs of the estimates under the global pooling (hour 18).

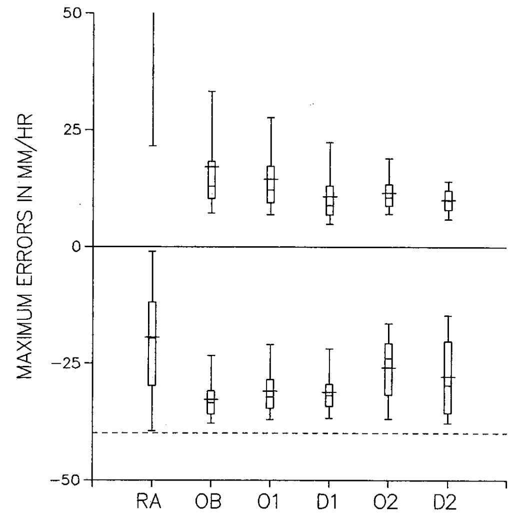

Fig. 10. An example showing the maximum errors of overestimation (top) and the maximum errors of underestimation (bottom). The dashed line corresponds to the maximum rainfall depth observed in the ground truth rainfall field (hour 18), and thus the maximum errors of underestimation are bounded by it. Note that the larger the maximum error of underestimation, the better.

between one standard deviation and two standard deviations, etc. For each area, the relative percent mean error was computed. For simplicity in this subsection, no distinction was made between "superiority" and "clear superiority." Thus, in this discussion, the term "superiority" implies either "superiority" or "clear superiority" as defined previously. Also, only the results from global pooling are given.

In general, all the kriging and cokriging estimators consistently overestimated the lower rainfall depth and consistently underestimated higher rainfall depth. The overall performance of O1 was better than OB: O1 was superior to OB at all levels except level 3. The overall performance of D1 was better than O1 (D1 was superior to O1 at all levels except level 2). Surprisingly, O2 was superior to D2 at all levels except level 1. Visual inspection showed that the inferiority of D2 to O2 at higher levels was partially due to higher variances of the RPMEs of the D2 estimates. When only the mean of the RPMEs were considered, D2 performed better than O1 at levels 1–3 but still did worse at levels 5 and 6.

In general, the kriging and cokriging estimators consistently performed better than the radar-only estimator at higher levels. At lower levels, however, the opposite was observed: RA was never inferior to any of the kriging or the cokriging estimators at levels 1 and 2. Figures 11 and 12 show examples of the relative performance among the estimators under global and marginal pooling, respectively.

It was again seen that the characteristics of the ground truth rainfall field affect the estimators. To illustrate this, a portion of the table used in the marginal comparison between D1 and RA at level 2 is shown in Table 4. When the gage density was low, D1 performed better than RA only at hour 21 where the coefficient of variation was the smallest and correlation distance the longest. When the gage density was medium, D1 performed better than RA only at the mature stage. When the gage density was high, D1 performed worse than RA only at the early developing and dissipating stages.

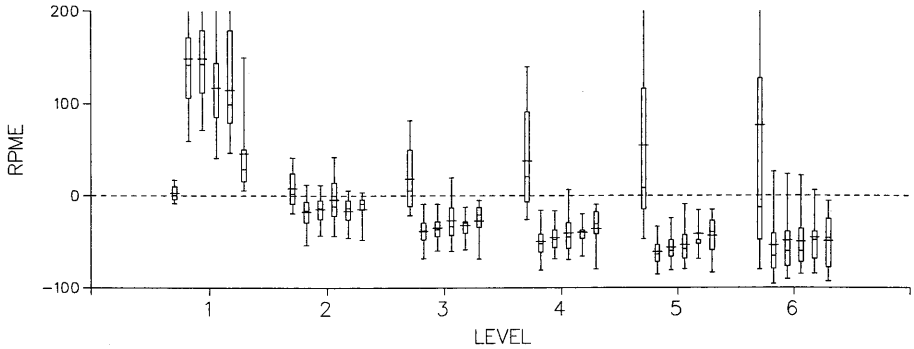

Fig. 11. An example showing the relative performance among the estimators under the RPME criterion (hour 9). At each level, the six box-whisker plots correspond to RA, OB, O1, D1, O2, and D2 from the left to the right.

A similar pattern was observed for the other kriging and cokriging estimators.

EXPERIMENT II SIMULATION

In this experiment, assumed ground truth rainfall fields were generated using the space-time rainfall model by *Valdes et al.* [1985]. Experiment II was structured very similarly to experiment I; therefore we describe only the steps that were different from those in experiment I. Unlike experiment I, the gage locations were fixed throughout the simulation. A total of 286 randomly scattered gage locations were generated only once over the 200 × 200 km area. The first 32 gage locations represented the low gage density. The first 32 and next 128 gage locations represented the medium gage density, and all 286 gage locations represented the high gage density. A single simulation run consisted of the following steps:

1. Generate an assumed ground truth field (one of the three storm types).
2. Generate a radar rainfall field (using one of the eight combinations of the values of the parameters).
3. Select a raingage network (one of the three networks).
4. Perform rainfall estimation.
5. Compute the estimation error statistics.
6. Go to step 3 until all three networks are used.
7. Go to step 2 until all eight combinations are used.

For all three storm types, only the rainfall fields corresponding to the first hour of the storm development were used. The reader is referred to *Seo* [1988] for the rainfall maps and the empirical semivariograms. Unlike the GATE data, the rainfall fields from the space-time rainfall model have the largest mean and variance and the longest correlation distance at the first hour of the storm development for all three storm types. The storms die off quickly, and too small a number of data points are available at later hours. A total to 10 simulation runs were made for each storm type.

In experiment I, the gage network configurations were allowed to vary from one radar rainfall field to another since we had only one realization of the assumed ground truth field (i.e., the hourly GATE data). In experiment II, however, a multiple number of realizations of the assumed ground truth field was available from the rainfall model; thus the gage network configurations were left unchanged as would normally be the case in a real world rain gage network.

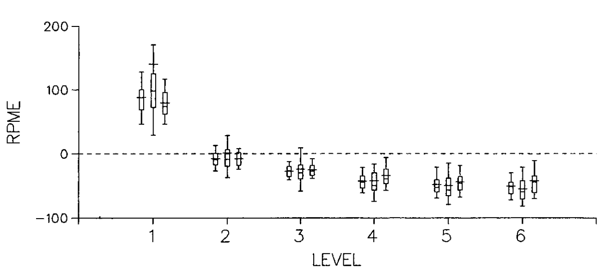

Fig. 12. RPMEs of the O1, D1, and U1 estimates when the gage density was high at hour 15. At each level, the three box-whisker plots correspond to O1, D1, and U1 from the left to the right.

EXPERIMENT II RESULTS

The average sample statistics of the assumed ground truth fields are shown in Figure 13 for each storm type. Probably, the two biggest differences between experiment I and experiment II were the assumed ground truth fields in experiment II were statistically more homogenous and had generally longer correlation distance. The results from experiment II probably present a more favorable picture for kriging and cokriging estimators than under real world situations. In particular, disjunctive kriging and cokriging estimators would be expected to excel under the structure of experiment II. Disjunctive kriging or cokriging, in the Gaussian domain, are equivalent to optimal linear estimation [*Journel and Huijbregts*, 1978] and should benefit, to a greater extent, from homogeneity in the mean.

The results given below were compiled following the same procedure used in experiment I, treating the storm types as the hours in experiment I. They thus reflect the overall performance of the estimators for all three storm types. The estimators included in the comparisons below are RA, OB, O1, O2, D1, and D2.

TABLE 4. Comparison Between D1 and RA at Level 2 Under the RPME Criterion

NG/hour	3	6	9	12	15	18	21	24
32	*	*	*	*	*	*	D1	*
160	*	*	*	†	D1	D1	D1	*
286	*	*	D1	D1	†	D1	D1	*

D1 denotes that D1 performed better than RA.
*D1 performed worse than RA.
†D1 and RA performed about equally.

Percent Mean Error

Globally, the O1 estimates were less biased than the OB estimates, as O1 was clearly superior to OB. As in experiment I, however, the relative unbiasedness of the O1 estimates depended on the error characteristics of the radar rainfall field: O1 was clearly superior to OB when SIGMAR was low or CORDIS was long, whereas O1 was clearly inferior to OB when BIAS was high or CORDIS was short. When the gage density was low, O1 was clearly superior to OB. When the gage density was medium or high, however, O1 was inferior to OB. The D1 estimates were as unbiased as the O1 estimates, as D1 and O1 were tied in rank. Marginally, D1 was clearly inferior to O1 only when gage density was low or high. Globally, D2 was clearly superior to O2. Marginally, D2 was inferior to O2 only when BIAS was high.

Globally, both O1 and D1 were clearly superior to RA. Marginally, both O1 and D1 were clearly inferior to RA only when there was no bias. When there was no bias, only D2 was comparable to RA, as was observed in experiment I.

Percent Root Mean Square Error

In both experiments I and II, better performances of O1 over OB, and D1 over O1 were clearly seen. Globally, O1 was clearly superior to OB. Throughout the marginal comparison, O1 was never inferior to OB. Globally and throughout the marginal comparison, D1 was clearly superior to O1. Globally, D1 was also superior to O2. D1 was clearly inferior to O2 only when the gage density was high.

Globally, both O1 and D1 were clearly superior to RA. When there was no bias, however, O1 was inferior to RA, but both O2 and D1 were clearly superior to RA. Figure 14 shows the mean PRMSEs of the estimates for all three storm types under global pooling. The improvement by O1 over OB is very small. D1 improves substantially over both OB and O1, and also improves over O2 for storm types 1 and 2. The results under the SMSE criterion were very similar to those of experiment I and therefore are not presented here.

Maximum Errors

In this subsection, only the results from the global pooling are given. As for the maximum error of overestimation, the relative performance among the estimators may be summarized by the following ranking (from the best group to the worst group): D2-O2, D1-O1, OB-RA. Similarly for the maximum error of underestimation, the ranking is D2-RA, D1-OB-O1, O2. When only the maximum error of underestimation is considered, the relative performance of the ordinary cokriging estimators was poorer than in experiment I, whereas the relative performance of the disjunctive cokriging estimators was better than in experiment I.

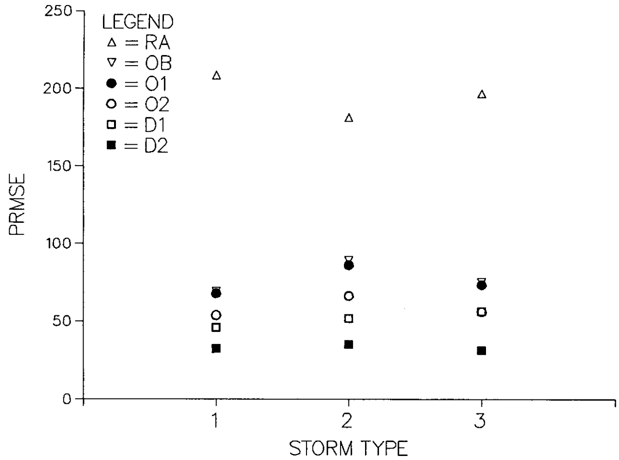

Fig. 14. Mean PRMSEs of the estimates under the global pooling.

Relative Percent Mean Error

As was observed in experiment I, no distinction was made between "superiority" and "clear superiority," and only the results from global pooling are given. The general performance patterns were similar to those in experiment I. The relative performance of O1 was poorer than in experiment I: O1 was superior to OB only at levels 1, 4, and 5. The relative performance of disjunctive cokriging estimators was better than in experiment I: D1 was superior to O1 at all levels, and D2 was never inferior to O2 at any level. All the kriging and cokriging estimators were inferior to RA only at level 1.

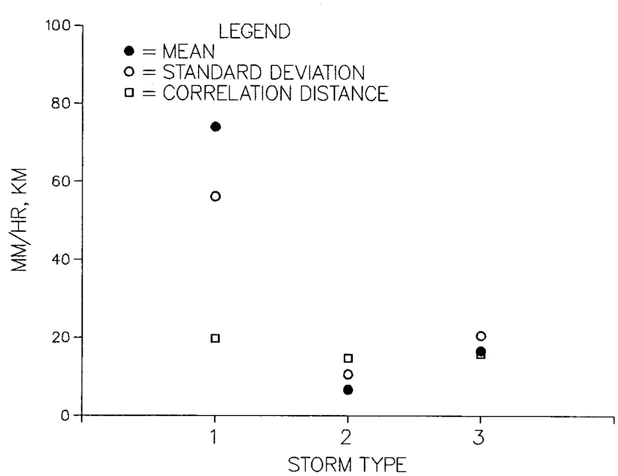

Fig. 13. Theoretical mean, standard deviation, and correlation distance of the ground truth rainfall for each storm type.

CONCLUSIONS AND FUTURE RESEARCH RECOMMENDATIONS

Under various conditions of rain gage network density and error characteristics of radar rainfall data, the radar-gage estimation using ordinary or disjunctive cokriging is shown to consistently provide rainfall estimates that are better, in the minimum-error-variance sense, than the gage-only or radar-only estimates. The improvement from ordinary cokriging is only marginal. Disjunctive cokriging, on the other hand, shows substantial improvement over ordinary cokriging and almost provides the potential level of performance obtainable in the ordinary cokriging framework. When the ground truth field closely satisfies the homogeneity assumptions, the margin of improvement by disjunctive cokriging is greater.

The consistency of the improvement, under various radar rainfall error characteristics, makes either ordinary or disjunctive cokriging an attractive tool in rainfall estimation. Even when radar rainfall data are bias-free, the radar-gage estimates from disjunctive cokriging are better than the radar-only estimates. The quality of radar rainfall data is not usually known a priori. If the radar rainfall data are known to be of good quality, there may be no advantage in merging radar rainfall data and rain gage measurements. Under a range of raingage network densities, the margin of improvement by gage-radar estimation using cokriging is greatest over gage-only estimation when the gage density is lowest (i.e., with the average gage density over the continental United States). It should be remembered that the assumption of perfectly known rain area used in the estimation scheme was more favorable to gage-only estimation than to radar-gage estimation.

On the average, radar-gage estimates obtained from ordinary cokriging are as unbiased as gage-only estimates obtained from ordinary kriging. The relative unbiasedness of merged radar-gage estimates, however, is seen to be affected by the error characteristics of the radar rainfall data, and when the gage network density is sufficiently high, the unbiasedness is better achieved by gage-only estimation. The unbiasedness property is seen to be better realized by ordinary cokriging than by disjunctive cokriging. When there is a high bias in the mean of the radar rainfall field, however, any of the cokriging estimators is capable of removing the adverse effect of the high bias.

In the mini-max sense (maximum errors of underestimation and overestimation), the radar-gage estimates are generally better than the gage-only estimates. When the homogeneity assumptions are well met, disjunctive cokriging shows substantial improvement over ordinary cokriging. The kriging and cokriging estimators tend to overestimate low rainfall depths and underestimate high rainfall depths. The merged radar-gage estimates are generally better than the gage-only estimates in predicting either low or high rainfall depths. Disjunctive cokriging, particularly when the homogeneity assumptions are well met, improves substantially over ordinary cokriging in predicting either low or high rainfall depths.

Universal cokriging is seen to offer little or no advantage over ordinary cokriging or disjunctive cokriging. It is computationally expensive, and structure identification is likely to be unsuccessful. However, it is seen, in a few cases, to provide better predictions of high rainfall depths when a nonconstant trend is present (see Figure 12).

In real world applications, the following guidance can be given concerning the choice between ordinary cokriging and disjunctive cokriging. When the rainfall data do not clearly indicate the presence of a nonconstant trend, disjunctive cokriging can be expected to perform significantly better than ordinary cokriging, but at the expense of a 10-fold increase in CPU time. When the rainfall data do indicate nonhomogeneity in the mean due to, for example, orographic effects, ordinary cokriging is an inexpensive alternative.

Unbiasedness and error variance of kriging and cokriging estimates are found to be affected by the coefficient of variation and correlation distance of the ground truth rainfall, respectively. If the rainfall field has a smaller coefficient of variation and a longer correlation distance than those seen in this work (e.g., stratiform rainfall field), a larger reduction in root mean square error may be expected.

The potential of cokriging in rainfall estimation is greatly reduced by uncertain second-order statistics. In cokriging, the second-order statistics are estimated only from the presently available data. Then, due to the sparsity of rain gages, the second-order statistics associated with gage rainfall will be estimated with large uncertainty. It is recommended that the Bayesian approach [e.g., *Kitanidis*, 1986] be used. By making use of the past measurements of rainfall, a priori information on the second-order statistics can be obtained, which can then be updated with the presently available measurements. In this way, both the past and the present measurements of rainfall are utilized. This approach has been investigated by one of the authors [*Azimi-Zonooz*, 1989], and issues such as the margin of improvement versus additional computational requirements will be reported in the near future.

Acknowledgments. This material is based upon work supported by the National Science Foundation under grant ECE-8419189. Additional support was provided by the Utah Water Research Laboratory, Utah State University, Logan, Utah.

REFERENCES

Azimi-Zonooz, A., Evaluation of nonlinear co-kriging and Bayesian estimation for stochastic interpolation of rainfall data from multiple sensors. Ph.D. dissertation, Dep. of Civ. and Environ. Eng., Utah State Univ., Logan, 1989.

Hildebrand, P. H., N. Towery, and M. R. Snell, Measurement of convective mean rainfall over small areas using high-density raingages and radar, *J. Appl. Meteorol.*, *18*, 1316–1326, 1979.

Hudlow, M. D., and B. L. Patterson, GATE radar-rainfall atlas, special report, Natl. Oceanic and Atmos. Admin., Washington, D. C., 1979.

Journel, A. G., and C. J. Huijbregts, *Mining Geostatistics*, Academic, San Diego, Calif., 1978.

Karr, A. G., Inference for stationary random fields given Poisson samples, *Adv. Appl. Probab.*, *18*, 406–422, 1986.

Kitanidis, P. K., Parameter uncertainty in estimation of spatial function: Bayesian analysis, *Water Resour. Res.*, *22*(4), 499–507, 1986.

Seo, D.-J., Stochastic interpolation of rainfall field using multiple sensors, Ph.D. dissertation, Dep. of Civ. and Environ. Eng., Utah State Univ., Logan, 1988.

Seo, D.-J., W. F. Krajewski, and D. S. Bowles, Stochastic interpolation of rainfall data from rain gages and radar using cokriging, 1, Design of experiments, *Water Resour. Res.*, *26*, 469–477, 1990.

Valdes, J. B., I. Rodriguez-Iturbe, and V. K. Gupta, Approximation of temporal rainfall from a multidimensional model, *Water Resour. Res.*, *21*(8), 1259–1270, 1985.

Wilson, J. W., and E. A. Brandes, Radar measurement of rainfall—A summary, *Bull. Am. Meteorol. Soc.*, *60*(9), 1048–1058, 1979.

A. Azimi-Zonooz and D. S. Bowles, Utah Water Research Laboratory, Utah State University, Logan, UT 84322.

W. F. Krajewski, Institute of Hydraulic Research, University of Iowa, Iowa City, IA 52242.

D.-J. Seo, Hydrologic Research Laboratory, National Weather Service, Silver Spring, MD 20910.

(Received October 5, 1988;
revised September 5, 1989;
accepted September 18, 1989.)