2.4 Improving Analysis of Heavy-to-Extreme Precipitation using Conditional Bias-Penalized Optimal Estimation

Ridwan Siddique[1,]*, Dong-Jun Seo[1], Yu Zhang[2], Dongsoo Kim[3]
[1]Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX
[2]Office of Hydrologic Development, National Weather Service, Silver Spring, MD
[3]Remote Sensing Application Division, National Climatic Data Center, Asheville, NC

1. INTRODUCTION

For its obvious importance, quantitative precipitation estimation (QPE) has been a topic of active research for over a century (Thiessen 1911). Whether it is based on gauge-only or multisensor estimation, QPE generally involves spatial prediction using statistical or dynamical-statistical models. Statistical models, by far the more widely used of the two to date, use optimal (in some sense of the word) estimation, of which various types of linear and nonlinear techniques are available (see e.g. Creutin and Obled (1982), Tabios and Salas (1985) and references therein). For example, the algorithms used operationally in the National Weather Service (NWS) for gauge-only and radar-gauge analyses in their Multisensor Precipitation Estimator (MPE, Seo et al. 2010) are variants of kriging and cokriging, respectively (Seo 1998a,b).

Real-time QPE demands accurate estimation particularly of large amounts as they represent greater hazards to lives and properties. In flood forecasting, what matters most for QPE is the ability to estimate large amounts of precipitation as accurately as possible over the range of spatiotemporal scales of aggregation associated with the size and response time of the basin. Kriging or its variants do produce, as theoretically expected, precipitation estimates that are unbiased and of minimum error variance in the unconditional sense. In the conditional sense, however, these so-called optimal estimation techniques very often severely underestimate heavy precipitation and overestimate light precipitation (Seo and Breidenbach 2002, Ciach et al. 2000, Habib et al. 2012). These results arise because, to achieve (un-conditional) minimum error variance, it is necessary to reduce the error variance associated with light to moderate precipitation, which occurs frequently and over large areas, even if it may increase the error variance associated with heavy precipitation, which occurs relatively rarely and generally over small areas. For accurate estimation of large amounts, however, it is more important to reduce conditional bias (CB), in particular Type-II CB, than to minimize unconditional error variance. QPE for flood forecasting is a prime example of that. In the above, Type-II CB is defined as $E[\hat{X} \mid X = x] - x$ where X, $\hat{X}$ and x denote the unknown truth, the estimate, and the realization of X, respectively (Joliffe and Stephenson 2003).

*Corresponding author address: Ridwan Siddique, Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX, USA; e-mail: ridwan.siddique@mavs.uta.edu

The proposed methodology is based on novel extension of classical optimal linear estimation theory in which, in addition to error variance, Type-II conditional bias (CB) is explicitly minimized. The resulting Fisher-like solution may also serve as an alternative or complementary observation equation for a range of Fisher solution-based static or dynamic filters, such as Kalman filter and its variants. When cast in the form of well-known kriging or its variants used in the NWS MPE, the proposed methodology yields a new kriging estimator, referred to herein as CB-penalized kriging (CBPK) (Seo 2012).

CBPK, however, gives negative estimates in areas of light precipitation. To address this, an extension of CBPK, referred to herein as Extended Conditional Bias Penalized Kriging (ECBPK), has been developed. Loosely speaking, ECBPK truncates negative values to zeros and introduces a scalar weight as a function of probability of precipitation (PoP) through which the total amount of precipitation is preserved. In this study, comparative evaluation of ECBPK is carried out with ordinary kriging (OK) (Journel and Huijbregts 1978), a variant of which is used in MPE for gauge-only estimation, for estimation of point and mean areal precipitation (MAP) through real-world and synthetic experiments, respectively.

2. METHODOLOGY

In this section, CBPK is briefly described and ECBPK is introduced as an extension of CBPK.

2.1 CONDITIONAL BIAS PENALIZED KRIGING (CBPK)

CBPK minimizes the sum of the error variance and the mean squared CB of the estimate:

$$
\begin{aligned}
J &= E_{Z_0^*,z_0}[(Z_0^* - Z_0)^2] + E_{z_0}\left[\left\{E_{Z_0^*}[Z_0^*|Z_0] - Z_0\right\}^2\right] \\
&= E_{z_i,z_0}\{\textstyle\sum_{i=1}^n \lambda_i(Z_i - m_i) - (Z_0 - m_0)\}^2] \\
&+\int\{E_{z_i}[\textstyle\sum_{i=1}^n \lambda_i(Z_i - m_i)|Z_0 = z_0] - (z_0 - m_0)\}^2 \\
&\quad \mathrm{x} f_{z_0}(z_0)dz_0 \qquad (1)
\end{aligned}
$$

Where Z_0^* denotes the simple kriging estimate of the random variable of interest at location u_0, Z_0, m_0 denotes the mean of Z_0, λ_i denotes the weight given to the observation at u_i of the random variable Z_i, m_i denotes the mean of Z_i, z_0 denotes the realization of Z_0, n denotes the number of neighbors, $f_{Z_0}(z_0)$ denotes the marginal probability density function (pdf) of Z_0 and the expectation operations are with respect to the variables subscripted. The weights, $\lambda_j, j=1,...,n$, may be obtained by solving the following CBPK system:

$$\sum_{j=1}^{n} \lambda_i(\rho_{ij} + \rho_{i0}\rho_{j0})\sigma_i\sigma_j = 2\rho_{i0}\sigma_i\sigma_0,$$
$$i = 1, \dots, n \quad (2)$$

where ρ_{ij} denotes the correlation between Z_i and Z_j, and σ_i denotes the standard deviation of Z_i. For further details, the reader is referred to Seo (2012).

2.2 EXTENDED CONDITIONAL BIAS-PENALIZED KRIGING (ECBPK)

While CBPK is superior to OK over the tail ends of the distribution, it is slightly inferior to OK over the mid-ranges. Also, if the distribution is skewed and the variable of interest is nonnegative, a significant fraction of the CBPK estimates may be significantly negative. A methodology is therefore needed to combine the OK and CBPK estimates so that the resulting estimate is close to the more accurate of the two, depending on the magnitude of the precipitation amount being estimated, and that the precipitation estimates are nonnegative. For that, we write the estimate, $E[Z_0 \mid Z_1 = z_1,...,Z_n = z_n]$, as follows:

$$E[Z_0 \mid \bullet] = \sum_{k=1}^{K} E[Z_0 \mid \bullet, Z_0 \in A_k]\Pr[Z_0 \in A_k \mid \bullet] \quad (3)$$

where $\{\bullet\}$ denotes the event $\{Z_1 = z_1,...,Z_n = z_n\}$ for brevity, A_k's denote the sub-ranges of the truth, the union of which encompasses the entire range of the truth, and Pr[ ] denotes the probability of occurrence of the event bracketed. We rewrite the conditional expectation in Eq.(3) as:

$$\begin{aligned} &E[Z_0 \mid \bullet, Z_0 \in A_k] \\ &= E[Z_0 \mid \bullet, Z_0 \in A_k, OK^*]\Pr[OK^* \mid \bullet, Z_0 \in A_k] \\ &+ E[Z_0 \mid \bullet, Z_0 \in A_k, CBPK^*]\Pr[CBPK^* \mid \bullet, Z_0 \in A_k] \end{aligned} \quad (4)$$

where OK^* or $CBPK^*$ denotes the event that the OK or CBPK estimate is more accurate than the CBPK or OK estimate, respectively. It can be shown that, under simplifying assumptions, the ECBPK estimate may be written as:

$$E[Z_0 \mid \bullet] \approx \gamma \; E[Z_0 \mid \bullet, CBPK^*] \quad (5)$$

where

$$\begin{aligned} &\gamma \\ &= \Pr[CBPK^* \mid Z_0 = 0] \\ &+ \Pr[Z_0 > 0 \mid \bullet](\Pr[CBPK^* \mid Z_0 > 0] - \Pr[CBPK^* \mid Z_0 = 0]) \end{aligned}$$

In actual implementation of ECBPK in this work, γ has been parameterized as a function of PoP. The larger or the smaller PoP is, the closer to unity or zero γ is, respectively. That is, the more likely precipitation is to occur, the larger the weight given to the CBPK estimate is.

3. EVALUATION

Comparative evaluation of ECBPK is carried out through real world and synthetic experiments. The evaluation experiments are described in this section.

3.1 REAL WORLD EXPERIMENTS

To comparatively evaluate ECBPK, several real world experiments were carried out in a cross validation mode using hourly and daily rain gauge data for a collection of heavy-to-extreme precipitation events in the U.S. They include events over the Arkansas-Red Basin River Forecast Center (ABRFC) service area in Oklahoma, the 2009 Southeast extreme event, and tropical storm events over the Lower Colorado River Authority (LCRA) service area in Texas. The number of nearest neighbors used is 30 throughout this work. For each event evaluated in this paper, the spatial correlation scale was calculated using the Stage IV data. For that, conditional (on occurrence of precipitation) correlograms were calculated in eight different directions (0^0, 26.6^0, 45^0, 63.4^0, 90^0, 116.6^0, 135^0, 153.4^0) which are then fitted to the exponential, Gaussian and spherical models (Journel and Huijbregts 1978). It was found that the exponential model provides the best fit and the correlation structure is largely isotropic (see Figure 1). As such, an average of all directional correlation parameters was used for both hourly and daily analysis.

3.2 SYNTHETIC EXPERIMENTS

For mean areal precipitation (MAP) analysis, the Stage IV data for the 2009 Southeast extreme event were used as truth. The synthetic rain gauge networks were then generated by randomly selecting grid boxes from the 150x150 HRAP domain. The gauge networks consist of 125, 500, 1000, 2000 and 4000 gauges. The gauge observations (i.e. the Stage IV estimates at the randomly selected gauge locations) were interpolated for all HRAP grid points using OK and ECBPK. The MAP estimates were then calculated for the square areas ranging from 4x4 km^2 to 128x128 km^2.

4. RESULTS

In this section, cross validation results from hourly and daily point precipitation analysis and MAP analysis results from the synthetic experiment are presented.

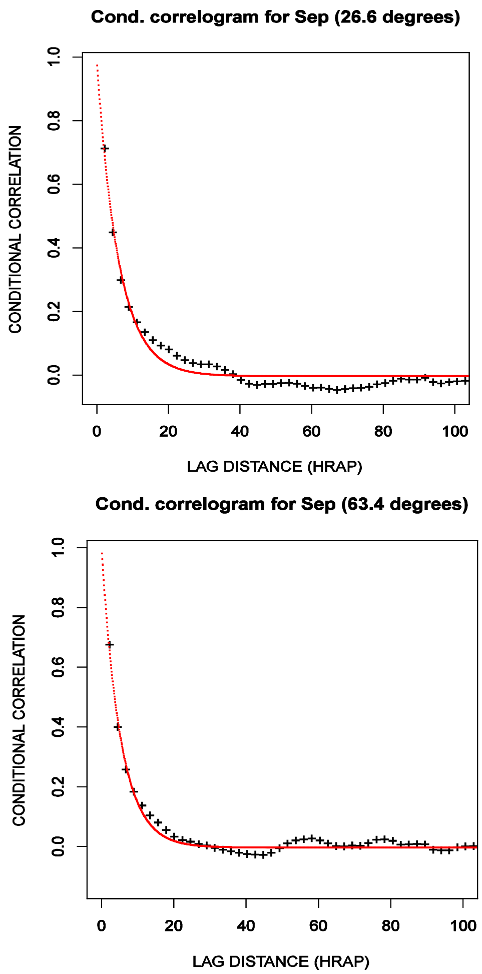


Figure 1: Directional correlograms (cross) and the exponential model fit (solid red) for hourly gauge analysis for the 2009 Southeast extreme precipitation event.

4.1. POINT PRECIPITATION ANALYSIS

For estimation of point precipitation, the OK and ECBPK estimates are compared using scatter plots and summary statistics. Figure 2 shows the scatter plots of estimated vs. observed hourly point precipitation over the ABRFC service area in Oklahoma. Note that ECBPK reduces conditional bias significantly.

Figure 3 shows the reduction in RMSE by ECBPK over OK for hourly and daily analyses of the 2009 Southeast extreme precipitation event and for Tropical Storm Erin in TX, respectively. In this figure, the reference is the OK estimates and the margin of improvement by ECBPK over OK is shown in terms of percent reduction in RMSE. In this work, in an attempt to quality-control the gauge data at least partially, we used the coefficent of variations (CV) and PoP of gauge data from individual gauges to identify bad gauges, which were then thrown out. Visual examination of the scatter plots indicates, however, that a large number of bad gauge reports still exist. It is suspected that the margin of improvement by ECBPK, particularly for hourly analysis, is significantly compromised by the bad gauge data.

Figures 4 shows the precipitation accumulation maps for the 2009 Southeast extreme precipitation event as estimated by OK and ECBPK via hourly analysis. For comparison, the corresponding Stage IV analysis is also shown. Note that ECBPK picks up very large and small precipitation amounts better than OK, and that ECBPK depicts the spatial pattern of precipitation better than OK.

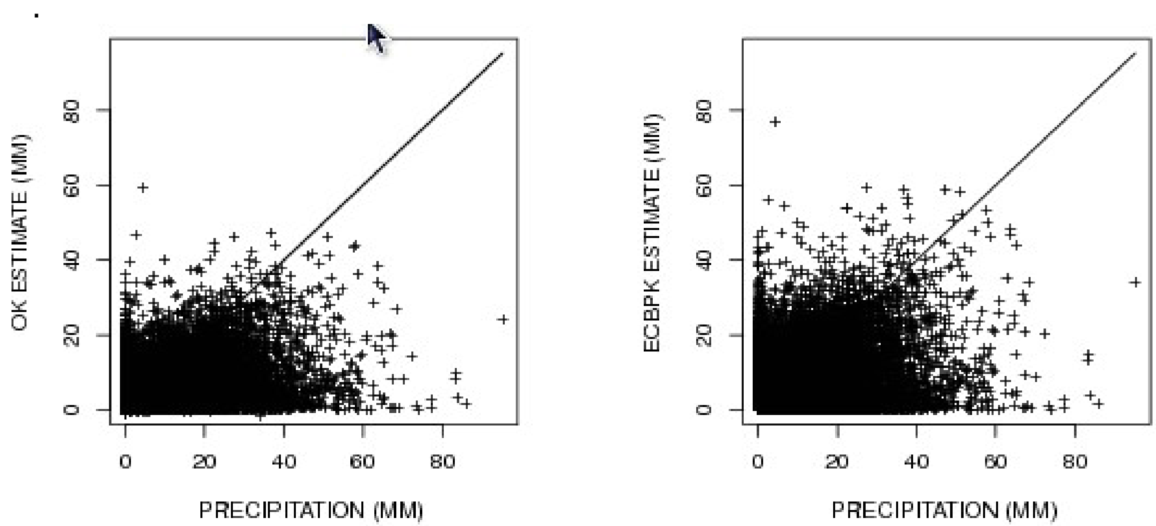
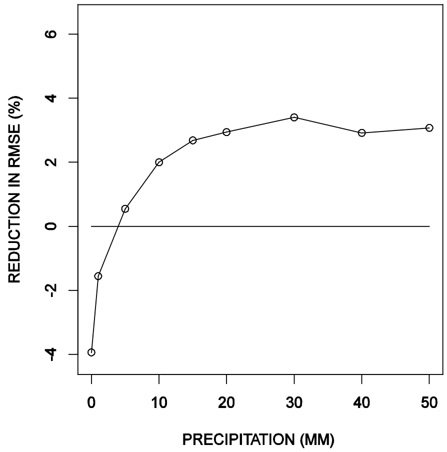
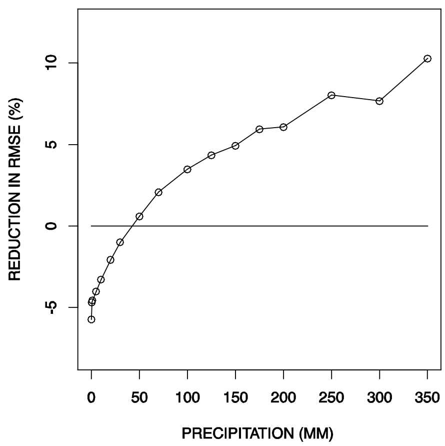


Figure 2: Scatter plots of observed vs. estimated hourly precipitation over the ABRFC service area in Oklahoma.

amounts, however, ECBPK produces more accurate estimates than OK even for large basins (see Figure 5).



(a)

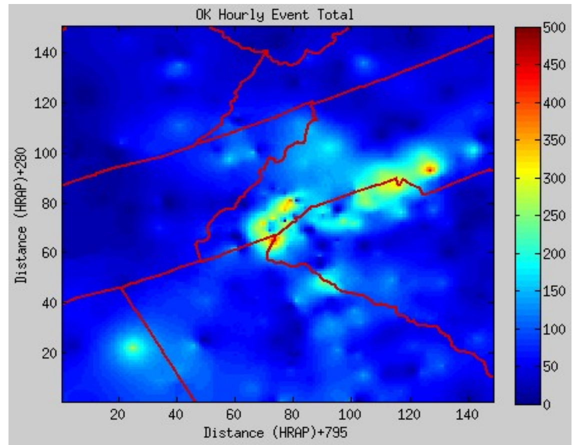


(b)

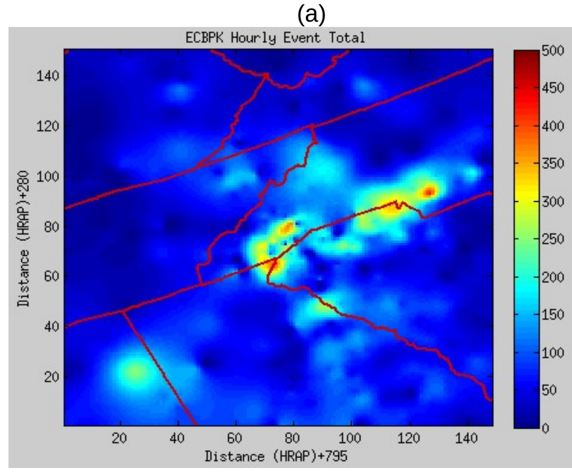
Figure 3. Percent reduction in RMSE by ECBPK over OK for a) hourly precipitation for the 2009 Southeast extreme precipitation event and b) for daily precipitation in Texas.

4.2. MAP ANALYSIS

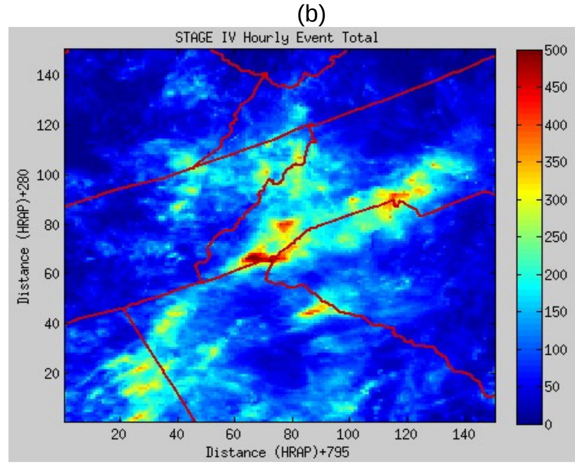
For MAP, hourly analysis is carried out for thirty different combinations of gauge network density and basin scale. The gauge network density ranged from 125 to 4,000 gauges in an area of 600 x 600 km^2 and the basin size ranged from 16 km^2 to 16,384 km^2. The densities of 500 and 1,000 gauges are comparable to the operational HADS and COOP gauge networks, respectively (Kim et al. 2010). For MAP, the margin of improvement in RMSE reduction by ECBPK increases with increasing basin scale up to 256 km^2 and increasing gauge network density. Beyond 256 km^2, the improvement starts to decrease. For larger basin sizes, the reduction is negative in the mean sense because the area of small-to-moderate precipitation increases as the basin size increases. For very large precipitation



(a)



(b)



(c)

Figure 4. Precipitation accumulation maps for a) OK analysis b) ECBPK analysis c) Stage IV data for the 2009 Southeast extreme event.

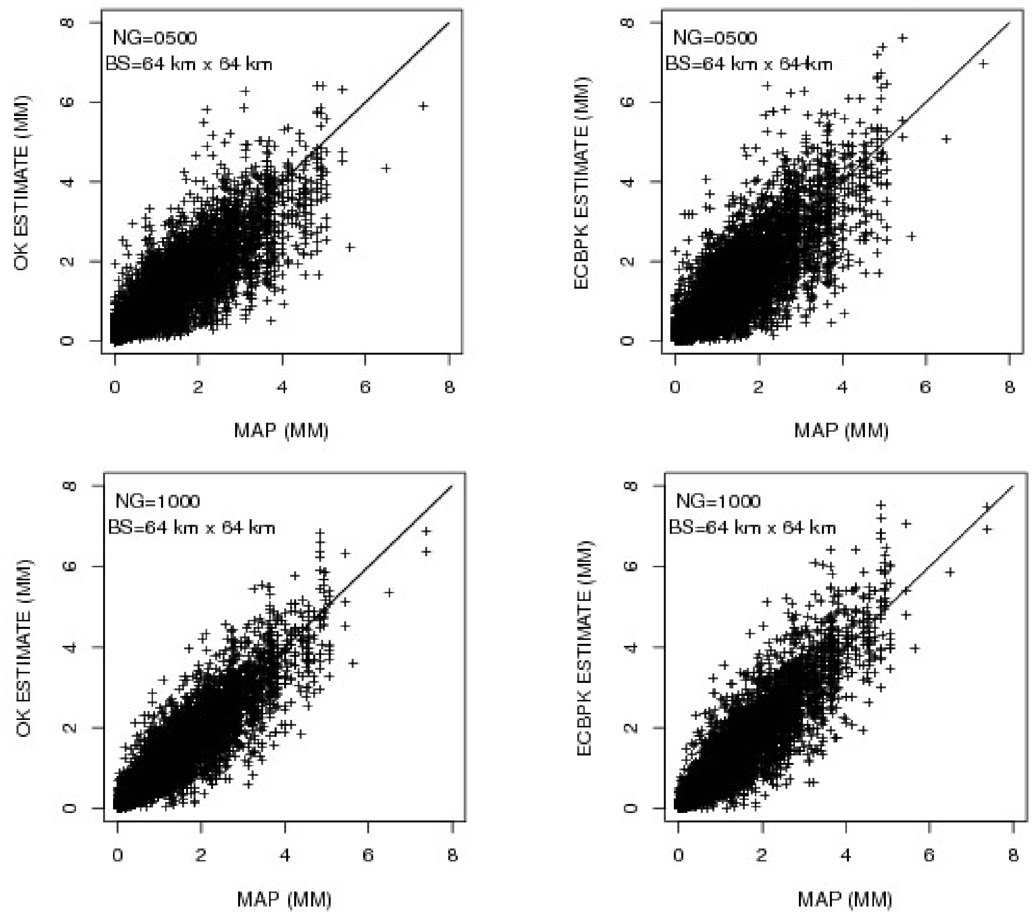


Figure 5: Scatter plots for hourly mean areal precipitation of 4096 sq. km basin for gauge densities 500 (top) and 1000 (bottom) gauges

5. CONCLUSIONS AND FUTURE RESEARCH RECOMMENDATIONS

For accurate estimation of heavy-to-extreme precipitation, reducing Type-II CB is important. In this paper, we describe and evaluate a new precipitation analysis technique, ECBPK, which significantly reduces Type-II CB. Compared to OK, a variant of which is used in MPE in NWS, ECBPK improves performance for estimation of large amounts of daily and hourly precipitation. For daily analysis, the margin of improvement is very significant (over 10% reduction in RMSE). For hourly analysis, the margin is smaller due possibly to lack of quality control of the rain gauge data used. For MAP analysis, ECBPK improves performance for smaller basins (64 km^2 to 512 km^2). While the improvement for larger basins is smaller in the mean sense, ECBPK significantly improves estimation for very large precipitation amounts.

For future research, we plan to pursue multivariate extension of ECBPK and CB-penalized linear filter. In the context of multisensor QPE, ECBPK is expected to provide improvement when the auxiliary variable(s) are not very skillful. Such application is hence particularly applicable when merging rain gauge data with satellite QPE, NWP analysis and/or cool-season radar QPE. In the context of linear filtering, one may expect the CB-penalized approach to improve filter performance when the observations and/or the dynamical model used is susceptible to CB.

6. REFERENCES

Ciach, G. J., M. L. Morrissey, and W. F. Krajewski, 2000. Conditional bias in radar rainfall estimation. J. Appl. Mete-or., 39, 1941-1946.

Creutin, J. D. and C. Obled, 1982. Objective analysis and mapping techniques for rainfall field: An objective comparison. Water Resour. Res. 18(2), 413-431.

Habib, E., Qin, L., D.-J. Seo, G. Ciach, and B. R. Nelson, 2012. Independent As-sessment of Incremental Complexity in the NWS Multi-Sensor Precipitation Es-timator Algorithms, Journal of Hydro-logic Engineering, doi: http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000638.

Journel, A. G., and Ch. J. Huijbregts, 1978. Mining geostatistics, Academic Press, 600pp.

Kim, D., B. Nelson, and D.-J. Seo, 2009. Characteristics of Reprocessed Hydro-meteorological Automated Data System (HADS) Hourly Precipitation Data. Weather and Forecasting, 24, 1287-1296.

Seo, D.-J., 1998a. Real-time estimation of rainfall fields using rain gage data under fractional coverage, J. Hydrol., 208, 25-36.

Seo, D.-J., 1998b. Real-time estimation of rainfall fields using radar rainfall and rain gage data, J. Hydrol., 208, 37-52.

Seo, D.-J. And J. P. Breidenbach, 2002. Re-al-time correction of spatially nonuni-form bias in radar rainfall data using rain gauge measurements, J. Hydrome-teorol., 3, 93-111.

Seo, D.-J., A. Seed, and G. Delrieu, 2010. Radar-based rainfall estimation, chapter to appear in AGU Book Volume on Rainfall: Microphysics, Measurement, Estimation, and Statistical Analyses, F. Testik and M. Gebremichael, Editors.

Seo D-J. 2012. Conditional bias-penalized kriging. Stochastic Environmental Re-search and Risk Assessment. DOI 10.1007/s00477-012-0567-z (online first),
http://www.springerlink.com/content/f5q475570q08252t/.

Tabios, G. Q. III and J. D. Salas, 1985. A comparative analysis of techniques for spatial interpolation of precipitation. Water Resour. Bull. 21(3), 365-380.

Thiessen, A. H., 1911. Precipitation Averag-es for Large Areas. Mon. Wea. Rev. 39(7), 1082-1089.