Journal of Hydrology 519 (2014) 1824–1835

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Improving real-time estimation of heavy-to-extreme precipitation using rain gauge data via conditional bias-penalized optimal estimation

Dong-Jun Seo [a,*], Ridwan Siddique [a,1], Yu Zhang [b], Dongsoo Kim [c]

[a] Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX, United States
[b] Office of Hydrologic Development, National Weather Service, Silver Spring, MD, United States
[c] Remote Sensing Application Division, National Climatic Data Center, Asheville, NC, United States

ARTICLE INFO

Article history:
Received 9 January 2014
Received in revised form 8 July 2014
Accepted 20 September 2014
Available online 28 September 2014
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Emmanouil N. Anagnostou, Associate Editor

Keywords:
Heavy-to-extreme precipitation
Rain gauges
Conditional bias
Optimal estimation

SUMMARY

A new technique for gauge-only precipitation analysis for improved estimation of heavy-to-extreme precipitation is described and evaluated. The technique is based on a novel extension of classical optimal linear estimation theory in which, in addition to error variance, Type-II conditional bias (CB) is explicitly minimized. When cast in the form of well-known kriging, the methodology yields a new kriging estimator, referred to as CB-penalized kriging (CBPK). CBPK, however, tends to yield negative estimates in areas of no or light precipitation. To address this, an extension of CBPK, referred to herein as extended conditional bias penalized kriging (ECBPK), has been developed which combines the CBPK estimate with a trivial estimate of zero precipitation. To evaluate ECBPK, we carried out real-world and synthetic experiments in which ECBPK and the gauge-only precipitation analysis procedure used in the NWS's Multisensor Precipitation Estimator (MPE) were compared for estimation of point precipitation and mean areal precipitation (MAP), respectively. The results indicate that ECBPK improves hourly gauge-only estimation of heavy-to-extreme precipitation significantly. The improvement is particularly large for estimation of MAP for a range of combinations of basin size and rain gauge network density. This paper describes the technique, summarizes the results and shares ideas for future research.

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 procedures. Statistical procedures, 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). More recently, artificial neural networks (Bellerby et al., 2000, Grimes et al., 2003; Hsu et al., 2007; Chiang et al., 2007) and support vector regression (Chen et al., 2011) have been added to the list of techniques for QPE.

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. While radar and multisensor QPE have been playing an increasingly large role in flood forecasting (Seo et al., 2010), it is expected that rain gauges remain as the primary source of QPE in many areas due to radar coverage gaps, lack of accuracy in remotely sensed estimates and latency in satellite-based estimates. 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

* Corresponding author at: Department of Civil Engineering, The University of Texas at Arlington, Box 19308, Rm 248E Nedderman Hall, 416 Yates St, Arlington, TX 76019-0308, United States. Tel.: +1 817 272 5063; fax: +1 817 272 2630.
E-mail address: djseo@uta.edu (D.-J. Seo).
[1] Present address: Department of Civil and Environmental Engineering, Pennsylvania State University, United States.

http://dx.doi.org/10.1016/j.jhydrol.2014.09.055
0022-1694/

overestimate light precipitation (Ciach et al., 2000; Seo and Breidenbach, 2002; Brown and Seo, 2010; Habib et al., 2012). These results arise because, to achieve (unconditional) 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 at least as important to reduce conditional bias (CB), in particular Type-II CB, as to minimize unconditional error variance. QPE for flood forecasting is a prime example of that. Type-I and -II CBs arise from Type-I and -II errors in the mean sense, respectively. A Type-I error is associated with a false alarm (e.g. crying wolf without a wolf in sight) whereas a Type-II error is associated with failing to raise an alarm (i.e. failing to see the wolf). Note that, whereas Type-I CB can be reduced by calibration (e.g. if the false alarm rate is too high, one may not cry wolf as often), Type-II CB cannot. In the above, Type-II CB is defined as $E[\overline{X} \mid X = x] - x$ where X, $\overline{X}$ and x denote the unknown truth, the estimate, and the realization of X, respectively (Joliffe and Stephenson, 2003). In the context of spatial estimation, Reducing Type-II CB amounts to improving discrimination in spatial prediction of large or small values. Unlike Type-I CB (i.e. $E[X \mid \overline{X} = \hat{x}] - \hat{x}$), which relates to calibration-refinement factorization (Murphy and Winkler, 1987), Type-II CB is not very amenable to statistical bias correction or post processing (Moskaitis, 2008).

Recently, Seo (2013) has demonstrated in the context of gauge-only QPE that the potential impact of reducing Type-II CB (hereafter referred to as CB for short) at sub-daily and mean areal precipitation (MAP) scales of $O(10^0)$–$O(10^3)$ km^2 may be substantial. The synthetic experiments (Seo, 2013) suggest that the margin of improvement for estimating heavy precipitation from reducing CB may be comparable to that from greatly increasing the density of the rain gauge network or, equivalently in the context of multisensor estimation, that from significantly improving the quality of the remotely sensed data or scale-compatible NWP precipitation analysis. Often, lack of performance by linear estimators has been attributed to their linear (as opposed to nonlinear) nature. Experimental and empirical evidences suggest, however, that the marginal improvement by nonlinear estimation is relatively small (see e.g. Azimi-Zonooz et al., 1989; Seo, 1996a,b), and that CB may be a much more important limiting factor than linearity in estimation of heavy-to-extreme precipitation.

Seo (2013) extends optimal linear estimation theory in which, in addition to error variance, CB is explicitly minimized. The resulting Fisher-like solution may also serve as an alternative 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 MPE, the proposed methodology yields a new kriging estimator, referred to as CB-penalized kriging (CBPK). For estimation of skewed nonnegative variables such as precipitation, however, CBPK yields estimates that may be significantly negative in areas of light to no precipitation. To address this, an extension of CBPK, referred to herein as extended CB-penalized kriging (ECBPK), has been developed. In this paper, we describe and comparatively evaluate ECBPK with the variant of ordinary kriging (OK) (Journel and Huijbregts, 1978) used in MPE for gauge-only estimation. The evaluation is carried out for estimation of point and mean areal precipitation (MAP) through real-world and synthetic experiments, respectively. The data sets used include a number of heavy-to-extreme precipitation events in Texas, Oklahoma and vicinity, and the Southeastern US. The new contributions of this work include development and evaluation of a new technique for improved real-time estimation of heavy-to-extreme precipitation, and improving understanding of errors in spatial optimal estimation of precipitation and their dependence on the magnitude and intermittency of precipitation, catchment scale and rain gauge network density.

This paper is organized as follows. Section 1 provides the context of the problem and motivation for the research. Section 2 describes the proposed technique. Section 3 describes comparative evaluation of the technique. Section 4 presents the results. Section 5 summarizes the main conclusions and future research recommendations.

2. Extended conditional bias-penalized kriging (ECBPK)

In this section, we provide a summary description of CBPK and describe ECBPK in some detail.

2.1. Conditional bias-penalized kriging (CBPK)

CBPK may be considered an extension of simple kriging (SK) in which the objective function is made not only of error variance but also of CB. The SK estimator (see e.g. Journel and Huijbregts, 1978) is given by:

$$Z_0^* = m_0 + \sum_{i=1}^{n} \lambda_i (Z_i - m_i) \tag{1}$$

where Z_0^* denotes the SK estimate of $E[Z_0] \mid Z_1 = z_1, \ldots, Z_n = z_n]$ of the random variable of interest at location u_0, m_0 denotes the mean of Z_0, λ_i are the weights assigned to Z_i, m_i denotes the a priori mean of Z_i and n denotes the number of neighbors used in estimation. In SK, the weights are obtained by minimizing the error variance of the estimate, J_{SK}:

$$\begin{aligned} J_{SK} &= E_{Z_0^*, Z_0}[(Z_0^* - Z_0)^2] \\ &= E_{Z_i, Z_0}\left[\left\{\sum_{i=1}^{n} \lambda_i (Z_i - m_i) - (Z_o - m_0)\right\}^2\right] \end{aligned} \tag{2}$$

where the expectation operations are with respect to the variables subscripted. In CBPK, the CB penalty term, or the unconditional expectation of CB squared, is added to the objective function as follows:

$$\begin{aligned} J_{CBPK} &= E_{Z_0^*, Z_0}[(Z_0^* - Z_0)^2] + \alpha \cdot E_{Z_0}\left[\left\{E_{Z_0^*}[Z_0^* \mid Z_o] - Z_o\right\}^2\right] \\ &= E_{Z_i, Z_0}\left[\left\{\sum_{i=1}^{n} \lambda_i (Z_i - m_i) - (Z_o - m_0)\right\}^2\right] + \alpha \\ &\quad \cdot \int \left\{E_{Z_i}\left[\sum_{i=1}^{n} \lambda_i (Z_i - m_i) \mid Z_0 = z_0\right] - (z_0 - m_0)\right\}^2 f_{Z_0}(z_0) dz_0 \end{aligned} \tag{3}$$

where z_0 denote the experimental values of Z_0, $f_{Z_0}(z_0)$ denotes the marginal probability density function (pdf) of Z_0 and α denotes the positive weight given to the CB penalty term. Experience thus far suggests that a reasonable choice of α is unity but it may be optimized for improved balanced performance between reducing error variance and reducing CB. In Eq. (3), we specify $E_{Z_i}[Z_i - m_i \mid Z_0 = z_0]$ using the Bayesian optimal estimator (Schweppe, 1973) as:

$$E[Z_i - m_i \mid Z_0 = z_0] = \rho_{i0} \frac{\sigma_i}{\sigma_0} (z_0 - m_0) \tag{4}$$

where ρ_{i0} denotes the correlation between Z_i and Z_0, and σ_i and σ_0 denote the standard deviation of Z_i and Z_0, respectively. Note that Eq. (4) is identical to the SK or the linear regression solution for estimating Z_i given Z_0. The CBPK system results from minimizing J_{CBPK} in Eq. (3) with respect to the weights, λ_i's: