Rainfall estimation using raingages and radar – A Bayesian approach: 1. Derivation of estimators

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

J. A. Smith
Dept. of Civil Engineering and Operations Research, Princeton University, NJ 28544, USA

Abstract: Procedures for estimating rainfall from radar and raingage observations are constructed in a Bayesian framework. Given that the number of raingage measurements is typically very small, mean and variance of gage rainfall are treated as uncertain parameters. Under the assumption that log gage rainfall and log radar rainfall are jointly multivariate normal, the estimation problem is equivalent to lognormal co-kriging with uncertain mean and variance of the gage rainfall field.

The posterior distribution is obtained under the assumption that the prior for the mean and inverse of the variance of log gage rainfall is normal-gamma 2. Estimate and estimation variance do not have closed-form expressions, but can be easily evaluated by numerically integrating two single integrals. To reduce computational burden associated with evaluating sufficient statistics for the likelihood function, an approximate form of parameter updating is given. Also, as a further approximation, the parameters are updated using raingage measurements only, yielding closed-form expressions for estimate and estimation variance in the Gaussian domain.

With a reduction in the number of radar rainfall data in constructing covariance matrices, computational requirements for the estimation procedures are not significantly greater than those for simple co-kriging. Given their generality, the estimation procedures constructed in this work are considered to be applicable in various estimation problems involving an undersampled main variable and a densely sampled auxiliary variable.

Key words: Co-kriging, parameter uncertainty, Bayesian estimation, radar rainfall, gage rainfall.

1 Introduction

Optimal linear estimation using an auxiliary variable (including co-kriging) has found its use in many hydrological applications. In many situations, however, second-order statistics required for estimation can only be known with large uncertainty due to lack of data. In some situations, a large number of data may be available for the auxiliary variable, but only a very small number of data for the main variable. A case in point is the problem of rainfall estimation using radar rainfall data and raingage measurements (Creutin et al. 1987; Krajewski 1987; Azimi-Zonooz et al. 1989; Seo et al. 1990a,b). Over a radar umbrella, the number of radar rainfall data, each datum corresponding to a radar bin, can easily exceed a thousand whereas the number of raingage measurements is typically in the order of 10-100. In such a case, second-order statistics for radar rainfall (auxiliary variable) can be estimated with confidence, and may be assumed perfectly known. On the other hand, second-order statistics involving gage rainfall (main variable) will suffer from large sampling errors, and cannot be assumed perfectly known.

In single-variable cases (e.g., ordinary kriging), when the number of data is too small
to obtain a reliable correlogram (or, semi-variogram), cross-validation is typically used. In this situation we choose a covariance function, perform estimation, compare actual estimation variance with predicted estimation variance, and repeat until a good agreement between the two variances is reached. When an auxiliary variable is involved, however, cross-validation can be an exhaustive undertaking. Worse yet, a closeness between the two variances does not necessarily mean that the chosen covariance function is indeed close to the actual covariance function. One may opt to forego cross-validation, and use the (cross-) correlograms (how unreliable they may be). In such a case, however, it has been shown that estimation variances are very often grossly erroneous (Seo et al. 1990b).

An alternative is to build estimation procedures in a Bayesian framework (Raiffa and Schlaifer 1961; Kitanidis 1986). It allows an explicit accounting of uncertainties associated with second-order statistics, and use of any a priori knowledge about them. In this work, we construct a set of such estimation procedures. The goal of the estimation procedure is to predict rainfall depth at an arbitrary location using raingage measurements and radar rainfall data (Seo et al. 1990a). We assume that mean and variance of gage rainfall are uncertain (but, correlation and cross-correlation functions are perfectly known) whereas all the second order statistics of radar rainfall are assumed perfectly known. Although we build estimators following the framework of the classical Bayesian estimation (Raiffa and Schlaifer 1961; Zellner 1971), our interpretation is that of a frequentist's in that the estimators are viewed as derived from random coefficient regression models (Johansen 1984). This point will be made clearer when we assess the prior in Part II of the paper (Seo and Smith, this issue).

This paper is Part I of a two-part series. In Part I, we describe the problem, describe parameter updating using raingage measurements and radar rainfall data, construct estimators, discuss issues, and give conclusions. In Part II (Seo and Smith, this issue), an actual application is given.

2 Problem description

The problem is to estimate rainfall depth at an arbitrary location, \( Z_{go} \), using raingage measurements, \( Z_g=(Z_{g1},...,Z_{gn})^T \), and radar rainfall data, \( Z_r=(Z_{r1},...,Z_{rm})^T \). Temporal scale of rainfall accumulation is about an hour. By a radar rainfall datum, we mean a spatial average over a radar bin of rainfall estimates from equivalent reflectivity factor (for details, see Part II, this issue). In addition to \( Z_g \) and \( Z_r \), past radar rainfall data and raingage measurements are also available. The subscript "g" in \( Z_{go} \) signifies that \( Z_{go} \) is the rainfall depth that would be measured by a raingage if one existed at that location. It is thus implicitly assumed that raingage measurements are more accurate than radar rainfall data. At this point, a distinction between gage rainfall and ground-truth rainfall is in order. Ideally we would like to estimate ground-truth rainfall rather than gage rainfall. Estimation of ground-truth rainfall, however, requires that the sensors' error structures be known. Also, one must have error-free ground-level rainfall measurements for validation purposes. For these reasons, we estimate gage rainfall rather than ground-truth rainfall. Our goal is then to evaluate \( E[Z_{go} | Z_g, Z_r] \) and \( Var[Z_{go} | Z_g, Z_r] \).

We treat only positive rainfall as the realization of rainfall process, and thus zero rainfall is excluded from analysis. Defining the area of positive rainfall, \( \Omega \), however, may not be straightforward since the sensors have limited measurement resolution and minimum detectable rainfall depth of the sensors may be different. In this work, a common minimum detectable rainfall depth is assumed, and the area of positive rainfall is set equal to the area of positive radar rainfall. Therefore, if there is a raingage measurement of positive rainfall outside of the radar rainfall field, that raingage measurement is excluded. Likewise, if there is a radar rainfall datum of positive rainfall but the coinciding raingage measurement reports no rainfall, that radar rainfall datum is excluded.