Conditional Bias in Radar Rainfall Estimation

GRZEGORZ J. CIACH AND MARK L. MORRISSEY

Environmental Verification and Analysis Center, University of Oklahoma, Norman, Oklahoma

WITOLD F. KRAJEWSKI

Iowa Institute of Hydraulic Research, The University of Iowa, Iowa City, Iowa

(Manuscript received 23 August 1999, in final form 7 December 1999)

ABSTRACT

The goal of this study is to improve understanding of the optimization criteria for radar rainfall (RR) products. Conditional bias (CB) is formally defined and discussed. The CB is defined as the difference between a given rain rate and the conditional average of its estimates. A simple analytical model is used to study the behavior of CB and its effect on the relationship between the estimates and the truth. This study shows the measurement errors of near-surface radar reflectivity and the natural reflectivity–rainfall rate variability can affect CB. This RR estimation error component is also compared with the commonly used mean-square error (MSE). A dilemma between the minimization of these two errors is demonstrated. Removing CB from the estimates significantly increases MSE, but minimizing MSE results in a large CB that manifests itself in underestimation of strong rainfalls.

1. Introduction

The goal of radar rainfall (RR) estimation is to provide approximate rainfall values that are as close to the physical truth as possible. Here, the *true rainfall* is defined as the amount of rainwater that falls on a specified area of the earth's surface in a specified time interval. The question is how to formalize rigorously the requirement of "closeness to the truth." The abundance of the existing reflectivity–rainfall rate (*Z–R*) relationships (Battan 1973) suggests that an unambiguous way to fulfill this apparently obvious demand might not exist. In this paper, it is shown that indeed there are at least two RR estimation performance criteria that are in opposition to each other. One is the commonly used mean-square error (MSE), and the other is a conditional bias (CB).

Generally, CB (it can also be called conditional mean error) can be described as a quantity that shows how the radar rainfall estimates differ, on the average, from a given true rainfall. Formally, we define it as a statistic that quantifies discrepancies between a true rain rate and the conditional average of its estimates, where the averaging is done only for the cases close to the given

true rain rate (the conditioning). Large CB can manifest itself in the severe underestimation of strong rainfalls. Demonstrating the behavior of this new RR error component and the role it plays in RR estimation is the main purpose of this study. We apply a conceptual model proposed by Ciach and Krajewski (1999) to illustrate how CB affects the relationship between the RR estimates and the truth and how CB and MSE compete with each other.

The CB problem is fairly well recognized in the statistical literature that discusses the "error-in-variable" effects (Carroll et al. 1995; Seber 1989; Fuller 1987). It appears when the least square regression is used directly to build a predictive relationship in a situation where the predictors are measured with substantial errors. The resulting predictions (by definition) minimize MSE but also systematically underestimate the outcomes in their higher range and overestimate them in their lower range. This systematic discrepancy is known in the statistical literature as an "attenuation" effect, because it can be viewed as significant reduction of the dynamic range of the response. Recently, Rosenfeld and Amitai (1998) encountered the conditional bias problem in the RR context. In their paper, they demonstrate the effect using disdrometer data and argue that their "window probability matching method" removes it. In our study, we attempt to formalize the CB description and to develop an analytical insight into the problem.

This paper is organized as follows. In section 2, we briefly define and discuss the model and the RR esti-

Corresponding author address: Grzegorz J. Ciach, Environmental Verification and Analysis Center, University of Oklahoma, Norman, OK 73069.
E-mail: gciach@ou.edu

mation formulas used in this study. Next, we define CB and demonstrate its effect on the RR estimates. In section 4, we compare and discuss the CB and MSE errors. A summary and conclusion section ends the study.

2. The model framework

The conceptual model of radar reflectivity and rain-rate measurements that we use here is a simplified version of the model described and discussed in detail in Ciach and Krajewski (1999). Let us consider a precipitation system in which the relationship between the true near-surface rain rate R_t and the true near-surface radar reflectivity Z_t can be described by a one-to-one functional dependency of power-law type:

$$Z_t = AR_t^b, \qquad (1)$$

where the coefficients A and b are called the relationship multiplier and exponent, respectively. The true reflectivity factor Z_t is a physical descriptor of the precipitation volume, as defined by Battan (1973). Both the true rain rate R_t and reflectivity Z_t are averaged over the same area over which the radar reflectivity is measured. Also, let the measured radar reflectivity Z_m, corrupted with errors, be associated with Z_t through the following equation:

$$Z_m = Z_t E_z, \qquad (2)$$

where E_z is the reflectivity measurement error. From Eqs. (1) and (2), we can obtain the "observation equation" of the modeled system,

$$Z_m = AR_t^b E_z, \qquad (3)$$

which binds the measured Z_m to the true rain rates and the errors. We define R_t and E_z as the following independent and lognormally distributed random variables:

$$R_t \sim \mathrm{LN}(1, \mathrm{CV}_r) \quad \text{and} \qquad (4a)$$

$$E_z \sim \mathrm{LN}(1, \mathrm{CV}_{ez}), \qquad (4b)$$

where CV_r are CV_{ez} are the coefficients of variation of the rain-rate process and the reflectivity measurement errors, respectively. Note that, in the above definitions, to simplify further notation and without any significant loss of generality, we assume that the ensemble means of the two random variables are equal to 1. For the R_t values, it is equivalent to converting them into dimensionless quantities by scaling the rain rates with their climatological average. This idealized model is designed to be analytically tractable and to describe qualitatively only those elements of reality that are crucial for the conditional bias question. As a consequence, its solutions can be obtained in a synthetic form of closed mathematical expressions. For a discussion of the model assumptions, see Ciach and Krajewski (1999).

One of the assumptions of the model used here is a fixed power-law Z_t–R_t relationship such as Eq. (1) between the true reflectivity and rain rate. However, our

analytical framework is, in fact, more flexible, and this assumption can be relaxed. We can incorporate a simple Z–R variability model into the multiplicative factor A in Eq. (1), assuming that the A parameter is also a lognormal random variable:

$$A = A_o E_a \quad \text{and} \qquad (5a)$$

$$E_a \sim \mathrm{LN}(1, \mathrm{CV}_{ea}), \qquad (5b)$$

where A_o is the average value of the Z_t–R_t relationship multiplier A, and the independent lognormally distributed random variable E_a describes the variability of this multiplier in a real precipitation system. Substituting definition (5a) into Eq. (3), one obtains the modified observation equation

$$Z_m = A_o R_t^b (E_a E_z), \qquad (6)$$

where the natural Z–R variability factor E_a and the reflectivity measurement error E_z are mixed in a single error factor. The product of E_a and E_z is also a lognormally distributed random variable, independent of R_t, describing the combined effect of the two uncertainty sources in our rainfall observation system. It suggests that the natural Z–R variability can be considered in a simplified way as an increment of the radar reflectivity measurement error, with all its consequences for the CB error discussed in the next sections.

The RR estimation method that we model here is based on a conversion of the radar-measured reflectivities Z_m of a precipitation system into radar-estimated rain rates R_r. For this conversion, we need to choose a Z_m–R_r relationship adjusted to a specific situation. The R_r values obtained this way are approximations of the true rain rates R_t averaged over the area over which the radar reflectivities are measured. Within our model framework, this predictive Z_m–R_r relationship is also a power-law function preserving the lognormality of the system. It can be represented in a traditional way as a Z–R relationship:

$$Z_m = \alpha R_r^\beta, \qquad (7)$$

which, for the RR estimation purpose, is rearranged to a mathematically equivalent form,

$$R_r = \alpha^{-1/\beta} Z_m^{1/\beta}, \qquad (8)$$

that is applied to perform the necessary Z_m-to-R_r conversion.

We can substitute the observation equation [Eq. (3)] into this predictive equation to express the dependence of the RR estimate on the true rainfall and the errors:

$$R_r = (A/\alpha)^{1/\beta} R_t^\gamma E_z^{1/\beta} = c R_t^\gamma E_z^{1/\beta}, \qquad (9)$$

where the denotations for the exponent ratio $\gamma = b/\beta$ and multiplier c are introduced to simplify further derivations. This equation is a basis for our further analysis of the behavior of the RR estimator [Eq. (8)]. Our focus is on the exponent parameters b and β that, as Eq. (9) shows, govern the way the radar estimates R_r are related
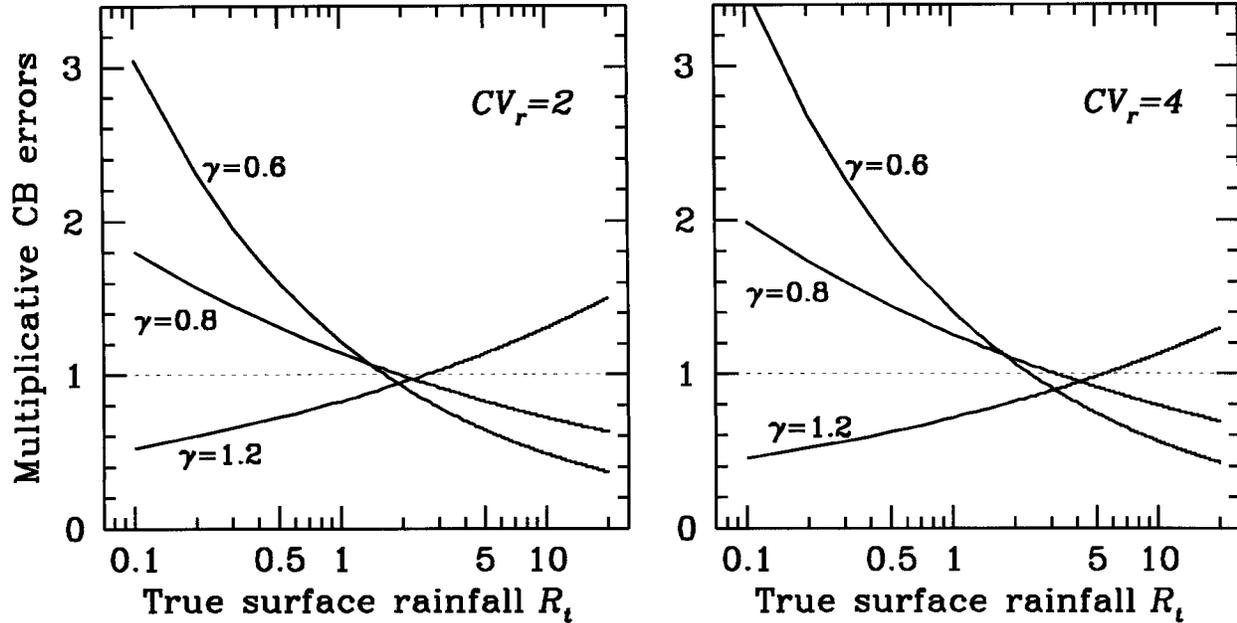
FIG. 1. Multiplicative conditional bias as a function of given true rain rate. Two cases of true-rain-rate variability coefficients are shown, $CV_r = 2$ (left panel) and $CV_r = 4$ (right panel). In each panel, conditional biases for three values of the exponent ratios $\gamma = b/\beta$ are presented (descriptions are at the corresponding curves).

to different intensities R_t of the true rain rate. First, however, let us consider the multiplier α. For any fixed β, α can be adjusted so that the overall bias of the RR estimation is removed, and the climatological averages of R_r and R_t are equal (Ciach et al. 1997). Representing these averages in our model with statistical expectations, we obtain:

$$E\{R_r\} = E\{R_t\}, \tag{10}$$

which leads to the following expression for the multiplier in Eq. (9):

$$c = \frac{1}{E\{R_t^\gamma\}E\{E_z^{1/\beta}\}}, \tag{11}$$

significantly simplifying all further derivations.

3. Definition of CB and its effect on RR estimates

The CB can be formally defined with either of the following two expressions:

$$CB_m(r_t) = \frac{E\{R_r | R_t = r_t\}}{r_t} \quad \text{or} \tag{12a}$$

$$CB_a(r_t) = E\{R_r | R_t = r_t\} - r_t, \tag{12b}$$

where r_t is a given R_t value (the conditioning), CB_m is the multiplicative conditional bias, and CB_a is its additive equivalence. Both expressions tell us how the average RR for a given R_t differs from this fixed true rain rate. Equation (12a) expresses it as a dimensionless coefficient, and Eq. (12b) expresses it in absolute rain-rate units. Both statistics can be useful in different sit-

uations, and we will utilize them in our further analysis. The CB_a averaged over all the r_t values is equal to zero. This fact is due to the unbiasedness condition [Eq. (10)] and the "double expectation theorem" [see, e.g., p. 6 in Bickel and Doksum (1977)]. Note, however, that the respective average CB_m is not equal to 1, because the expectation of a ratio is not the same as the ratio of expectations.

In the previous section, we already have established the components necessary to analyze the properties of the CB within our model. They can be applied to express it explicitly as a function of given true rain rate and the model parameters. For the multiplicative version, after substituting Eqs. (9) and (11), and after several rearrangements, we obtain

$$CB_m(r_t) = \frac{r_t^{\gamma-1}}{E\{R_t^\gamma\}} = (CV_r^2 + 1)^{\gamma(1-\gamma)/2}r_t^{\gamma-1}. \tag{13}$$

Note that for the Z_m–R_r conversion exponent β equal to the model assumed b, that is, for $\gamma = 1$, the exponents in Eq. (13) vanish and $CB_m(r_t) = 1$ for all r_t values. In this, and only this, case we obtain conditionally unbiased rain-rate estimates

$$R_r = cR_t E_z^{1/b} \tag{14}$$

that relate to the true rainfall in a linear way. For all other values of β, the average RR estimate conditioned on a specified rain rate r_t differs from this truth. This nonlinear departure is a function of r_t and can be described by $CB_m(r_t)$ not only within this model but also in analysis of results based on real data, provided that we can estimate it.
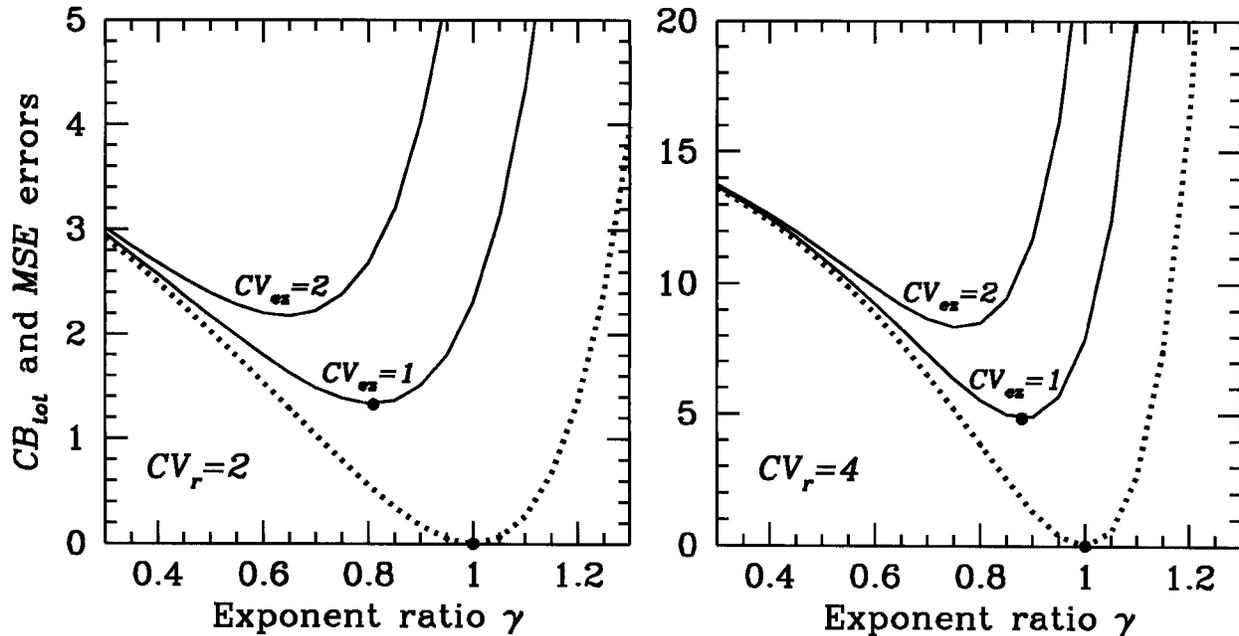
FIG. 2. Behavior of the total conditional bias and MSE as a function of the exponent ratio γ. The CB_{tot} is represented by the dotted curves. The solid curves show MSE for two values of the variation coefficient CV_{ez} of the reflectivity measurement errors (descriptions are at the corresponding curves). Two cases of true-rain-rate variability coefficients are considered, $CV_r = 2$ (left panel) and $CV_r = 4$ (right panel).

Equation (9) and the overall nonbias condition Eq. (10) imply that, if we choose $\beta > b$ (i.e., $\gamma < 1$), then the R_r values obtained based on Eq. (8) underestimate high rain rates and overestimate low rain rates. Statistical literature on the error-in-variable problem (e.g., Fuller 1987) calls this effect attenuation. Note that statistical meaning of the term attenuation (pertaining to the excessive decrease of the estimation's dynamic range) must not be confused with its use in radar meteorology, where it means the extinction of electromagnetic waves in heavy precipitation. For $\beta < b$ it works in the opposite direction, causing on the average an "enhancement" of the dynamic range of RR estimates in comparison with the truth. Equation (13) shows that the magnitude of the effect depends on the actual rain-rate intensity r_t, the climatological rain-rate variability CV_r, and the difference between the predictive and the physical Z–R exponents.

In Fig. 1, we demonstrate the behavior of the multiplicative conditional bias defined by Eq. (12a), as a function of the true rain rate, for a few cases of the assumed model parameters. One can see that using β, which is significantly larger than the physical Z–R exponent b (the case of $\gamma = 0.6$), leads, on the average, to underestimation of the high rainfalls by a factor of about 2. To maintain zero overall bias, this underestimation is compensated by the 2 to 3 times overestimation of the weak rainfalls. As a result, the dynamic range of the intensities is much smaller for the RR estimates than for the true rain rates. A comparison of the left and right panels in Fig. 1 shows that this result is

not affected much by the climatological variability of the rain rates. If the predictive exponent is smaller than b (e.g., $\gamma = 1.2$), the opposite effect occurs.

4. Comparison of CB and MSE

As we have already stated, we want the predictive Z_m–R_r relationship [Eq. (7)] to result in RR estimates that are as close to the true rain rates as possible. This goal can be achieved through an appropriate adjustment of its parameters α and β. An obvious component of this closeness is the overall nonbias requirement [Eq. (10)] that can be fulfilled by an adjustment of the α multiplier, as shown in section 2. The remaining differences between R_r and R_t can be evaluated in different ways. One meaningful criterion is the multiplicative conditional bias $CB_m(r_t)$ defined by Eq. (12a). We used it to describe the systematic differences between the RR estimators for low versus high rainfalls. Using Eq. (12b), one can construct a statistic that can be applied to compare CB with MSE. The MSE is a traditional statistic, often used to characterize the random estimation errors. Within our model framework, MSE can be simply defined as follows.

$$\text{MSE} = \text{Var}\{R_r - R_t\} = E\{(R_r - R_t)^2\}, \quad (15)$$

where the variance Var is equal to the mean square of the RR–truth difference from the unbiasedness condition [Eq. (10)].

To compare MSE in a meaningful way with the CB effect, we must first express the conditional bias using

a statistic that is compatible with MSE. The multiplicative and additive CBs discussed before are functions of given true rain rate. On the other hand, the MSE for given parameters of the system is a number and cannot be compared directly with a function. Thus, we need to reduce the function to a measure that characterizes globally its magnitude. The appropriate measure is a quantity that we can call a "total additive conditional bias," given by

$$\text{CB}_{\text{tot}} = \text{Var}\{\text{CB}_a(R_t)\} = E\{\text{CB}_a(R_t)^2\}, \quad (16)$$

where the definition of the additive CB is given by Eq. (12b). Again, the variance is equal to the mean square because the expectation of the $\text{CB}_a(R_t)$ is equal to zero, as stated in the previous section. The CB_{tot} measures the conditional bias effect synthetically for all of the r_t values, is an analytically tractable quantity, and, being also a variance, is the same kind of statistic as MSE. Formally, it is possible to define many global measures of the CB error. However, the above advantages make the CB_{tot} a good choice for our purposes.

To compare the behavior of MSE and CB_{tot}, we first need to derive explicit expressions for the two statistics as functions of the Z_m–R_r exponent β and other parameters of our model. For MSE, we expand Eq. (15) and substitute Eq. (9) to obtain

$$\text{MSE} = c^2 E\{R_t^{2\gamma} E_z^{2/\beta}\} - 2cE\{R_t^{\gamma+1} E_z^{1/\beta}\} + E\{R_t^2\}. \quad (17)$$

This equation, after replacing c with Eq. (11) and several rearrangements, yields

$$\text{MSE}(\gamma) = [(\text{CV}_r^2 + 1)(\text{CV}_{ez}^2 + 1)^{1/b^2}]\gamma^2 - 2(\text{CV}_r^2 + 1)^\gamma + \text{CV}_r^2 + 1, \quad (18)$$

which, for any coefficients of variation CV_r and CV_{ez} and exponent b, is a function of the argument $\gamma = b/\beta$.

We now proceed with the total conditional bias defined with Eq. (16). The additive bias definition Eq. (12a), after substituting Eqs. (9) and (11), can be expanded as follows:

$$\text{CB}_a(r_t) = cr_t^\gamma E\{E_z^{1/\beta}\} - r_t = \frac{r_t^\gamma}{E\{R_t^\gamma\}} - r_t. \quad (19)$$

This equation, after substituting to Eq. (16) and several rearrangements, leads to the following result:

$$\text{CB}_{\text{tot}}(\gamma) = (\text{CV}_r^2 + 1)^{\gamma^2} - 2(\text{CV}_r^2 + 1)^\gamma + \text{CV}_r^2 + 1, \quad (20)$$

which is similar to Eq. (18), except that it is free of the error term.

Analysis of functions $\text{MSE}(\gamma)$ and $\text{CB}_{\text{tot}}(\gamma)$ shows that they are both concave and have single minima. For $\gamma = 0$ (i.e., in the limit of large β values), they have the same value and $\text{MSE}(0) = \text{CB}_{\text{tot}}(0) = \text{CV}_r^2$. This result is natural because, in this case, the RR estimates are equal to the climatological average rain rate, and the error of such predictions is the same as the rainfall variability. For large values of γ, both functions go to infinity. The argument values minimizing $\text{MSE}(\gamma)$ and $\text{CB}_{\text{tot}}(\gamma)$, however, differ significantly. They can be found by differentiating Eqs. (18) and (20) and setting the derivatives equal to zero. The MSE has its minimum for the following value of the β parameter:

$$\beta_{\text{ms}} = b\left[1 + \frac{\ln(\text{CV}_{ez}^2 + 1)}{b^2 \ln(\text{CV}_r^2 + 1)}\right], \quad (21)$$

which agrees with the result in Ciach and Krajewski (1999) obtained based on different considerations. On the other hand, minimum of the total CB error [Eq. (20)] is for $\beta_{\text{cb}} = b(\gamma = 1)$, and, at this point, the conditional bias simply vanishes, that is, $\text{CB}_{\text{tot}}(1) = 0$. The minimum for the total conditional bias function is, of course, in perfect agreement with the results in the previous section regarding the multiplicative bias.

Note that the β_{ms} value that optimizes MSE is always bigger than b. For given physical parameters b and CV_r of the modeled precipitation system, the difference between β_{ms} and $\beta_{\text{cb}} = b$ minimizing the two error variances depends on the level of the reflectivity measurement errors described by CV_{ez}. Figure 2 demonstrates the relative behavior of the MSE and CB error expressions derived above, as a function of the γ parameter, for a few cases of the model parameters. In the left panel, one can see that, for moderate reflectivity measurement errors ($\text{CV}_{ez} = 1$), β_{ms} is about 22% larger than the β_{cb} exponent value, whereas, for large errors ($\text{CV}_{ez} = 2$), the difference is almost 40%. For the model with larger climatological rain-rate variability ($\text{CV}_r = 4$ in the right panel), the effect is slightly smaller.

From Fig. 2 one can immediately see the dilemma that arises when the β exponent of the predictive Z_m–R_r conversion is tuned to make the RR estimates as close as possible to the truth. Inherently, one cannot simultaneously minimize both the MSE and CB_{tot} errors. Minimization of MSE (the most common statistical optimization method) leads to substantial conditional bias distortions. On the other hand, choosing an RR estimator that has no conditional bias ($\text{CB}_{\text{tot}} = 0$ for $\gamma = 1$) can only be done at the cost of a significant increase of MSE. For moderate reflectivity measurement errors, MSE is larger than its minimum by a factor of about 1.7. However, for higher measurement errors, MSE is about 3 times larger for the conditionally unbiased estimates than at the minimum point and is even bigger than MSE at $\gamma = 0$ (climatological average prediction).

Note that, in this framework, there is no conflict of the overall unbiasedness condition and CB_{tot} (or MSE). This fact is most likely also true in real data setups. Not all performance criteria compete, even if they pertain to different aspects of a product. The knowledge of which criteria are in conflict and which can be adjusted independently is necessary to control the optimization process properly.

5. Summary and conclusions

We investigated a new radar rainfall estimation performance criterion, called the conditional bias error, using a simple analytical model. After a short presentation of the model, we formally defined two useful descriptors of CB. We then demonstrated the behavior of this quantity and showed how it affects the relationship between the RR estimates and the true rain rates. We also compared the conditional bias with the commonly used mean square estimation error, pointing out an acute dilemma between reduction of CB and minimization of MSE. We showed that the natural variability of the physical Z–R relationship can add to the CB effect in the same way as the reflectivity measurement errors.

Our study shows that nonzero CB in the RR products indicates that the RR values are not linearly related to the corresponding true rainfall values and that CB_{tot} is a general and convenient global measure of the magnitude of this distortion that can also be used for real data analysis. It is an undesirable systematic effect that one would like to remove from the RR products. However, the cost of removing CB from the RR estimates can be high in terms of the MSE increase. On the other hand, minimizing MSE can result in strong underestimation of heavy rainfalls and overestimation of weak rain rates. This dilemma is an inevitable consequence of the uncertainties involved in the RR estimation process and cannot be solved with the uncertainties present. Finding a compromise between the two extreme RR estimation strategies most likely requires careful recognition of the prevailing priorities in a specific application of the hydrological radar–rainfall products. Such a discussion is beyond the scope of this paper, and the tradeoff between MSE and CB is considered here only in a descriptive way as a fact of nature. However, in our opinion, one should be aware of the conditional bias effects and be able to quantify them in order to make conscious decisions about the RR estimation schemes used in practice. Practical problems of efficient estimation of the CB errors using uncertain reference data will be one of the topics of our future research.

We hope that this study will promote better understanding and more rigorous definitions of the optimality criteria for the RR estimation/validation procedures. It may also be viewed as the next step toward a parsimonious model of the RR error structure and the efficient estimation of its parameters. Pursuing these difficult goals will also be a subject of our future research effort.

REFERENCES

Battan, L. J., 1973: *Radar Observation of the Atmosphere.* University of Chicago Press, 324 pp.

Bickel, P. J., and K. A. Doksum, 1977: *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden–Day, 492 pp.

Carroll, R. J., D. Ruppert, and L. A. Stefanski, 1995: *Measurement Error in Nonlinear Models.* Chapman and Hall, 305 pp.

Ciach, G. J., and W. F. Krajewski, 1999: Radar–rain gauge comparisons under observational uncertainties. *J. Appl. Meteor.,* **38,** 1519–1525.

——, ——, E. N. Anagnostou, M. L. Baeck, J. A. Smith, J. R. McCollum, and A. Kruger, 1997: Radar rainfall estimation for ground validation studies of the Tropical Rainfall Measuring Mission. *J. Appl. Meteor.,* **36,** 735–747.

Fuller, W. A., 1987: *Measurement Error Models.* John Wiley and Sons, 440 pp.

Rosenfeld, D., and E. Amitai, 1998: Comparison of WPMM versus regression for evaluating Z–R relationships. *J. Appl. Meteor.,* **37,** 1241–1249.

Seber, G. A. F., 1989: *Nonlinear Regression.* John Wiley and Sons, 768 pp.