

Evaluation of Some Ground Truth Designs for Satellite Estimates of Rain Rate

EUNHO HA

Department of Statistics, Yonsei University, Wonju, Korea

GERALD R. NORTH

Department of Atmospheric Sciences, Texas A&M University, College Station, Texas

CHULSANG YOO

Department of Environmental Engineering, Korea University, Seochang, Korea

KYUNG-JA HA

Department of Atmospheric Sciences, Pusan National University, Pusan, Korea

(Manuscript received 11 December 2000, in final form 22 June 2001)

ABSTRACT

In this paper point gauge measurements are analyzed as part of a ground truth design to validate satellite retrieval algorithms at the field-of-view spatial level (typically about 20 km). Even in the ideal case the ground and satellite measurements are fundamentally different, since the gauge can sample continuously in time but at a discrete point, while a satellite samples an area average but a snapshot in time. The design consists of comparing a sequence of pairs of measurements taken from the ground and from space. Since real rain is patchy, that is, its probability distribution has large nonzero contributions at zero rain rate, the following ground truth designs are proposed. Design 1 uses all pairs. Design 2 uses the pairs only when the field-of-view satellite average has rain. Design 3 uses the pairs only when the gauge has rain. For the nonwhite noise random field having a mixed distribution, the authors evaluate each design theoretically by deriving the ensemble mean and the mean-square error of differences between the two systems. It is found that design 3 has serious disadvantage as a ground truth design due to its large design bias. It is also shown that there is a relationship between the mean-square error of design 1 and design 2. These results generalize those presented recently by Ha and North for the Bernoulli white noise random field. The strategy developed in this study is applied to a real rain rate field. For the Global Atmospheric Program (GARP) Atlantic Tropical Experiment (GATE) data, it is found that by combining 50 data pairs (containing rain) of the satellite to the ground site, the expected error can be reduced to about 10% of the standard deviation of the fluctuations of the system alone. For the less realistic case of a white noise random field, the number of data pairs is about 100. Hence, the use of more realistic fields means that only about half as many pairs are needed to detect a 10% bias.

1. Introduction

Since rainfall is highly variable in both space and time, measurements from existing sparse networks of rain gauges may not be representative of rainfall generally. Satellite rainfall measurements, by contrast, are attractive because they promise information about rainfall rates on a nearly global basis. Several satellites are orbiting to measure precipitation especially in remote areas like the tropical oceans (see Simpson et al. 1988; Wilheit 1991; Theon et al. 1992; Thiele 1992). The

Tropical Rainfall Measuring Mission (TRMM), which was launched in November 1997, has already carried out several years of measurements of rainfall. Among the many issues concerning TRMM satellite data processing (Simpson et al. 1988; McConnell and North 1987; Shin and North 1988; Kedem et al. 1990; North et al. 1991), the ground truth problem is one of the most critical because its solution will provide information about comparative accuracy of gauge (or radar) and satellite measurements.

There are various forms of ground validation for precipitation. The ground-based or airplane-based radars will be the ultimate "bearer" of ground truth, but the algorithms for radar backscatter estimation of rain rate remain controversial. The advantage of the point gauge ground truth is that it does not introduce the inevitably

Corresponding author address: Dr. Gerald R. North, Department of Atmospheric Science, Texas A&M University, MS 3150, College Station, TX 77843-3150.
E-mail: northhead@ariel.tamu.edu

controversial algorithms associated with estimating the rain rate from other “ground truth” measurements such as that derived from radar. We thus present here an analysis of validating the satellite estimations with point gauge measurements.

A problem in the satellite estimation of rain rate is that the measurement taken by the satellite sensor is fundamentally different from the point gauge measurement. This is because the satellite measures a snapshot in time (actually about a 5–10-min average) of an area average over its field of view (FOV), while the point gauge measures precipitation nearly continuously in time. For an individual measurement pair of satellite and gauge, there is likely to be a large random difference between the two. North et al. (1994) and Ha and North (1994) developed a ground validation strategy that reduces the random errors by taking enough simultaneous pairs of measurements.

Since the probability distribution of real rain has a large contribution at zero rain rate (usually greater than 90%), the measurement pairs can be (no rain, no rain), (rain, no rain), and (rain, rain), where the first entry is the satellite measurement and the second one is the gauge measurement. For this reason, it is necessary for us to decide whether to throw out no-rain measurements from the stream of data pairs. Based on the method of throwing out the no-rain measurements, Ha and North (1999) proposed the following three ground truth designs.

- Design 1 uses all data pairs of the two measurements (gauge, satellite) from all visits.
- Design 2 throws out all the data pairs when the FOV has no rain. This design uses only data pairs (rain, no rain) and (rain, rain).
- Design 3 throws out all the data pairs when the gauge has no rain. This design uses only data pairs (rain, rain).

Ha and North (1999) derived the error distribution and statistics (mean and mean-square error) with a spatially white noise Bernoulli random field for each of the designs mentioned above. The major finding in Ha and North (1999) was that design 3 cannot be used as a ground truth design due to its large design bias. It was also shown that there is a relationship between the mean-square error of design 1 and design 2 for the Bernoulli random field. Since actual rain is not a Bernoulli random field, the results in Ha and North (1999) may have limited applicability to the real situation.

In this study, a spatially non-white noise homogeneous random field having the mixed distribution is used as the rain-rate model. This means that it is raining or not raining at all with prescribed probability. This model is more general in its ability to describe the real rain compared to the simple Bernoulli model in Ha and North (1999). We derive the ensemble mean and the mean-square error for each design and thus we evaluate which scheme is the most appropriate as a ground truth design.

At the same time, we examine the relationship between the ground truth designs proposed here.

To gain a better understanding of the scheme developed in this paper, we check the applicability of the ground truth design developed in this study for the non-white noise random field using Global Atmospheric Research Program (GARP) Atlantic Tropical Experiment (GATE) data and evaluate each of the three designs proposed in this paper.

2. Definitions

Consider a random field $\psi(\mathbf{r}, t)$ defined in the $\mathbf{r} = (x, y)$ plane and along the time axis t . Let the ensemble average of $\psi(\mathbf{r}, t)$ be $\langle \psi(\mathbf{r}, t) \rangle$ and its variance at a point in space be σ^2 . The random variable $\psi(\mathbf{r}, t)$ is assumed to be weakly statistically homogeneous in space and time; that is, the lagged covariance is a function only of $\xi = |\mathbf{r} - \mathbf{r}'|$ and $\tau = |t - t'|$. Since rain rates are inherently patchy, a mixed distribution such as the mixed lognormal distribution is sometimes used as a model for the distribution of rain rates (Kedem et al. 1990) because this distribution includes the possibility of the no-rain phenomenon characteristic of real rain. That is, the rain rate has a mixed distribution such as the mixed lognormal distribution at a point in space and time.

The random variable $\psi(\mathbf{r}, t)$ has a positive probability $1 - p$ for the event $\{\psi(\mathbf{r}, t) = 0\}$, but otherwise $P[\bar{\psi}(\mathbf{r}, t) = r] = 0, r > 0$.

More precisely, let G be the cumulative distribution function of $\psi(\mathbf{r}, t)$, $P[\psi(\mathbf{r}, t) \leq r]$. Then it can be represented as a combination of two increasing functions H and F :

$$G(r) = (1 - p)H(r) + pF(r), \quad (1)$$

where

$$H(r) = 0, \quad r < 0, \quad (2)$$

$$H(r) = 1, \quad r \geq 0, \quad (3)$$

and F is a continuous distribution function such that $F(r) = 0, r \leq 0$, with a density $f(r) = F'(r) \geq 0, r > 0$.

For the problem posed here there is one gauge located at some fixed site \mathbf{r}_g . The satellite passes over the site, and one of the FOVs in the swath along the ground track covers the gauge. The satellite measurement based on R , which is the FOV for a visit, is

$$\Psi_s = \frac{1}{A} \int_R \psi(\mathbf{r}, t) d^2\mathbf{r}, \quad (4)$$

where R is the FOV of this visit and A is the area of R . The instantaneous precipitation rate from this gauge measurement is

$$\Psi_g = \psi(\mathbf{r}_g, t) = \frac{1}{A} \int_R \psi(\mathbf{r}, t) K(\mathbf{r}) d^2\mathbf{r}, \quad (5)$$

where $K(\mathbf{r}) = A\delta(\mathbf{r} - \mathbf{r}_g)$.

For a satellite such as TRMM, the location of the centroid of the FOV with respect to the position of the point gauge can be considered as uniformly distributed in a space from one visit to another. Equivalently, we take the centroid to be fixed and consider the gauge location to be a random variable uniformly distributed over the FOV:

$$P(\mathbf{r}_g) = \begin{cases} \frac{1}{A} & \text{when } \mathbf{r}_g \in R \\ 0 & \text{otherwise.} \end{cases}$$

In practice, the microwave radiometer estimates the rain rate by measuring the upwelling radiation from an atmospheric column. In the ideal case this represents a column average of the rain rate over the projection of the FOV onto the ground for the sensor. Such a measurement is not compared with the instantaneous rain rate at the surface but rather some kind of time average at the surface, since it takes raindrops several minutes to fall from the top of the column to the surface. This is very fortunate since instantaneous rain rates are notoriously variable in space and the effect of time averaging is to smooth considerably the spatial variability of the field. Hence, the comparison can be made between the satellite estimate for an FOV and a few-minute time-averaged measurement from a rain gauge. In this paper, the ground and satellite measurements can be instantaneous or time-averaged measurements and are usually denoted by Ψ_g and Ψ_s .

Two measurements are taken from the n th visit, Ψ_s^n and Ψ_g^n , where the subscripts denote the satellite and gauge, respectively. As outlined in the introduction, design 1 makes use of the measurements (Ψ_s^n, Ψ_g^n) of all visits. Design 2 uses only the measurements (Ψ_s^n, Ψ_g^n) for $\Psi_s^n > 0$ and design 3 uses (Ψ_s^n, Ψ_g^n) for $\Psi_g^n > 0$. We typically denote the satellite and gauge measurement for each design (Ψ_{si}, Ψ_{gi}) where $i = 1, 2, 3$ corresponds to the design index.

The error ϵ_{di}^n for the n th data pair is formed by $\epsilon_{di}^n = \Psi_{si}^n - \Psi_{gi}^n$. The mean-square error, which is usually used as an index of the accuracy in comparing the ground measurement to the satellite measurement, is defined by

$$\langle \epsilon_{di}^2 \rangle = \langle (\Psi_{si} - \Psi_{gi})^2 \rangle. \quad (6)$$

If the members of the sequence of pairs of measurements are statistically independent [autocorrelation times for a typical area the size of an FOV are only an hour or so, whereas revisit intervals for a satellite are an order of magnitude longer; e.g., Bell et al. (1990)], we can sharpen the histogram of the difference between the satellite measurement and the gauge measurement by adding independent measurements of several visits. The mean of the average error for N data pairs is the same as the mean of error for one data pair. The variance of average error for N independent data pairs is

$$\langle \epsilon_N^2 \rangle = \frac{\langle \epsilon^2 \rangle}{N}. \quad (7)$$

The advantage of this strategy is that the error can be as small as we please by adding new measurements (i.e., by taking N sufficiently large).

3. Evaluation of some ground truth designs

a. Preliminary remarks

The ground truth design must satisfy two conditions to detect the *retrieval bias* of the retrieval algorithm we want to check. 1) The error $\epsilon_{di} = \Psi_{si} - \Psi_{gi}$ must have no bias; that is, $\langle \epsilon_{di} \rangle = \langle \Psi_{si} - \Psi_{gi} \rangle = 0$. If the error has this bias, which we call *design bias*, the retrieval bias and the design bias are combined and thus it is difficult to evaluate the retrieval bias we want to detect. 2) For the ground truth designs that satisfy $\langle \epsilon_{di} \rangle = 0$, the mean-square error $\langle \epsilon_{di}^2 \rangle$ of the difference between the ground measurement and the perfect satellite measurement should be comparable to or less than the size of biases in the satellite rain-rate algorithm we are trying to test.

The ensemble mean of the error for design 2 and design 3 can be written as follows:

$$\langle \epsilon_{d2} \rangle = \langle \Psi_{s2} - \Psi_{g2} \rangle = \langle \Psi_{s1} - \Psi_{g1} | \Psi_{s1} > 0 \rangle, \quad (8)$$

and

$$\langle \epsilon_{d3} \rangle = \langle \Psi_{s1} - \Psi_{g1} | \Psi_{g1} > 0 \rangle, \quad (9)$$

where $\langle Y | X > x \rangle$ denotes the conditional mean of Y , given that $X > x$. The mean-square error for design 2 and design 3 can be written as

$$\begin{aligned} \langle \epsilon_{d2}^2 \rangle &= \langle (\Psi_{s2} - \Psi_{g2})^2 \rangle \\ &= \langle (\Psi_{s1} - \Psi_{g1})^2 | \Psi_{s1} > 0 \rangle, \quad \text{and} \end{aligned} \quad (10)$$

$$\langle \epsilon_{d3}^2 \rangle = \langle (\Psi_{s1} - \Psi_{g1})^2 | \Psi_{g1} > 0 \rangle. \quad (11)$$

Equations (8) and (10) imply that one needs the information about the *conditional* random field given $\Psi_{s1} > 0$ to compute the statistics (ensemble mean and the mean-square error) for design 2. The same story holds for design 3 because this design considers the conditional random field given $\Psi_{g1} > 0$. Ha and North (1999) derived some statistics (ensemble mean and mean-square error) based on the conditional random field for the white noise Bernoulli random field and evaluated the ground truth designs proposed in this paper. Even though they used the most simple random field (spatially white noise Bernoulli random field), it was not an easy task to compute the necessary statistics for evaluating the designs based on the conditional random field. In general, it is difficult to have the spectral density or any other statistics of the conditional random field compared to obtaining those of the *unconditional* random field when the random field has the mixed distribution.

The error statistics for one ground truth design can be obtained from the error statistics of other designs if there is a relationship of the statistics (mean, variance, and mean-square error) among the ground truth designs.

We thus here seek the relationship of the statistics among the ground truth designs and provide a method of computing the statistics for design 2 and design 3 using the statistics of design 1.

When deriving statistical properties of the ground truth designs, we will use the well-known theorem for conditional expectation (Parzen 1962, p. 43). The theorem can be expressed as

$$\langle h(Y) \rangle = \sum_{\text{over } x} \langle h(Y) | X = x \rangle P(X = x). \quad (12)$$

This theorem says that the unconditional statistic is equal to the sum of the conditional statistic of all the disjoint events multiplied by their associated probability.

b. The bias of the error

Since the mean of the design error must be zero to unambiguously detect the retrieval bias, we thus first compute the ensemble mean of the error for each ground truth design to see which ground truth design is good for the validation of the satellite measurements. Since the random field is assumed to be homogeneous, the gauge and satellite measurements for design 1 have the same mean $\langle \Psi_{s1} \rangle = \langle \Psi_{g1} \rangle = \langle \psi(\mathbf{r}, t) \rangle$. The ensemble mean of error ϵ_{d1} is $\langle \epsilon_{d1} \rangle = 0$ and thus design 1 has no bias.

With the aid of the theorem, the gauge mean of design 1 can be written as

$$\begin{aligned} \langle \Psi_{g1} \rangle &= \langle \Psi_{g1} | I(\Psi_{s1}) = 1 \rangle P[I(\Psi_{s1}) = 1] \\ &\quad + \langle \Psi_{g1} | I(\Psi_{s1}) = 0 \rangle P[I(\Psi_{s1}) = 0] \\ &= \langle \Psi_{g1} | \Psi_{s1} > 0 \rangle P(\Psi_{s1} > 0) \\ &\quad + \langle \Psi_{g1} | \Psi_{s1} = 0 \rangle P(\Psi_{s1} = 0), \end{aligned} \quad (13)$$

where the notation

$$I(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x = 0 \end{cases} \quad (14)$$

is used to apply the theorem conveniently. When the perfect satellite measurement Ψ_{s1} has no rain, that is when $\Psi_{s1} = 0$, the gauge measurement also has no rain and thus $\langle \Psi_{g1} | \Psi_{s1} = 0 \rangle$ is zero. The relationship between the ensemble mean of the gauge measurements for design 1 and design 2 is

$$\langle \Psi_{g1} \rangle = \langle \Psi_{g2} \rangle P(\Psi_{s1} > 0) = \langle \Psi_{g2} \rangle P_s, \quad (15)$$

where P_s is used to denote the probability that the satellite measurement has rain inside the FOV. It is not easy to compute P_s for the non-white noise random field, but P_s can be written as $P_s = 1 - (1 - p)^A$ for the white noise random field, where A is the area of FOV. The probability P_s actually depends on the probability of rain p , the FOV size, and the correlation structure of the random field. Note that when the size of the FOV is large and/or the probability of rain is large, P_s

will be large. The satellite measurement of design 1 can be written as

$$\langle \Psi_{s1} \rangle = \langle \Psi_{s2} \rangle P_s. \quad (16)$$

The ground truth design 2 has no bias because

$$\begin{aligned} \langle \epsilon_{d2} \rangle &= \langle \Psi_{s2} - \Psi_{g2} \rangle = \langle \psi(\mathbf{r}, t) \rangle / P_s - \langle \psi(\mathbf{r}, t) \rangle / P_s \\ &= 0. \end{aligned} \quad (17)$$

By conditioning Ψ_{g1} on the random field in the theorem, the ensemble mean of gauge and satellite measurements of design 1 can be written as

$$\begin{aligned} \langle \Psi_{g1} \rangle &= \langle \Psi_{g1} | \Psi_{g1} > 0 \rangle P(\Psi_{g1} > 0) \\ &\quad + \langle \Psi_{g1} | \Psi_{g1} = 0 \rangle P(\Psi_{g1} = 0) = p \langle \Psi_{g3} \rangle, \end{aligned} \quad (18)$$

and

$$\langle \Psi_{s1} \rangle = p \langle \Psi_{s3} \rangle + \langle \Psi_{s1} | \Psi_{g1} = 0 \rangle (1 - p). \quad (19)$$

The notation $\langle \psi_+(\mathbf{r}, t) \rangle$ will be used as the conditional ensemble mean of $\psi(\mathbf{r}, t)$ when it rains; that is, $\langle \psi_+(\mathbf{r}, t) \rangle = \langle \psi(\mathbf{r}, t) | \psi(\mathbf{r}, t) > 0 \rangle$. The ensemble mean of the error for design 3 can be written as

$$\begin{aligned} \langle \epsilon_{d3} \rangle &= \langle \Psi_{s3} - \Psi_{g3} \rangle \\ &= \langle \psi_+(\mathbf{r}, t) \rangle - \frac{1 - p}{p} \langle \Psi_{s1} | \Psi_{g1} = 0 \rangle - \langle \psi_+(\mathbf{r}, t) \rangle \\ &= -\frac{1 - p}{p} \langle \Psi_{s1} | \Psi_{g1} = 0 \rangle \\ &= -\frac{1 - p}{p} \langle \psi(\mathbf{r}, t) | \psi(\mathbf{r}_g, t) = 0 \rangle. \end{aligned} \quad (20)$$

This result implies that $\langle \epsilon_{d3} \rangle$ can be zero only when the random field $\psi(\mathbf{r}, t)$ is all zero. However, since this is unrealistic for a real rain-rate field, design 3 always has a negative design bias. If the probability distribution of $\psi(\mathbf{r}, t)$ conditional on $\psi(\mathbf{r}_g, t) = 0$ for $\mathbf{r} \neq \mathbf{r}_g$ is known, the quantity $\langle \psi(\mathbf{r}, t) | \psi(\mathbf{r}_g, t) = 0 \rangle$ and thus the magnitude of the design bias can be calculated, but it is hard to know it. However, for a white noise random field we can compute the magnitude of the bias easily because the conditional distribution of $\psi(\mathbf{r}, t)$ given $\psi(\mathbf{r}_g, t) = 0$ is equal to the distribution of $\psi(\mathbf{r}, t)$ itself for $\mathbf{r} \neq \mathbf{r}_g$. Taking into account the random location of the gauge site over the FOV, we can compute the bias of design 3 for white noise random filed as

$$\langle \epsilon_{d3} \rangle = -(1 - p) \langle \psi_+(\mathbf{r}, t) \rangle. \quad (21)$$

Note that the bias formula in (21) is valid only for the ‘‘point’’ gauge measurement. Since the rain gauge actually measures the rain rate over the small spatial coverage $G \subset R$, not at a point \mathbf{r}_g , the bias for design 3 can be written as

$$\langle \epsilon_{d3} \rangle = -(1 - p) \langle \psi_+(\mathbf{r}, t) \rangle \left(1 - \frac{l}{A} \right), \quad (22)$$

where l is the area of G . This formula generalizes what Ha and North (1999) showed for the white noise Bernoulli random field. The bias of design 3 for the white noise random field is neatly factored into three parts, $(1 - p)\langle\psi_+(\mathbf{r}, t)\rangle$, and $1 - l/A$. The bias here has negative values and the magnitude (absolute value) of the bias increases when 1) the probability of rain p is small, 2) the rain rate conditional on rain $\langle\psi_+(\mathbf{r}, t)\rangle$ is large, and 3) the FOV size A is large. As discussed in section 2, the design bias for the ground truth design must be zero in order to detect the retrieval bias. Therefore, ground truth design 3, which throws out all visits when the gauge has no rain, has serious disadvantages.

c. Mean-square error

Section 3b showed that design 3 cannot be used as a ground truth design, so henceforth only design 1 and design 2 are considered as ground truth designs, whose mean-square errors will be calculated.

As discussed in section 3a, it is not easy to compute the mean-square error for design 2 [mse(d2), hereafter] based on the random field conditional on $\Psi_s > 0$. In this section, we thus try to give a method of computing mse(d2) using the mean-square error for design 1 [mse(d1), hereafter], which is the mean-square error for the unconditional random field. The spectral formalism (North and Nakamoto 1989) can be used to compute the mse(d1) for the random field. By the theorem in section 3a, the mse(d1) can be written as

$$\begin{aligned} \langle\epsilon_{d1}^2\rangle &= \langle(\Psi_{s1} - \Psi_{g1})^2 | \Psi_{s1} > 0\rangle P(\Psi_{s1} > 0) \\ &+ \langle(\Psi_{s1} - \Psi_{g1})^2 | \Psi_{s1} = 0\rangle P(\Psi_{s1} = 0) \\ &= \langle\epsilon_{d2}^2\rangle P_s \\ &+ \langle(\Psi_{s1} - \Psi_{g1})^2 | \Psi_{s1} = 0\rangle (1 - P_s). \end{aligned} \quad (23)$$

Since $(\Psi_{s1} - \Psi_{g1})^2$ is zero when Ψ_{s1} is zero, we have the relationship between mse(d1) and mse(d2) as

$$\langle\epsilon_{d1}^2\rangle = \langle\epsilon_{d2}^2\rangle P_s. \quad (24)$$

This equation shows that mse(d2) is equal to the mse(d1) divided by the probability that the satellite measurement has rain inside the FOV. Although Eq. (24) provides a way to compute the mse(d2) without using the conditional random field, the approach can be problematic because it is difficult to find P_s .

d. Number of visits

The dimensionless mean-square error $W^2(di)$, ($i = 1, 2$), which is defined as the mean-square error divided by the variance of gauge measurements, is used in this paper as an index of accuracy in estimating the ground measurement by the satellite measurement:

$$W^2(di) = \frac{\langle\epsilon_{di}^2\rangle}{\sigma_{gi}^2}, \quad i = 1, 2, \quad (25)$$

where σ_{gi}^2 ($i = 1, 2$) is the variance of the gauge measurement for design 1 and design 2, respectively. The σ_{g1}^2 is chosen as a normalization because it is a convenient quantity that can be measured independently. It is also natural that we would want our histogram of errors to be narrow compared to this histogram describing the local climatological variances. Using the theorem in section 3a again, we have

$$\langle\Psi_{g1}\rangle = \langle\Psi_{g2}\rangle P_s, \quad \text{and} \quad (26)$$

$$\begin{aligned} \langle\Psi_{g1}^2\rangle &= \langle\Psi_{g1}^2 | \Psi_{s1} > 0\rangle P(\Psi_{s1} > 0) \\ &+ \langle\Psi_{g1}^2 | \Psi_{s1} = 0\rangle P(\Psi_{s1} = 0) \\ &= \langle\Psi_{g2}^2\rangle P_s. \end{aligned} \quad (27)$$

Equations (26) and (27) gives the way of computing the variance of gauge measurements σ_{g2}^2 for design 2 based on the variance of gauge measurements σ_{g1}^2 for design 1:

$$\begin{aligned} \sigma_{g1}^2 &= \langle\Psi_{g1}^2\rangle - \langle\Psi_{g1}\rangle^2 = \langle\Psi_{g2}^2\rangle P_s - \langle\Psi_{g2}\rangle^2 P_s^2 \\ &= \sigma_{g2}^2 P_s - \left(1 - \frac{1}{P_s}\right) \langle\psi(\mathbf{r}, t)\rangle^2. \end{aligned} \quad (28)$$

As discussed in section 2, the dimensionless mean-square error of $N(di)$ independent visits is obtained by

$$\frac{W^2(di)}{N(di)}, \quad i = 1, 2. \quad (29)$$

The number of data pairs, $N(d1)$ for design 1 and $N(d2)$ for design 2, to achieve the given tolerance level of detecting the retrieval bias can be computed by Eq. (29). By letting $W^2(d1)/N(d1) = W^2(d2)/N(d2)$, we have the relationship between $N(d1)$ and $N(d2)$ as

$$N(d2) = \frac{N(d1)\sigma_{g1}^2}{\sigma_{g1}^2 - \left(\frac{1}{P_s} - 1\right) \langle\psi(\mathbf{r}, t)\rangle^2}. \quad (30)$$

From the equation, it can be seen that 1) the probability that the satellite measurement has rain P_s increases, $N(d2)$ will become close to $N(d1)$, and 2) if the mean rain rate increases, the difference between $N(d2)$ and $N(d1)$ will be large.

Because design 2 is designed to throw away all visits where the satellite measurement has no rain, it is worthwhile to ask how many visits, say $N_{\text{visits}}(d2)$, we need to have $N(d2)$ qualifying data pairs, which is the number of measurement pairs (satellite, gauge) we use when we apply design 2. In other words, $N_{\text{visits}}(d2)$ is the number of data pairs we use plus the number of visits (data pairs) we throw away for design 2. The total expected number of visits necessary to detect the retrieval bias becomes $N_{\text{visits}}(d2) = N(d2)/P_s$. Note that the design 1 uses all visits and thus the total expected number $N_{\text{visits}}(d1)$ is equal to the number of data pairs $N(d1)$. Therefore, if the satellite visits the gauge site once a day, it will take

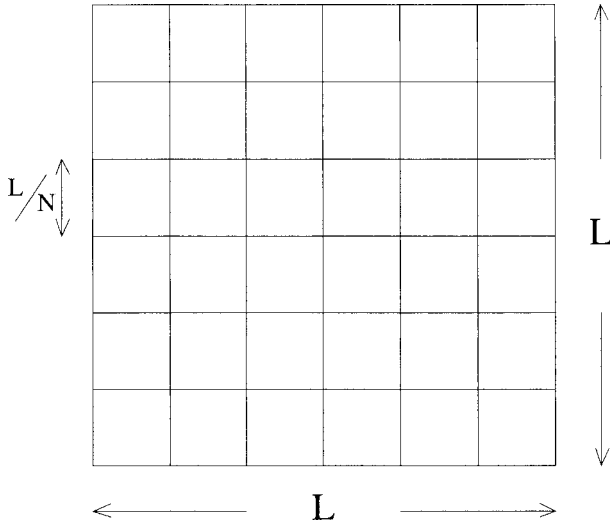


FIG. 1. Schematic diagram for the FOV of L km on a side.

$N_{\text{visits}}(d1) = N(d1)$ days for design 1 and $N_{\text{visits}}(d2) = N(d2)/P_s$ days for design 2.

4. Numerical examples

In this section, for the applicability of the ground truth design developed in this study, we use GATE data and evaluate each ground truth design.

The GATE phase I data used in this study were collected for the 19-day interval (28 June–16 July) and consist of 1716 fields, which covers a $400 \text{ km} \times 400 \text{ km}$ field at 4-km resolution, approximately every 15 min (with gaps). Arkell and Hudlow (1977) composited the radar measurements and presented an atlas of radar echos every 15 min. Patterson et al. (1979) converted the radar measurements to instantaneous rain rates averaged over 4 km by 4 km pixels. The data in the central 280-km box within the GATE area are used in this study. The structure of GATE phase I data can be written in the following matrix time series:

$$\psi_i(i, j), \quad i = 1, \dots, 70; \quad j = 1, \dots, 70; \\ t = 1, \dots, 1716, \quad (31)$$

where i , j , and t are the subscripts in the directions of north–south (NS) and west–east (WE), and for time, respectively, and $\psi_i(i, j)$ is the rain rate at i , j , t . We consider the GATE data to be like the rain rates observed by a point gauge and the perfect satellite. For the purpose of data analysis, we split each 280-km spatial rain rate field into a sequence of the nonoverlapping $L \times L$ km² area that is considered as the FOV. Figure 1 shows a schematic diagram of such an FOV of L km on a side.

The satellite passes over the site where the gauge is located and one of the FOVs ($L = N \times 4$ km across, nominally $20 \text{ km} \times 20 \text{ km}$) in the swath along the ground track covers the gauge. The satellite measurement is then defined as

$$\Psi_s = \frac{1}{N^2} \sum_i^N \sum_j^N \psi_i(i, j). \quad (32)$$

It is essential to mention here that a point measurement in the GATE dataset is actually an area-average rain rate over a $4 \text{ km} \times 4 \text{ km}$ pixel. As discussed before, the satellite measures the column average of rain rate, which is equivalent to a time average of a few minutes at the surface. Because the act of area averaging is also to smooth the variability of the field similar to the operation of time averaging, we here consider the $4 \text{ km} \times 4 \text{ km}$ area average of the gauge measurement over the tile as like a few-minute average of the gauge measurement. The rain gauge measurement is then defined as

$$\Psi_g = \psi_i(i_g, j_g), \quad (33)$$

where (i_g, j_g) is the fixed location of the rain gauge.

For each FOV with $L = N \times 4$ km on a side, we create the database of the satellite measurements

$$\{\Psi_s^k; k = 1, 2, \dots, [70/N] \times [70/N] \times 1716\}, \quad (34)$$

where $[x]$ gives the greatest integer less than or equal to x . When we cannot use the whole tiles in the GATE dataset, we use the data in the central $[70/N] \times 4$ km box.

Since the centroid of the FOV is located randomly with respect to the position of the gauge for each point, we equivalently take the centroid to be fixed and thus consider \mathbf{r}_g to be randomly located within an FOV. Therefore, we generate uniform random numbers using the following uniform distribution and select the location of the rain gauge:

$$P(\mathbf{r}_g) = \begin{cases} \frac{1}{N^2}, & \text{when } \mathbf{r}_g = \mathbf{r}_{(1,1)}, \mathbf{r}_{(1,2)}, \dots, \mathbf{r}_{(N,N)}, \\ 0, & \text{otherwise.} \end{cases}$$

We then create the database of the gauge measurements:

$$\{\Psi_g^k; k = 1, 2, \dots, [70/N] \times [70/N] \times 1716\}. \quad (35)$$

Using the database $\{(\Psi_s^k, \Psi_g^k); k = 1, 2, \dots, [70/M] \times [70/M] \times 1716\}$, we create the histogram of satellite measurements, gauge measurements, and the error as shown in Fig. 2. Figure 2 indicates that the distribution of rain rate is the mixed distribution with skewed continuous part.

The ensemble mean of satellite measurements, gauge measurements, and the errors are provided in Table 1 to investigate the theoretical results in section 3. In Table 1, we can see that there is no bias for design 1 or design 2. But, as we showed theoretically already, design 3 has a design bias. This result says that the ground truth design 3 has a serious disadvantage as a ground truth design. The absolute value of the bias for design 3 increases as the width of the FOV increases. It was shown

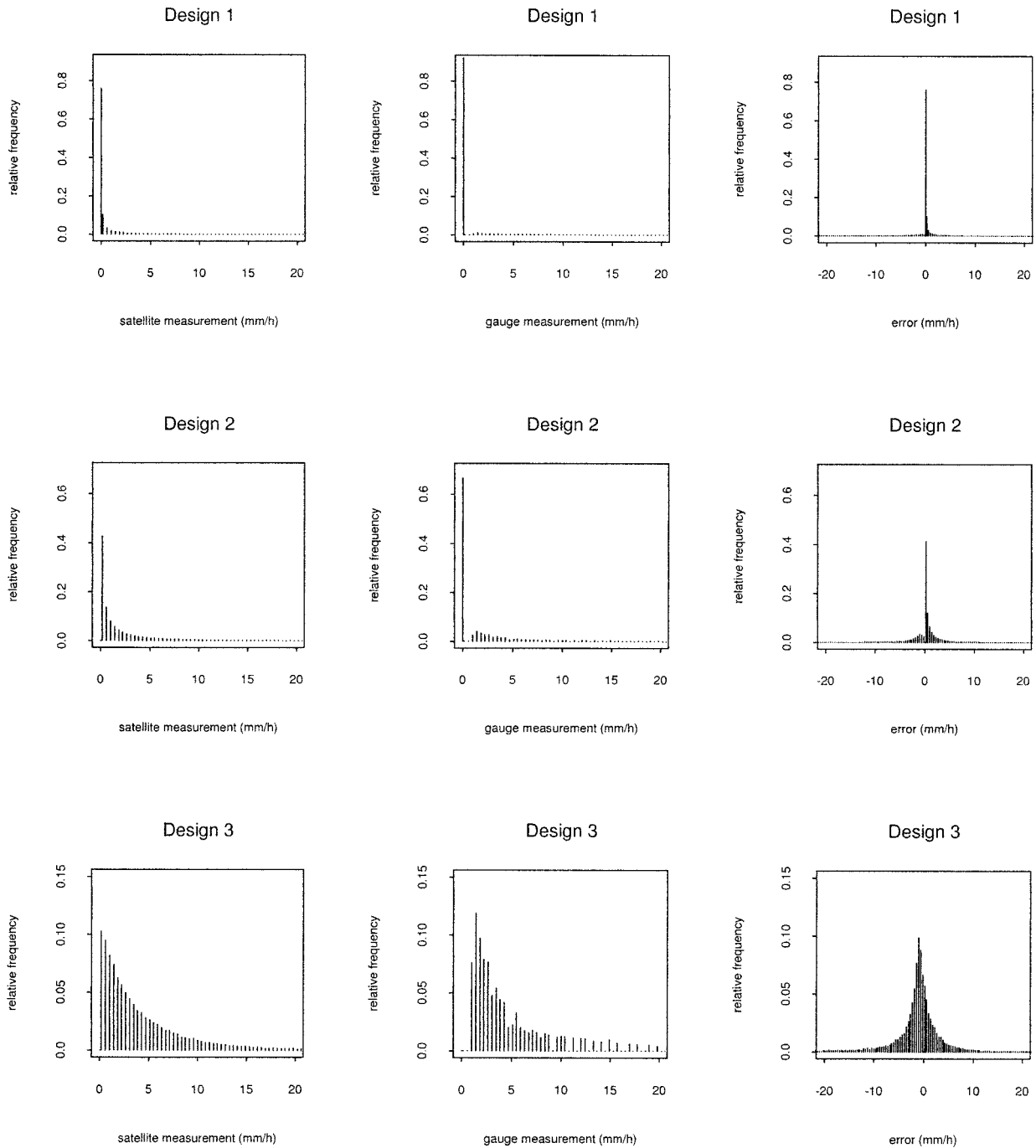


FIG. 2. The histogram of the satellite and gauge measurements, and the errors. The GATE data and the FOV of size 20 km \times 20 km (resolution is 4 km) are used.

in section 3 that the $mse(d2)$ is equal to the $mse(d1)$ divided by the probability that the satellite measurement has rain inside the FOV. Figure 3 provides the estimated probability (relative frequency) P_s that the satellite measurement has rain inside the FOV based upon GATE data. The P_s is linearly increasing as the width of the FOV increases and the linear regression equation is $P_s = 0.0555 + 0.00876 \times (\text{width of FOV})$. The coefficient

of determination for the regression equation is $R^2 = 0.995$. The variances of the gauge measurement for design 1 and design 2 are computed as $\sigma_{g1}^2 = 7.835$ and $\sigma_{g2}^2 = 65.5$, respectively.

Table 2 gives the dimensionless root-mean-square error (drmse, hereafter), which is the square root of the dimensionless mean square error of design 1 and design 2 for a single visit. The drmse of design 2 is a little

TABLE 1. Ensemble mean of the satellite measurements $\langle\psi_{st}\rangle$, gauge measurements $\langle\psi_{gt}\rangle$, and error (=satellite measurement – gauge measurement) $\langle\epsilon_{dt}\rangle$ with GATE data.

Width of footprint (km)	Design 1			Design 2			Design 3		
	$\langle\psi_{s1}\rangle$	$\langle\psi_{g1}\rangle$	$\langle\epsilon_{d1}\rangle$	$\langle\psi_{s2}\rangle$	$\langle\psi_{g2}\rangle$	$\langle\epsilon_{d2}\rangle$	$\langle\psi_{s3}\rangle$	$\langle\psi_{g3}\rangle$	$\langle\epsilon_{d3}\rangle$
4	0.474	0.474	0.000	5.895	5.895	0.000	5.895	5.895	0.000
8	0.473	0.474	-0.001	3.839	3.844	-0.005	5.917	6.547	-0.630
12	0.467	0.466	0.001	2.874	2.870	0.004	5.864	6.861	-0.997
16	0.460	0.461	-0.001	2.302	2.311	-0.009	5.885	7.221	-1.336
20	0.474	0.472	0.002	1.981	1.973	0.008	5.358	6.904	-1.546
24	0.444	0.444	0.000	1.646	1.643	0.003	5.771	7.535	-1.764
28	0.474	0.461	0.013	1.531	1.487	0.041	5.776	7.587	-1.811
32	0.429	0.431	-0.002	1.290	1.295	-0.005	5.824	7.998	-2.174
36	0.422	0.422	0.000	1.168	1.168	0.000	5.779	8.145	-2.366
40	0.474	0.476	-0.002	1.175	1.180	-0.005	5.965	8.402	-2.437

greater than the drmse of design 1 for any size of FOV. Table 2 also gives the number of visits to achieve 10% of the standard deviation of the gauge measurement. We take our nominal FOV to have a size of $20\text{ km} \times 20\text{ km}$ because the footprint size of the TRMM Microwave Imager (TMI) (19.4-GHz channel) might be thought of as having a nominal 25-km resolution. For the typical $20\text{ km} \times 20\text{ km}$ FOV, the number of data pairs is $N(d1) = 50$ for design 1 and $N(d2) = 56$ for design 2 to detect 10% bias. As we explained in section 3, the number of visits $N_{\text{visits}}(d2)$ necessary to detect 10% of the variability of gauge measurement is $N_{\text{visits}}(d2) = N(d2)/P_s$. With Table 2 and Fig. 3, the number of visits for design 2 of GATE data can be obtained. For example, for the 20-km FOV, the expected number of visits is $N_{\text{visits}}(d2) = N(d2)/P_s = 56.3/0.24 \approx 234$. The expected number of visits to detect the 10% retrieval bias for design 1 is $N_{\text{visits}}(d1) = N(d1) \approx 50$ because design 1 uses the data pairs from all visits. It is notable that the total number of visits of design 1 is smaller than that of design 2 to detect the retrieval bias with the same tolerance level. This result seems to suggest that design 1 may be better than design 2, but this may not be so. Because the rain

rate is patchy, the error (=satellite measurement – gauge measurement) has many zeros (see Fig. 2; the probability is about 0.9) and thus the mse(d1) can be so small due to the characteristic of the patchy rain. For the extreme rain field that always has no rain, the mse(d1) is zero. Even though this extreme case is trivial, it seems to us that the mse(d1) is not as appropriate measure to evaluate the accuracy of the ground truth design of satellite rain rate.

It is interesting to compare our result to the model study in North et al. (1994). They used the noise-forced diffusive rain model tuned to GATE data and obtained the number of visits $N = 60$ to detect the retrieval bias with 10% level. Since the noise-forced diffusive rain model always provides continuous rain fields, this model does not make any difference between design 1 and design 2. That is, for the noise-forced diffusive model, the probability that the satellite has rain is always $P_s = 1$. The number of visits $N = 60$ for the noise-forced diffusive is quite close to $N(d1) = 50$ and $N(d2) = 56$ with real GATE data.

For the white noise mixed lognormal random field using the statistics tuned to GATE, we also analytically

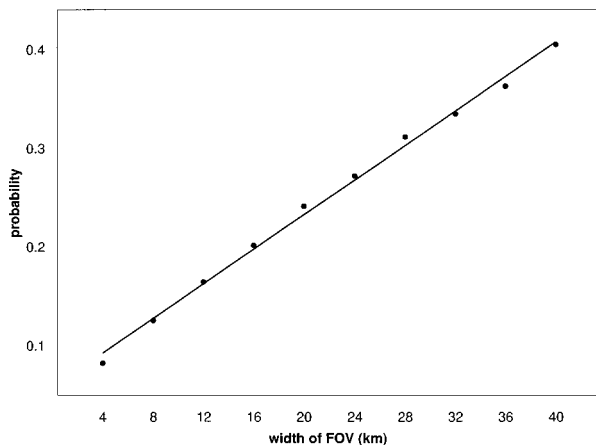


FIG. 3. The estimated probability P_s that the satellite measurement has rain inside the FOV with GATE data. The linear regression equation is $P_s = 0.0555 + 0.00876 \times (\text{width of the FOV})$.

TABLE 2. The number of data pairs $N(d1)$ for design 1 and $N(d2)$ for design 2 to achieve about 10% of the standard deviation of the gauge measurement are given, as well as the dimensionless root-mean-square error $W(di)$ of one data pair. The dimensionless mean-square error is defined as $W^2(di) = \langle\epsilon_{dt}^2\rangle/\sigma_{gt}^2$, $i = 1, 2$. The number of visits $N_{\text{visits}}(d2)$ for design 2 can be computed by $N_{\text{visits}}(d2) = N(d2)/P_s$, where P_s is in Fig. 3. These calculations were made with GATE data.

Width of footprint (km)	Design 1		Design 2	
	$W(d1)$	$N(d1)$	$W(d2)$	$N(d2)$
8	0.496	24.6	0.553	30.6
12	0.607	36.8	0.665	44.2
16	0.668	44.6	0.718	51.5
20	0.709	50.2	0.751	56.3
24	0.719	51.7	0.784	61.5
28	0.743	55.2	0.800	63.9
32	0.745	55.5	0.813	66.1
36	0.762	58.1	0.832	69.2
40	0.809	65.4	0.836	70.0

derived the statistics and evaluated each design proposed in this paper. It was found that design 3 has bias and the number of visits to detect 10% bias is about 96 for both design 1 and design 2 with $p = 0.1$. Ha and North (1999) found that the number of data pairs to detect 10% bias is about $N(d1) = 96$ and $N(d2) = 97$ for the white noise Bernoulli random field with $p = 0.1$. Remember that the probability that satellite measurement has rain is $P_s = 1 - (1 - p)^A$ for the white noise random field. Because this probability is quite close to 1 if the probability of rain p is small and/or the FOV size is large, it is expected that the numbers $N(d1)$ and $N(d2)$ are almost the same for the white noise random field. Therefore, the expected number of visits $N_{\text{visits}}(d2)$ is almost the same as the number of data pairs $N(d2)$ used in design 2. This result says that the white noise random field requires twice the number of measurement pairs for more realistic rain fields when we use design 2. However, the number of visits for realistic rain fields $N_{\text{visits}}(d2) = 234$ is more than twice the number of visits $N_{\text{visits}}(d2) = 96$ for white noise random fields.

5. Summary and conclusions

In this paper we have considered ground truth designs based on point gauge measurements to validate satellite measurements. Based upon properties encountered with real rain, we modeled the non-white noise homogeneous random field having a mixed distribution as a rain-rate field. Because either or both measurements (satellite, gauge) may have no rain, three ground truth designs based on the method of throwing out the no-rain measurements were proposed. Design 1 uses data pairs from all visits. Design 2 uses data pairs only when the FOV average has rain. Design 3 uses data pairs only when the gauge has rain.

It was theoretically shown that the satellite measurement is an unbiased estimator of the gauge measurement for design 1 and design 2. However, design 3 has a serious disadvantage as the ground truth design because it exhibits a large design bias. The efficiencies of design 1 and design 2 are indexed by the mean-square error (difference) between the satellite and gauge estimates. It was derived that the mean-square error for design 2 is equal to the mean-square error for design 1 divided by the probability that the satellite measurement has rain inside the FOV. This fact gives us a way to compute the mean-square error for design 2 without using the conditional random field. The theoretical results were confirmed with the GATE data. These results generalize what Ha and North (1999) showed for the white noise Bernoulli random field.

With the GATE data having an FOV width of 20 km, we have found that for design 1 and design 2 the number of data pairs necessary to distinguish a bias of 10% is of the order of 50, which is almost half of number of data pairs for a white noise random field. Since design 2 rejects data pairs when the satellite measurement has

no rain, about 230 overpasses are required. Since a TRMM FOV will include a given gauge about once per day near the equator, this suggests that 8–10 months of data should be adequate. The retrieval of rain rate gives a beam-filling error (Chiu et al. 1990; Ha and North 1995), which is composed of a retrieval bias and the random error with ensemble mean zero. It will thus take more than 230 overpasses to validate the satellite measurements with point gauge measurements due to the beam-filling error. Since GATE data are characteristic of precipitation in the ITCZ where it is most intense, the results may not fully apply outside these areas.

Acknowledgments. The first author (EH) wishes to thank the Yonsei University Research Foundation for its support. The second author (GRNI) thanks the NASA TRMM program for its support.

REFERENCES

- Arnell, R., and M. Hudlow, 1977: *GATE International Meteorological Radar Atlas*. NOAA, Washington, DC, 222 pp.
- Bell, T. L., A. Abdullah, R. L. Martin, and G. R. North, 1990: Sampling error for satellite-derived tropical rainfall: Monte Carlo study using a space-time stochastic model. *J. Geophys. Res.*, **95**, 2195–2205.
- Chiu, L. S., G. R. North, D. A. Short, and A. McConnell, 1990: Rain estimation from satellites: Effect of finite field of view. *J. Geophys. Res.*, **95**, 2177–2185.
- Ha, E., and G. R. North, 1994: Use of multiple gauges and microwave attenuation of precipitation for satellite verification. *J. Atmos. Oceanic Technol.*, **11**, 629–636.
- , and —, 1995: Model studies of the beam-filling errors for rain-rate retrieval with microwave radiometers. *J. Atmos. Oceanic Technol.*, **12**, 268–281.
- , and —, 1999: Error analysis for some ground validation designs for satellite observations of precipitation. *J. Atmos. Oceanic Technol.*, **16**, 1949–1957.
- Kedem, B., L. S. Chiu, and G. R. North, 1990: Estimation of mean rain rate: Application to the satellite observations. *J. Geophys. Res.*, **95**, 1965–1972.
- McConnell, A., and G. R. North, 1987: Sampling errors in satellite estimates of tropical rain. *J. Geophys. Res.*, **92** (D8), 9567–9570.
- North, G. R., and S. Nakamoto, 1989: Formalism for comparing rain estimation designs. *J. Atmos. Oceanic Technol.*, **6**, 985–992.
- , S. S. P. Shen, and R. B. Upson, 1991: Combining rain gauges with satellite measurements for optimal estimates of area-time averaged rain rate. *Water Resour. Res.*, **27**, 2785–2790.
- , J. B. Valdes, E. Ha, and S. P. Shen, 1994: The ground-truth problem for satellite estimates of rain rate. *J. Atmos. Oceanic Technol.*, **11**, 1035–1041.
- Parzen, E., 1962: *Stochastic Processes*. Holden-Day, 324 pp.
- Patterson, V. L., M. D. Hudlow, P. J. Pytlowany, F. P. Richards, and J. D. Hoff, 1979: GATE radar rainfall processing system. NOAA Tech. Memo. EDIS 26, Washington DC, 158 pp.
- Shin, K. S., and G. R. North, 1988: Sampling error study for rainfall estimates by satellite using a stochastic model. *J. Appl. Meteor.*, **27**, 1218–1231.
- Simpson, J., R. F. Adler, and G. R. North, 1988: A proposed Tropical Rainfall Measuring Mission (TRMM) satellite. *Bull. Amer. Meteor. Soc.*, **69**, 278–295.
- Theon, J. S., T. Matsuno, T. Sakata, and N. Fugono, 1992: *The Global Role of Tropical Rainfall*. A. Deepak, 280 pp.
- Thiele, O. W., 1992: Ground truth for rain measurement from space. *The Global Role of Tropical Rainfall*, J. S. Theon et al., Eds., A. Deepak, 245–260.
- Wilheit, T. T., A. T. C. Cheng, and L. S. Chiu, 1991: Retrieval of monthly rainfall indices from microwave radiometric measurements using probability distribution functions. *J. Atmos. Oceanic Technol.*, **8**, 118–136.