

Error Analysis for Some Ground Validation Designs for Satellite Observations of Precipitation

EUNHO HA

Department of Statistics, Yonsei University, Wonju, Korea

GERALD R. NORTH

Department of Meteorology, Texas A&M University, College Station, Texas

(Manuscript received 12 August 1998, in final form 15 February 1999)

ABSTRACT

In this paper point gauges are used in an analysis of hypothetical ground validation experiments for satellite-based estimates of precipitation rates. The ground and satellite measurements are fundamentally different since the gauge can sample continuously in time but at a discrete point, while the satellite samples an area average (typically 20 km across) 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 has a large nonzero contribution 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, and design 3 uses the pairs only when the gauge has rain. The error distribution of each design is derived theoretically for a Bernoulli spatial random field with different horizontal resolutions. It is found that design 3 cannot be used as a ground-truth design due to its large design bias. The mean-square error is used as an index of accuracy in estimating the ground measurement by satellite measurement. It is shown that there is a relationship between the mean-square error of design 1 and design 2 for the Bernoulli random field. Using this technique, the authors derive the number of satellite overpasses necessary to detect a satellite retrieval bias, which is as large as 10% of the natural variability.

1. Introduction

Several satellite programs are under way to measure precipitation especially in remote areas such as the tropical oceans (see Simpson et al. 1988; Wilheit et al. 1991; Theon et al. 1992). This paper concerns itself with the ground validation program of the Tropical Rainfall Measuring Mission (TRMM), which was launched in November 1997 (Simpson et al. 1988) and has now generated over a year of data from several sensors. One problem inherent in all such missions is that of ground validation. We present here an analysis of validating the satellite estimations with point gauge measurements. This is a complex comparison because the two sensors are measuring different quantities: 1) the point gauge measures precipitation nearly continuously in time at a point on the surface, while 2) the satellite measures a snapshot in time (actually about a 5–10-min average) of an area average over its field of view (FOV) at the surface (typically 20 km across). In comparing two si-

multaneous measurements the point gauge may be located at some random position within the FOV. While in the long run these two measurements should agree, there is likely to be a large random (zero mean) difference between the two because of the different space–time sampling configurations. It should be possible, however, by taking enough simultaneous pairs of measurements to compare them and check for bias in the satellite retrieval algorithm.

There are various forms of ground validation for precipitation. One is comparing horizontal precipitation patterns from ground-based or airplane-based radars. This method is very useful in judging the qualitative patterns of the precipitation, but because of the uncertainty in the conversion of radar reflectivity to surface rain rate, it cannot alone answer the question of whether rain rates derived from the satellite are biased. Hence, we turn to a more certain estimate of the rain rate at the ground (the point gauge) but surrender the certainty of matching sampling designs.

We proceed in our purely theoretical study by asking about the size of the random errors incurred in comparing the two designs, satellite and point gauge. We make the null hypothesis that both measurements are correct and see what the distribution of differences is

Corresponding author address: Dr. Gerald R. North, Department of Meteorology, Texas A&M University, 1204 Eller O&M Building, College Station, TX 77843-3150.
E-mail: northed@ariel.tamu.edu

due solely to the differences in sampling design. In principle, we can reduce these kinds of random errors by collecting more and more pairs and averaging. We seek the number of such pairs that would allow us to see a real bias in the satellite measurement if it should occur. The question is will a few such pairs suffice to reveal a significant bias or not. As we shall see in this somewhat idealized study it takes many months of data to reduce the random error enough to detect a bias of 10%. This figure comes from assuming an FOV of about 20 km and an autocorrelation distance of a few kilometers. Larger autocorrelation lengths will lead to smoother fields and therefore a smaller number of pairs to detect bias at the 10% level.

Our approach is to take a hypothetical FOV and subdivide it into M subareas, which we call tiles. We then construct a random field for the rain rate, which is somewhat like real rain. The random field is constructed by drawings from a Bernoulli distribution where values in neighboring tiles are independent. This effectively means that the horizontal correlation length is about half a tile width. This model of rain rates is of course highly simplified but provides some insight into the problem since it allows mostly analytical results that can be interpreted easily.

Thiele (1992) presents an extensive discussion of all the issues relating to ground truth and indicates several areas of active research. We have visited this problem in earlier studies (North et al. 1994; Ha and North 1994) in which we developed the ground validation strategy that we extend in this paper. The main difference between our earlier studies and the present one is the use of the Bernoulli random field, which allows us to have areas of rain and no rain. This is an important distinction between continuous models of the rain-rate field because actual rain is of this type and most of the time it is not raining. Furthermore, one might want to examine strategies in which no-rain measurements are tossed out of the stream of data pairs. Quantitatively, the strategy may be stated as follows: We consider two measurements taken from the n th visit, the satellite measurement Ψ_s^n , and a ground measurement Ψ_g^n . We can think of one of the measurements as an estimator of the other. If we use the mean-square error as an index of our accuracy in estimating the ground measurement by satellite measurement, the mean-square error is

$$\epsilon_1^2 = \langle (\Psi_s^n - \Psi_g^n)^2 \rangle, \quad (1)$$

where the ensemble mean $\langle \cdot \rangle$ includes all the possible realizations of the rain-rate field along with all the uniformly distributed locations of the centroid of the FOV holding the ground configuration fixed. If the members of the sequence of pairs of measurements are statistically independent, then the average mean-square error for N visits is

$$\epsilon_N^2 = \frac{\epsilon_1^2}{N}. \quad (2)$$

As we can see from (2), the advantage of this strategy is that we can easily add new measurements and thus we can make the error as small as we please (i.e., where N is sufficiently large).

In this paper, the point gauge is used as the ground-truth measurement to validate satellite precipitation retrieval algorithms at the FOV spatial level (typically about 20 km).

Because the probability distribution of real rain has a large nonzero contribution at zero rain rate (usually greater than 90%), many of the visits will lead to (no rain, no rain) measurement pairs or perhaps (no rain, rain) pairs, where the second entry is the FOV average. For this reason, we can consider the following three ground-truth designs based on the point gauge measurement.

- Design 1 uses all visits even though either of the two measurements (gauge, satellite) may have no rain.
- Design 2 throws out all the visits when the FOV average has no rain. Note that when the FOV has no rain, the gauge also has no rain.
- Design 3 throws out all the visits when the gauge has no rain. Note that the FOV average can have rain even though the gauge has no rain.

We use a spatial white noise Bernoulli random field as the rain-rate model. This means that for each tile within the FOV it is either raining at (fixed) rate r or it is not raining at all. For a particular overpass of the satellite and for a given tile the probability of rain being nonzero is p . While this is a fairly crude as a model of rain fields, simplicity is important here to establish the main principles involved in the problem. We derive the probability density function, ensemble mean, and the variance of gauge measurements for each design. At the same time, we examine the relationship between ground-truth designs proposed here and evaluate each design.

To keep the numbers specific and roughly relevant to TRMM we choose some parameters to have nominal values such as the FOV width to be 20 km. The finest resolution taken for the tiles within the FOV used in simulating the random field is 4 km, which is the same resolution as Global Atmospheric Research Program Atlantic Tropical Experiment (GATE) radar data. This resolution is also consistent with the fact that the satellite actually measures 5–10-min averages of the rain rate. Such time averaging is thought to be roughly equivalent to spatial averaging to about 4 km.

2. Definitions

Consider a random field $\psi(\mathbf{r}, t)$ defined in the $\mathbf{r} = (x, y)$ plane and along the time axis t . As a typical experiment we envision a point gauge located at some fixed site \mathbf{r}_g . The satellite measurement based on R , which is the FOV for a visit, is

$$\Psi_s = \frac{1}{A} \int_R \psi(\mathbf{r}, t) d^2\mathbf{r}, \quad (3)$$

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

$$\Psi_g = \psi(\mathbf{r}_g, t). \quad (4)$$

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, when we refer to the ground and satellite measurements, their measurements can be time averaged or instantaneous measurements that are usually denoted by Ψ_g and Ψ_s .

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. We have two measurements 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 uses the measurements (Ψ_s^n, Ψ_g^n) of all visits. Design 2 uses only the measurements (Ψ_s^n, Ψ_g^n) for $\Psi_g^n > 0$, 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. We form the difference between the satellite measurement and ground measurement and call this the error $\epsilon_{di}^n = \Psi_{si}^n - \Psi_{gi}^n$ for the n th visit. The mean-square error, which is usually used as an index of the accuracy in estimating the ground measurement by satellite measurement, is defined by

$$\epsilon_{di}^2 = \langle (\Psi_{si} - \Psi_{gi})^2 \rangle. \quad (5)$$

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 in addition to a retrieval bias, there is a bias due to the design itself, which we refer to as *design bias*, we could mistakenly attribute an error to the retrieval algorithm that is actually inherent in the design. (In reality both types of bias are likely to be present). 2) The mean-square error $\langle \epsilon_{di}^2 \rangle$ of the difference between the ground measurement and the perfect satellite measurement should be com-

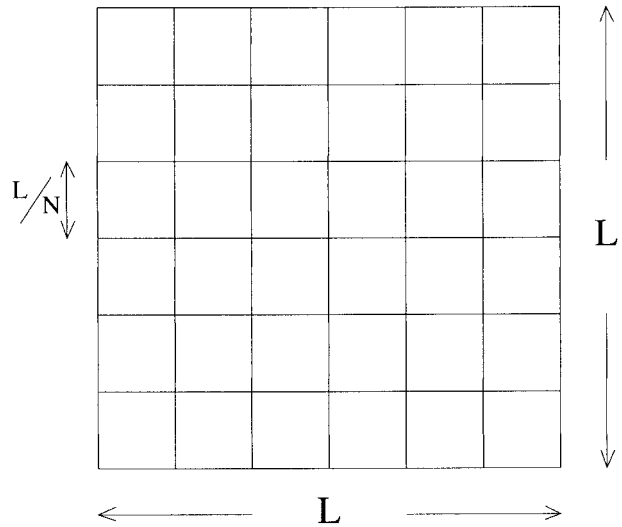


FIG. 1. Schematic diagram for FOV of L km on a side.

parable to or less than the size of biases in the satellite rain-rate algorithm we are trying to test.

We can write the mean of the error for design 2 and design 3 as follows:

$$\begin{aligned} \langle \epsilon_{d2} \rangle &= \langle \Psi_{s2} - \Psi_{g2} \rangle \\ &= \langle \Psi_{s1} - \Psi_{g1} | \Psi_{s1} > 0 \rangle, \quad \text{and} \quad (6) \end{aligned}$$

$$\langle \epsilon_{d3} \rangle = \langle \Psi_{s1} - \Psi_{g1} | \Psi_{g1} > 0 \rangle, \quad (7)$$

where $\langle Y | X > x \rangle$ denotes the conditional mean of Y , given that $X > x$. We can also express the mean-square error for design 2 as

$$\langle \epsilon_{d2}^2 \rangle = \langle (\Psi_{s2} - \Psi_{g2})^2 \rangle = \langle (\Psi_{s1} - \Psi_{g1})^2 | \Psi_{s1} > 0 \rangle. \quad (8)$$

We partition the $L \times L$ km² FOV into $M = N^2$ tiles to treat the random field effectively as a multivariate vector. Figure 1 shows a schematic diagram of such an FOV of L km on a side. The satellite measurement then can be written as

$$\Psi_s = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \psi(i, j), \quad (9)$$

where $N = L/l$ is the number of subdivisions that have been chosen for the partitioning of the FOV and $\psi(i, j)$ represents the area-average rain rate in an l km \times l km grid square, which we call a tile. For the fixed FOV size, if we increase the number of tiles $M = N^2$, then the resolution (roughly equivalent to autocorrelation length) l will be smaller. Conversely, if we fix the resolution l , the FOV size will be large when the number of tiles M increases. In our study we will present numerical results in which case l is fixed and is given the same spatial averaging (4 km) as the gridded GATE data. 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.

The precipitation rate from the gauge measurement is

$$\Psi_g = \psi(x_g, y_g), \tag{10}$$

where (i_g, j_g) is the position of the gauge site. For convenience of notation, we will use $\psi(\mathbf{r}_k)$ ($k = 1, 2, \dots, M$) and $\psi(\mathbf{r}_g)$ instead of $\psi(i, j)$ ($i, j = 1, 2, \dots, N$) and $\psi(i_g, j_g)$, respectively. The satellite and gauge measurements are then

$$\Psi_s = \frac{1}{M} \sum_{k=1}^M \psi(\mathbf{r}_k), \tag{11}$$

$$\Psi_g = \psi(\mathbf{r}_g). \tag{12}$$

When we compute the mean and the mean-square error, we have to consider the position of the centroid of the FOV. Since the centroid of the FOV is located randomly with respect to the position of the gauge for each point, we can take the centroid to be fixed and thus consider \mathbf{r}_g to be randomly located within an FOV. Therefore, we have the uniform distribution as a probability distribution of \mathbf{r}_g :

$$P(\mathbf{r}_g) = \begin{cases} \frac{1}{M}, & \text{when } \mathbf{r}_g = \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M, \\ 0, & \text{otherwise.} \end{cases}$$

3. Ground-truth designs for Bernoulli random field

We use the Bernoulli distribution as the distribution of rain rate to compare the ground-truth designs described above because this distribution describes the no-rain phenomenon of real rain in its simplest possible form. Let x of the M tiles be raining with rain rate r , and in the rest of the $M - x$ tiles there is no rain. We assume that the probability of rain rate in an individual tile is p and one tile is independent of the other. In this section, we present some numerical values for the 4-km resolution (as in GATE radar data). This means that when we increase the number of tiles M , the FOV increases. In order to change the tile size for fixed FOV size and to see the effect of the tile size using the results in this section, one needs to adjust the probability of rain p and rain rate r because it is dependent on the resolution.

a. The error distributions

Since the random field here is a spatially white noise Bernoulli random field, we can easily obtain the distribution of the gauge and satellite measurements for design 1:

$$P(\Psi_{g1} = x) = \begin{cases} p, & \text{when } x = r, \\ 1 - p, & \text{when } x = 0. \end{cases} \tag{13}$$

$$P\left(\Psi_{s1} = \frac{rx}{M}\right) = \binom{M}{x} p^x (1 - p)^{M-x}, \tag{14}$$

$$x = 0, 1, \dots, M.$$

The distribution of the error $\epsilon_{d1} = \Psi_{s1} - \Psi_{g1}$ can be derived as

$$P\left(\epsilon_{d1} = \frac{rx}{M} - r\right) = \binom{M-1}{x-1} p^x (1 - p)^{M-x},$$

$$x = 1, 2, \dots, M - 1,$$

$$P(\epsilon_{d1} = x) = p^M + (1 - p)^M,$$

$$x = 0,$$

$$P\left(\epsilon_{d1} = \frac{rx}{M}\right) = \binom{M-1}{x} p^x (1 - p)^{M-x},$$

$$x = 1, 2, \dots, M - 1.$$

Design 2 uses only the visits that have rain in the FOV. Thus, we can derive the distribution of ground and satellite measurements for design 2 using the distribution of Ψ_{g1} and Ψ_{s1} conditional on $\Psi_{s1} > 0$. Since we assumed that the random field is white noise, the probability that there is rain inside the FOV is $P(\Psi_{s1} > 0) = 1 - (1 - p)^M$. Note that when the random field is not white noise, the probability $P(\Psi_{s1} > 0)$ cannot be written in this form. For convenience of notation, we will use $P_M = P(\Psi_{s1} > 0) = 1 - (1 - p)^M$ hereafter. The distributions for design 2 are derived as

$$P(\Psi_{g2} = x) = P(\Psi_{g1} = x | \Psi_{s1} > 0)$$

$$= \begin{cases} \frac{(1-p)[1 - (1-p)^{M-1}]}{P_M}, & \text{when } x = 0, \\ \frac{p}{P_M}, & \text{when } x = r; \end{cases} \tag{15}$$

$$P\left(\Psi_{s2} = \frac{rx}{M}\right) = P(\Psi_{s1} = x | \Psi_{s1} > 0)$$

$$= \frac{1}{P_M} \binom{M}{x} p^x (1 - p)^{M-x},$$

$$x = 1, 2, \dots, M. \tag{16}$$

We can derive the error distribution $P(\epsilon_{d2}) = P(\epsilon_{d1} | \Psi_{s1} > 0)$,

$$P\left(\epsilon_{d2} = \frac{rx}{M} - r\right) = \frac{\binom{M-1}{x-1} p^x (1-p)^{M-x}}{P_M},$$

$$x = 1, 2, \dots, M-1,$$

$$P(\epsilon_{d2} = x) = \frac{p^M}{P_M}, \quad x = 0,$$

$$P\left(\epsilon_{d2} = \frac{rx}{M}\right) = \frac{\binom{M-1}{x} p^x (1-p)^{M-x}}{P_M},$$

$$x = 1, 2, \dots, M-1.$$

Design 3 uses only the visits when the gauge has rain. The distributions of ground and satellite measurement for design 3 are

$$P(\Psi_{g3} = x) = P(\Psi_{g1} = x | \Psi_{g1} > 0)$$

$$= \begin{cases} 1, & \text{when } x = r, \\ 0, & \text{when } x = 0; \end{cases} \quad (17)$$

$$P\left(\Psi_{s3} = \frac{rx}{M}\right) = P(\Psi_{s1} = x | \Psi_{g1} > 0)$$

$$= \frac{\binom{M-1}{x-1} p^{x-1} (1-p)^{M-x}}{P_M},$$

$$x = 1, 2, \dots, M. \quad (18)$$

The distribution of error $\epsilon_{d3} = P(\epsilon_{d1} | \Psi_{s1} > 0)$ can be derived as

$$P\left(\epsilon_{d3} = \frac{rx}{M} - r\right) = \frac{\binom{M-1}{x-1} p^{x-1} (1-p)^{M-x}}{P_M},$$

$$x = 2, 3, \dots, M.$$

The probability density functions for ground-truth designs proposed in this study are in Fig. 2. The 4 mm h⁻¹ rain rate conditional on rain and 20 km × 20 km FOV size (resolution is 4 km and number of tiles M is 25) is used to make Fig. 2. We can see in Fig. 2 that the distribution of the error is bimodal and this phenomenon is even more clear when the probability of rain rate p is 0.5. We also see that the center of the error distribution for design 3 is far from zero. This indicates that the ensemble mean of the error for design 3 is not zero and thus has negative bias (design bias).

b. Mean of the error

We know that the mean of the satellite and ground measurement for design 1 is $\langle \Psi_{s1} \rangle = rp$ and $\langle \Psi_{g1} \rangle = rp$. Thus, the mean of the error for design 1 is

$$\langle \epsilon_{d1} \rangle = \langle \Psi_{s1} - \Psi_{g1} \rangle = 0. \quad (19)$$

This shows that the satellite measurement is an unbiased estimator of the ground measurement for design 1. Using the distributions derived in section 3a, we can compute that the mean of the satellite measurement is $\langle \Psi_{s2} \rangle = rp/P_M$, and the mean of ground measurement is $\langle \Psi_{g2} \rangle = rp/P_M$ for design 2. We thus know that the satellite measurement is also an unbiased estimator of the ground measurement for design 2 because we have

$$\langle \epsilon_{d2} \rangle = \langle \Psi_{s2} - \Psi_{g2} \rangle = 0. \quad (20)$$

Using the distributions of satellite and ground measurement for design 3 in section 3a, we can compute the mean of the satellite and gauge measurements:

$$\langle \Psi_{s3} \rangle = \frac{r}{M} [(M-1)p + 1], \quad (21)$$

$$\langle \Psi_{g3} \rangle = r. \quad (22)$$

We thus have

$$\langle \epsilon_{d3} \rangle = \langle \Psi_{s3} - \Psi_{g3} \rangle = -r(1-p) \left(1 - \frac{1}{M}\right). \quad (23)$$

This equation shows that the satellite measurement is not an unbiased estimator of the ground measurement for design 3. We can see that the bias in Eq. (23) has negative values and thus the satellite measurement underestimates the ground measurement when we use design 3 as our ground-truth design. The magnitude (absolute value) of the bias for design 3 is given in Fig. 3. The magnitude of the bias increases when 1) the rain rate conditional on rain is large, 2) the probability of rain is small, and 3) the FOV size is large. Note that when the probability of rain p is 1, the ground-truth designs in our study are all equivalent. We can see in Eq. (23) that if $p = 1$, the bias disappears. As discussed in section 1, the satellite measurement should be an unbiased estimator of the ground measurement in order to detect a retrieval bias in the algorithm converting microwave temperature to rain rate. In this sense, we have shown that design 3 cannot be used as the ground-truth design.

c. Mean-square error for a single visit

In section 3b, we showed that design 3 cannot be used as a ground-truth design, so we henceforth only consider design 1 and design 2 as ground-truth designs and compute the mean-square error for each.

The mean-square error for design 1 [mse(d1), hereafter] is computed as follows:

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

After some tedious calculations, we can derive mse(d1) as

$$\langle \epsilon_{d1}^2 \rangle = r^2 p (1-p) \left(1 - \frac{1}{M}\right). \quad (25)$$

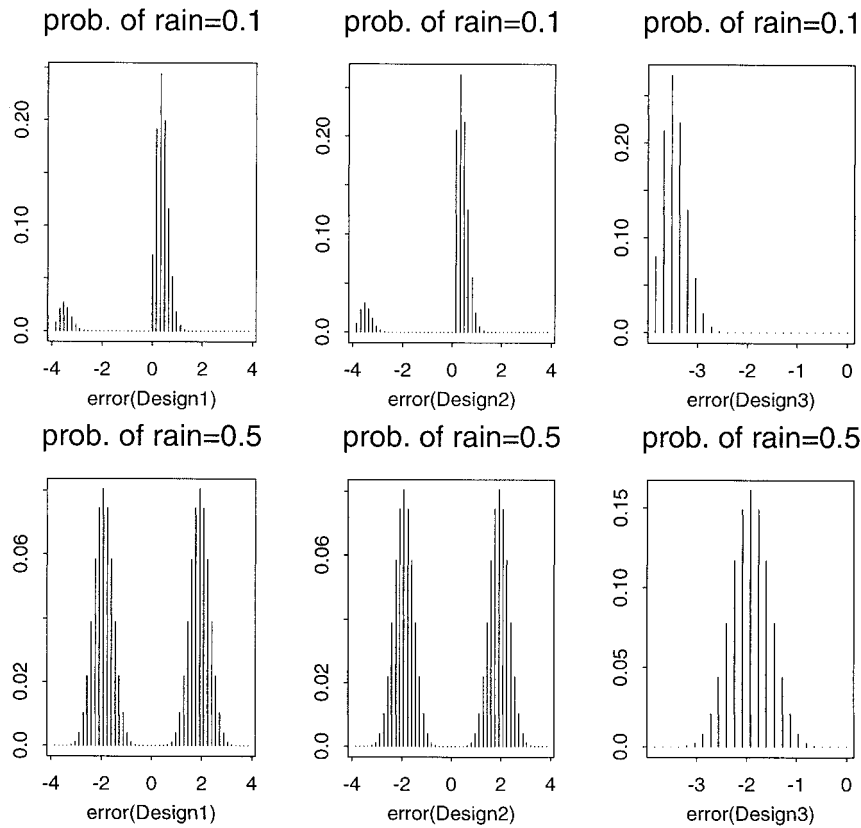


FIG. 2. The probability density function of the error (=satellite measurement – gauge measurement) for Bernoulli random field. A rain rate of 4 mm h⁻¹ and FOV size 20 km × 20 km (resolution is 4 km) are used. The probability of rain is $p = 0.1$ and $p = 0.5$.

The mean-square error for design 2 [mse(d2), hereafter] is computed as

$$\begin{aligned} \langle \epsilon_{d2}^2 \rangle &= \langle (\Psi_{d2} - \Psi_{g2})^2 \rangle = \langle (\Psi_{d1} - \Psi_{g1})^2 | \Psi_{s1} > 0 \rangle \\ &= r^2 p (1 - p) \left(1 - \frac{1}{M} \right) / P_M. \end{aligned} \tag{26}$$

Note that $P_M = 1 - (1 - p)^M$ is the probability that the satellite measurement has rain inside the FOV. From Eqs. (25) and (26), we can see that

$$\langle \epsilon_{d1}^2 \rangle = \frac{\langle \epsilon_{d2}^2 \rangle}{P_M}. \tag{27}$$

When we compute the mse(d2), we need information about the conditional random field, which is a random field on the condition that the satellite measurement has rain inside the FOV. But (27) gives us the way to compute mse(d2) without using the conditional random field. We can compute mse(d2) if we have mse(d1), which is computed from the random field, and if we know the probability that the satellite measurement has rain $P(\Psi_{s1} > 0) = P_M$. We already know that the probability of the satellite measurement having rain is $P(\Psi_{s1} > 0) = 1 - (1 - p)^M$ for a white noise random field.

However, it will not be easy to compute $P(\Psi_{s1} > 0)$ for a nonwhite noise (spatially) random field.

The values of rmse(d1) and rmse(d2), which are the square root-mean-square errors for design 1 and design 2, are given in Fig. 3. Note that rmse(d1) and rmse(d2) increase if the probability of rain p increases where $0 < p < 0.5$. When the FOV increases, rmse(d1) increases but rmse(d2) decreases. This result derives from the probability that the satellite measurement, if it has rain, will increase if the FOV size increases. When the rain rate r increases, rmse(d1) and rmse(d2) increase. In Fig. 4, note that rmse(1) and rmse(2) are almost the same for a rain rate $r = 4$, the probability of rain $p = 0.1$, and typical FOV size (20 km across). The values $r = 4$ and $p = 0.1$ are chosen to match rain-rate statistics from GATE data.

Ultimately we want to know the ratio of the mean-square error to the variance of gauge measurement

$$W^2(di) = \frac{\langle \epsilon_{di}^2 \rangle}{\sigma_{gi}^2}, \quad i = 1, 2, \tag{28}$$

where $\langle \epsilon_{di}^2 \rangle$ is the mean-square error and σ_{gi}^2 is the variance of the gauge measurement for design 1 and design 2. This measure gives us a sense of how the error var-

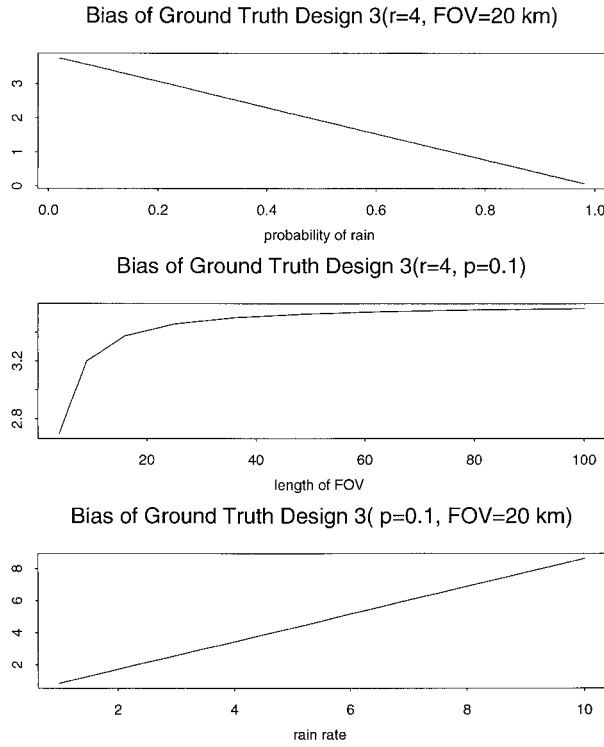


FIG. 3. The absolute value of the bias of the error (=satellite measurement - gauge measurement) for Bernoulli random field when the resolution is 4 km.

iance for the design compares to the climatological variance for precipitation measured by the point gauge. We refer to $W^2(di)$ as the dimensionless mean-square error. We choose σ_{gi}^2 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 differences to be narrow compared to this histogram describing the local climatological variance.

The variance of gauge measurement for design 1 is

$$\sigma_{g1}^2 = \langle (\Psi_{g1} - \langle \Psi_{g1} \rangle)^2 \rangle = r^2 p (1 - p). \quad (29)$$

Thus, the dimensionless mean-square error for design 1 is

$$W^2(d1) = \frac{\langle \epsilon_{d1}^2 \rangle}{\sigma_{g1}^2} = \left(1 - \frac{1}{M} \right). \quad (30)$$

Using the distribution of the gauge measurements for design 2 in section 3a, we can compute the variance of the gauge measurement for design 2:

$$\begin{aligned} \sigma_{g2}^2 &= \langle (\Psi_{g2} - \langle \Psi_{g2} \rangle)^2 \rangle = \langle \Psi_{g2}^2 \rangle - \langle \Psi_{g2} \rangle^2 \\ &= r^2 \frac{P}{P_M} \left(1 - \frac{P}{P_M} \right). \end{aligned} \quad (31)$$

We thus have dimensionless mean-square error for design 2:

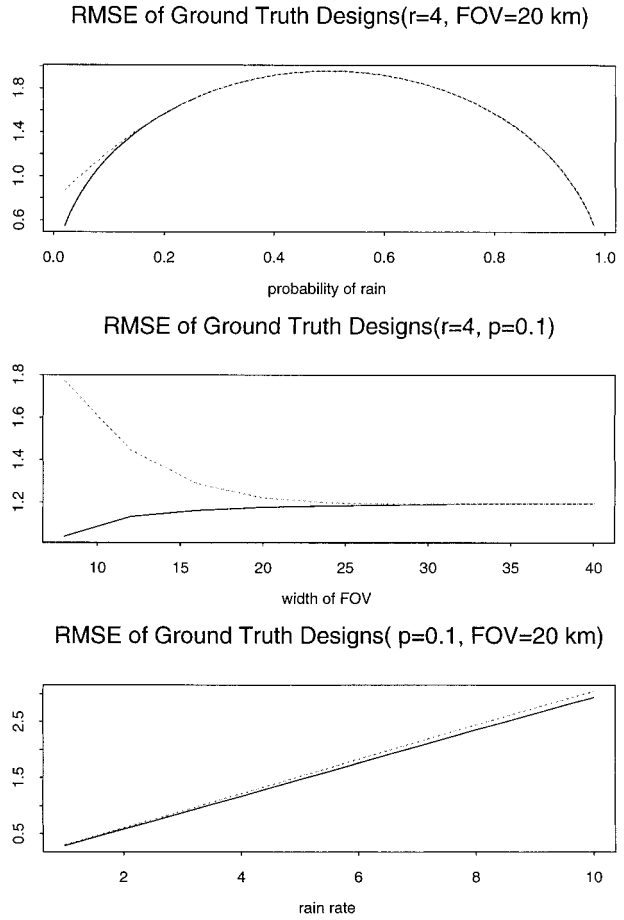


FIG. 4. The root-mean-square errors for design 1 and design 2 of Bernoulli random field when the resolution is 4 km.

$$W^2(d2) = \frac{\langle \epsilon_{d2}^2 \rangle}{\sigma_{g2}^2} = \frac{(1 - p)(1 - 1/M)}{1 - p/P_M}. \quad (32)$$

We can see from Eqs. (30) and (32) that the rain rate does not affect the values of $W^2(d1)$ and $W^2(d2)$. When the probability of the rain p is 1, the probability $P_M = 1 - (1 - p)^M = 1$ and thus the $W^2(d1)$ and $W^2(d2)$ are equal. The square roots of $W^2(d1)$ and $W^2(d2)$ are given in Fig. 5. For the values of $r = 4$, $p = 0.1$, and FOV size (20 km across), the $W^2(d1)$ and $W^2(d2)$ are almost the same.

d. Mean-square error for N visits

The error between a point and the satellite measurement for a single visit is likely to contain a large component of random error (see Fig. 2). 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

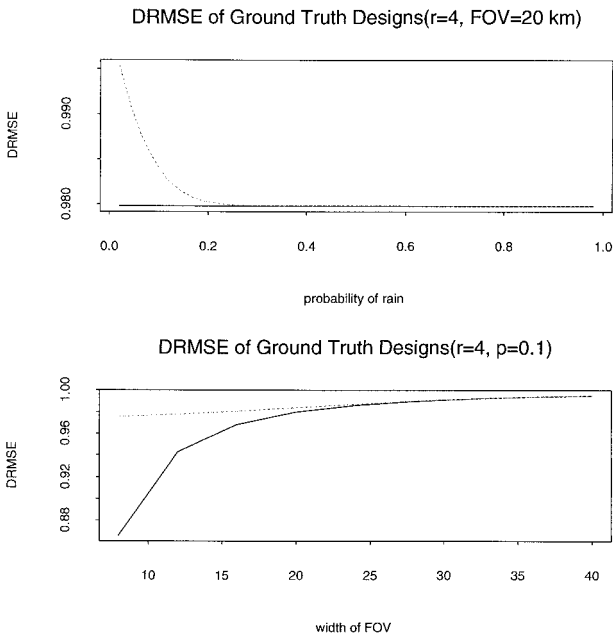


FIG. 5. The dimensionless root-mean-square errors for design 1 and design 2 of the Bernoulli random field when the resolution is 4 km. Key: design 1, solid line; design 2, dotted line.

the gauge measurement by adding independent measurements of several visits. The mean of the average error for N visits is the same as the mean of error for one visit. The variance of average error for N visits is

$$\langle \epsilon_N^2 \rangle = \frac{\langle \epsilon^2 \rangle}{N}. \tag{33}$$

By the central limit theorem (see Wilks 1995 for reference), we also know the distribution of the average error converges to the normal distribution when the number of visits N is sufficiently large, even though the error distribution for a single visit is further from a normal distribution (see Figs. 2 and 6). Figure 6 shows schematically the error distribution of one visit and the average error distribution of N independent visits. From (33), we can easily make the error as small as we want by adding a new measurement.

The dimensionless mean-square error for N visits is then

$$W_N^2(d) = \frac{W^2(di)}{N}, \quad i = 1, 2. \tag{34}$$

Let N_1 and N_2 be the number of visits to achieve the same tolerance level of detecting the retrieval bias using design 1 and design 2, respectively. We can derive the relationship between N_1 and N_2 using

$$\frac{W^2(d1)}{N_1} = \frac{W^2(d2)}{N_2}. \tag{35}$$

Using Eqs. (30) and (32), we thus have

$$N_2 = \frac{1 - p}{1 - \left(\frac{p}{P_M}\right)} N_1. \tag{36}$$

When the probability of rain p is small (usually less than 10%), the number of visits N_1 and N_2 are almost the same. Remember that N_1 is the number of all visits and N_2 is the number of visits when the satellite has rain.

We have already found that $W^2(d1)$ and $W^2(d2)$ are almost the same for the values $p = 0.1$, $r = 4$, and FOV size 20 km. For these values, the number of visits to achieve 10% of the standard deviation of the gauge measurement is $N_1 = 96$ for design 1 and is $N_2 = 97$ for design 2. Since the probability that the satellite measurement has rain $P_M = 1 - (1 - p)^M$ is almost 1 for $p = 0.1$, we obtained almost the same number of visits to achieve 10% tolerance level.

4. Summary and conclusions

In this paper we have considered three ground-truth designs based on the point gauge measurement to validate the satellite measurement. Design 1 uses data pairs from all visits. Design 2 uses visits only when the FOV average has rain. Design 3 uses the visits only when the gauge has rain. The ground-truth designs we proposed in this paper are based upon properties encountered with real rain.

We modeled the rain-rate field by subdividing the FOV into M tiles and took the random rain variable for each tile to be rain at a fixed rate r with probability p , or no rain to be $1 - p$. The number of tiles in a fixed FOV size corresponds in some sense to the autocorrelation length of the rain field. Once this model is adopted all comparisons between the point gauge (randomly located in the FOV) and the FOV average corresponding to the satellite measurement can be calculated.

We found that the error distribution is bimodal for design 1 and design 2 but not for design 3. We have 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. We have shown that the mean-square error of design 2 is equal to the mean-square error of 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 of design 2 without using the conditional random field. The dimensionless mean-square errors of design 1 and design 2 are almost the same for the typical rain statistics.

Our major finding is that for an FOV width of 20 km and autocorrelation length of a few kilometers, the number of measurement pairs (containing rain) necessary to distinguish a bias of 10% is of the order of 100. This means

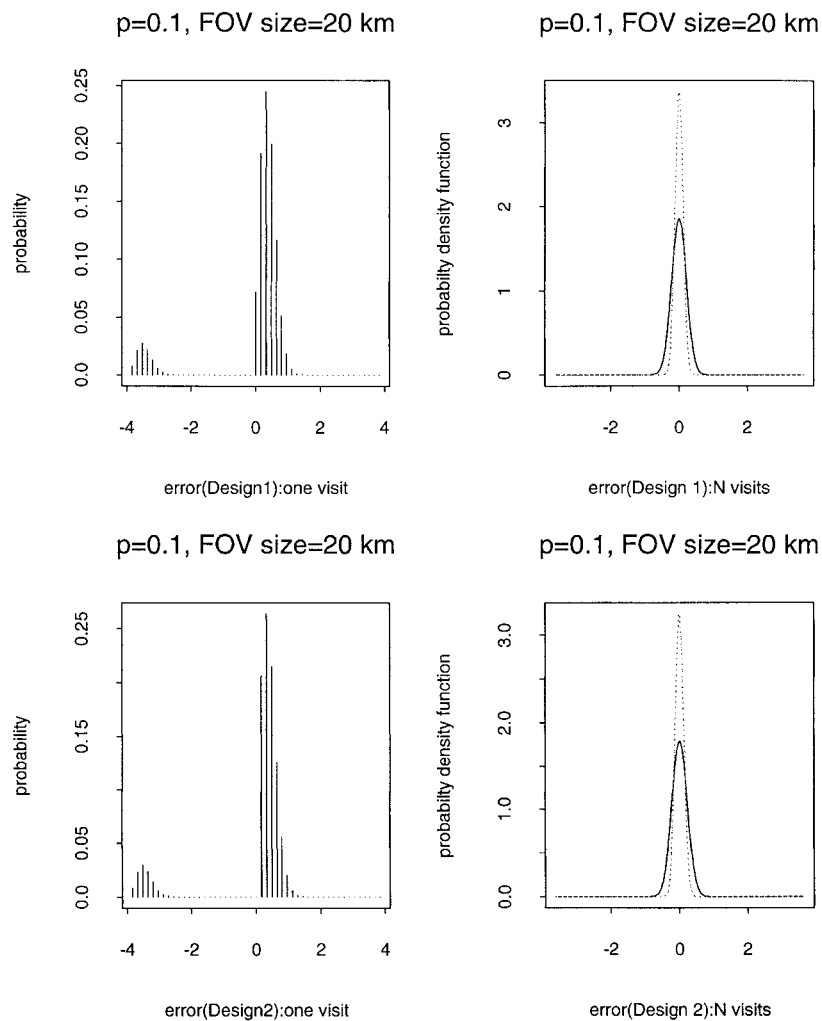


FIG. 6. The schematic distribution of the error for N visits. The Bernoulli random field with rain rate $r = 4 \text{ mm h}^{-1}$, the probability of rain $p = 0.1$, and the FOV size $20 \text{ km} \times 20 \text{ km}$ (resolution is 4 km) are used. Key: $N = 30$ visits, solid line; $N = 100$ visits, dotted line.

many months of data will need to be taken for a single-point gauge to detect such a bias with any confidence.

Our model of rain is too simplified and needs to be made more realistic at the cost of clarity. The Bernoulli field is too uniformly “speckly” compared to real rain, which consists of “patches” of speckly areas. We are presently looking at fields that have this more realistic property.

Acknowledgments. The first author (E.H.) wishes to thank the Korea Research Foundation for its support. The second author (G.R.N.) thanks the NASA TRMM program for its support.

REFERENCES

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.
 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.
 North, G. R., 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.
 Simpson, J., R. F. Adler, and G. R. North, 1988: A proposed rainfall measuring mission (TRMM) satellite. *Bull. Amer. Meteor. Soc.*, **69**, 278–295.
 Theon, J. S., T. Matsuno, T. Sakata, and N. Fugono, Eds., 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, T. Matsuno, T. Sakata, and N. Fugono, Eds., A. Deepak, 245–260.
 Wilheit, T. T., 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.
 Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 623 pp.