Formalism for Comparing Rain Estimation Designs

GERALD R. NORTH AND SHOICHIRO NAKAMOTO

Climate System Research Program, College of Geosciences, Texas A&M University, College Station, Texas

(Manuscript received 10 September 1988, in final form 4 May 1989)

ABSTRACT

Space–time averages of rain rates are needed in several applications. Nevertheless, they are difficult to estimate because the methods invariably leave gaps in the measurements in space or time. A formalism is developed which makes use of the frequency-wavenumber spectrum of the rain field. The mean square error of the estimate is expressed as an integral over frequency and two-dimensional wavenumber of an integrand consisting of two factors, a design-dependent-filter multiplied by the space–time spectrum of the rain rate field. Such a formalism helps to separate the design issues from the peculiarities of rain rate random fields. Two cases are worked out in detail: a low orbiting satellite which takes cell-wide snapshots at discrete intervals and a network of raingages which are gappy in space but continuous in time.

1. Introduction

Area–time averages of rain rates are needed in several applications in atmospheric science. For example, regional latent heat release rates which are important as drivers of the atmospheric general circulation can be monitored by recording area-average rain rates. In hydrology, time-smoothed basin-wide averages are needed in operational hydroelectric power management applications. Obtaining estimates of such area-time averages is difficult, since the area-time volume cannot be densely probed by rain rate sensors. Usually one must be content with certain sparce sampling designs on the area-time volume. The problem is exacerbated by the complicated and largely unknown statistical structure of rain rate fields. These evolving fields may be considered as space-time stochastic processes, but the statistics are hardly of the type routinely encountered in most applications. For example, they exhibit an extreme deviation from Gaussian—instead the probability density function at a point is a spike at zero rain rate and a continuum at finite rain rates; this combination is sometimes referred to as a mixed distribution (Kedem et al. 1989). Most of the probability is concentrated in the spike—it is raining only a few percent of the time at a point (or over a small area) in space. To further complicate matters, the continuum part of the probability density function is reasonably close to lognormal, a distribution known to have a rather fat tail (Kedem and Chiu 1987).

Corresponding author address: Dr. Gerald R. North, Climate System Research Program, College of Geosciences, Texas A&M University, College Station, TX 77843-03146.

The point to point autocorrelation structure of rain rate fields is also problematic. Studies from the GATE [GARP (Global Atmospheric Research Program) Atlantic Tropical Experiment] rain dataset (Hudlow and Paterson 1979) show that the correlations are not of a simple short range type, but rather tend to be power-law dependent over a range from 4 km separation out to possibly 70 km (Bell 1987; Bell et al. 1989). The autocorrelation times for area average rain rates depend strongly upon the magnitude of the area, being nearly a linear increase with length of the cell (Bell 1987; North 1987).

There is clearly a need for better measurements of the statistical properties of rain rate fields to help in arriving at sensible sampling designs, but we are left at this time with pitifully few observations of the type necessary to answer all of our design questions. Unfortunately, the need to acquire area-time datasets over the whole globe for the foregoing applications, especially those requiring large expenditures for space-based observational programs, force us to take up the problem even with incomplete knowledge of rain rate statistics in space and time. We have particularly in mind the meeting of recently discussed requirements of the atmospheric dynamics community for a precipitation dataset that is of about 15% accuracy for month averages and for areas of about 500 by 500 km (Simpson et al. 1988). For example, we wish to know how many raingages are necessary in such an averaging cell to ensure that the sampling errors (and instrument errors) will not exceed the tolerances set by the users. There is also the possibility of measuring the precipitation rates from low orbiting satellites that pass over the averaging cell. The sampling scheme of such a spaceborne sensor is very different from the gages, since the satellite returns to the cell after finite intervals (close to 12 hours) and takes a snapshot of the entire cell (or some fraction of its area). Hence, this design is gappy in time but continuous in space—quite the opposite of the gage design which is gappy in space but continuous in time.

In this study we present a formalism that allows the computation of sampling errors because of space-time gaps through an expression that is essentially a design-dependent filter that acts upon the space-time spectrum of the rain rate field. Previous works making use of space-time random fields of rain are summarized by Bras and Rodriguez-Iturbe (1985). Our spectral approach cleanly separates the problem of design from that of the statistics of the rain rate field. We can try the design filter on prototype rain field spectra to get a clearer idea of the inherent problems and find, for example, what aspects of rain rate statistics contribute most to our sampling problems. Another clarification coming from our study is the clear identification of the scale aspects of the problem. For example, there are three length scales to the problem (length of the edge of the averaging cell, distance between point gages and inherent length scales in the rain field). Similarly, there are three times in the problem (duration of the cell averages, interval between satellite overpasses and inherent time scales in the rain rate field). By isolating these factors and studying their individual influences, we shed light on the impact of an otherwise confusing array of parameters.

Earlier studies of satellite data using spectral methods include that of Salby (1982). Several relevant studies of sampling errors for raingage networks have been conducted (e.g., Silverman et al. 1981; Gabriel 1981; Zawadski 1973; Huff 1970).

The plan of the paper is first, to show our spectral formalism with the satellite overpasses and the raingages as explicit examples of filters; second, to illustrate the formalism with an idealized stochastic rain field model that permits analytical approximations to the errors for these designs. Of course we are aware of the simplicity of our rain field model which has inherent length and time scales, but ease of analytical manipulation for this model has motivated its use here. A similar class of models was used by Cahalan et al. (1982) to model cloudiness fluctuations. Bell (1987) has introduced a much more sophisticated rain field model, making use of all the statistical properties of the fields as learned from GATE. He has used it in some applications similar to those contemplated here (Bell et al. 1988). An interesting feature of his model is that it is rather closely imitated by a simpler first order Markov model in a number of satellite sampling simulations (Laughlin 1981; Shin and North 1988). This last result gives us some reason to think that our model results are actually better than one might have thought a priori.

2. Spectral formalism

Consider a random field $\psi(\mathbf{r}, t)$ defined in the x, y plane and along the time axis, t. Let the ensemble average of $\psi(\mathbf{r}, t)$ be zero and its variance at a point in space-time be σ^2. In most of what follows we will be interested in space-time averages of $\psi(\mathbf{r}, t)$,

$$\Psi \equiv \frac{1}{TL^2} \int_{-L/2}^{L/2} \int_{-L/2}^{L/2} \int_{-T/2}^{T/2} \psi(\mathbf{r}, t)dtdxdy. \quad (1)$$

We assume the field $\psi(\mathbf{r}, t)$ is statistically homogeneous and isotropic in space and stationary in time; its ensemble mean can be taken as zero. More precisely, if the measurement can be taken as unbiased, we could subtract away the long-term or ensemble mean. The ensemble mean of Ψ, denoted $\langle \Psi \rangle$, vanishes. The variance of Ψ can be computed from

$$\langle \Psi^2 \rangle = \frac{1}{(L^2T)^2} \int\int\int\int\int\int \langle \psi(\mathbf{r}', t')\psi(\mathbf{r}, t) \rangle \times d^2\mathbf{r}' d^2\mathbf{r}\, dtdt' \quad (2)$$

where the integrals run over the volume of the space-time box. The integrand in (2) is the space-time lagged covariance. For spatially homogeneous and temporally stationary fields this function is a function only of $\boldsymbol{\xi} = \mathbf{r} - \mathbf{r}$, and $\tau = t - t'$. We denote the lagged covariance by

$$\langle \psi(\mathbf{r}', t')\psi(\mathbf{r}, t) \rangle = \sigma^2 \rho(\boldsymbol{\xi}, \tau) \quad (3)$$

where $\rho(\boldsymbol{\xi}, \tau)$ is the space-time lagged autocorrelation. The restriction to homogeneity and stationarity is not thought to be serious in the type of study envisioned here. It is a common approximation to use especially over homogeneous oceanic regions (e.g., Bell 1987). We define the normalized wavenumber-frequency spectrum of the space-time field by the Fourier transform (FT) of the autocorrelation

$$S(\boldsymbol{\nu}, f) = \int\int\int \rho(\boldsymbol{\xi}, \tau) e^{2\pi i(\boldsymbol{\xi}\cdot\boldsymbol{\nu} + f\tau)} d^2\boldsymbol{\xi} d\tau \quad (4)$$

where the integrals run over all space. It follows that

$$\int\int\int S(\boldsymbol{\nu}, f) d^2\boldsymbol{\nu} df = \rho(0, 0) = 1. \quad (5)$$

Also of interest is the FT of the field itself

$$\tilde{\psi}(\boldsymbol{\nu}, f) = \int\int\int e^{2\pi i(\boldsymbol{\nu}\cdot\mathbf{r} + ft)} \psi(\mathbf{r}, t) d^2\mathbf{r}\, dt \quad (6)$$

and its FT inverse

$$\psi(\mathbf{r}, t) = \int\int\int e^{-2\pi i(\boldsymbol{\nu}\cdot\mathbf{r} + ft)} \tilde{\psi}(\boldsymbol{\nu}, f) d^2\boldsymbol{\nu} df \quad (7)$$

where again all three limits are from minus to plus infinity. Using the above Fourier transform pair, we can express the normalized spectrum in terms of the inverse FT as follows:

$$\langle \tilde{\psi}(\nu', f')\tilde{\psi}(\nu, f)\rangle = \sigma^2\delta^{(2)}(\nu' - \nu)\delta(f' - f)S(\nu, f). \quad (8)$$

In deriving the last formula, we made use of the well-known formula:

$$\int_{-\infty}^{\infty} e^{ikx}dx = 2\pi\delta(k). \quad (9)$$

After some rearrangements, we may construct the variance of the space-time average:

$$\langle \Psi^2 \rangle = \sigma_V{}^2 = \sigma^2 \iiint G(\pi\nu_1 L)^2 G(\pi\nu_2 L)^2 G \times (\pi f T)^2 S(\nu, f) d^2\nu dt \quad (10)$$

where $G(\pi x)^2 \equiv [\sin(\pi x)/\pi x]^2$ is familiar from physical optics and in time series analysis is called the Bartlett filter (Blackman and Tukey 1959). It is plotted in Fig. 1. An interpretation of (10) is that the space-time box averaging of the field is represented as a three-dimensional low-pass filter acting on the space–time power spectrum of the field.

Also of interest is the spatial average of the random process over the cell area L^2. The variance of the spatial average can be represented in a similar form:

$$\sigma_A{}^2 = \sigma^2 \iiint G(\pi\nu_1 L)^2 G(\pi\nu_2 L)^2 S(\nu, f) d^2\nu df. \quad (11)$$

Clearly, $\sigma_V{}^2 \leqslant \sigma^2$ and $\sigma_A{}^2 \leqslant \sigma^2$, since $\iiint S(\nu, f) \times d^2\nu df = 1$ and $G^2 \leqslant 1$, $S \geqslant 0$. Note that S has units *(length)*^{2}*time.*

An intent of this paper is to analyze systematically the errors incurred in estimating Ψ by various statistical measurement designs. For example, one design of interest is that of a low altitude satellite that passes over the area L^2, taking an instantaneous area average, then returning at intervals Δt until the entire period T is covered. The snapshot area-means are then averaged to form an estimate of the space-time average over the volume L^2T. Other sampling designs that need to be combined with this one, or compared with it, easily come to mind, and we will discuss some of these later. However, we present next the satellite design example explicitly.

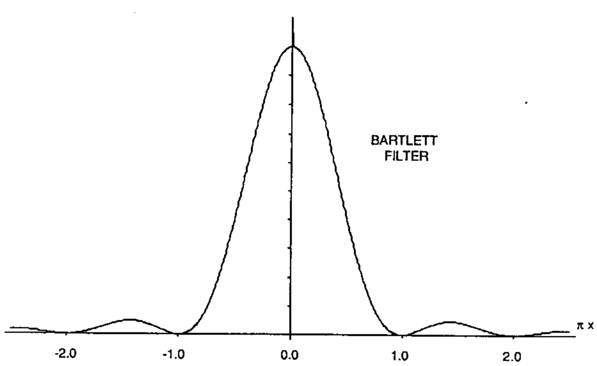

FIG. 1. The Bartlett filter $\left[\frac{\sin(\pi x)}{\pi x}\right]^2$.

Another approach to the problem of satellite sampling errors is to construct a random field that mimics real data and conduct simulations of the evolution of the field with overflights by an imaginary satellite with precise orbital parameters specified. Such studies have been conducted by Bell (1987), Shin and North (1988) and Bell et al. (1989). The advantage of the Monte Carlo simulation procedure is that very accurate information about overpasses can be injected into the program. The advantage of the present formalism is clarification of the underlying problems involved in the design issue. Hence, precision is not the goal of the present undertaking; rather, we wish to understand better the origin of the sampling errors.

3. Mean squared errors and sampling designs

Consider the satellite overflight case discussed previously. The estimator may be written

$$\Psi_S = \frac{1}{NL^2}\sum_{n=0}^{N-1}\iint \psi\left(\mathbf{r}, \left(n + \frac{1}{2}\right)\Delta t\right)d^2\mathbf{r} \quad (12)$$

where N is the total number of visits in the averaging interval ($T = N\Delta t$) and the integral is over the box (Fig. 2). Equation (12) can also be written

$$\Psi_S = \frac{1}{L^2T}\iiint \Delta t \sum_{n=0}^{N-1} \delta\left(t - \left(n + \frac{1}{2}\right)\Delta t\right) \times \psi(\mathbf{r}, t)d^2\mathbf{r}dt \quad (13)$$

or

$$\Psi_S = \frac{1}{L^2T}\iiint K(\mathbf{r}, t)\psi(\mathbf{r}, t)d^2\mathbf{r}dt \quad (14)$$

where

$$K(\mathbf{r}, t) = \Delta t \sum \delta\left(t - \left(n - \frac{1}{2}\right)\Delta t\right)$$

is a kernel representative of the discrete time sampling design.

Now consider the mean square error made in estimating Ψ using this design

$$\epsilon^2 = \langle(\Psi - \Psi_S)^2\rangle \quad (15)$$

which can be written

$$\epsilon^2 = \frac{1}{L^4T^2}\iiiint\iint \langle \psi(\mathbf{r}', t')\psi(\mathbf{r}, t)\rangle \times [1 - K(\mathbf{r}', t')][1 - K(\mathbf{r}, t)]d^2\mathbf{r}'dt'd^2\mathbf{r}dt. \quad (16)$$

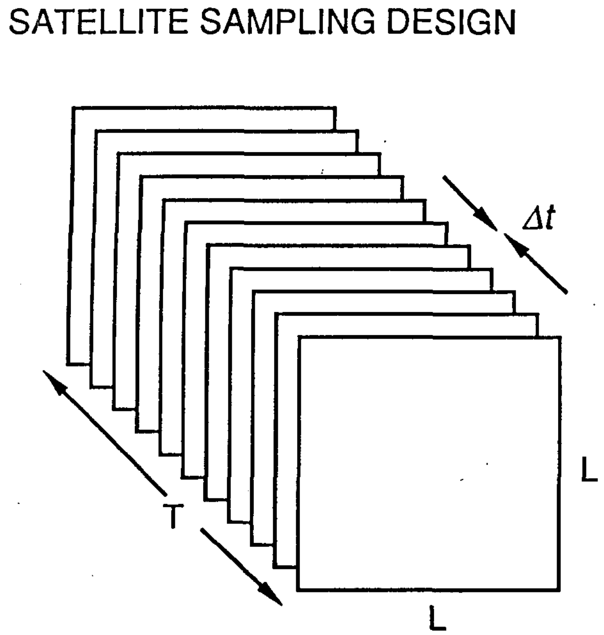

FIG. 2. A schematic diagram of the satellite design. The instrument passes over the averaging cell at intervals Δt, making a complete snapshot of the entire cell. The snapshots are then aggregated to form an estimate of the space-time average.

Using the Fourier representation, we may write

$$\epsilon^2 = \iiint |H(\nu, f)|^2 S(\nu, f) d^2\nu df. \quad (17)$$

This is the desired form which indicates the error as a filtered integral over all frequencies and wavenumbers of the space-time spectrum of the geophysical process. Note that the information about the design and the properties of the rain rate field are clearly factored.

Specifically in the case of the satellite design:

$$|H_S(\nu, f)| = G(\pi\nu_1 L)G(\pi\nu_2 L)G(\pi f T) \times \left[1 - \frac{1}{G(\pi f \Delta t)}\right]. \quad (18)$$

Analysis of (17) and (18) is quite interesting; this will be pursued in section 4. Let us first consider another very important design, that of the network of raingages.

We can work through the same formalism for the regular array of raingages as we were able to develop for the satellite snapshot case. In this case we picture a rectangular array of point samplers in space but, in this instance, continuous in time (Fig. 3). The analog to (12) is

$$\Psi_G = \frac{1}{TM^2} \int \sum_{n_1=0}^{M-1} \sum_{n_2=0}^{M-1} \psi\left(\left(n_1 + \frac{1}{2}\right)\Delta l, \left(n_2 + \frac{1}{2}\right)\Delta l, t\right) dt. \quad (19)$$

The analogous formulas to (13) and (14) are easily constructed from (19) and the raingage equivalent to the design filter (18) becomes

$$|H_G(\nu, f)| = G(\pi f T)G(\pi\nu_1 L)G(\pi\nu_2 L) \times \left[1 - \frac{1}{G(\pi\nu_1\Delta l)G(\pi\nu_2\Delta l)}\right]. \quad (20)$$

4. A simple rain field example

In this section we introduce a random field generated by a simple stochastic model. The field is admittedly simplistic compared to real two-dimensional evolving rain fields. For example, the field has inherent length and time scales with a "red" type spectrum in space and time. The field is also Gaussian which is quite different from real rain. (Actually, the formalism above does not depend upon Gaussian statistics). Nevertheless, we feel that it is important to work with a concrete example and use our field as such. Later, comparisons will be speculated upon for the case of real rain.

Consider the model rain field $\psi(\mathbf{r}, t)$ generated by the process governed by

$$\tau_0 \frac{\partial\psi}{\partial t} - \lambda_0^2 \nabla^2 \psi + \psi = F(\mathbf{r}, t) \quad (21)$$

where τ_0 is a time scale and λ_0 is a length scale, both inherent to the field; F is a noise forcing function which might, for example, be white in both space and time

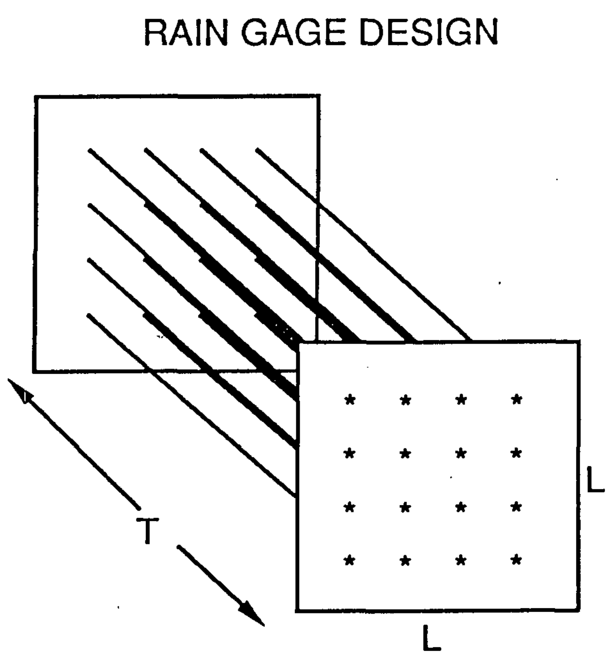

FIG. 3. A schematic diagram of the raingage design. The gages are situated in a discrete array in the cell and make measurements continuously in time. The averages are then aggregated to form the space-time average.

(up to some cutoff frequency and wave number). The motivation for the model is weak but worth stating: Rain rates are generated and destroyed by a random uncorrelated source in space and time; the field of rain rates is modified by decay of the whole field at time scale τ_0; finally, rain rates are displaced from one point to another by a simple down gradient diffusion process.

The FT of $\psi(\mathbf{r}, t)$ is easily computed directly from (6) and (21):

$$\tilde{\psi}(\nu, f) = \frac{\tilde{F}(\nu, f)}{2\pi i\tau_0 f + (1 + 4\pi^2\lambda_0^2\nu^2)}. \quad (22)$$

The spectrum is computed from (7)

$$S(\nu, f) = \frac{\alpha}{4\pi^2\tau_0^2 f^2 + (1 + 4\pi^2\lambda_0^2\nu^2)^2} \quad (23)$$

where

$$\alpha = \frac{8\pi\tau_0\lambda_0^2}{\ln(1 + 4\pi^2\lambda_0^2\nu_c^2)^2} \quad (24)$$

is a normalization factor which ensures that (5) holds. Note here that to ensure convergence a cutoff ν_c has been introduced; this latter can be thought of as the smallest scale excited by the white spatial driving noise.

The spectrum (23) is a well-behaved function, smooth and maximum at the origin and tapering off in all directions in f and $|\nu|$. When we compare (23) to tropical rain data, we find that reasonable values for τ_0 and λ_0 are about 12 hours and 40 km respectively (Bell 1987; Bell et al. 1989; Shin and North 1988; cf. also Kedem et al. 1989). If our space-time averaging box is taken to be 30 days by 500 km, we see that the Bartlett filters $G(\pi\nu L)^2$ and $G(\pi f T)^2$ have effectively fallen to zero (as $|\nu|$ and f increase) before $S(\nu, f)$ differs appreciably from its value at the origin, α. We may then approximate $S(\nu, f)$ by α and calculate, for example, the value of σ_A^2 from (11). Noting that

$$\int_{-\infty}^{\infty} G(\pi x L)^2 dx = \frac{1}{L} \quad (25)$$

we find

$$\sigma_A^2 \approx \sigma^2 \frac{\alpha}{2\tau_0 L^2} \quad (26)$$

which is a convenient phenomenological factor for absorbing such dependences as the cutoff parameter, since σ_A^2 is an observable quantity easily related to measurements. One interesting fact revealed in (26) is that the variance of area averages is inversely proportional to the box area in units of the fundamental length scale λ_0. When a length scale exists in the field such as the present, we can interpret this finding in terms of a simple standard error of estimating the mean: the variance is inversely proportional to the number of independent samples—there being of the order of one independent sample in each box whose side is λ_0 in length.

Now we come to the problem of estimating the errors for the satellite design on this random field. The important factor in $|H(\nu, f)|^2$ to be considered is

$$G(\pi f T)^2 \left[1 - \frac{1}{G(\pi f \Delta t)}\right]^2$$

which can be written

$$B(f) = \frac{\sin^2(\pi f N \Delta t)}{N^2 \sin^2(\pi f \Delta t)} [1 - G(\pi f \Delta t)]^2, \qquad N\Delta t = T. \quad (27)$$

This form is easily analyzed if it is noted that

$$\lim_{N\to\infty} \frac{\sin^2 N\pi\Delta t f}{N \sin^2 \pi\Delta t f} = \frac{1}{\Delta t} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{\Delta t}\right) \quad (28)$$

which is a series of spikes at $f = n/\Delta t$, for all positive and negative integers n; it is known as the Dirac comb (Blackman and Tukey 1959) and is shown in Fig. 4 for the case $N = 10$. The second factor in (27) is unity at each of the spikes except zero, where it vanishes; this function is shown in Fig. 5. Thus, the error becomes to a very good approximation

$$\epsilon^2 \approx \frac{\sigma^2\alpha}{L^2} \frac{1}{N\Delta t} \int df \sum_{n\neq 0} \frac{\delta(f - n/\Delta t)}{1 + 4\pi^2 f^2 \tau_0^2}. \quad (29)$$

In the special case of interest, $N\Delta t = T$, $\Delta t \approx \tau_0$, $N = 60$. The product of (27) and the spectrum are schematically depicted in Fig. 6, but for $N = 10$. Note that the spikes are six times narrower for $N = 60$. This figure shows clearly how well the series converges. Using (26), we may write

$$\epsilon^2 \approx \sigma_A^2 \frac{1}{\pi^2} \frac{\Delta t}{T} \frac{\Delta t}{\tau_0}. \quad (30)$$

Inserting the numbers: $\epsilon \approx 0.04\sigma_A$; in other words, the

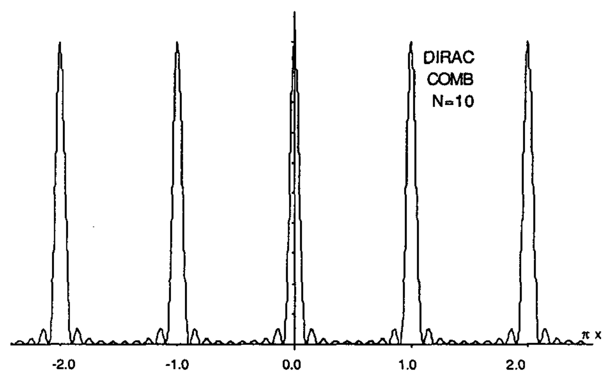

FIG. 4. A plot of the Dirac comb given by Eq. (28) for $N = 10$. Note that the width of the spikes is inversely proportional to N. Hence in more realistic cases the spikes are much narrower, but preserve the unit area under each.

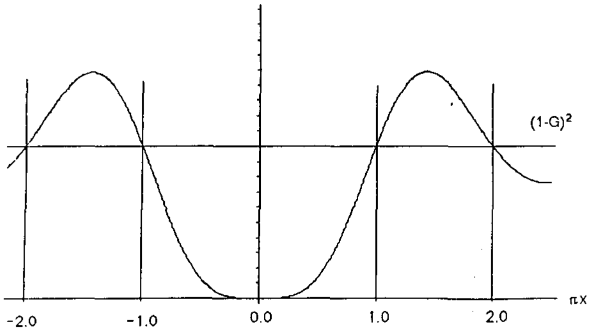

FIG. 5. The high-pass filter part of $B(f)$ from Eq. (27). Note that this factor obliterates the Dirac comb at the origin but preserves it at all other integer values.

error is expected to be about 4% of the area average rain rate standard deviation. We can compare this with a similar study by Laughlin (1981) which did not use a spectral formalism, but used a first order autoregressive process to model area average rain rates. Laughlin used an area 280 by 280 km^2 for which the autocorrelation time was 6 hours; hence we should use $\tau_0 = 0.5\Delta t$, instead of $\tau_0 = \Delta t$, and for easy comparison, we must also quote the result as a percentage of the mean rain rate. In Laughlin's case $\sigma_A/\mu = 1.3 \pm 0.1$, based upon GATE data. Scaling our result accordingly, we arrive at an error of 10.4% which is in exact agreement. In recent studies by Bell et al. (1989) and Shin and North (1988), the case of 500 by 500 km boxes was used and a value of $\sigma_A/\mu = 1.00$ was used by extrapolation from GATE data; furthermore, the larger area leads to $\Delta t/\tau_0 = 1.0$. The result is a sampling error of 4%. This is slightly misleading since in the real satellite case, the visits to the cell do not cover the entire cell on each visit, but rather are only partial intersections of the satellite swath with the cell. Bell et al. (1989) and Shin and North (1988) showed that for realistic orbits and crossings, this approximately doubles the sampling error.

It is interesting to return to the interpretation of (29). Consider the meaning of the sum contributing to mean square error. Note that the sum consists of contributions from the power spectrum at harmonics (Fig. 5) of the frequency $1/\Delta t$, twice the so-called Nyquist frequency. The smaller the spectrum is at frequencies above this important value, the smaller will be the error. In fact, if the spectrum is zero above some critical frequency that is below the Nyquist frequency, there will be no error; this is equivalent to the famous "sampling theorem," well known in communications theory (e.g., Blackman and Tukey 1959). The conclusion is that the errors will be quite small ($\approx$4%) when the satellite revisit time is the same as or smaller than the autocorrelation time for the time series generated area average of the field.

Consider the satellite revisit case further. If Δt is 12 hours, then the spikes of Fig. 6 are at the semidiurnal time scale and harmonics of it. Why should the error depend on the spectrum at the semidiurnal frequency and its harmonics? From Fig. 7 an attempt is made to

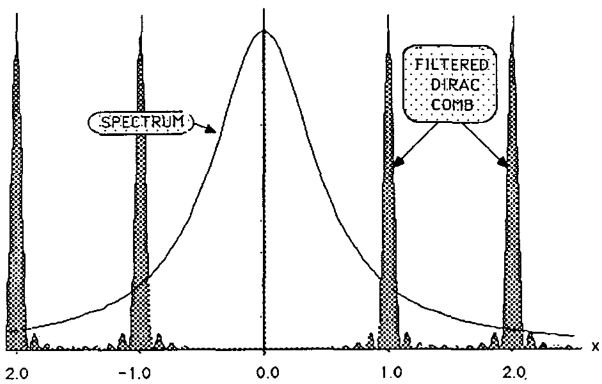

FIG. 6. The filter $B(f)$ with a plot of the temporal spectrum for zero spatial wavenumber. Note that the filter picks up contributions to the error from the temporal spectrum at twice the Nyquist frequency and all of the harmonics of it. Since the spectrum falls off as the square of frequency, the series of contributions picked up by the Dirac comb converges rapidly. In reference to the case of a satellite returning every 12 hours, $x = 1$ corresponds to the semidiurnal frequency.

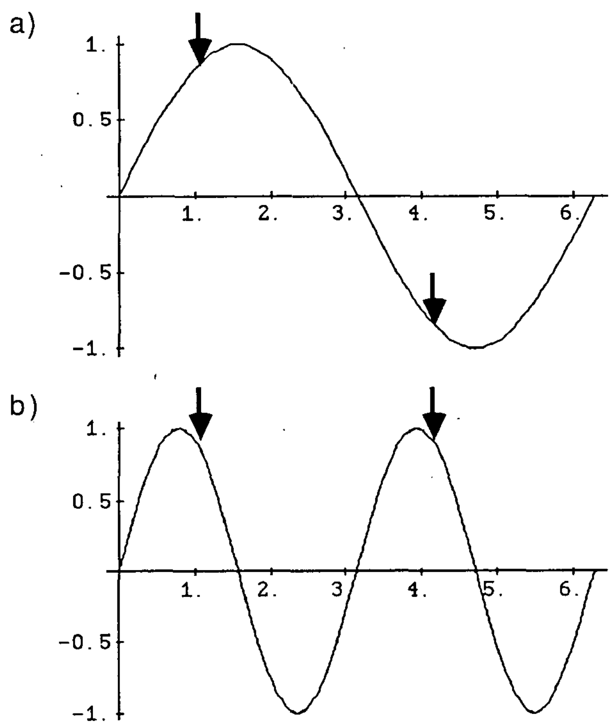

FIG. 7. Schematic illustration of errors incurred in estimating the mean (zero) when diurnal and semidiurnal cycles occur in the satellite case of return visits every 12 hours. (a) Variance (spectral power) at the diurnal cycle leads to cancellation of the errors made in the two visits. (b) Variance in the semidiurnal component leads to systematic error since the two overestimates reinforce.

answer this question intuitively. In the upper panel, we see that measurements of the diurnal Fourier component at half period intervals lead to a large value on the first visit but an off setting low value on the next pass. On the other hand, the semidiurnal component will lead to reinforcing contributions on both passes. If the semidiurnal cycle is deterministic, the error is a bias; if the semidiurnal component is phase randomized, the error will be random but of the magnitude specified by (30).

5. Raingage example

We wish now to evaluate the mean square error for the case of raingages represented by (19) and (20). The same types of manipulations that lead to (27) and the same limit (28) can be applied here. We find that

$$|H_G(\nu,f)|^2 = \frac{G(\pi f T)^2}{L^2} \sum_{n_1,n_2\neq 0} \delta\left(\nu_1 - \frac{n_1}{\Delta l}\right) \times \delta\left(\nu_2 - \frac{n_2}{\Delta l}\right). \quad (31)$$

Inserting into the error formula (17),

$$\epsilon^3 = \sigma^2 \int \sum_{n_1,n_2\neq 0} S\left(\frac{n_1}{\Delta l}, \frac{n_2}{\Delta l}, f\right) G(\pi T f)^2 df \quad (32)$$

which for the simple model mentioned above becomes

$$\epsilon^2 = \sigma_A{}^2 \frac{8\tau_0}{T} \sum_{n_1,n_2=1}^{\infty} \frac{1}{[1 + 4\pi^2(\lambda_0/\Delta l)^2 (n_1{}^2 + n_2{}^2)]^2}. \quad (33)$$

Evaluating this series out to several tens of terms suggests that to achieve the 4% sampling errors found before for monthly averages, we need to have a raingage about every 108 km (assuming $\lambda_0 \approx 40$ km). This means there should be more than 20 gages in every 500 km box for error parity between the two designs.

Lest the reader adopt (33) and the sparce gage network it implies without further consideration, we need to point out that this rain field model has an exaggerated suppression of variance at high wavenumbers—the very variance that might not be picked up because of gaps between gages. For example, a model without diffusion but rather pure horizontal advection of the (temporally red noise) rain would replace the diffusion term with $\tau_0 \mathbf{V}_0 \cdot \nabla \psi$, where $\mathbf{V}_0$ is the advection velocity which we may take to be along the x axis. This model was used by Cahalan et al. (1982) to fit cloud fluctuation data and by Gupta and Waymire (1987) as a model of rain fields to examine the Taylor hypothesis. From the advection model we find that the power at low temporal frequencies falls off only as the second power of wavenumber in the x direction and is flat in the y direction (white). This latter forces us to use a cutoff wavenumber which for definiteness we take (conservatively) as corresponding to a wavelength of 10 km. When we take the advection velocity to be 5 km h^{-1}, we may evaluate the error formula and find that the gages must be of the order of 30 km apart to have the same 4.5% errors encountered in the perfect satellite case. Now the required density of gages is prohibitively large (nearly 300) for the 500 km averaging cells. If we relax the requirement to 8% errors, the length between gages can be estimated since it scales as the ⅔ power of the error. In this case the distance between gages enlarges by a factor of 1.59. In the purely advective noise case, we have then a gage separation of 43 km.

The important lesson to be drawn from the raingage design study is that the shape of the wavenumber spectrum at low frequencies and high wavenumbers is very important in its contribution to sampling errors for discrete gage networks. By high wavenumbers we mean at those corresponding to twice the spatial Nyquist wavenumber and at harmonics of it. This is physically reasonable since this is the part of the variance missed by this kind of design. Since we do not as yet know what this spectrum is for real rain, we are not yet in a position to know how many gages are necessary to give monthly area averages as good as those from a single satellite crossing every 12 hours.

6. Conclusions

We have presented a formalism that should prove useful in future studies pertaining to the evaluation of sampling strategies for geophysical fields. The formalism is simple and intuitive: i.e., factoring the dependence of the physical phenomenon (space-time spectrum) away from the dependence on the measurement design (spectrum filter). We demonstrated the formalism with a very simple class of rain field models; namely, a diffusive model and one which employed advected spatial and temporal red noise. We find small sampling errors (about 5%) for a satellite design that makes cell-wide snapshot evaluations of the averaging cell about once every autocorrelation time for the large-scale field. Comparing to a spatially sparse but temporally continuous raingage design is more difficult, since at this time we need to have a better estimate of the low frequency, high wavenumber portion of the rain field spectrum. However, present estimates suggest that a gage array of one every 30 to 100 km will be necessary to have the same sampling errors as a satellite system.

We believe the formalism presented here can be used in several studies involving rain parameter estimation. For example, one can take into account the actual satellite visiting sequence and its partial covering of cells on individual visits by suitable random variable techniques. One can also take into account the measurement errors associated with individual instruments such as thermal noise in microwave radiometers. The formalism can be extended to evaluate designs for estimating diurnal harmonic amplitudes, as well as various low frequency oscillatory phenomena in the global atmospheric system. A similar formalism can be used in modeling the ground truth process in satellite retrievals. Finally, optimum weighting strategies can be developed for combining designs into grander systems. We anticipate that some restrictions such as the spatial homogeneity assumption can be relaxed with the use of location dependent spatial correlation statistics and their associated expansion functions (see e.g., North 1984).

REFERENCES

Bell, T. L., 1987: A space-time stochastic model of rainfall for satellite remote-sensing studies. *J. Geophys. Res. Atmos.*, **92**(D8), 9631–964.

——, A. Abdullah, R. Martin and G. R. North, 1989: Monte Carlo study of sampling errors for satellite-derived tropical rainfall using space-time stochastic model. *J. Geophys. Res.*, in press.

Blackman, R. B., and J. W. Tukey, 1959: *The Measurement of Power Spectra.* Dover Publications, 190 pp.

Bras, R. L., and I. Rodrigues-Iturbe, 1985: *Random Functions and Hydrology.* Addison-Wesley, 559, pp.

Cahalan, R. F., D. A. Short and G. R. North, 1982: Cloud fluctuation statistics. *Mon. Wea. Rev.*, **110**, 26–42.

Gabriel, K. R., 1981: Gage density and variability of rainfall estimates: A complement to Silverman, Rogers and Dahl's study. *J. Appl. Meteor.*, **20**, 1537–1542.

Gupta, V., and E. Waymire, 1987: On Taylor's hypothesis and dissipation in rainfall. *J. Geophys. Res.*, **92**, 9657–9660.

Hudlow, M., and V. Paterson, 1979: GATE Radar Rainfall Atlas. NOAA Special Rep. Government Printing Office, Washington, DC 158 pp.

Huff, F. A., 1970: Sampling errors in measurement of mean precipitation, *J. Appl. Meteor.*, **11**, 35–44.

Kedem, B., and L. Chiu, 1987: On the lognormality of rain rate. *Proc. Nat. Acad. Sci.*, **84**, 901–905.

——, —— and G. R. North, 1989: Estimation of mean rain rate: Application to satellite observations. *J. Geophys. Res.*, in press.

Laughlin, C., 1981: On the effect of temporal sampling on the observation of mean rainfall. *Precipitation Measurements from Space, NASA/Workshop Report,* Greenbelt, Md. D. Atlas and O. Thiele, Eds. [Available from Goddard Space Flight Center, Greenbelt, MD 20771.], 61–66.

North, G. R., 1984: Empirical orthogonal functions and normal modes. *J. Atmos. Sci.*, **41**, 879–887.

——, 1987: Sampling studies for satellite estimation of rain. *Tenth Conference on Probability and Statistics*, Edmonton, Canada, *Amer. Meteor. Soc.*,

Salby, M. L., 1982: Sampling theory for asynoptic satellite observations. Part I: Space-timespectra, resolution, and aliasing. *J. Atmos. Sci.* **39**, 2577–2600.

Shin, K.-S., and G. R. North, 1988: Sampling error study for rainfall estimates by satellite using a stochastic model. *J. Appl. Meteor,* **11**, 1218–1231.

Silverman, B. A., L. K. Rogers and D. Dahl, 1981: On the sampling variance of raingage networks. *J. Appl. Meteor.*, **20**, 1468–1478.

Simpson, J., R. Adler and G. R. North, 1988: On some aspects of a proposed tropical rainfall measuring mission (TRMM). *Bull. Amer. Meteor. Soc.*, **69**, 278–295.

Zawadski, I. I., 1973: Errors and fluctuations of raingage estimates of areal rainfall. *J. Hydrol.*, **18**, 243–255.