Ann. Inst. Statist. Math.
Vol. 48, No. 2, 201–213 (1996)

THE THRESHOLD METHOD FOR ESTIMATING TOTAL RAINFALL

SHIGERU MASE

Division of Mathematical and Information Sciences, Faculty of Integrated Arts and Sciences, Hiroshima University, Kagamiyama 1-3, Higashi-Hiroshima 739, Japan

(Received June 3, 1994; revised August 4, 1995)

Abstract. The threshold method estimates the total rainfall F_G in a region G using the area B_G of the subregion where rainfall intensity exceeds a certain threshold value c. We model the rainfall in a region by a marked spatial point process and derive a correlation formula between F_G and B_G. This correlation depends not only on the rainfall distribution but also on the variation of number of raining sites, showing the importance of taking account of the spatial character of rainfall. In the extreme case where the variation of number of raining sites is dominant, the threshold method may work regardless of rainfall distributions and even regardless of threshold values. We use the lattice gas model from statistical physics to model raining sites and show a huge variation in the number of raining sites is theoretically possible if a phase transition occurs, that is, physically different states coexist. Also, we show by radar observation datasets that there are huge variations of raining sites actually.

Key words and phrases: Threshold method, meteorology, rainfall, radar, marked spatial point process, linear regression, lattice gas model, Gibbsian process, potential, phase transition, non-ergodicity.

1. Introduction

It is widely recognized that meteorological conditions of tropical regions have world wide influences and meteorological data from these areas is important in understanding many problems. Nevertheless, observation in tropical regions (including land and sea) is frequently difficult and making the use of satellites is the only practical possibility. The TRMM (Tropical Rainfall Measuring Mission) program, a Japan-USA joint national project, is planning to send a first satellite equipped with a rainfall radar in 1997 to collect rainfall data in tropical regions. The threshold method is one of the algorithms which will be used in the TRMM program to estimate rainfall intensity from measurement of the rainfall radar of the satellite.

The threshold method has its basis on an empirical fact as observed in Chiu's (1988) systematic analysis of the GATE datasets. (The Garp Atlantic Tropical

Experiment, GATE, was the first major international experiment of the Global Atmospheric Research Program, GARP. It was conducted over the tropical Atlantic ocean and adjacent land areas under the joint auspices of the World Meteorological Organization, WMO, and the International Council of Scientific Unions, ICSU, from June through September 1974, see Patterson *et al.* (1979).) He pointed out an almost linear relation between total rainfall in a region and the area of the subregion where rainfall intensity exceeds a certain threshold value. For a suitable choice of the threshold value, the square R^2 of the correlation coefficient can be very close to 1 ($R^2 = 0.98$ for GATE phase I dataset and $R^2 = 0.97$ for GATE phase II dataset). Although such high correlations are only possible for carefully chosen threshold values, we can also notice a curious fact that the R^2-value is always fairly high irrespective of threshold values. For example, $R^2 = 0.78, 0.89, 0.98, 0.95, 0.85$ for GATE phase I dataset and $R^2 = 0.71, 0.85, 0.97, 0.96, 0.87$ for GATE phase II dataset for threshold values 0, 1, 5, 10, 20 mm/hour (optimal threshold values are both 5 mm/hour). The threshold method is attractive because it allows us to estimate rainfall intensity even if we cannot measure lower values of rainfall intensity precisely, which is usual in radar measurement.

Observation as in Chiu's study has been reconfirmed in various other rainfall datasets and corresponding optimal threshold values are given. Also, several theoretical studies have been done to explain why the threshold method works and to give a method of deciding the optimal threshold value, see, e.g., Braud *et al.* (1993), Shimizu (1992), Shimizu *et al.* (1993), Short *et al.* (1993), and their references. These studies are based mainly on the modeling of rainfall intensity distribution by mixtures of the form $(1 - \lambda)\delta + \lambda\rho$ where δ is the unit mass at the origin, λ is the probability of rain, and ρ is the conditional distribution of rainfall intensity under the condition of rain. A possible criticism of them is that they do not take account of the apparent spatial character of the problem. In this research, we propose a framework of modeling the problem as a spatial phenomenon.

Our model, the *marked spatial point process model*, consists of two spatial processes. The first is the point process X which indicates random positions of sites (discretized as lattice points) where it rains. The second is an accompanying random field S which represents (potential) rainfall intensities (instantaneous or cumulative in a unit area and a unit period) observable at raining sites. We will give the formula of the theoretical correlation between the total rainfall F_G within a base region G and the area B_G of sites in G where rainfall intensities exceed a given threshold value. This formula is seen to have a spatial character, that is, it depends not only on the distribution ρ of rainfall but also on the spatial variation of number of raining sites, showing the importance of taking account of the spatial character of rainfall. An important consequence is that, in an extreme case where a variation of the number of raining sites is dominant, the rainfall distribution and even the threshold value may be of secondary importance. It is our opinion that this explains why the correlation is always fairly high regardless of threshold values.

In order to show that such a large variation of numbers of raining sites is possible at least theoretically, we model the raining site process by the lattice gas model which is well-known in statistical physics. Although there is no easily verifiable general condition, it is known that a huge variation of number of raining sites can occur if a phase transition exists, that is, physically different raining phases coexist. This is also known to be equivalent to the non-ergodicity of raining sites process.

Finally, we analyze a radar observation dataset, the TRMMGT dataset measured by the NOAA/TOGA C-band radar. This analysis actually shows variations of number of raining sites which are far larger than those the independent raining sites assumption predicts.

The main purpose of this research is the spatial formulation of the problem and the theoretical explanation why the threshold method works so fine. The actual estimation procedure is not discussed. One may use a simple linear regression once a strong correlation is observed.

2. Preliminary

Let $\mathbb{Z}^2$ be the planar lattice. Each $\mathbf{i} \in \mathbb{Z}^2$ is called a *site*. Each (finite or infinite) subset of $\mathbb{Z}^2$ is called a *configuration* and the space of all configurations is denoted by $\mathcal{C}$. For each $\mathbf{h} \in \mathbb{Z}^2$ the map $\tau_{\mathbf{h}}$ is the translation by the vector $\mathbf{h}$. The translated set $\tau_{\mathbf{h}}G$ is denoted by $G_{\mathbf{h}}$. The area, that is, the number of sites in G is denoted by $|G|$.

Let $X = \{X_i\}$ be a point process defined on the lattice $\mathbb{Z}^2$. The point process X is a set-valued random function and indicates the random positions where some phenomenon (rain in our case) happens. The process X is *simple* if $X_i \neq X_j$ whenever $i \neq j$. It is *stationary* if distributions of numbers $|X \cap G_{\mathbf{h}}|$, $\mathbf{h} \in \mathbb{Z}^2$, are the same for every finite G. Let $S = \{S(\mathbf{i})\}$, $\mathbf{i} \in \mathbb{Z}^2$, be an accompanying random field. Each $S(\mathbf{i})$ means the random quantity which is *potentially* associated with the site $\mathbf{i}$. That is, if it rains at a site $\mathbf{i}$, it results in the rainfall $S(\mathbf{i})$ during a certain unit period of observation. Combining X and S, we construct our basic model $\{(X_i, R_i)\}$, where $R_i = S(X_i)$. In the terminology of the marked point process theory, R_i is called the *mark* attached to the *point* X_i.

Fix a finite set G and two arbitrary functions $b(x)$ and $f(x)$ and define following quantities

$$\begin{aligned} A_G &= |X \cap G|, \\ B_G &= \sum_i 1_G(X_i) b(R_i), \\ F_G &= \sum_i 1_G(X_i) f(R_i). \end{aligned}$$

In the threshold method A_G is the area of the subregion of G where it rains and F_G is the sum of quantities $f(R_i)$ within the region G. For example, if $f(x) = x$, F_G is the total rainfall in the region G during a unit period.

We always assume the simpleness of X, the stationarity of both X and S, and the finiteness of expectations $\mathbb{E}\{b^2(S(\mathbf{0}))\}$ and $\mathbb{E}\{f^2(S(\mathbf{0}))\}$ throughout the paper. Let us list basic assumptions used in the sequel.

(A) X and S are independent to each other.

(B) X and S are independent to each other and S is an independent random field.

(C) Both X and S are independent random processes and are independent to each other.

Remark 1. The random field S is independent if random variables $\{S(\mathbf{i})\}$ are mutually independent. The point process X is independent if it is equal to the set $\{\mathbf{i}; Y(\mathbf{i}) = 1\}$ of an independent random field $\{Y(\mathbf{i})\}$ taking only two values $\{0,1\}$. The assumption (B) is equivalent to say that we attach iid replicates R_i to each X_i of X.

3. Variance formula

Let ρ be the common distribution of random variables $\{S(\mathbf{i})\}$. Because of the assumed stationarity of X, we have

$$\mathbb{E}\{A_G\} = \lambda|G|$$

where the constant $\lambda = \mathbb{E}\{A_{\{\mathbf{i}\}}\}$ is the *intensity* of the process X, that is, the mean number of points per one site. Also we have

$$\begin{aligned}(3.1)\qquad \mathbb{E}\{F_G\} &= \mathbb{E}\left\{\mathbb{E}\left\{\sum_i 1_G(X_i)f(S(X_i)) \mid X\right\}\right\}\\ &= \mathbb{E}\left\{\left(\int f d\rho\right)\sum_i 1_G(X_i)\right\}\\ &= \left(\int f d\rho\right)\lambda|G|.\end{aligned}$$

Next we consider the correlation of B_G and F_G. Since

$$B_G F_G = \sum_i 1_G(X_i)b(R_i)f(R_i) + \sum_{i\neq j} 1_G(X_i)1_G(X_j)b(R_i)f(R_j),$$

we can show

$$\begin{aligned}(3.2)\quad \mathbb{E}\{B_G F_G\} &= \mathbb{E}\{\mathbb{E}\{B_G F_G \mid X\}\}\\ &= \int bf d\rho\cdot\lambda|G| + \sum_{\mathbf{h}\neq\mathbf{0}} \mathbb{E}\{b(S(\mathbf{0}))f(S(\mathbf{h}))\}|G\cap G_{-\mathbf{h}}|\lambda_2(\mathbf{h})\\ &= \sum_{\mathbf{h}} |G\cap G_{\mathbf{h}}|\mathbb{E}\{b(S(\mathbf{0}))f(S(\mathbf{h}))\}\lambda_2(\mathbf{h})\end{aligned}$$

where we set

$$\lambda_2(\mathbf{h}) = \mathbb{E}\{A_{\{\mathbf{0}\}}A_{\{\mathbf{h}\}}\}.$$

Note that $\sum_{\mathbf{h}} |G\cap G_{-\mathbf{h}}|\lambda_2(\mathbf{h}) = \mathbb{E}\{A_G^2 o\}$ and $\lambda_2(\mathbf{0}) = \lambda$. Therefore we have the first result.

PROPOSITION 3.1. *The covariance of B_G and F_G is given by*

$$\begin{aligned}\operatorname{Cov}\{B_G, F_G\} &= \sum_{\mathbf{h}} |G \cap G_{-\mathbf{h}}| \lambda_2(\mathbf{h}) \operatorname{Cov}\{b(S(\mathbf{0})), f(S(\mathbf{h}))\} \\ &\quad + \left(\int b d\rho \int f d\rho\right) \operatorname{Var}\{A_G\}.\end{aligned}$$

In particular, the variance of F_G is given by

$$\begin{aligned}\operatorname{Var}\{F_G\} &= \sum_{\mathbf{h}} |G \cap G_{-\mathbf{h}}| \lambda_2(\mathbf{h}) \operatorname{Cov}\{f(S(\mathbf{0})), f(S(\mathbf{h}))\} \\ &\quad + \left(\int f d\rho\right)^2 \operatorname{Var}\{A_G\},\end{aligned}$$

and we have a similar expression for $\operatorname{Var}\{B_G\}$. *If Assumption* (B) *is supposed*

$$\begin{aligned}\operatorname{Cov}\{B_G, F_G\} &= \operatorname{Cov}\{b(S(\mathbf{0})), f(S(\mathbf{0}))\} \cdot \lambda |G| \\ &\quad + \left(\int b d\rho \int f d\rho\right) \operatorname{Var}\{A_G\}.\end{aligned}$$

If, further, Assumption (C) *is supposed*

$$\begin{aligned}\operatorname{Cov}\{B_G, F_G\} &= \operatorname{Cov}\{b(S(\mathbf{0})), f(S(\mathbf{0}))\} \cdot \lambda |G| \\ &\quad + \left(\int b d\rho \int f d\rho\right) \lambda(1-\lambda)|G|.\end{aligned}$$

PROOF. The first result follows from relations (3.1) and (3.2). Others are easy. Note that $A_G = \sum_{\mathbf{i} \in G} Y(\mathbf{i})$ in the last case. □

4. Threshold method

Let W and Z be two correlated random variables. Suppose the observation of W is impossible or difficult, while that of Z is tractable or easier. Hence, we are interested in the estimation of W using Z. If we use the linear estimator $\beta Z + \alpha$, the theoretical estimation error is given by

$$\mathbb{E}\{|W - \beta Z - \alpha|^2\}.$$

To get the scale independent error, we consider the estimation error

$$\mathbb{E}\left\{\left|\frac{W}{\sqrt{\operatorname{Var}\{W\}}} - \beta \frac{Z}{\sqrt{\operatorname{Var}\{Z\}}} - \alpha\right|^2\right\}.$$

The minimum of this expression is seen to be equal to

$$\operatorname{Err} = 1 - |\operatorname{Corr}\{W, Z\}|^2.$$

Therefore, if the correlation of W and Z is close to 1, we are encouraged to estimate W linearly using Z.

In the threshold method, we want to estimate F_G using B_G where $f(x) = x$ and $b(x) = 1_c(x) = 1_{[c,\infty)}(x)$ with some threshold value $c > 0$. Note that coefficients α and β depend on the region G in general. The square of the correlation of B_G and F_G is given by the results of the preceding section. We use the following symbols

$$\begin{aligned}
m_1 &= \mathbb{E}\{S(\mathbf{0})\} = \int m d\rho(m), \\
m_2 &= \mathrm{Var}\{S(\mathbf{0})\} = \int m^2 d\rho(m) - m_1^2, \\
n_1(c) &= \mathbb{E}\{1_c(S(\mathbf{0}))\} = \int_c^\infty d\rho(m), \\
n_2(c) &= \mathrm{Var}\{1_c(S(\mathbf{0}))\} = n_1(c) - n_1^2(c), \\
k_2(c) &= \mathrm{Cov}\{S(\mathbf{0}), 1_c(S(\mathbf{0}))\} = \int_c^\infty m d\rho(m) - m_1 n_1(c).
\end{aligned}$$

If Assumption (B) is supposed,

$$\mathrm{Err} = 1 - \frac{|k_2(c)\lambda|G| + m_1 n_1(c)\,\mathrm{Var}\{A_G\}|^2}{\{m_2\lambda|G| + m_1^2\,\mathrm{Var}\{A_G\}\}\{n_2(c)\lambda|G| + n_1^2(c)\,\mathrm{Var}\{A_G\}\}}.$$

If Assumption (C) is valid,

$$\mathrm{Err} = 1 - \frac{|k_2(c) + m_1 n_1(c)(1-\lambda)|^2}{\{m_2 + m_1^2(1-\lambda)\}\{n_2(c) + n_1^2(c)(1-\lambda)\}}.$$

Note that the last expression is independent of G.

From these expressions, we can see that whether or not we can estimate F_G linearly using B_G does indeed depend not only on S but also on X. An important consequence is that the correlation may be close to 1 regardless of both the distribution of S and the threshold value c in an extremal case where, say, $\mathrm{Var}\{A_G\}$ is of order $O(|G|^2)$ and the area $|G|$ is large enough. Of course, this is only possible if the process X is dependent. We believe that this explains the empirical fact that R^2-values can be fairly large regardless of threshold values.

Remark 2. A radar can scan a very wide region instantaneously. On the other hand, one cloud cluster may sometimes be as big as the total scan field, see the last section. Therefore, it seems natural that an estimation of regression coefficients α and β should be based on observations on different periods over the same fixed wide region G. Of course, regression coefficients should be chosen for a specified region and a specified season. As long as the R^2-value is large, the knowledge of λ or ρ may be of secondary importance. Several papers discussed how to estimate the best threshold value from estimated λ and ρ under the assumption (C). If, further, we estimate $\mathrm{Var}\{A_G\}$, λ_2 and so on from data, we may estimate the best threshold value from our formulas under assumptions (B) or (A).

PROPOSITION 4.1. *Assume Assumption* (A). *Let $\mathcal{B}_{\mathbf{i}}$ be the σ-algebra generated by $S(\mathbf{i})$ and let $\alpha(n)$ be the mixing coefficient defined by*

$$\alpha(n) = \sup\{|\mathbb{P}\{A \cap B\} - \mathbb{P}\{A\}\mathbb{P}\{B\}| : A \in \mathcal{B}_{\mathbf{i}}, B \in \mathcal{B}_{\mathbf{j}}, |\mathbf{i} - \mathbf{j}| > n\}.$$

If there is a constant $d > 0$ such that

$$\sum_{n=1}^{\infty} n\alpha(n)^{d/(2+d)} < \infty \tag{4.1}$$

and $S(\mathbf{0})^{2+d}$ is integrable, then we can show

$$\begin{aligned}
\operatorname{Var}\{F_G\} &= O(|G|) + \left(\int_0^\infty m d\rho(m)\right)^2 \operatorname{Var}\{A_G\}, \\
\operatorname{Var}\{B_G\} &= O(|G|) + \left(\int_c^\infty d\rho(m)\right)^2 \operatorname{Var}\{A_G\}, \\
\operatorname{Cov}\{B_G, F_G\} &= O(|G|) + \left(\int_c^\infty d\rho(m) \int_0^\infty m d\rho(m)\right) \operatorname{Var}\{A_G\}.
\end{aligned}$$

PROOF. Let $W_{\mathbf{i}}$ be measurable with respect to $\mathcal{B}_{\mathbf{i}}$ and $\mathbb{E}\{|W_{\mathbf{i}}|^{2+d}\}$ be finite and bounded. Then it is known that the sum $\sum_{\mathbf{h}} |\operatorname{Cov}\{W_{\mathbf{0}}, W_{\mathbf{h}}\}|$ is finite under the assumption (4.1), see, e.g., Bolthausen (1982). Note that $|\lambda_2(\mathbf{h})| \leq 1$. Hence

$$\left|\sum_{\mathbf{h}} |G \cap G_{-\mathbf{h}}| \lambda_2(\mathbf{h}) \operatorname{Cov}\{S(\mathbf{0}), S(\mathbf{h})\}\right| \leq |G| \sum_{\mathbf{h}} |\operatorname{Cov}\{S(\mathbf{0}), S(\mathbf{h})\}|.$$

Other relations can be proved similarly. □

Remark 3. From this result, we can observe that $\operatorname{Corr}\{B_G, F_G\}$ is asymptotically equal to 1 if $\operatorname{Var}\{A_G\}/|G| \to \infty$ as $G \uparrow \mathbb{Z}^2$, that is, the threshold method does work fine if $|G|$ is large enough. If, further, $\operatorname{Var}\{A_G\} = O(|G|^2)$,

$$\min_{\alpha,\beta} \mathbb{E}\left\{\left|\frac{1}{|G|}F_G - \beta\frac{1}{|G|}B_G - \alpha\right|^2\right\} = \frac{\operatorname{Var}\{F_G\}}{|G|^2} \operatorname{Err} \to 0,$$

as $|G| \to \infty$, that is, $F_G/|G|$ can be estimated linearly using by $B_G/|G|$.

PROPOSITION 4.2. *Assume Assumption* (C) *and also assume that ρ has a continuous density function. The minimum of the best linear estimation error of the threshold method is attained by the solution c_0 of the following equation*

$$(\lambda m_1 - c)(n_2(c) + n_1^2(c)(1 - \lambda)) = (\lambda n_1(c) - 0.5)(k_2(c) + m_1 n_1(c)(1 - \lambda)).$$

The minimum is equal to

$$1 - \frac{4(\lambda m_1 - c_0)^2}{(2\lambda n_1(c_0) - 1)^2} \times \frac{n_2(c_0) + n_1^2(c_0)(1 - \lambda)}{m_2 + m_1^2(1 - \lambda)}.$$

PROOF. Straightforward. □

5. Lattice gas model and coexistence of phases

Our conjecture assumes a huge variation of numbers of raining sites. Although the analysis of the TRMMGT database in the next section suggests it strongly, it is worthwhile to show that it is possible at least theoretically. We will borrow a model, the *lattice gas model*, from statistical physics and relevant results. But it is not our intention to suggest that this particular model is a decent model of the rainfall phenomenon. Nevertheless, we believe, two characteristic features which this model is capable to express, that is, the formation of clusters and the coexistence of different physical conditions, seem to be reasonable in actual rainfall phenomenon. Although all results of this section are straightforward consequences of existing theories, it is worthwhile to state them in some details.

Let μ be a constant, called a *chemical potential*, and let $\phi(\mathbf{i}) : \mathbb{Z}^2\backslash\{\mathbf{0}\} \mapsto \mathbb{R}$ be an even function, called a *pair interaction potential function*. We assume the reader is familiar with the definition of lattice gas model, that is, the Gibbsian model on the configuration space $\mathcal{C} = \{0,1\}^{\mathbb{Z}^2}$, and basic concepts such as stationarity and ergodicity of Gibbsian distribution, see, e.g., Ruelle (1969, 1978) or Georgii (1988). The totality of stationary Gibbsian distributions corresponding to (μ, ϕ) is denoted by $\mathcal{G} = \mathcal{G}_{\mu,\phi}$. If the condition

$$\sum_{\mathbf{i}\neq\mathbf{0}} |\phi(\mathbf{i})| < \infty$$

holds, it is known $\mathcal{G} \neq \emptyset$. But it is also known $\mathcal{G}$ may contain more than two different distributions. This phenomenon is known as a *phase transition* in statistical physics and implies that physically different states can coexist.

Ergodic Gibbsian distributions correspond to physically *pure* states. It is known that the totality $\mathcal{G}^*$ of stationary ergodic Gibbsian distributions is non-empty and coincides with the set of extremal points of the convex set $\mathcal{G}$. Each $\mathbb{P} \in \mathcal{G}$ can be expressed uniquely as a mean of a probability measure on $\mathcal{G}^*$ equipped with a suitable measurable space structure

$$\mathbb{P} = \int_{\mathcal{G}^*} \mathbb{Q}d\gamma(\mathbb{Q}). \tag{5.1}$$

The following characterization of ergodic Gibbsian distributions is known, see e.g., Ruelle (1978). Let H be a continuous function on $\mathcal{C}$ and set

$$\langle H\rangle_G = \frac{1}{|G|}\sum_{\mathbf{h}\in G} H \circ \tau_{\mathbf{h}}.$$

A sequence of finite sets $G = \{G_n\}$ tends to infinity in the sense of van Hove, written as $G \nearrow \infty$, if $|G_n| \to \infty$ and $|G_n \circ \tau_{\mathbf{h}} \backslash G_n|/|G_n| \to 0$ for each $\mathbf{h}$,

PROPOSITION 5.1. (Ruelle (1978), Chapter 3) *A stationary Gibbsian distribution $\mathbb{P}$ is ergodic iff*

$$\lim_{G\nearrow\infty} \mathrm{Var}\{\langle H\rangle_G\} = 0$$

for every continuous H. *If a Gibbsian distribution* $\mathbb{P}$ *is not ergodic, has the decomposition* (5.1), *and* $\mathbb{Q}\{\xi_{\mathbf{0}}\}$ *is not constant* $\gamma(\mathbb{Q})$*-a.s., then*

$$\lim_{G\nearrow\infty} \mathrm{Var}\{\langle\xi_{\mathbf{0}}\rangle_G\} = \lim_{G\nearrow\infty} \mathrm{Var}\{A_G\}/|G|^2 > 0.$$

This result shows that, if the assumption of Proposition 5.1 holds, $\mathrm{Var}\{A_G\}$ is of order $|G|^2$. Hence, the threshold method works if $|G|$ is large enough.

One question remains here. What conditions on μ and ϕ are necessary to have $\#\mathcal{G}_{\mu,\phi} > 1$? Although there have been many studies, we have still only partial understanding of phase transition phenomena. This is in contrast with our daily experience where coexistence of several different phases (such as ice, water, and vapor) is fairly common. Existing results are almost all mere existence theorems and the following seems to be most general. This is a particular case of Georgii ((1988), Theorem 16.21) and its subsequent comments, see also Exercise 2 of Ruelle ((1978), Chapter 3).

PROPOSITION 5.2. (Georgii (1988), Theorem 16.21) *Let* (μ_0, ϕ_0) *be a potential such that* $\sum_{\mathbf{i}\neq\mathbf{0}} |\phi_0(\mathbf{i})| < +\infty$. *Then there exists a potential* (μ_1, ϕ_1) *with* $\mu_1 \geq 0$, $\phi_1 \leq 0$ *satisfying* $\sum_{\mathbf{i}\neq\mathbf{0}} |\phi_1(\mathbf{i})| < +\infty$, *and two members* $\mathbb{Q}_1$ *and* $\mathbb{Q}_2$ *of* $\mathcal{G}_{\mu_0+\mu_1,\phi_0+\phi_1}$ *such that* $\int \xi_{\mathbf{0}} d\mathbb{Q}_1 \neq \int \xi_{\mathbf{0}} d\mathbb{Q}_2$. *Moreover, we can assume, for each fixed* $r > 0$, *that* $\phi_1(\mathbf{i}) = 0$ *if* $|\mathbf{i}| \leq r$.

Remark 4. In general, ϕ_1 may not be of finite range, that is, $\{\mathbf{i}; \phi_1(\mathbf{i}) \neq 0\}$ may be an infinite set. Therefore, remote sites may interact with each other. We can assume that both $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are ergodic. Roughly speaking, this proposition implies that a phase transition may occur if

1) there is an attractive tendency among remote raining sites, that is, raining sites tend to form clusters, while

2) the number of raining sites remains moderate.

Remark 5. The lattice gas model can be considered as a process on $\{0,1\}^{\mathbb{Z}^2}$. The Gibbsian model on the state space $\{-1,1\}^{\mathbb{Z}^2}$ is well-known in statistical physics as the *Ising model* and considered as a model of magnets or crystals. Assume that the chemical potential $\mu = 0$ and that the interaction is of the form with a positive constant J

$$\phi(\mathbf{i}) = \begin{cases} J & \text{if } \mathbf{i} = (0,\pm1), (\pm1, 0), \\ 0 & \text{otherwise.} \end{cases}$$

Then it is known that there are two different elements $\mathbb{P}_+$, $\mathbb{P}_-$ of $\mathcal{G}^*$ such that every $\mathbb{P} \in \mathcal{G}$ is expressed as their convex linear combination iff $J > J_c$, see Prum and Fort (1991). In particular, there is a phase transition. The critical parameter J_c is given as the root of $\sinh 2J_c = 1$. Since, under the transformation $x = \pm1 \mapsto y = (x+1)/2 = 0, 1$, this Ising model is equivalent to a lattice gas model, we have an example of phase transition models.

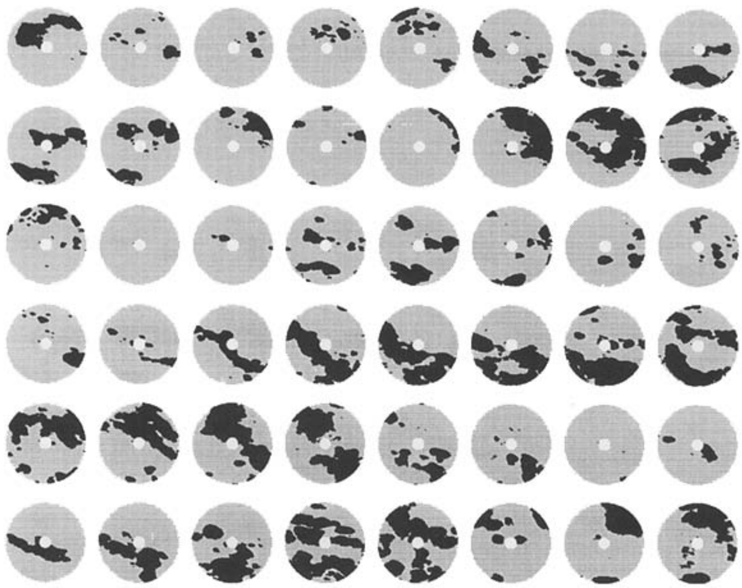

Fig. 1. Transition (row-by-row) of raining sites from 1988.28.00:00 to 1988.31.22:00 for every two hours (dataset II).

Remark 6. We note that, even if $\mathrm{Var}\{A_G\}$ is not of order larger than $O(|G|)$, the threshold method still works if the finite limit $\lim_{G \nearrow +\infty} \mathrm{Var}\{A_G\}/|G|$ exists and is relatively large. A condition which guarantees the existence of this asymptotic variance is known. The basic tools are Dobrushin's uniqueness condition and Föllmer's covariance estimate. See Föllmer (1982) for details. But the asymptotic variance is always difficult to be evaluated numerically and a Monte Carlo estimate is the last resource.

Remark 7. A phase transition implies a mixture of rain conditions with different rainfall characteristics, variances in particular. On this account, it may be enlightening to refer to Shimizu *et al.* (1993). In this paper, authors noted two meteorological types of rainfall, that is, the stratiform and the convective environments. For example, a low area-average rain rate with low variance suggests a stratiform rain régime, while the same area average with a high variance suggests a convective environment.

6. Analysis of the TRMMGT database

In this last section, we analyze the TRMMGT database. This database contains rainfall maps as measured by the NOAA/TOGA C-band radar which was located near Darwin, Australia. Each data, the unit in mm, consists of a 141×141 pixel array with areal resolution of 2 km $\times$ 2 km. Sites beyond 140 km and closer than 20 km from the radar are missing and the actual number of observed sites is $15,380$. We use two hourly data sequences which cover periods Julian days the

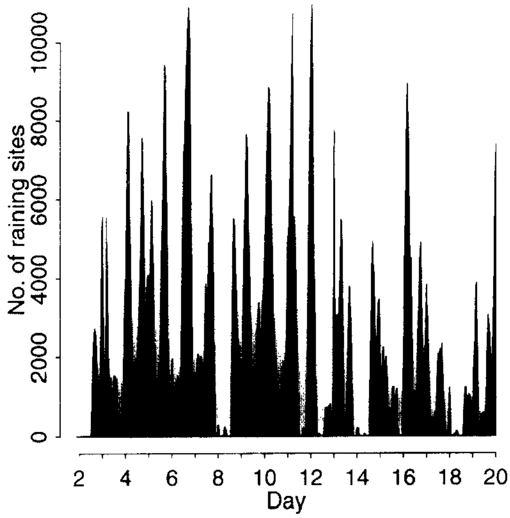

Fig. 2. Hourly time series of number of raining sites from 1988.01.20:00 to 1988.19.23:00 (dataset I).

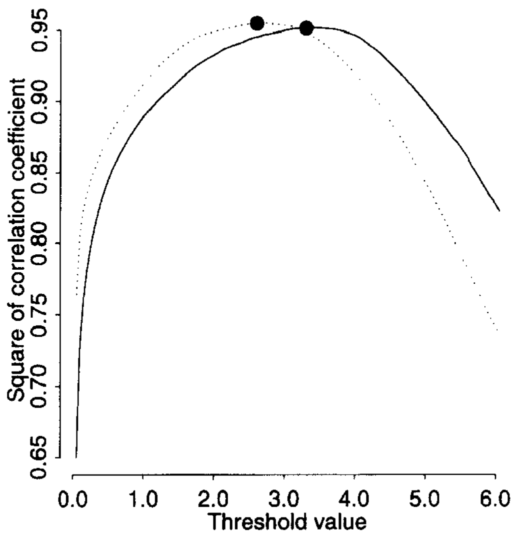

Fig. 3. R^2-values for threshold values $c = 0(0.05)6$ for dataset I (straight line) and II (dotted line). Largest values are marked.

first to 19th (Dataset I) and 26th to 47th (Dataset II) in 1988. Note that Darwin was in the wet season. Figure 1 is an animation of the transition of raining sites for 4 days from the dataset II. Figure 2 shows the hourly time series of numbers of raining sites of the dataset I. They clearly show that both the area and the position of raining sites change very quickly and, therefore, cause a huge variation of A_G.

Table 1. Sample intensities $\hat{\lambda}$ and sample variances V of A_G for TRMMGT database. V^* is $15,380 \times \hat{\lambda}(1-\hat{\lambda})$.

	data size	$\hat{\lambda}$	V^*	V
Dataset I	436	0.1452	1910	$6330 \times 10^3 = 3314 \times V^*$
Dataset II	515	0.2157	2601	$1154 \times 10^4 = 4437 \times V^*$

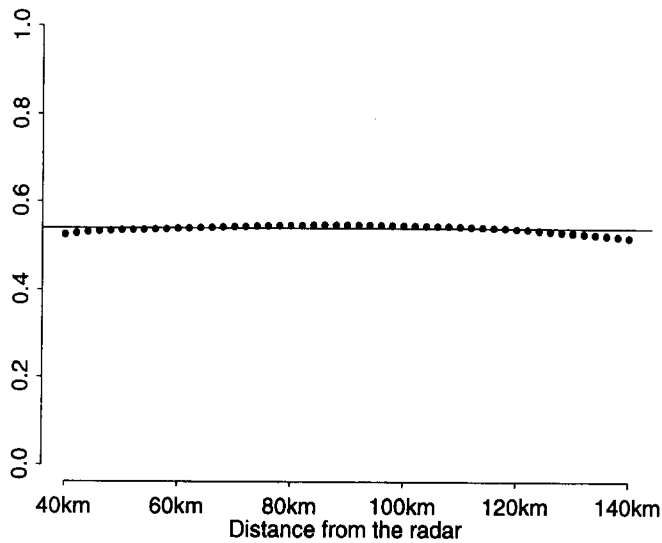

Fig. 4. Graph of $\sqrt{\mathrm{Var}(A_G)}/|G|^{0.912}$ for varying radii (dataset II). The horizontal line $y = 0.54$ is added for comparison.

Figure 3 is graphs of R^2-values for threshold values $c = 0.0(0.05)6.0$. We can reconfirm Chiu's discovery that there is a very high correlation between F_G and B_G. Also note that correlations are fairly high irrespective of threshold values. Best threshold values are seen to be 3.3 mm/hour (Dataset I) and 2.6 mm/hour (Dataset II). Corresponding R^2-values are 0.955 and 0.952 respectively.

Sample intensities $\hat{\lambda}$ and sample variances V of A_G based on hourly data are tabulated in Table 1. If the independent raining site assumption is valid, V should be close to $V^* = 15,380 \times \hat{\lambda}(1-\hat{\lambda})$. Clearly the independent raining site assumption is rejected. Finally, Fig. 4 is the graph of $\sqrt{\mathrm{Var}\{A_G\}}/|G|^{0.912}$ for varying radii (dataset II). It shows that $\sqrt{\mathrm{Var}(A_G)} \approx 0.54|G|^{0.912}$, a fact which strongly suggests the presence of a phase transition. For the dataset I we have $\sqrt{\mathrm{Var}(A_G)} \approx 0.80|G|^{0.842}$.

From these table and figures, we can see a huge variation of numbers of raining sites which we have conjectured to be the main reason why the threshold method works so fine.

Acknowledgements

The author would like to express his thanks to Dr. K. Shimizu, the Science University of Tokyo, who first introduced him the threshold method. He also would like to express his thanks to Dr. D. B. Wolff, the Goddard Space Flight Center of NASA, who edited the TRMMGT database and made it publicly accessible. The author owes many helpful comments to two referees.

REFERENCES

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields, *Ann. Probab.*, **10**, 1047–1050.

Braud, I., Creutin, J. D. and Barancourt, C. (1993). The relation between the mean real rainfall and fractional area where it rains above a given threshold, *Journal of Applied Meteorology*, **32**, 193–202.

Chiu, L. S. (1988). Rain estimation from satellites: Areal rainfall-rain area relations, Paper 7.14, Reprints, *Third Conf. on Satellite Meteorology*.

Föllmer, H. (1982). A covariance estimate for Gibbs measures, *J. Funct. Anal.*, **46**, 387–395.

Georgii, H.-O. (1988). *Gibbs Measures and Phase Transitions*, Walter de Gruyter, Berlin.

Patterson, V. L., Hudlow, M. D., Pytlowany, P. J., Richards, F. P. and Hoff, J. D. (1979). GATE radar rainfall processing system, NOAA Technical Memo., EDIS 26.

Prum, B. and Fort, J. C. (1991). *Stochastic Processes on a Lattice and Gibbs Measures*, Kluwer, Dordrecht.

Ruelle, D. (1969). *Statistical Mechanics: Rigorous Results*, Benjamin, New York.

Ruelle, D. (1978). *Thermodynamic Formalism: The Mathematical Structures of Classical Equilibrium Statistical Mechanics*, Addison-Wesley, London.

Shimizu, K. (1992). The estimation of area rain rate characteristics by the threshold method, *Journal of Japan Applied Statistical Society*, **21**, 133–151 (in Japanese).

Shimizu, K, Kedem, B. and Short, D. A. (1993). Single- and double-threshold methods for estimating the variance of area rain rate, *Journal of the Meteorological Society of Japan*, **71**, 673–683.

Short, D. A., Shimizu, K. and Kedem, B. (1993). Optimal thresholds for the estimation of area rain-rate moments by the threshold method, *Journal of Applied Meteorology*, **32**, 182–192.