POINT PROCESSES AND INFERENCE FOR RAINFALL FIELDS

MICHAEL J. PHELAN*
Princeton University

ABSTRACT

We treat the role of point processes in the statistical analysis of rainfall fields. We look in specific at Poisson-based models of the Le Camian type. These represent rainfall fields as a smoothing transformation of a Poisson random measure. We provide a number of examples of particular interest to hydrologists who analyze rainfall fields and we outline strategies for their assessment. Here we find a place for novel applications of statistical techniques, such as simulation, method of moments, nonlinear regression, and image and life history analysis.

1. **Introduction** Rainfall fields refer to the observed pattern of ground-level intensity of rain falling from precipitating cloud systems. Reflecting the nonequilibrium thermodynamic origins of such systems, these patterns manifest highly-structured distributions of rainfall over space. Our discussion here concerns modeling and inference for the temporal evolutions of these distributions.

The temporal evolutions of rainfall fields proceed from the changing thermodynamic conditions governing the phenomenon of precipitation. As part of a family of processes involved in global atmospheric circulation, precipitation is a process of complex geophysical origins. To model the dynamics of rainfall fields is obviously beyond the scope of present endeavors. In substitute, we treat phenomenological models of rainfall fields, that aim to describe the phenomenon as observed.

Nicolis and Prigogine (1977) note that the *observable* behavior of many nonequilibrium thermodynamic systems is well represented by a stochastic process. We take this view in modeling rainfall fields with a spatial stochastic process. In particular, we look at models that represent such fields as a smoothing transformation of a Poisson random measure. The general framework for these was introduced in Le Cam (1961). Nonetheless, it was not until after the GATE experiment of 1974 that Le Cam's work began to have impact. From this observational study of tropical rainfall over the Atlantic Ocean, meteorologists reported new understandings of the organization and structure of precipitating cloud

*Partially supported by Office of Naval Research Grant N00014-89-J-1830

systems; see for example Houze and Hobbs (1982). Since that time, Le Cam's work has influenced recent efforts to model rainfall fields including those in Smith and Karr (1985), Waymire, Gupta, and Rodriguez-Iturbe (1984), Rodriguez-Iturbe, Cox, and Eagleson (1986), Rodriguez-Iturbe, Cox, and Isham (1987, 1988), and Cox and Isham (1988).

We later recall the Le Camian representation of rainfall fields and discuss its principal characteristics. In providing a number of examples, we interpret the primitives in these representations in terms of the geophysical constructs of the raincell and the cloud cluster. These constructs have informed the discussions of rainfall events by meteorologist, and so provide the motivation behind the examples. As examples of particular interest to hydrologists who analyze rainfall fields, we include the WGR model in Waymire, et al. (1984) and the random-disks model in Cox and Isham (1988).

The complexity of these models and the phenomenon they represent compels the need for flexible strategies in their assessment. One such strategy, which we recall below, is outlined in Rodriguez-Iturbe, et al. (1988). This is a practical approach to parametric estimation that blends formal statistical procedures with informal physical judgements in promoting physically meaningful answers. To it, we add our own emphasis on the sample path analysis of rainfall fields as implemented in Phelan and Goodall (1989). This approach uses nonlinear regression and image analysis in data-analytic assessments of the ability of the model to represent observed rainfall fields. We emphasize that for any strategy to be successful it must stimulate the dialogue among observation, modeling, and inference.

This dialogue will often lead to fresh scientific inquiries as well as to novel applications of statistical techniques. To illustrate, we anchor our final discussion in the question of aging in rainfall fields. Here we introduce the aging function, which is a cumulative hazard function appearing as parameter in many models of rainfall. We address the problem of nonparametric estimation of such functions from observed rainfall fields, where we argue for the application of the techniques of life history analysis. In specific, we introduce a martingale-type estimator as a nonparametric estimator of the aging function in a class of point process models for rainfall fields. Finally, we integrate our approach to nonparametric estimation with the strategy for parametric estimation described above, and thus enhance its flexibility as an assessment regime.

2. Poisson Random Measures and Le Camian Representations The work in Le Cam (1961) develops a general framework for the stochastic description of precipitation using point processes and their transformations. Here we recall the main idea behind such descriptions and provide examples. Our intention is to draw attention to what we call Le Camian representations of rainfall fields.

Convective Raincells and Point Processes. In describing precipitating cloud systems, meteorologists refer to a geophysical construct known as the convective raincell. Briefly, the convective raincell denotes a localized region of high-intensity rainfall induced by convection or the exchange of heat from the earth to the atmosphere. In practice, we typically observe composites of such cells at a scale of about 4 km. At maturity, they may cover a surface area of 16 km^2 with rainfall at an intensity reaching 8 mm/hr. In the tropics for example, mature raincells are observed in groups called cloud clusters. They are usually connected by a layer of stratiform cloud. Aided by the winds, such precipitating systems horizontally traverse the lower atmosphere while depositing their water contents on the surface of the earth. We refer the reader to for example Houze and Hobbs (1982) for further details of the phenomenology of raincells in mesoscale descriptions of precipitative events.

In the importance of the convective raincell, Le Cam recognized a role for point processes in his stochastic descriptions of precipitation. The principal idea is that the occurrence of convective cells be identified with the atoms of a point process. In specific, let $(\Omega, \boldsymbol{H}, P)$ denote a probability space and $(E, \boldsymbol{E})$ a measurable space. Now let M and N denote random measures on $(E, \boldsymbol{E})$. We refer the reader to for example Karr (1986) for further mathematical details. We assume that N is a Cox process having directing measure M. Here in the context of modeling mesocale precipitation, N is the primitive that models the occurrence of convective raincells. For Le Cam, the specification of the space $(E, \boldsymbol{E})$ and the directing measure M is determined by the geophysical context and the design of the modeler.

We fix ideas and take E to be a product space of the form $T \times S \times \chi$. Here T denotes a subset of $\boldsymbol{R}$ and a time-index set, S denotes a subset of $\boldsymbol{R}^2$ and a planar region affixed to the surface of the earth, and χ denotes a vector space used in characterizing raincells by their water content, dispersion, velocity, etc. We assume that M is the fixed measure μ satisfying

$$\mu(dt, dx, dz) = c\,dt\,dx\,\pi(dz), \quad (t \in T,\ x \in S,\ z \in \chi)\,, \tag{1}$$

where c is a positive constant and π denotes a probability measure on χ equipped with its measurable sets. Note that N is thus a Poisson random measure on $T \times S \times \chi$ with mean measure μ.

We say that the atoms of N identify convective raincells. That is, fix ω in Ω and suppose that (t, x, z) is an atom of $N(\omega)$ where $t \in T$, $x \in S$, and $z \in \chi$. Here (t, x) localizes the time and place of birth of a raincell. The vector z endows the raincell with a set of distinguishing characteristics such as its velocity and water content. According to equation (1), the model places raincells at random over time and space, whereas their characteristics vary independently over χ with common distribution π.

Remark. We mentioned earlier that convective raincells tend to occur in clusters. Le Cam addressed this issue in his proposal to construct the directing measure M to reflect the observed organizational structure of storm systems. For example, M may be chosen so that N is a cluster point process that clusters raincells in the manner observed in tropical storms. We refer the reader to Le Cam (1961) for the details of this part of the framework. In doing so, we nevertheless note that clustering among raincells is an important yet open issue in phenomenological modeling of rainfall. □

Rainfall Fields and Le Camian Representations Rainfall fields refer to the observed pattern of ground-level intensity of rain falling from a precipitating cloud system. Le Cam suggested that these be represented by a smoothing transformation of the point process N. In describing these representations, we proceed with our ideas fixed as above.

Our mathematical setting is the space $T \times S \times \chi$ equipped with its measurable sets, where we have defined the Poisson random measure N that generates raincells and their characteristics. We aim to model the temporal evolution of the intensity or rate of rain falling over S. For each $t \in T$ and $x \in S$, let $k(t, x)$ denote a positive, measurable function defined on $T \times S \times \chi$. Let $k = \{k(t,x), t \in T, x \in S\}$ denote the corresponding family of functions, where we assume k satisfies

$$\int_{T \times S \times \chi} \mu(du, dy, dz)\, k^2(t, x, u, y, z) < \infty, \tag{2}$$

for every $t \in T$, $x \in S$. Here μ is the mean measure belonging to N as specified at equation (1). Next, we introduce a spatial stochastic process $R = (R(t, x))$, $t \in T$, $x \in S$, where $R(t, x)$ satisfies

$$R(t,x) = \int_{T\times S\times \chi} N(du,\ dy,\ dz)\, k(t,x,u,y,z)\,, \tag{3}$$

Here we interpret $R(t,x)$ as the intensity of rain falling on position x at time t, so that the process R models the temporal evolution of the intensity of rain falling over S. Note that condition (2) guarantees the existence of at least the mean and variance of this process.

In the general framework of Le Cam, k is an example of a smoothing kernel, and R is thus an example of a smoothing transformation of N. A rainfall intensity process so obtained, we shall call a *Le Camian representation* of the rainfall field. Such representations appear in the examples below, where we add specific interpretations to the items introduced here.

Examples We illustrate the main ideas above and provide examples of particular interest to hydrologists who analyze rainfall fields. For simplicity in doing so, we omit the presence of a clustering mechanism, but we refer the reader to remark (1) below.

*a) The WGR Mode*l. This model is developed in Waymire, et al. (1984). Let T, S, and χ denote the sets $\mathbb{R}$ $\mathbb{R}^2$ and $\mathbb{R}_+$ respectively. We assume that N is a Poisson random measure on $T\times S\times\chi$ equipped with its Borel sets. For the purpose of interpretation, fix ω and suppose that $(t,\ xz)$ is an atom of $N(\omega)$. We then say that t and x is the time and place of birth of a raincell having water content or total intensity z. If the mean measure μ belonging to N satisfies equation (1), then π specifies the distribution of the raincells' water contents.

The WGR model includes a Le Camian representation of the rainfall intensity process. In particular, let a and b denote fixed, positive scalars and let v_0 denote a fixed vector in the plane. Here the smoothing kernel k satisfies: for each $t\in T$ and $x\in S$,

$$k(t,x,u,y,z) = z1(u\le t)\, ae^{-a(t-u)}\sqrt{2/\pi}\ be^{-bd(x,\,y+(1-u)v_0)}\,,$$

for every $u\in T$, $y\in S$, and $z\in\chi$. Here $d(\cdot,\cdot)$ denotes the squared Euclidean distance between points in S. Under the smoothing kernel thus defined, the rainfall intensity process $R = R(t,x))$, $t\in T$, $x\in S$ is represented as in equation (3).

For each t and x, $R(t,\ x)$ denotes the total intensity of rain falling at $(t,\ x)$. It is the sum of the contributions of all the raincells born before time t. That is, fix ω and suppose that $(u,\ y,\ z)$ is an atom or raincell of $N(\omega)$. According to equation (3), $k(t,x,u,y,z)$ is the contribution of that raincell to the total intensity $R(\omega;\ t,$

x). In particular, by the action of k, its contribution is proportional to its total intensity z. To understand the proportion, we interpret $t - u$ as the age and $y + (t-u)v_0$ as the position of the raincell at time t. As it ages, the raincell thus propagates along the plane with velocity v_0. We therefore see that, as a function of t, the raincell's contribution to $R(\omega; t, x)$ decays exponentially with its age and its Euclidean distance to x. Here we call a the rate of aging and b the coefficient of disperson. □

b) The Random-Disks Model. This model is developed in Cox and Isham (1988). As the name suggested, it envisions the coverage of a precipitating raincell to be a random disk in the plane. The initial placement of the disk is specified at birth, when the raincell is characterized by a radius, water content, duration, and planar velocity. A raincell becomes active at birth and it remains so for its duration. The ensemble of active raincells migrates s across the plane; the individual motions being at uniform rate along that raincell's velocity. In doing so, they deposit their water contents uniformly over their identifying disks and uniformly over their durations. Thereafter, the raincell vanishes.

We formalize the qualitative description above. Let T, S, and χ denote the sets R ,R^2 , $R^3{}_+ \times R^2$, respectively. We assume that N is a Poisson random measure on $T \times S \times \chi$ equipped with its Borel sets. For the purpose of interpretation, fix ω and suppose that $(t,x,(r,w,z,v))$ is an atom of $N(\omega)$, where $t \in T$, $x \in S$, $r, w, z \in R_+$, and $v \in R^2$. We then say that t and x is the time and place of birth of a raincell having radius or dispersion r, water content w, duration z, and velocity v. If the mean measure μ belonging to N satisfies equation (1), then t specified the joint distribution of the raincell' redii, water contents, durations, and velocities.

For each time t and site x, let $R(t,x)$ denote the total intensity of rain falling at (t,x). According to the model, this is given by the sum of the contributions of those active raincells that cover the site x. Here we give its Le Camian representations. Let k denote the smoothing kernel satisfying: for each $t \in T$ and $x \in S$,

$$k(t,x,u,y,r,w,z,v) = w1(u \leq t)\, 1\, (d(x,y + (t - u)v) \leq r)1(z \geq t - u),$$

for every $u \in T$, $y \in S$, $r,w,z \in R_+$, and $v \in R^2$. Now we define $R(t,x)$ according to equation (3), and thus obtain the desired representation of the rainfall field $R = (R(t,x), t \in T, x \in S$. In leaving the specifics of the interpretation of the action of k to the reader, we ask him to verify that the representation of R given here is (stochastically) equivalent to that of equaiton (8) in Cox and Isham (1988). □

Remarks

(1) *Clustering.* The models described above admit generalizations involving higher-order clustering. This entails constructing a directing measure M that subordinates the point process N. For example, M itself may be a smoothing transformation of an initial point process. The action of the initial process is to generate cluster potentials, so that M may direct N to cluster raincells in the vicinity of these. Such a mechanism may well model the appearance of the cloud clusters mentioned earlier. For example, this device is used in Cox and Isham (1988) to induce either Bartlett-Lewis or Neyman-Scott type cluster processes.

(2) *Geometry and Kinematics.* The choice of smoothing kernel reflects or imposes a choice of geometry and kinematics for the underlying storm system. In the geometry of the WGR model, for example, each raincell is identified with a continuum of Gaussian surfaces. This has obvious ramifications for the spatial distribution of rain falling from that cell. In contrast, the geometry of the random-disks model distributes rain over a migrating coverage process, so that each raincell deposits water uniformly over a random set in the plane. Nevertheless, the two models maintain essentially the same kinematics. The raincells move at uniform rate along a specified velocity. The difference lies in that they do so in lockstep fashion in WGR, whereas random-disks allows for individual variations in these motions. □

3. Strategies in Assessment Stochastic descriptions of precipitation are useful from a number of viewpoints. On the one hand, they offer a ready framework to guide the hand of the data analyst. They suggest ways of examining observations and discovering relationships. On the other hand, they can be an integral part of scientific inference about precipitation or a workable tool in the engineering of water resources. Hydrologists, for example, use stochastic models in forecasting and simulating precipitation as part of the hydrological cycle. They may require only that such descriptions preserve averages or provide realistic patterns in rainfall activity. Nevertheless, in assessing stochastic descriptions of precipitation, we do well to remember the physical origins of the phenomenon in designing strategies and in judging the descriptive capabilities of the model.

In this section, we describe two broad strategies for the assessment of point process models of rainfall fields. The first is basically a parametric approach as described in Rodriguez-Iturbe, Cox, and Isham (1988). The method is well suited for assessing a model's ability to preserve averages in rainfall activity. The second strategy is a data-analytic approach to the sample path analysis of rainfall fields as implemented in Phelan and Goodall (1989). As such, it is well suited for assessing a model's ability to provide realistic patterns of rainfall. The two approaches are complementary, and together they offer a practical strategy for assessing Le Camian representations of rainfall fields.

1. The Method of Primary Features. This method is both a method of fitting parameters in a rainfall model and of assessing the fitted model. Respecting the earnest conceptual and computational problems entailed by the likelihood approach to parametric estimation from rainfall fields, the procedure recommended here involves solving a system of simultaneous equations drawn from a set of primary features of the data. Typically, only a subset of such features are needed to fit a finite dimensional parameter. Once fitted, therefore, the model is assessed against the subset of primary features that remain.

We fix ideas and consider the random-disks model as described in the previous section. There the mean measure μ depends on two parameters; the rate c and the distribution π. Recall that π denotes the distribution of the raincells' characteristics including their radii, water contents, durations, and velocities. For the present illustration, we assume that π satisfies

$$\pi(dr, dw, dz, dv) = ae^{-ar}\, be^{-bw}\, de^{-dz}\, drdwdz\, Q(dv), \tag{4}$$

for every $r,w,z \in R_+$, and $v \in R^2$. Here a, b, and d denote fixed parameters and Q satisfies

$$Q(dv) = (\beta(\beta\| v \|)^{\alpha-1}\, e^{-\beta\| v \|} / \Gamma(\alpha))\delta_0,$$

where $\| v \|$ denotes the length of v, α and β are fixed, positive parameters, and δ_0 is the direct measure concentrated at the fixed vector v_0 of unit length in the plane. This choice of π entails the independence of the raincells' characteristics. Moreover, the radii, water contents, and durations are exponentially distributed with parameters a, b, and d, respectively. Finally, the velocities are concentrated along the direction v_0, while their magnitudes are Gamma distributed with shape parameter α and scale parameter β. We have therefore specified a model of 7 parameters, namely a,b,c,d,α,β, and v_0, which we hereafter denote by the symbol θ.

The problem is to estimate θ from a record of observed rainfall activity. The method of primary features resolves the problem as follows. Typically, the data will consist of partial observations from the rainfall intensity process R, perhaps drawn from a network of rain gauges placed in a catchment. Let **G** denote a set of primary features of the data. We require that the corresponding model features be available either theoretically or by simulation. As examples of features in **G**, we take means, variances, and spatial-temporal autocorrelations of rainfall intensity, as well as derived distributions of say dry periods, level crossings, or extreme values in rainfall activity. Next, let **F** denote a subset of **G** to be used in fitting θ. We therefore require that **F** contain at least as many features as the dimension of our parameter. Now for each $f \in F$, let $D(f)$ denote the value of f as determined by the data. Similarly, let $M(\theta;f)$ denote the value of f as determined by the model. For example, suppose f denotes the proportion of time it is dry at a fixed site. Then $D(f)$ is the observed fraction of time it is dry there, and $M(\theta;f)$ is $\exp(-2c\pi/da^2)$ as determined by equaiton (2) in Cox and Isham (1988). The method of primary features is to solve for θ by solving the system of simultaneous equations satisfying

$$D(f) = M(\theta;f)\,,\; f \in F, \tag{5}$$

provided there is a unique solution.

The choice of features to fill **F** will be guided by many criteria. The foremost is the observational scheme, as we require data-based values for every feature in **F**. Otherwise, we look for features that will provide a sensitive and balanced test of the model's capabilities. For example, we select features that are relatively uncorrelated and have small sampling variability. Also, we select features that are available at a range of scales, this reflecting the physical importance of scaling principles in rainfall fields. Overall, we aim to achieve a nice blend of formal statistical thinking with informal physical judgment. For example, we may incorporate auxiliary information and appeal to physical measurements of the ambient winds to determine v_0. These measurements may then be used to help determine the equations above, say, by fixing an initial value to the mean α/β in the distribution of raincells' velocities.

In the method of primary features, the fitted model is assessed in two ways. The first is cross-validation, where the model-based values of the features in **F** are compared to data-based values as computed from a hold-out sample. Alternatively or in addition, if data-based values of the features in **F** are available over a range of scales of aggregation, say hourly, daily, and monthly amounts of rainfall, then model-based values are assessed across this range. In either case, we assess the capability of the model to reproduce the set of primary features used in fitting θ. On the other hand, the second way in which the model is assessed is to consider those primary features in $\mathbf{F}^c$, where $\mathbf{F}^c$ denotes the complement of **F** in **G**. Here we assess the capability of the model to extrapolate to the set of primary feature not used in fitting θ. For example, suppose the distribution of extreme values at a fixed site is derived (at least approximately) from the model and placed among the features in $\mathbf{F}^c$. Then a formal or diagnostic test of the model is made by comparing observed extremes to the extremes expected by the model evaluated at the value of the fitted parameter.

We believe that the method of primary features is a workable and flexible strategy for the assessment of point process models of rainfall fields. We find that it is particularly successful at selecting models that well preserve averages in rainfall activity. This is due in part to the computational ease of using various statistical moments as primary features. Nevertheless, it is of certain value to the community of hydrologists and for illustrations of the methodology, we refer the reader specifically to Rodriguez-Iturbe, Cox and Isham (1987, 1988).

II. A Method of Sample Path Analysis. This method aims to probe the representational capabilities of stochastic descriptions of precipitation, such as the Le Camian representations of rainfall fields. The method proceeds through a sample path analysis of an observed rainfall field in the context of a particular representation. The basic exploratory principal is that the observed data may be viewed as the composition of a fitted representation with a residual. Here the fitted representations are probed for their ability to provide realistic patterns of rainfall and the residuals are probed for systematic departures from the patterns desired. The procedures involved range from nonlinear regression and image analysis to time-series analysis and cross-validation against independent analyses of similar or related rainfall events.

We fix ideas and consider the WGR model as described in the previous section. We do, however, introduce a modest generalization. First, let χ denote the set $R^3_+ \times R^2$. We assume that N is a cluster point process on $T \times S \times \chi$, although the exact clustering regime is not central to this treatment. Here the raincells are endowed with four random characteristics; namely water content, aging rate, dispersion, and velocity. Thus the water contents are random as before, but the parameters a, b, and v_0 appearing in the WGR smoothing kernel are additionally taken to be random variables. This leaves the functional form of the smoothing kernel itself as the parameter determining the Le Camian representation in the model.

The problem is to assess the choice of smoothing kernel from a record of observed rainfall fields. Typically, the data consist of a time series of radar images drawn from the partial observation of the rainfall intensity process R. As an example of such data, we take the GATE rainfall fields as converted by Hudlow and Patterson (1979). These fields refer to a time series of radar images depicting derived rates of rain falling over a disk-shaped region of the Atlantic Ocean during the summer of 1974. The issue is thus the extent to which the rainfall intensity derived from the WGR model produces sample paths that represent the spatial evolution exhibited in this time series.

To fix ideas further, let $Y = (Y(t,x)), t \in I$ and $x \in D$ denote the time series of observed rainfall rates. Here I and D denote respectively the summer period of 1974 and the observational disk over the Atlantic Ocean. We view the data Y as a composition satisfying

$$Y(t, x) = R(t, x) + Z(t, x),$$

for every $t \in I$ and $x \in D$. Here the $R(t, x)$ denote the WGR representations of $Y(t,x)$ (see equation (3)) and the $Z(t,x)$ denote the corresponding residuals. The principal strategy of the present method of assessment is to probe the sample paths of the triplet of processes Y, R, and Z for evidence of the descriptive capabilities of the model.

The method of sample path analysis proceeds as follows:

(1) *The Fit.* The first step is to obtain a fitted version of the model-based representation. Recall that R is a function of the point process N and the smoothing kernel k. Since our focus is set upon assessing the choice of k and N cannot be observed directly, we look for a suitable state estimator $\hat{N}$, say, of N. A general theory for handling such problems is described in section 3.3 of Karr (1986). As a practical approach, we recommend finding a least-squares fit of N by minimizing the square-error loss between R and Y over the range of observation. In our case, this involves estimating the placements and the characteristics of those raincells generating Y. This may be handled using nonlinear regression. That is, we introduce a vector-valued "parameter" consisting of those atoms of N whose placements lie in $I \times D$; namely the placements and the characteristics of all the existing raincells perceived to be generating rainfall over the window of observation. We then solve for the minimizing value of that vector using the Gauss-Newton procedure described in Bates and Watts (1988).

(2) *Image Analysis.* Let $\hat{N}$ denote a suitable fit of the raincells and their characteristics. On substituting $\hat{N}$ for N in equation (3), we obtain a fitted ver-

sion $\hat{R}$, say, of the model-based representation of Y. Of course, the patterns of rainfall manifest by $\hat{R}$ are determined by the choice of the WGR kernel, and thus these patterns can be explored for the representational capabilities entailed in this choice. Generally speaking, we look for agreement in shape, organization, magnitude, and temporal evolution between the fields in the fitted image and those in the images of data. This may be handled informally by comparing animations of the two time series or by detailed data-analysis of individual images. Of course, we recommend doing both.

Next, we look at the least-squares residuals $\hat{Z}$ obtained by subtracting $\hat{R}$ from Y. This yields a time series of images depicting departures in the data from the fitted WGR representation. Here we examine the temporal evolution of the distribution of residuals with particular regard to their spatial homogeneity. That is, we look for reasonably Gaussian (i.e. structureless) distributions without apparent anomalies; this may be handled using normal probability plots. Secondly and perhaps more importantly, we examine the images of residuals for their adherence to spatial homogeneity using animation or detailed analysis of individual images. For example, the presence of inhomogeneities may reflect an inability of the model to properly fit peak intensities located at the raincells' centers or an inability to reproduce the stratiform precipitation supported by many cloud systems. By such an examination, we ascertain when in the life cycle of a storm system the WGR model gives an adequate description of precipitation and, moreover, when it fails to do so.

(3) *Decoupling and Time Series Analysis.* As a further exploration of the data and the model, we recommend that the analysis above be decoupled in time. That is, we are analyzing a time series of spatial distributions of rainfall over D, where, for each $t \in I$, $Y(t) = (Y(t,x))_{x \in D}$ denotes the observed distribution at time t. Here we recommend analyzing $Y(t)$, separately for every t, using the techniques above. In this sense, we decouple the analysis in time.

The aim of the decoupled analysis is for example to partially separate the choice of geometry from the choice of kinematics entailed in the choice of the smoothing kernel (see remark (2) of the previous section). Thus, for each t in the decoupled analysis, we decouple the WGR representation as follows. Let N_t denote a point process on $S \times R_+^2$. Here, if (x_t, z_t, b_t) denotes an atom of N_t for some ω, then x_t denotes the raincell's position and z_t and b_t denote its intensity and coefficient of dispersion. Now we introduce the decoupled smoothing kerned k' satisfying

$$k'(y;x_t,z_t,b_t) = z_t \sqrt{2/\pi}\ b_t\ e^{-b_t d(y,x_t)},$$

for every $y \in S$. Note that the decoupled kernel k' preserves only the spatial component or geometry in the starting WGR kernel. We have, for example, absorbed the exponential aging term into the time-dependent intensity z_t.

The decoupled analysis proceeds by obtaining a fit $\hat{N}_t$, say, to N_t, a fitted representation $\hat{R}_t$ of $Y(t)$, and the corresponding residual $\hat{Z}_t$. Here $\hat{R}_t$ and $\hat{Z}_t$ are put to the individual image analyses mentioned above, again emphasizing representational capability and spatial homogeneity. In addition, the $\hat{N}_t$ provide a times series of fitted raincells and their characteristics. These may be analyzed as a multivariate time series for consistency with the hypothesized model and historical analyses of similar rainfall events. In the case of the former, one can examine the choice of exponential aging as well as the chosen kinematic mechanism describing the motions of the raincells.

In closing, we remark that the analyses above depend to some extent upon subjective evaluations as well as a variety of informal physical judgements. For example, detecting raincells in radar images is to some degree a subjective operation, but this can be achieved with the help of sound data-analytic definitions of raincells as drawn from the meteorological literature. In fact, without the use of such informal physical judgements one may not be able to fix suitable starting values in the nonlinear regressions cited above. Nevertheless, some of these judgements may ultimately find support in the stability of the resulting least squares fits. Finally, we believe that the method of sample path analysis is a strategy that is complementary to the method of primary features for assessing Le Camian representations of rainfall fields. For an illustration of the methodology, we refer the reader to the analysis in Phelan and Goodall (1989).

4. *Aging Functions and Their Nonparametric Estimation.* The purpose of this section is to discuss the role of aging in rainfall fields. Our principal aim is to introduce the aging function, which is a cumulative hazard function appearing as parameter in many models of rainfall. We define a nonparametric estimator of the aging function and sketch its properties. In closing, we suggest integrating the estimator into the method of primary features as described in the previous section.

Aging in Rainfall Fields. The process of aging in convective raincells appears to be integral to the evolution of rainfall fields. For example, aging has been found empirically to effect the propagation of storms, the structure of turbulence in rainfall fields, as well as the shape of the spatial distribution of rainfall intensity.

For example, Houze (1981) discusses the importance of the aging of older raincells in the creation of new raincells at the leading edge of a storm system. He describes the resulting interplay between downdrafts and updrafts that appear to serve the propagation of the system. On the other hand, Zawadski (1973) drew a connection between the decay of the empirical autocovariance function of an observed rainfall field and Taylor's hypothesis of turbulence in such fields. Briefly, he showed that the expected duration of a convective raincell appears to define a cutoff for the approximate validity of this hypothesis for rainfall fields. Finally, using the method of sample path analysis outlined above, Phelan and Goodall (1989) assessed the WGR model against GATE rainfall fields. We found that the model fit the data reasonably well during the earlier stages of a storm, but that as the convective raincells that were supporting the storm neared the end of their effective lives, the model lost its ability to describe the patterns of observed rainfall. It thus appears that the geometry and kinematics of an aging storm is notably different than that of a developing or mature storm.

The importance of aging in these and other treatments of observed rainfall fields suggests the need to better understand the process of aging in convective raincells. One way to do this is in the context of a statistical investigation of the durations of such cells. The remainder of our discussion illustrates this approach by treating the problem of survival analysis of convective raincells from partially observed rainfall fields.

Aging Functions in Rainfall Models. An aging function is simply a cumulative hazard function belonging to a positive random variable. Nevertheless, we prefer the simpler term aging function. One reason for this preference lies in the role such functions play in the story of aging in rainfall fields; a fuller treatment of this role is to be found in Phelan (1989). We define an aging function.

Definition *Aging function.* The function $A = (A(t))$, $t \geq 0$ is said to be an aging function if A is an increasing, continuous function satisfying $A(0) = 0$ and $A(t)$ tends to infinity with t. □

Note that we have chosen to restrict our treatment to continuous aging functions for the mathematical convenience of this choice below.

We use aging functions here to parametrize the distribution of positive random variables. In specific, let X denote a positive random variable of distribution ψ. The random variable X is said to have aging function $A = A(t))$, $t \geq 0$ if

$$1 - \psi(t) = e^{-A(t)} \text{ or } \psi(dt) = A(dt)e^{-A(t)}, \quad t \geq 0,$$

The aging function is thus minus the natural logarithm of the survival curves belonging to X, which, in the context of survival analysis, is also called the cumulative hazard function.

We give an example. Let X denote a positive random variable having Pareto distribution of shape parameter b and scale parameter c. The aging function belonging to X then satisfies

$$A(t) = b\ln(1 + ct/b), \quad t \geq 0.$$

This aging function entails that X has a long-tailed distribution relative to, say, the distribution of an exponential random variable. Nevertheless, as b tends to infinity, $A(t)$ tends to ct for every t, where the latter is the aging function belonging to an exponential random variable of parameter c.

Aging functions appear as statistical parameters in many stochastic models of rainfall fields. As an illustration, we fix ideas and introduce the aging function to the random-disks model as described in sections 3 and 4. Recall that N is a Poisson random measure on $T \times S \times \chi$, where χ is the product space $R^3_+ \times R^2$. Here we simplify the model in assuming that the mean measure μ belonging to N satisfies

$$\mu(dt, dx, dr, dw, dz, dv) = dtdxdre^{-r}\, dwe^{-w}\, dze^{-z}\, \delta_0,$$

for every $t \in T, x \in S, r,w,z \in R_+$, and $v \in R^2$, where δ_0 is the direct measure concentrated at the fixed velocity v_0. The raincells thus move lockstep along the fixed velocity v_0. Moreover, their radii, water contents, and durations are independent exponential random variables of unit parameter.

Next, let $A = (A(t))$, $t \geq 0$ denote an aging function and let f denote the measurable mapping satisfying

$$f(t, x, r, w, z, v) = (t, x, r, w, A^{-}(z), v),$$

for every $t \in T, x \in S, r,w,z \in R_+$, and $v \in R^2$. Here A^{-} denotes the functional inverse of A. Now consider the point process $\tilde{N}$ given by $N o f^{-1}$, where f^{-1} denotes the inverse image of f. According to definition 1.36 in Karr (1986), $\tilde{N}$ is a measurable mapping of N by f. Moreover, $\tilde{N}$ is a Poisson random measure on

$T \times S \times \chi$. Its mean measure $\tilde{\mu}$ is $\mu o f^{-1}$ and satisfies

$$\tilde{\mu}(dt,dx,dr,dw,dz,dv) = dtdxdre^{-r}dwe^{-w}A(dz)\,e^{-A(z)}\delta_0,$$

for every $t \in T, x \in S, r, w, z \in R_+$, and $v \in R^2$. This follows easily from the nature of the transformation *f*. In specific, fix ω and suppose (*t*, *x*, *r*, *w*, *z*, *v*) is an atom of *N*(ω). Then, by definition, the point (*t*, *x*, *r*, *w*, $A^{-}(z)$, *v*) is an atom of $\tilde{N}(\omega)$. Now, it is well known that the mapping $z \rightarrow A^{-}(z)$ transforms the exponentially distributed durations among the atoms of *N* to durations among the atoms of $\tilde{N}$ having aging function *A*. This, or course, implies the desired result.

According to the argument above, $\tilde{N}$ generates raincells whose durations have aging function *A*. Obviously, we can enrich the mapping *f* by introducing transformations on the remaining characteristics of the raincells, but this would distract from our present purpose. That is, we view *A* as an infinite dimensional parameter in the model above and address its nonparametric estimation.

Nonparametric Estimation. We consider the nonparametric estimation of the aging function *A* appearing in the random-disks model described above. In this consideration, we describe a suitable observational scheme that allows for the partial observation of a set of convective raincells and their activities. We then define a martingale-type estimator of *A* that is to be drawn from these partial observations.

We imagine that we are positioned at a fixed weather radar station, such as the one presently operating off Darwin, Australia. Over a period of time, storm systems pass in the vicinity of our instruments, which record the activities of the convective raincells. We specifically assume that our instruments track the trajectories of the raincells as they pass through the field of view. An objective method for doing so is described in Rosenfield (1987).

More formally, let $T_0 \subset T$ denote an interval of time and let $S_0 \subset S$ denote a disk of fixed radius in the plane. Here T_0 denotes the period of observation and S_0 denotes the field of view of our instruments, so that $T_0 \times S_0$ denotes the coverage window of our observational scheme. Our scheme is to observe the trajectory of a raincell in S_0, for every raincell that enters S_0 during the period T_0.

We assume that raincells and their characteristics are being generated by the

Poisson random measure $\tilde{N}$ on $T \times S \times \chi$, where $\tilde{N}$ is described above. We identify $\tilde{N}$ with the countable sequence $(T_n, X_n, R_n, W_n, Z_n, v_0)$, $n \geq 0$ of $T \times S \times \chi$ - valued random variables. For the raincell having label n, (T_n, X_n) denotes its time and place of birth in $T \times S$, R_n, W_n, and Z_n respectively denote its radius, water content, and duration, and v_0 denotes its fixed velocity. According to the kinematics of the random-disks model, the trajectory of the raincell traces a line segment L_n, say, in S. According to our observational scheme, we observe that part of L_n contained in S_0. Figure 1 depicts the four contingencies in the partial observation of L_n, depending on the position of the raincell at birth and at death.

More formally, for each n, let L_n denote the line segment in S traced by the trajectory of the raincell labeled n Then L_n satisfies

$$L_n = \{x: x \in S, x = X_n + tv_0 \text{ or some } t \in [0, Z_n]\}.$$

Let $l(L_n)$ denote the length of L_n. Clearly, we have

$$Z_n = l(L_n)/\|v_0\|,$$

for every n, provided $\|v_0\| > 0$. Thus, by observing the trajectory of the raincell, we observe its duration. The observational scheme, however, allows only the partial observation of the L_n, and thus it allows only the partial observation of the Z_n. That is, we observe only the subtrajectory $S_0 \cap L_n$, see figure 1, and therefore we observe only the censored duration $\tilde{Z}_n$ satisfying

$$\tilde{Z}_n = l(S_0 \cap L_n)/\|v_0\|.$$

According to the model, the Z_n have aging function $A = (A(t))$, $t \geq 0$. Here we propose an estimator of A to be drawn from the $\tilde{Z}_n$; the censored durations. For each n, let δ_n denote the indicator $1(Z_n = \tilde{Z}_n)$; this indicator is one for any raincell whose trajectory is fully contained in S_0. Let M denote the random variable satisfying

$$M = \sum_{n \geq 0} 1([T_n, T_n + Z_n \cap T_0 \neq \phi, S_0 \cap L_n \neq \emptyset).$$

Here M is the random number of raincells whose trajectories enter our field of view. If for example T_0 is a bounded interval, then M is finite almost surely.

Now let (N_t), $t \geq 0$ and (R_t), $t \geq 0$ denote the stochastic processes satisfying

$$N_t = \sum_{n=1}^{M} 1(\tilde{Z}_n \leq t,\ \delta_n = 1) \text{ and } R_t = \sum_{n=1}^{M} 1(\tilde{Z}_n \geq t).$$

These processes are respectively the analogue of the survival counting process and the risk process referred to in the analysis of censored survival data; see for example Andersen and Borgan (1985). Finally, we define the stochastic process $\tilde{A} = (\tilde{A}_t)$, $t \geq 0$ satisfying

$$\tilde{A} = \int_0^t \frac{dN_s}{R_s}.$$

Here we propose $\tilde{A}$ as a nonparametric estimator of the aging function A.

Properties of the Estimator $\tilde{A}$. The estimator $\tilde{A}$ is in essence a Nelson-Aalen estimator of the aging function A. It may be justified as a nonparametric maximum likelihood estimator (NMLE) or as a martingale estimator. In the case of the former, one writes down the appropriate likelihood for the censored durations as a function of the infinite dimensional parameter A. Then one appeals to a generalized principle of maximum likelihood to derive $\tilde{A}$ as the NMLE in the manner of argument given for example in Johansen (1978).

On the other hand, $\tilde{A}$ is also a martingale estimator. That is, let $G = \left(\underline{\underline{G}}_t \right)$, $t \geq 0$ denote the internal history generated by the counting process (N_t), $t \geq 0$. Then the process formed of the difference between $\tilde{A}$ and A, namely $\tilde{A} - A = (\tilde{A} - A(t))$, $t \geq 0$, is a semimartingale relative to G. Although such estimators have been studied extensively in for example Andersen and Borgan (1985), their application to the analysis of rainfall fields is a novel one.

Whether one views $\tilde{A}$ as an NMLE or a martingale estimator, the martingale property is useful in developing its statistical properties. For example, we consider the asymptotic distribution of $\tilde{A}$ as the observational period T_0 grows unboundedly. In this consideration, we expect that $\tilde{A}$ - A (suitably normalized) converges weakly to a Gaussian process. Since $\tilde{A}$ is a martingale estimator, this expectation can be proved by appeal to the appropriate martingale central limit theorem.

A complete treatment of the statistical properties of $\tilde{A}$ are to appear in a Phelan (1990). There we treat a broader range of observational schemes and stochastic models of greater generality than the one described above.

The estimator $\tilde{A}$ and The Method of Primary Features. In order to integrate the estimator $\tilde{A}$ into the method of primary features, we recommend that the aging function A be chosen as one of the primary features. That is, we identify A with the model-based value of the "aging-function feature." Typically A or $A(\theta)$ will depend on one or two parameters in the model, such as the shape and scale parameter determining the Pareto aging function exhibited above. Next, we identify the data-based value of the aging-function feature with the nonparametric estimator $\tilde{A}$. Finally, we incorporate the pair $(A(\theta),\tilde{A})$ directly into either the estimation or the post-estimation assessment procedure developed in the method of primary features. On the one hand, $\tilde{A}$ may be used to determine the parameter θ in the simultaneous equation shown at (5). Otherwise, suppose an estimate $\hat{\theta}$ of θ is obtained from (5) without the help of $\tilde{A}$. We then use $\hat{\theta}$ to form the model-based parametric estimator $A(\hat{\theta})$ of $A(\theta)$ and we compare $A(\hat{\theta})$ against $\tilde{A}$ as a partial, nonparametric assessment of the fit of the model.

5. Concluding Remarks. We argued for the use of spatial stochastic processes as phenomenological models of rainfall fields. In this spirit, we treated the stochastic descriptions in Le Cam (1961). We described Le Camian representations of rainfall fields and cited two examples of interest to hydrologists who analyze such fields.

We referred to the need of flexible strategies in the assessment of Le Camian representations of rainfall. In doing so, we outlined two strategies of assessment; one called the method of primary features and the other a method of sample path analysis. We argued that the two offer complementary approaches to the problem of assessment. We remark here that each strategy is made possible by the increase in sophistication of both data and data-handling capability available today.

We pointed out that the dialogue among observation, modeling, and inference leads to fresh scientific inquiries as well as to novel applications of statistical technique. In illustration, we introduced a nonparametric estimator of an aging function in a rainfall model by way of a martingale estimator typically found in the literature on life history analysis. We argued for the integration of this estimator into the method of primary features.

We remark, finally, that the Le Camian framework for the stochastic description of precipitation remains to be fully explored. For example, little is understood of how well these models satisfy the evolution equations that describe the physical dynamics of precipitation. Notwithstanding, more work is needed in the theory and application of such descriptions in the statistical analysis of observed rainfall fields.

References

[1] Andersen, P. K. and Borgan, O. (1985). "Counting process models for life history data: a review." *Scand. J. Statist.* **12**, 97-158.

[2] Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*, John Wiley & Sons, Inc., New York.

[3] Cox, D. R. and Isham, V. S. (1988). "A simple spatial-temporal model of rainfall," *Proc. R. Soc. Lond.*, A **915**, 317-328.

[4] Houze, R. A. (1981). "Structure of atmospheric precipitation systems: a global survey," *Radio Science*, **16**, 671-689.

[5] Houze, R. A. and Hobbs, P. V. (1982). "Organization and structure of precipitating cloud systems," *Adv. Geophys.*, **24**, 225-315.

[6] Hudlow, M. D. and Patterson, P. (1979). GATE Radar Rainfall Atlas, NOAA Special Report.

[7] Johansen, S. (1978). "The product limit estimator as maximum likelihood estimator." *Scand. J. Statist.* **5**, 195-199.

[8] Karr, A. (1986). *Point Processes and Their Statistical Inference*, Marcel Dekker, Inc., New York.

[9] Le Cam, L. M. (1961). "A stochastic description of precipitation," in *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.), 3, 165-186, Berkely, CA.

[10] Nicolis, G. and Prigogine, I. (1977). *Self-Organization in Nonequilibrium Systems*, John Wiley and Sons, New York.

[11] Phelan, M. J. (1991). "Aging functions and their nonparametric estimation in point process models of rainfall fields," to appear in *Statistics in Environmental and Earth Sciences*, P. Guttorp and A. T. Walden, eds., Edward Arnold Limited, Kent, England.

[12] Phelan, M. J. and Goodall, C. R. (1989). "An assessment of a generalized WGR model for GATE rainfall. *J. Geophys. Res., Atmos.*

[13] Rodriguez-Iturbe, I., Cox, D. R. and Eagleson, P. S. (1986). "Spatial modelling of total storm rainfall, "*Proc. R. Soc. Lond. A*, **403** 27-50.

[14] Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987). "Some models for rainfall based on stochastic point processes, "*Proc. R. Soc. Lond. A*, **410**, 269-288.

[15] Rodriguez-Iturbe, I., cox, D. r. and Isham, V. (1988). "A point process model for rainfall: further developments, "*Proc. R. Soc. Lond. A.*, **417**, 283-298.

[16] Rosenfeld, (1987). "Objective method for tracking convective cells as seen by radar," *J. Atmos. and Oceanic Tech.*, **4**, 422.

[17] Smith, J. A. and Karr, A. F. (1985). "Parameter estimation for a model of space-time rainfall," *Water Resource Res.*, **21**, 1251-1257.

[18] Waymire, E., Gupta, V. K., and Rodriguez-Iturbe, I. (1984). "Spectral theory of rainfall intensity at the meso-β scale," *Water Resource Res.*, **20**, 1453-1465.

[19] Zawadzki, I. I. (1973). "Statistical properties of precipitation patterns," *J. Appl. Meteorol.*, **121**, 459-472.

Figure 1. Four contingencies in the partial observation of L_n. Here S_0 denotes a disk in the plane, v_0 points in the horizontal direction, x denotes the position of the raincell at birth, and 0 denotes its position at death.