Proc. R. Soc. Lond. A **410**, 269–288 (1987)
Printed in Great Britain

Some models for rainfall based on stochastic point processes

By I. Rodriguez-Iturbe[1], D. R. Cox[2], F.R.S., and Valerie Isham[3]

[1] *Graduate Program in Hydrology and Water Resources, Simon Bolivar University and Instituto Internacional de Estudios Avanzados, Caracas, Venezuela*
[2] *Department of Mathematics, Imperial College, London SW7 2BZ, U.K.*
[3] *Department of Statistical Science, University College London, London WC1E 6BT, U.K.*

(*Received* 8 *May* 1986)

Stochastic models are discussed for the variation of rainfall intensity at a fixed point in space. First, models are analysed in which storm events arise in a Poisson process, each such event being associated with a period of rainfall of random duration and constant but random intensity. Total rainfall intensity is formed by adding the contributions from all storm events. Then similar but more complex models are studied in which storms arise in a Poisson process, each storm giving rise to a cluster of rain cells and each cell being associated with a random period of rain. The main properties of these models are determined analytically. Analysis of some hourly rainfall data from Denver, Colorado shows the clustered models to be much the more satisfactory.

1. Introduction

The object of this paper is to describe some probabilistic models for the behaviour in time of the precipitation at a fixed point in space. The aim is to reduce the description of the rainfall process to the specification of a small number of parameters from which other properties of the natural process can be deduced via the mathematical model.

One of the key issues concerns the accommodation of effects with a number of time scales. For instance, one might have a model which fitted quite adequately the main features of hourly rainfall. From such a mathematical model it would then, in principle, be possible to deduce the properties after aggregation to daily or weekly totals. Goodness of fit of the main features of the aggregated model is in no way assured! The capability of representing effects on different time scales within one model is, we consider, an important one and the later models discussed in this paper are specifically designed to possess this capability via the use of the rain cell as the basic element in the mathematical structure.

In the present paper hourly rainfall data from Denver, Colorado, are used to compare the fits of the different models. A more extensive analysis of that data is being prepared separately.

The models are inevitably idealized and some of the detailed assumptions are made, at least partly, for mathematical convenience. Whereas the models thus do not represent detailed physical mechanisms, it is intended that the parameters have a physical interpretation.

All the models are based in the first instance on rectangular pulses and therefore generate realizations with an unrealistic appearance. In §5, we indicate how, by superimposing a 'jitter' on the process, more realistic forms can be obtained. Whereas second-moment properties can be obtained for arbitrary pulse shapes, reasonable economy in parametrization and the absence of clear patterns in cell distributions in time suggest the use of rectangular pulses, modified in the way outlined, as the primary representation.

There is an extensive literature on stochastic models for rainfall. For some general comments on the connection with point processes, see Waymire & Gupta (1981). Stern & Coe (1984) have a model for daily rainfall in which wet and dry days occur in a Markov chain with seasonally dependent transition probabilities and in which the amounts of rain per wet day have a gamma distribution with seasonally dependent parameters. Also LeCam (1961) discusses in some generality a spatial-temporal process in which the pulses of rain are regarded as falling instantaneously in time in a clustered point process. These models are substantially different from those of the present paper, in particular because of their relative inability to represent phenomena on different time scales.

2. Poisson model with rectangular pulses

2.1. *Specification of the model*

We start with a brief description of the simplest model built from rectangular pulses associated with a Poisson process. That is, with each point of a Poisson process of rate λ per unit time is associated a rectangular pulse of random duration, L, and random 'depth', X, representing rainfall intensity (millimetres per unit time). The total intensity, $Y(t)$, is the sum of all such contributions 'active' at t. It is assumed that the different intensities and durations are mutually independent and independent of the Poisson process.

The 'points' in the Poisson process are not intended to have any very specific physical interpretation; in particular, different points may generate overlapping pulses which correspond to a single 'storm'. The model has been studied and used by Rodriguez-Iturbe *et al.* (1984).

We denote the mean, variance and probability density function of the depth, X, of a single pulse by $\mu_X, \sigma_X^2, f_X(x)$ the corresponding quantities for the duration, L, by $\mu_L, \sigma_L^2, f_L(l)$ and the survivor function of L by $\mathscr{F}_L(l)$. An important special case, at least roughly concordant with data, is when both these distributions are exponential, with densities, respectively, $\xi\, e^{-\xi x}$ and $\eta\, e^{-\eta l}$. We call this the exponential case.

Note that $\mu_X = \xi^{-1}$ and $\mu_L = \eta^{-1}$; it is often convenient to write $\lambda = \rho\eta$, $\xi^{-1} = \psi\eta$, where ρ is dimensionless and ψ has the dimensions of length.

Another special case is useful to represent long-term effects. In this the cell durations have a long-tailed distribution. One convenient form is the Pareto density

$$\eta(1+\eta l/\omega)^{-\omega-1},$$

with survivor function

$$(1+\eta l/\omega)^{-\omega}, \tag{2.1}$$

and mean $\mu_L = \omega/\{\eta(\omega-1)\}$. Here ω is a dimensionless parameter; as $\omega \to \infty$ we recover the exponential density. The form results from assigning a gamma mixing density to the rate parameter of an exponential distribution.

An alternative form showing qualitatively similar behaviour has

$$\mathscr{F}_L(l) = \begin{cases} 1 & (\eta l < \omega), \\ (\eta l/\omega)^{-\omega} & (\eta l \geqslant \omega). \end{cases} \tag{2.2}$$

2.2. *Some second-order properties*

Let $Y(t)$ be the total intensity at time t.

The second-order properties of $Y(t)$ can be derived in various ways of which the simplest is probably to write

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u)\,\mathrm{d}N(t-u), \tag{2.3}$$

where $X_u(\tau)$ is the random depth of the pulse originating at time u measured a time τ later and where $\{N(t)\}$ counts occurrences in the Poisson process of pulse origins. In particular, for the rectangular pulses considered here, we have that

$$X_{t-u}(u) = \begin{cases} X & \text{with probability} \quad \mathscr{F}_L(u), \\ 0 & \text{with probability} \quad 1-\mathscr{F}_L(u). \end{cases} \tag{2.4}$$

It follows directly that

$$\left.\begin{aligned} E\{Y(t)\} &= \lambda \int_0^{\infty} E\{X_t(v)\}\,\mathrm{d}v = \lambda \mu_X \mu_L, \\ \operatorname{var}\{Y(t)\} &= \lambda(\mu_X^2+\sigma_X^2) \int_0^{\infty} \mathscr{F}_L(v)\,\mathrm{d}v = \lambda(\mu_X^2+\sigma_X^2)\,\mu_L, \end{aligned}\right\} \tag{2.5}$$

and that $c_Y(\tau) = \operatorname{cov}\{Y(t),\ Y(t+\tau)\}$ is given by

$$c_Y(\tau) = \lambda(\mu_X^2+\sigma_X^2) \int_{\tau}^{\infty} \mathscr{F}_L(v)\,\mathrm{d}v. \tag{2.6}$$

In the exponential case these results reduce to

$$E\{Y(t)\} = \psi\eta\rho, \quad \operatorname{var}\{Y(t)\} = 2\psi^2\pi^2\rho, \tag{2.7}$$

$$c_Y(\tau) = 2\psi^2\eta^2\rho\,\mathrm{e}^{-\eta\tau}. \tag{2.8}$$

In the Pareto case (2.1), $\omega > 1$,

$$c_Y(\tau) = \frac{\lambda(\mu_X^2+\sigma_X^2)\,\eta}{\omega(\omega-1)}\left(1+\frac{\eta\tau}{\omega}\right)^{-\omega+1}, \tag{2.9}$$

so that for sufficiently small values of ω slow decay of the autocovariance is induced, although only by having some very long periods of uninterrupted rain.

It is not difficult to show that the number R of active pulses at an arbitrary time has a Poisson distribution of mean $\lambda\mu_L$, so that the marginal distribution of $Y(t)$ is that of

$$X_1+\ldots+X_R,$$

where the X_j are independent copies of the random variable X specifying the depth of a pulse. Thus the Laplace transform $M_Y(s) = E(e^{-sY})$ is given by

$$M_Y(s) = \exp\{\lambda\mu_L M_X(s) - \lambda\mu_L\}, \tag{2.10}$$

where $M_X(s)$ is the Laplace transform of the density of X.

The corresponding cumulant generating function, $K_Y(s)$, is thus

$$K_Y(s) = \lambda\mu_L M_X(s) - \lambda\mu_L,$$

so that the rth cumulant of Y, $\kappa_r(Y)$, say, is given by

$$\kappa_r(Y) = \lambda\mu_L E(X^r) \tag{2.11}$$

generalizing and extending (2.6).

In the exponential case the marginal distribution of $Y(t)$ can be found directly on inversion of (2.10) as

$$e^{-\rho}\,\delta(y) + \left(\frac{\rho}{\psi\eta y}\right)^{\frac{1}{2}} e^{-\rho}\, e^{-y/(\psi\eta)}\, I_1\left\{\left(\frac{4\rho y}{\psi\eta}\right)^{\frac{1}{2}}\right\}, \tag{2.12}$$

where $I_n(\cdot)$ is the modified Bessel function of order n and $\delta(\cdot)$ is a Dirac delta function specifying an atom of probability at zero.

2.3. *Aggregated process*

Now rainfall data are available only in aggregated form, so that we consider the cumulative rainfall totals in disjoint time intervals of some fixed length h, say:

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(v)\,dv. \tag{2.13}$$

The second-order properties of the aggregated process (Rodriguez-Iturbe *et al.* 1984) are now obtained as

$$\left.\begin{aligned} E(Y_i^{(h)}) &= hE\{Y(t)\}, \\ \operatorname{var}(Y_i^{(h)}) &= 2\int_0^h (h-u)\,c_Y(u)\,du, \\ \operatorname{cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) &= \int_{-h}^{h} c_Y(kh+v)(h-|v|)\,dv. \end{aligned}\right\} \tag{2.14}$$

In particular, in the exponential case,

$$\left.\begin{aligned} &E(Y_i^{(h)}) = h\psi\eta\rho,\ \operatorname{var}(Y_i^{(h)}) = 4\psi^2\eta\rho(h\eta - 1 + e^{-h\eta}), \\ &\operatorname{cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) = 2\psi^2\rho(e^{h\eta} + e^{-h\eta} - 2)\,e^{-k\eta}\ (k \geqslant 1). \end{aligned}\right\} \tag{2.15}$$

To obtain higher cumulants is possible, in principle, but involves more complicated calculations and will not be discussed here.

The probability that an interval of length h is dry is the probability that the initial point is dry multiplied by the probability of no cell arriving in an interval h and is thus $e^{-\lambda\mu_L}\,e^{-\lambda h}$.

The distribution of the number of consecutive such dry periods is geometric with parameter $e^{-\lambda h}$. Note that this last result depends only on λ and that the form of the distributions of X and L does not enter.

2.4. *Fitting*

When this model is used in a hydrological context, it is sometimes fitted by subjective isolation of the component 'pulses' followed by evaluation of the corresponding parameters. This is likely to be a poor procedure, at least when there is appreciable overlap of the components and a preferable procedure is as follows.

The autocovariance of rainfall is strongly influenced by the form of the distribution of L. In particular, if the autocovariance is essentially exponential, an exponential distribution for L is indicated and the lag one autocorrelation determines η. The geometric distribution of dry periods can be used to check on the model and to estimate λ and finally the moments of rainfall and in particular the mean and variance can be used to estimate properties of the distribution of pulse depths, X. Note particularly that in the exponential case, with just three parameters to be estimated and at least four or five features reasonably estimable there is good scope for testing the adequacy of the model, especially when several levels of aggregation are used.

Formal methods of estimation based on the likelihood of the data under the model are not feasible in the present context.

2.5. *Scale of fluctuation*

For processes for which $|c_Y(\tau)|$ decreases sufficiently rapidly with τ, one convenient measure of the timescale of fluctuations (Taylor 1921; Vanmarcke 1983) is obtained from (2.14) for large h, when

$$\operatorname{var}\{Y_i^{(h)}/h\} = 2h^{-1}\int_0^\infty c_Y(u)\,\mathrm{d}u.$$

If for large h, $Y_i^{(h)}/h$ is equivalent to the mean of h/α independent observations then we must have

$$\alpha = 2\int_0^\infty c_Y(u)\,\mathrm{d}u/c_Y(0). \tag{2.16}$$

In the exponential case $\alpha = 2/\eta$.

2.6. *Some further properties*

We now discuss briefly some further properties. The process is easily generalized to the multisite case. An exponential version of such a scheme is able to preserve mean, variance and lag-one correlation at a series of k sites where rainfall is considered at a given level of aggregation. Moreover, agreement can be preserved between lag-zero or lag-one cross correlations between all pairs of sites, again at a given level of aggregation.

The processes considered above are all reversible in time, so that in particular the first differences of aggregated data are symmetrically distributed.

Next, it is possible to obtain the level-crossing properties of the process; we consider the exponential case for simplicity. In fact, the probability of an upcrossing of a level b in $(t, t+\mathrm{d}t)$ is the sum of two terms one referring to a jump from the dry state, the other to a jump from a non-dry state. Thus the upcrossing rate ν_b^+ is

$$\left.\begin{aligned} \nu_b^+ &= \lambda\{\exp(-\rho-b\xi)+\mathrm{e}^{-b\xi}\int_{0+}^{b} \mathrm{e}^{\xi y} f_Y(y)\,\mathrm{d}y\} \\ &= \lambda\exp(-\rho-b\xi)\, I_0\{(4\rho b\xi)^{\frac{1}{2}}\}. \end{aligned}\right\} \tag{2.17}$$

Finally we can examine in the exponential case the distribution of wet periods, taken in continuous time, i.e. evaluated for the process $\{Y(t)\}$. Note first that from the known probability that $Y(t) = 0$ and the expression of this as

$$E(T_\mathrm{D})/\{E(T_\mathrm{D})+E(T_\mathrm{W})\},$$

where T_D and T_W are dry and wet periods, it follows that

$$E(T_\mathrm{W}) = (\mathrm{e}^{\rho}-1)/\lambda, \tag{2.18}$$

where $\rho = \lambda/\eta$. In particular if $\rho \ll 1$, $E(T_\mathrm{W}) \sim \eta^{-1}$, the mean length of a single pulse.

To examine the distribution of T_W, we suppose that at $t = 0$, one pulse is active and we define a discrete state Markov process via R, the number of active pulses, defining $R = 0$ to be an absorbing state.

Let $p_r(t)$ denote the probability that, at time t, r pulses are active, so that $p_0(t) = P(T_\mathrm{W} \geqslant t)$ is the cumulative distribution of a wet period. Then

$$\left.\begin{aligned} &\mathrm{d}p_0(t)/\mathrm{d}t = \eta p_1(t), \\ &\mathrm{d}p_1(t)/\mathrm{d}t = -(\lambda+\eta)\, p_1(t)+2\eta p_2(t), \\ &\mathrm{d}p_r(t)/\mathrm{d}t = -(\lambda+r\eta)\, p_r(t)+(r+1)\,\eta p_{r+1}(t)+\lambda p_{r-1}(t) \quad (r = 2, 3, \ldots). \end{aligned}\right\} \tag{2.19}$$

We introduce a generating function and take Laplace transforms so that

$$\left.\begin{aligned} G^*(z, s) &= \int_0^\infty G(z, t)\,\mathrm{e}^{-st}\,\mathrm{d}t, \\ G(z, t) &= \sum_{r=1}^{\infty} p_r(t)\, z^r. \end{aligned}\right\} \tag{2.20}$$

Then it follows from (2.19) that $G^*(z, s)$ satisfies

$$(\partial/\partial z)\{G^*(z, s)\,\mathrm{e}^{-\rho z}(1-z)^{s/\eta}\}+\eta^{-1}\{z-sp_0^*(s)\}\,\mathrm{e}^{-\rho z}\,(1-z)^{s/\eta-1} = 0,$$

so that

$$G^*(z, s) = \mathrm{e}^{\rho z}(1-z)^{-s/\eta}\int_0^z \eta^{-1}\{sp_0^*(s)-v\}\,\mathrm{e}^{-\rho v}(1-v)^{s/\eta-1}\,\mathrm{d}v.$$

Now the normalizing condition for the distribution of R is, after Laplace transformation,

$$p_0^*(s)+G^*(1, s) = s^{-1}. \tag{2.21}$$

In order that $G^*(z, s)$ is finite as $z \to 1$, we must have

$$\int_0^1 \eta^{-1}\{sp_0^*(s)-v\}\,\mathrm{e}^{-\rho v}(1-v)^{s/\eta-1}\,\mathrm{d}v = 0.$$

After some integration by parts, we obtain

$$sp_0^*(s)=\left\{\int_0^1 \mathrm{d}v(1-v)^{s/\eta}\,\mathrm{e}^{-\rho v}(1-\lambda v/\eta)\right\}\left\{1-\lambda\eta^{-1}\int_0^1 \mathrm{d}v(1-v)^{s/\eta}\,\mathrm{e}^{-\rho v}\right\}^{-1}. \quad (2.22)$$

In fact, this is precisely $f^*_{T\mathrm{W}}(s)$, the Laplace transform of the density of T_{W}.

The formula (2.18) for the expected value is now recovered on differentiation. Further

$$E(T^2_{\mathrm{W}})=\frac{2\,\mathrm{e}^{\rho}}{\eta^2}\int_0^1 \mathrm{d}u\,\mathrm{e}^{\rho u}(1+\rho u\ln u). \quad (2.23)$$

If ρ is small

$$\eta^2E(T^2_{\mathrm{W}})=2\{1+1.3355\rho+O(\rho^2)\},$$

whereas if ρ is large

$$\eta^2E(T^2_{\mathrm{W}})\sim 2\rho^{-2}\,\mathrm{e}^{2\rho}.$$

Table 1 gives $C_\rho=\frac{1}{2}\rho^2\eta^2\,\mathrm{e}^{-2\rho}E(T^2_{\mathrm{W}})$ for a few intermediate values of ρ.

TABLE 1. FUNCTION C_ρ TO DETERMINE MEAN SQUARE OF WET PERIOD

ρ	1	1.2	1.4	1.6	1.8	2	3	4	5	10
C_ρ	0.4848	0.6081	0.7228	0.8264	0.9179	0.9971	1.2334	1.2944	1.2802	1.1301

If ρ is small we have on expansion that

$$f^*_{T\mathrm{W}}(s)=\frac{\eta}{s+\eta}-\frac{\rho\eta^2 s}{(s+\eta)^2(s+2\eta)}+O(\rho^2), \quad (2.24)$$

and that

$$f_{T\mathrm{W}}(t)=\eta\,\mathrm{e}^{-\eta t}+\rho\eta(2\,\mathrm{e}^{-2\eta t}+\eta t\,\mathrm{e}^{-\eta t}-2\,\mathrm{e}^{-\eta t})+O(\rho^2), \quad (2.25)$$

where the first term is the density of a single pulse.

2.7. *Empirical results*

A large body of hourly rainfall data from Denver, Colorado has been analysed in the light of the models of the present paper. A detailed account will appear elsewhere. Here we give a few typical results to illustrate and motivate the discussion.

The particular data are for the period 15 May–16 June for the 28 years 1949–76. They have been analysed on an hourly basis and also aggregated into $h = 6$, 12, 24 h totals. Table 2 summarizes a few of the key comparisons, involving also a model to be discussed in §4. Two Poisson models have been fitted, one with exponentially distributed pulse durations and the other with the Pareto distribution. For both the pulse depths were assumed exponentially distributed, but this does not have a critical effect on the following discussion.

Estimation was performed for the exponential model by fitting mean, variance and lag-one autocorrelation at one level of aggregation, and for the Pareto distribution by fitting also the probability of zero rain. Four fits are reported here corresponding to the 1 h rainfalls and the 6 h aggregates; fitting based on 12 and 24 h totals confirms the conclusions. Note that the fits are based on one level of aggregation but that the comparisons are made at all levels.

Table 2. Rainfall at Denver (15 May–16 June; 1949–76): observed and fitted values under Poisson and cluster-based models†

level of aggregation/h	mean/mm	var./mm²	skewness	autocorrelation, lag 1	2	3	probability of zero rain
1	*0.088*	*0.403*	*10.95*	*0.48*	*0.32*	*0.27*	*0.94*
		0.403	10.04	0.48	0.14	0.14	0.94
		0.259	7.92	0.81	0.58	0.42	0.93
		0.403	.	0.49	0.21	0.12	0.93
		0.333	.	0.63	0.37	0.26	0.94
		0.409	11.42	0.50	0.24	0.17	0.95
6	*0.531*	*5.970*	*6.03*	*0.33*	*0.13*	*0.06*	*0.86*
		4.93	5.44	0.08	0	0	0.80
		5.91	5.74	0.33	0.05	0.01	0.85
		5.55	.	0.18	0.04	0.02	0.73
		5.97	.	0.32	0.12	0.07	0.81
		6.04	5.88	0.33	0.15	0.08	0.90
12	*1.063*	*16.88*	*4.62*	*0.23*	*0.07*	*−0.03*	*0.78*
		10.63	4.11	0.04	0	0	0.66
		15.73	4.31	0.16	0	0	0.76
		13.12	.	0.07	0	0	0.55
		15.79	.	0.24	0.08	0.04	0.68
		16.02	4.25	0.27	0.07	0.02	0.86
24	*2.125*	*41.61*	*3.34*	*0.16*	*−0.03*	*−0.04*	*0.64*
		22.03	2.77	0.02	0	0	0.44
		36.54	2.94	0.07	0.01	0	0.61
		29.31	.	0.07	0.01	0	0.31
		39.15	.	0.18	0.05	0.03	0.47
		40.52	3.05	0.16	0.01	0	0.78

† First line, observed, italicized. Second and third lines, fitted by Poisson exponential model with 1 h and with 6 h data. Fourth and fifth lines, fitted by Poisson Pareto model with 1 h and with 6 h data. Sixth line, fitted by Bartlett–Lewis exponential model with 6 h and 12 h data.

The qualitative conclusions seem clear. The model with exponential durations produces autocorrelations that decay much too rapidly even at the time period used for fitting and it gives a poor fit at other levels of aggregation. The Pareto model does much better at the level of aggregation used for fitting but again the fits at other levels of aggregation are not good. A further important point is that in both cases the estimates of important parameters depend very appreciably on the level of aggregation used; see table 2.

We thus conclude that the models based directly on the Poisson process are of limited usefulness, essentially because of their inability to take account of phenomena over a range of timescales. In the remainder of the paper we turn to models based on clustered point processes which offer much greater flexibility and physical realism.

3. Cluster-based models

3.1. *Introduction*

Although the Poisson-based model outlined in §2 gives a reasonably flexible model for the process at a particular level of aggregation, it is very limited when a range of time scales is contemplated. As we have seen, relatively long-term correlations can be included by having a long-tailed distribution for the pulse durations, but this system is incapable of representing long periods of rainfall interspersed with short dry sections. Lengths of dry periods are independently exponentially distributed.

To have the flexibility for representing more than one timescale and for appreciably greater physical realism, we turn to models based on clustered point processes rather than on the Poisson process. That is, with each point of a clustered point process we associate a rectangular pulse, as in §2.

So far as second-order properties of the process $\{Y(t)\}$ are concerned calculation of the covariance density of the underlying point process leads to a relatively simple expression for the variance and autocovariance (Cox & Miller 1965, ch. 9). For more detailed study a particular form for the clustering of the point process is needed.

Two such processes are natural candidates for further consideration, the Neyman–Scott process and the Bartlett–Lewis process; see, for example, Cox & Isham (1980, §3.4). The difference between the two is relatively subtle. The key difference is explained below and concerns the precise way in which cells are distributed in time. It is very unlikely that empirical analysis of data can be used to choose between them.

In §4 we concentrate on a special version of the Bartlett–Lewis process which lends itself to analytical discussion.

Both processes are based on an initial Poisson process of storm origins. With each origin is associated a random number of cells; natural candidates for the distribution of the number of cells are the Poisson distribution and the geometric distribution, it being a matter of convention whether a cell is always attached to the storm origin. In the Neyman–Scott process the positions of these cells are determined by a set of independent and identically distributed random variables where the location of this distribution is given by the storm origin. In the Bartlett–Lewis process the intervals between successive cells are independent and identically distributed, in the special version we shall consider with an exponential distribution.

In §3.2 we outline a discussion based on a special form of the Neyman–Scott process.

It is likely that the rainfall models based on clustering processes are not reversible in time except in very special cases and that the particular models discussed in this paper are never reversible, although this aspect has not been studied in detail.

3.2. *Model based on a Neyman–Scott process*

We suppose that storm origins occur in a Poisson process of rate λ, and that a random number $C(C \geqslant 1)$ of cell origins is associated with each storm. The cell origins are independently displaced from the storm origin by distances which are exponentially distributed with parameter β, no cell origin being located at the storm origin. As before, a rectangular pulse (or cell) is associated independently with each cell origin, its duration and depth being independent random variables and the duration having an exponential distribution with parameter η. The clusters of cells associated with distinct storms are independent. It will sometimes be convenient to use the dimensionless parameter $\rho = \lambda/\eta$.

It is straightforward to determine the second-order properties of the depth process, $Y(t)$, from the integral representation

$$Y(t) = \int_0^\infty X_{t-u}(u)\,\mathrm{d}N(t-u)$$

of (2.3), where now $N(t)$ counts occurrences in the Neyman–Scott process of cell origins. Note that the intensity of $N(t)$ is $\lambda\mu_C$, where μ_C denotes the mean number, $E(C)$, of cells per storm. In particular, the mean of the depth process can be represented as the product of the rate at which cell origins occur, the mean length of a cell and the mean depth of a cell, that is,

$$E\{Y(t)\} = \lambda\mu_C\eta^{-1}\mu_X = \rho\mu_C\mu_X. \tag{3.1}$$

The autocovariance at lag τ is given by

$$\begin{aligned} c_Y(\tau) &= \operatorname{cov}\{Y(t), Y(t+\tau)\} \\ &= \int_0^\infty\int_0^\infty E\{X_{t-u}(u)\,X_{t+\tau-v}(v)\}\operatorname{cov}\{\mathrm{d}N(t-u), \mathrm{d}N(t+\tau-v)\}, \end{aligned} \tag{3.2}$$

where $\operatorname{cov}\{\mathrm{d}N(t_1), \mathrm{d}N(t_2)\}$ can be expressed (Cox & Isham 1980, §2.5) in terms of the conditional intensity function $h(\cdot)$ for the Neyman–Scott process as

$$c(u) = \operatorname{cov}\{\mathrm{d}N(t), \mathrm{d}N(t+u)\} = \lambda\mu_C\{\delta(u) + h(u) - \lambda\mu_C\}\,\mathrm{d}t\,\mathrm{d}u \tag{3.3}$$

and where δ denotes the Dirac delta function.

Now for a Neyman–Scott process, the conditional intensity function is given by

$$\begin{aligned} h(u) &= \lambda\mu_C + \mu_C^{-1}E\{C(C-1)\}\int_0^\infty \beta\,\mathrm{e}^{-\beta x}\,\mathrm{e}^{-\beta(x+u)}\,\mathrm{d}x \\ &= \lambda\mu_C + \tfrac{1}{2}\mu_C^{-1}E\{C(C-1)\}\,\beta\,\mathrm{e}^{-\beta u}, \end{aligned}$$

(Cox & Isham 1980, p. 78), so that

$$c(u) = \lambda\mu_C[\delta(u) + \tfrac{1}{2}\mu_C^{-1}E\{C(C-1)\}\,\beta\,\mathrm{e}^{-\beta u}].$$

Also,

$$X_{t-u}(u) = \begin{cases} X & \text{with probability} \quad \mathrm{e}^{-\eta u} \\ 0 & \text{with probability} \quad 1-\mathrm{e}^{-\eta u}. \end{cases}$$

Thus it follows from (3.2) that

$$c_Y(\tau) = \rho\{\mu_C\, E(X^2) + \tfrac{1}{2}\mu_X^2\, E(C^2-C)\,\beta^2/(\beta^2-\eta^2)\}\,\mathrm{e}^{-\eta\tau} - \tfrac{1}{2}\rho\mu_X^2\, E(C^2-C) \times \beta\eta\,\mathrm{e}^{-\beta\tau}/(\beta^2-\eta^2). \quad (3.4)$$

Setting $\tau = 0$ in (3.4), we obtain the variance of the depth process,

$$\operatorname{var}\{Y(t)\} = \rho\mu_C\, E(X^2) + \tfrac{1}{2}\rho\mu_X^2\, E(C^2-C)\,\beta/(\beta+\eta). \quad (3.5)$$

Note that if each storm is certain to consist of a single cell only, then these results reduce to those obtained for the Poisson model of §2.

We can now use the autocovariance function $c_Y(\tau)$ given in (3.4) to deduce the second order properties of the aggregated process $Y_i^{(h)}$, where $Y_i^{(h)}$ is the cumulative rainfall over an interval of length h. Specifically, with $Y_i^{(h)}$ defined as in (2.13),

$$\left.\begin{aligned} E(Y_i^{(h)}) &= \rho\mu_C\,\mu_X\, h, \\ \operatorname{var}(Y_i^{(h)}) &= \rho\eta^{-2}(\eta h - 1 + \mathrm{e}^{-\eta h})\{2\mu_C\, E(X^2) + E(C^2-C)\,\mu_X^2\,\beta^2/(\beta^2-\eta^2)\} \\ &\quad - \rho\eta(\beta h - 1 + \mathrm{e}^{-\beta h})\, E(C^2-C)\,\mu_X^2/\{\beta(\beta^2-\eta^2)\}, \end{aligned}\right\} \quad (3.6)$$

and for $k \geqslant 1$,

$$\left.\begin{aligned} &\operatorname{cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) = \\ &\rho\eta^{-2}(1-\mathrm{e}^{-\eta h})^2\,\mathrm{e}^{-\eta(k-1)h}\{\mu_C\, E(X^2) + \tfrac{1}{2}E(C^2-C)\,\mu_X^2\,\beta^2/(\beta^2-\eta^2)\} \\ &\quad - \rho\eta(1-\mathrm{e}^{-\beta h})^2\,\mathrm{e}^{-\beta(k-1)h}\,\tfrac{1}{2}E(C^2-C)\,\mu_X^2/\{\beta(\beta^2-\eta^2)\}. \end{aligned}\right\} \quad (3.7)$$

In this section, no assumptions have been made about the form of the distribution of C. One obvious possibility is to take C to have a geometric distribution, that is,

$$P(C=j) = \mu_C^{-1}(1-\mu_C^{-1})^{j-1} \quad (j = 1, 2, \ldots),$$

in which case $E(C^2-C) = 2\mu_C(\mu_C-1)$ in the above equations. An alternative distribution for C is the Poisson distribution. Because we have assumed that C is strictly positive we might take $C-1$ to have a Poisson distribution, in which case $E(C^2-C) = \mu_C(\mu_C+2)$; another possibility would be for C to have a truncated Poisson distribution so that C is forced to be positive. These choices have no effect on the general form of the autocovariance function.

4. Models based on a Bartlett–Lewis process

4.1. *Specification of the model*

In some ways a model based on a special form of the Bartlett–Lewis process is particularly accessible to mathematical analysis and may thus be the form of clustered process of most direct use in applications.

Suppose then that storm origins arise in a Poisson process of rate λ. Each origin is followed by a Poisson process of rate β of cell origins; after a time, exponentially distributed with rate γ, the process of cell origins terminates. We suppose without loss of generality that there is a cell associated with the storm origin. It is convenient to introduce the dimensionless parameters $\kappa = \beta/\eta$, $\phi = \gamma/\eta$. The number, C, of cells per storm thus has a geometric distribution of mean

$$\mu_C = 1+\kappa/\phi.$$

We suppose that rectangular pulses are associated with each cell origin as before; we take the duration of each cell to be exponentially distributed.

Note particularly that the storm may continue beyond the cell termination point and that there may be periods of zero intensity in the 'middle' of the storm.

It is possible to find the second-order properties of the process by evaluating the covariance density of the underlying point process and then using the integral representation (3.2). In fact, however, to obtain further properties, in particular the distribution of dry periods, we shall argue less directly.

4.2. *A single storm trajectory*

We begin by considering the trajectory of a single storm, starting from a storm origin at $t = 0$. Until the process of cell origins terminates, the storm will be called 'live', after this point some storms may continue to be active. The storm ends when all these cells are complete. Thus, for $r = 0, 1, \ldots$, we define

$$p_r(t) = P\{\text{storm live, } r \text{ cells active at time } t\},$$

$$q_r(t) = P\{\text{storm terminated, } r \text{ cells active at time } t\},$$

where

$$p_r(0) = \delta_{r1} = \begin{cases} 1 & r = 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$q_r(0) = 0.$$

These functions must satisfy the differential equations

$$\mathrm{d}p_r(t)/\mathrm{d}t = -(\kappa+\phi+r)\,\eta p_r(t) + (r+1)\,\eta p_{r+1}(t) + \kappa\eta p_{r-1}(t),$$

$$\mathrm{d}q_r(t)/\mathrm{d}t = -r\eta q_r(t) + \phi\eta p_r(t) + (r+1)\,\eta q_{r+1}(t),$$

for $r = 0, 1, \ldots$, where we define $p_{-1}(t) \equiv 0$. If we write

$$G_P^*(z, s) = \sum_{r=0}^{\infty} \int_0^{\infty} z^r\, \mathrm{e}^{-st}\, p_r(t)\, \mathrm{d}t,$$

$$G_Q^*(z, s) = \sum_{r=0}^{\infty} \int_0^{\infty} z_r\, \mathrm{e}^{-st}\, q_r(t)\, \mathrm{d}t,$$

then these generating functions satisfy the equations

$$\partial G_P^*(z, s)/\partial z - \{\kappa + (s+\phi\eta)\,\eta^{-1}\,(1-z)^{-1}\}\, G_P^*(z, s) + z\eta^{-1}(1-z)^{-1} = 0, \tag{4.1}$$

$$\partial G_Q^*(z, s)/\partial z - s\eta^{-1}(1-z)^{-1} G_Q^*(z, s) + \phi(1-z)^{-1} G_P^*(z, s) = 0, \tag{4.2}$$

with boundary condition $G_P^*(1, s) + G_Q^*(1, s) = s^{-1}$. Thus we find

$$G_P^*(z, s) = \eta^{-1}\, \mathrm{e}^{-\kappa(1-z)} \int_0^1 t^{\phi+s/\eta-1}(1-(1-z)\,t)\, \mathrm{e}^{\kappa(1-z)t}\, \mathrm{d}t \tag{4.3}$$

$$G_Q^*(z, s) = \eta^{-1} \int_0^1 \mathrm{d}v\, \phi v^{s/\eta-1}\, \mathrm{e}^{-\kappa(1-z)\,v} \int_0^1 \mathrm{d}t\, t^{\phi+s/\eta-1}\{1-(1-z)\,vt\}\, \mathrm{e}^{\kappa(1-z)\,vt}, \tag{4.4}$$

where

$$q_0^*(s) = G_Q^*(0,s) = \eta^{-1}\int_0^1 \mathrm{d}v\,\phi v^{s/\eta-1}\,\mathrm{e}^{-\kappa v}\int_0^1 \mathrm{d}t\,t^{\phi+s/\eta-1}(1-vt)\,\mathrm{e}^{\kappa vt} \tag{4.5}$$

is the Laplace transform of the probability, $q_0(t)$, that the storm is complete before time t.

We are now in a position to find some important properties of the single storm trajectory, that is, the time during which the storm is active. In particular, the expected duration, μ_{T} say, of the trajectory is given by

$$\begin{aligned}\mu_{\mathrm{T}} &= \sum_{r=0}^{\infty}\int_0^{\infty} p_r(t)\,\mathrm{d}t + \sum_{r=1}^{\infty}\int_0^{\infty} q_r(t)\,\mathrm{d}t = G_P^*(1,0)+G_Q^*(1,0)-q_0^*(0)\\ &= \phi\eta^{-1}\int_0^1 \mathrm{d}v\int_0^1 \mathrm{d}t\,v^{-1}t^{\phi-1}\{1-(1-vt)\,\mathrm{e}^{-\kappa v(1-t)}\}.\end{aligned} \tag{4.6}$$

This must be obtained numerically, but it is straightforward to expand μ_{T} on the assumption that both κ and ϕ are small, and to find that, to third order in κ and ϕ,

$$\mu_{\mathrm{T}} \approx (\phi\eta)^{-1}\{1+\phi(\kappa+\phi)-\tfrac{1}{4}\phi(\kappa+\phi)(\kappa+4\phi)+\tfrac{1}{72}\phi(\kappa+\phi)\ (4\kappa^2+27\kappa\phi+72\phi^2)\}.$$

Note that $(\phi\eta)^{-1}$ is simply the expected time until the process of cell origins terminates. The next terms in the expansion add higher-order corrections to allow for the effect of incomplete cells overlapping the termination point.

Suppose now that we consider a storm that is active at a particular instant in time and let R be the number of active cells at that instant. Then

$$P(R=r) = \mu_{\mathrm{T}}^{-1}\int_0^{\infty}\{p_r(t)+q_r(t)\}\,\mathrm{d}t \quad (r \geqslant 1),$$

$$P(R=0) = \mu_{\mathrm{T}}^{-1}\int_0^{\infty} p_0(t)\,\mathrm{d}t,$$

so that R has probability generating function

$$G_R(z) = \mu_{\mathrm{T}}^{-1}\lim_{s\to 0}\{G_P^*(z,s)+G_Q^*(z,s)-q_0^*(s)\}.$$

The moments of this distribution can be found by differentiating $G_R(z)$ and setting $z=1$, thus

$$\left.\begin{aligned} E(R) &= \mu_R = \frac{\kappa+\phi}{\mu_{\mathrm{T}}\eta\phi},\\ E(R^2) &= \frac{(\kappa+\phi)(\kappa+\phi+1)}{\mu_{\mathrm{T}}\eta\phi(\phi+1)},\\ E(R^3) &= \frac{\kappa^2\{2\phi(\phi+3)-\kappa(\phi-3)(\phi+2)\}}{\mu_{\mathrm{T}}\eta\phi(\phi+1)(\phi+2)(\phi+3)}+3E(R^2)-2E(R).\end{aligned}\right\} \tag{4.7}$$

4.3. *Superposition of storm trajectories*

Because the number of storms active at a particular instant must have a Poisson distribution, with mean $\lambda\mu_{\mathrm{T}}$, it follows that at such an instant the total number of active cells arising from all active storms has the probability generating function

$$\exp\{-\lambda\mu_{\mathrm{T}}(1-G_R(z))\}.$$

In particular, we find that the probability, π_0, that there is no rainfall at an arbitrary time instant is given by

$$\pi_0 = \exp\{-\lambda\mu_{\mathrm{T}}(1-G_R(0))\} = \exp\{-\lambda\mu_{\mathrm{T}}+\lambda G_P^*(0,0)\}, \tag{4.8}$$

where, from (4.3),
$$G_P^*(0,0) = \eta^{-1}\,\mathrm{e}^{-\kappa}\int_0^1 t^{\phi-1}(1-t)\,\mathrm{e}^{\kappa t}\,\mathrm{d}t,$$

which must be obtained numerically, but for small κ and ϕ can be approximated, to second order in κ and ϕ, by

$$G_P^*(0,0) \approx (\eta\phi)^{-1}\{1-\kappa-\phi+\tfrac{3}{2}\kappa\phi+\phi^2+\tfrac{1}{2}\kappa^2\}.$$

Thus,
$$\pi_0 \approx \exp\{-\rho(1+\kappa/\phi)(1-\tfrac{1}{2}\kappa)\}.$$

Table 3 is a short table of values of $\eta\phi G_P^*(0,0)$ from which the adequacy of approximations, the one above and corresponding ones for large κ and/or ϕ can be judged.

We are now in a position to deduce the distribution of the total depth, Y, at an arbitrary instant. For the Laplace transform of the density of Y is

$$E(\mathrm{e}^{-\zeta Y}) = \exp[-\lambda\mu_{\mathrm{T}}\{1-G_R(M(\zeta))\}],$$

where $M(\zeta) = E(\mathrm{e}^{-\zeta X})$ is the corresponding function for the depth, X, of a single cell. Perhaps the simplest way to obtain the first few moments of Y is by differentiating its cumulant generating function, $-\lambda\mu_{\mathrm{T}}\{1-G_R(M(\zeta))\}$. Thus,

$$\left.\begin{aligned} E(Y) &= \lambda\mu_{\mathrm{T}}\mu_R\mu_X,\\ \operatorname{var}(Y) &= \lambda\mu_{\mathrm{T}}E(R^2)\mu_X^2+\lambda\mu_{\mathrm{T}}\mu_R\sigma_X^2,\\ E[\{Y-E(Y)\}^3] &= \lambda\mu_{\mathrm{T}}\{E(R^3)-\mu_R\}\mu_X^3+3\lambda\mu_{\mathrm{T}}\{E(R^2)-\mu_R\}\mu_X\sigma_X^2+\lambda\mu_{\mathrm{T}}\mu_R E(X^3).\end{aligned}\right\} \tag{4.9}$$

In the special case when X is exponentially distributed, these moments simplify to

$$E(Y) = \lambda\mu_{\mathrm{T}}\mu_R\mu_X,$$

$$\operatorname{var}(Y) = \lambda\mu_{\mathrm{T}}\{E(R^2)+\mu_R\}\mu_X^2,$$

$$E[\{Y-E(Y)\}^3] = \lambda\mu_{\mathrm{T}}\{E(R^3)+3E(R^2)-1\}\mu_X^3.$$

Note that an alternative way of expressing $E(Y)$, mentioned in §3.2, is that

$$E(Y) = \lambda\mu_X\,E\,(\text{number of storms per cell})\,E\,(\text{cell length}),$$

which gives us another derivation of the result in (4.7) that

$$\mu_{\mathrm{T}}\mu_R = (\kappa+\phi)/(\eta\phi) = \eta^{-1}\mu_C.$$

TABLE 3. VALUES OF FUNCTION $\eta\phi G_P^*(0,0)$

κ	ϕ... 0.2	0.4	0.6	0.8	1.0	2.0	4.0
0.2	0.695	0.605	0.536	0.482	0.438	0.301	0.187
0.4	0.581	0.514	0.461	0.419	0.385	0.274	0.175
0.6	0.485	0.437	0.398	0.366	0.339	0.249	0.164
0.8	0.406	0.372	0.343	0.319	0.299	0.227	0.155
1.0	0.341	0.318	0.297	0.280	0.264	0.207	0.146
2.0	0.145	0.149	0.150	0.150	0.149	0.135	0.109
4.0	0.031	0.041	0.048	0.053	0.057	0.066	0.066

Suppose now that rather than considering the depth at a single instant, we look at the process over an arbitrary time interval $[0, h]$ of length h. In particular, what is the probability that the depth is zero throughout this interval? For this, a single storm which is active at $t = 0$, must be live but have no active cells at the beginning of the interval. The probability that no cells become active during the interval is then

$$\int_0^h \gamma\, e^{-\gamma u}\, e^{-\beta u}\, du + e^{-\gamma h}\, e^{-\beta h} = \{\gamma + \beta\, e^{-(\beta+\gamma)h}\}/(\beta+\gamma),$$

where these terms correspond to the storm terminating during and after the end of the interval respectively. Because the number of storms active at $t = 0$ has a Poisson distribution with mean $\lambda\mu_T$, and each such storm has no active cells at $t = 0$ with probability

$$\mu_T^{-1} \int_0^\infty p_0(t)\, dt = G_P^*(0,0)/\mu_T,$$

it follows that the probability of zero depth throughout $[0, h]$ is given by

$$\begin{aligned}\varpi_h &= e^{-\lambda h}\left\{\sum_{r=0}^{\infty} \frac{e^{-\lambda\mu_T}(\lambda\mu_T)^r}{r!}\left(\frac{G_P^*(0,0)}{\mu_T}\right)^r\left(\frac{\gamma+\beta\, e^{-(\beta+\gamma)h}}{\beta+\gamma}\right)^r\right\}\\ &= \exp\{-\lambda(h+\mu_T) + \lambda G_P^*(0,0)\,(\gamma+\beta\, e^{-(\beta+\gamma)h})/(\beta+\gamma)\}. \qquad (4.10)\end{aligned}$$

The probability ϖ_h is of course a quantity of interest when the rainfall data is aggregated over intervals of width h. Note that comparing (4.10) with (4.8), we have $\varpi_0 = \pi_0$ as expected.

It is now straightforward to obtain the distribution of the length, T_D, of a dry period. For if T_D has distribution function F_D with mean μ_D then the survivor function of the forward recurrence time from an arbitrary dry instant to the end of the dry period is well known to be

$$\mu_D^{-1} \int_h^\infty \{1 - F_D(x)\}\, dx.$$

In terms of the previous argument this survivor function is simply ϖ_h/π_0. Thus, from (4.10)

$$\mu_D^{-1}\{1 - F_D(h)\} = \{\lambda + \lambda\beta\, e^{-(\beta+\gamma)h}\, G_P^*(0,0)\}\,\varpi_h/\pi_0$$

so that, setting $h = 0$,

$$\mu_D^{-1} = \lambda\{1 + \beta G_P^*(0,0)\}$$

and, for small κ and ϕ

$$\mu_{\mathrm{D}} \approx \lambda^{-1}\{1+\kappa\phi^{-1}(1-\kappa-\phi+\tfrac{3}{2}\kappa\phi+\phi^2+\tfrac{1}{2}\kappa^2)\}^{-1}.$$

We can also deduce the mean length, μ_{W} say, of a wet period, which may occur either within or between storms. For, the probability, π_0, of zero depth at an arbitrary instant satisfies

$$\pi_0 = \mu_{\mathrm{D}}/(\mu_{\mathrm{D}}+\mu_{\mathrm{W}}),$$

so that $\mu_{\mathrm{W}} = \mu_{\mathrm{D}}(\pi_0^{-1}-1)$, where π_0 has already been derived in (4.8).

4.4. *The autocovariance structure*

So far, we have considered the properties of the Bartlett–Lewis model for the storm process at a single time instant. We now consider the autocovariance function, $c_Y(\tau)$, for the process and use the method applied to the Neyman–Scott model in §3.2. In particular, (3.2) and (3.3) apply with $\{N(t)\}$ counting occurrences of cell origins in the Bartlett–Lewis process and h being the corresponding conditional intensity function. Now, the conditional intensity function h in (3.3) is the limiting rate at which points occur in $(u, u+\mathrm{d}u)$, conditionally upon there being a point at the origin. For the Bartlett–Lewis process there are two possibilities. The cell origin in $(u, u+\mathrm{d}u)$ is either in the same storm as the cell origin at zero, or belongs to a different storm. In the former case the rate of cell origins is β, as long as the storm is live, and this has probability $\mathrm{e}^{-\gamma u}$. In the latter case the rate of cell origins is simply $\lambda\mu_C$. Thus

$$h(u) = \lambda\mu_C+\beta\,\mathrm{e}^{-\gamma u} \tag{4.11}$$

and, from (3.3)

$$\operatorname{cov}\{\mathrm{d}N(t),\, \mathrm{d}N(t+u)\} = \lambda\mu_C\{\delta(u)+\beta\,\mathrm{e}^{-\gamma u}\}\;\mathrm{d}t\,\mathrm{d}u.$$

Then, substituting this in (3.2) and using the fact that, as before,

$$X_{t-u}(u) = \begin{cases} X & \text{with probability} \quad \mathrm{e}^{-\eta u} \\ 0 & \text{with probability} \quad 1-\mathrm{e}^{-\eta u}, \end{cases}$$

we find that

$$c_Y(\tau) = \rho\mu_C\{E(X^2)+\kappa\phi(\phi^2-1)^{-1}\mu_X^2\}\mathrm{e}^{-\eta\tau}-\rho\mu_C\,\kappa(\phi^2-1)^{-1}\mu_X^2\,\mathrm{e}^{-\gamma\tau}. \tag{4.12}$$

Thus, the autocovariance is a sum of exponential terms, with decay parameters η and γ, which are the reciprocals of the mean cell length and mean interval between cells respectively.

If $\tau = 0$, then (4.12) becomes

$$\operatorname{var}\{Y(t)\} = c_Y(0) = \rho\mu_C\{E(X^2)+\kappa(\phi+1)^{-1}\mu_X^2\}, \tag{4.13}$$

which agrees with (4.9) when moments of R from (4.7) are substituted.

Note that the autocovariance function could also have been obtained from first principles. In outline, the argument is as follows. We can write

$$c_Y(\tau) = \operatorname{cov}\{Y(t),\, Y(t+\tau)\} = \lambda\int_0^\infty E\{H(u)\,H(u+\tau)\}\,\mathrm{d}u, \tag{4.14}$$

where $H(u)$ is the total depth of rainfall at time u, generated by a single storm originating at time zero. Clearly, $H(u)$ is a sum of contributions from $R(u)$ different cells, where $R(u)$ is the number of cells active a time u after the beginning of the storm. Thus

$$P\{R(u)=r\}=p_r(u)+q_r(u) \quad (r \geqslant 0),$$

where $p_r(u)$ and $q_r(u)$ were discussed in §4.2.

Then, given $R(u)=r$, the number S of these r cells which also contribute to $H(u+\tau)$ has a binomial distribution with index r and parameter $\mathrm{e}^{-\eta\tau}$. Therefore

$$E\{H(u)\,H(u+\tau)|R(u),R(u+\tau),S\} = R(u)\,R(u+\tau)\,\mu_X^2+S\sigma_X^2,$$

$$E\{H(u)\,H(u+\tau)\}=[\operatorname{cov}\{R(u),\ R(u+\tau)\}+E\{R(u)\}\,E\{R(u+\tau)\}]\,\mu_X^2 + E\{R(u)\}\,\mathrm{e}^{-\eta\tau}\,\sigma_X^2,$$

where $$\operatorname{cov}\{R(u),R(u+\tau)\}=\mathrm{e}^{-\eta\tau}\operatorname{var}\{R(u)\}.$$

We can now use the results of §4.2 to deduce the mean and variance of $R(u)$ and the mean of $R(u+\tau)$, and thus determine $c_Y(\tau)$.

4.5. *The aggregated process*

If, as before, we denote the cumulative rainfall over disjoint time intervals of a fixed length h by $Y_i^{(h)}$ then we can deduce the second-order properties of the aggregated process from (2.14). In particular

$$\left.\begin{aligned}
E(Y_i^{(h)}) &= h\lambda\mu_{\mathrm{T}}\mu_R\,\mu_X = h\rho\mu_C\,\mu_X,\\
\operatorname{var}(Y_i^{(h)}) &= 2\rho\mu_C\{E(X^2)+\beta\mu_X^2/\gamma\}\,h/\eta+2\rho\mu_C\,\mu_X^2\,\beta\eta(1-\mathrm{e}^{-\gamma h})/\{\gamma^2(\gamma^2-\eta^2)\}\\
&\quad -2\rho\mu_C\{E(X^2)+\beta\gamma\mu_X^2/(\gamma^2-\eta^2)\}(1-\mathrm{e}^{-\eta h})/\eta^2,\\
\operatorname{cov}(Y_i^{(h)},Y_{i+k}^{(h)}) &= \rho\mu_C\{E(X^2)+\beta\gamma\mu_X^2/(\gamma^2-\eta^2)\}(1-\mathrm{e}^{-\eta h})^2\,\mathrm{e}^{-\eta(k-1)h}/\eta^2\\
&\quad -\rho\mu_C\,\mu_X^2\,\beta\eta(1-\mathrm{e}^{-\gamma h})^2\,\mathrm{e}^{-\gamma(k-1)h}/\{\gamma^2(\gamma^2-\eta^2)\}.
\end{aligned}\right\} \tag{4.15}$$

4.6. *Empirical results*

Again only very brief numerical results will be given here. Some of the results from fitting the data of §2.7 are summarized in the sixth line of table 2. We give results only for the Bartlett–Lewis process; broadly similar although slightly less satisfactory results are available for the Neyman–Scott process and will be reported fully elsewhere.

The fit is based on the mean, variances and lag-one autocorrelations at $h=6$, 12 h with informal adjustment to achieve good agreement on all properties. The estimates are

$$\lambda=0.00796\ \mathrm{h}^{-1},\ \eta=1.7\ \mathrm{h}^{-1},\ \mu_X=2.99\ \mathrm{mm\ h}^{-1},\ \ \beta=0.6\ \mathrm{h}^{-1},\ \mu_C=6.33.$$

An important feature is that virtually identical parameter estimates are obtained on fitting at other levels of aggregation. The distribution of cell depth has again been taken as exponential. It seems likely that the discrepancies between observed and fitted probabilities of zero rain given in table 2, and corresponding departures in the extreme values, not reported here, could be resolved by allowing the cell depths to have a longer tailed distribution, such as a gamma distribution of negative index. This is under investigation.

We regard the most important feature, giving some justification for the extra complexity of the models and results of §§3 and 4, to be the ability to produce results over a range of time scales; this does, of course, remain to be confirmed with more extensive data.

5. Insertion of irregularity

All the models discussed above are constructed from rectangular pulses. The reasons for this are the need for economical parametrization and the desire for mathematical tractability. Note, however, that some results can be found for the second moment properties of $\{Y(t)\}$ when the pulse shapes are arbitrary, provided that the shapes attached to different origins are mutually independent.

One simple way of making the trajectories of rainfall more realistically irregular is by the introduction of a high-frequency 'jitter'. For this suppose that the rainfall process $\{\tilde{Y}(t)\}$ is related to one of the previous models $\{Y(t)\}$ by

$$\tilde{Y}(t) = \mathrm{e}^{Z(t)}\, Y(t). \tag{5.1}$$

Here $\{Z(t)\}$ is a stationary gaussian process of mean μ_Z variance σ_Z^2 and autocovariance function $c_Z(\tau)$. Further $\{Z(t)\}$ is independent of $\{Y(t)\}$. We choose the exponential form in (5.1) to ensure positivity and that $\tilde{Y}(t) = 0$ if and only if $Y(t) = 0$, so that properties of dry periods are unaffected. Sometimes it is convenient to take $\{Z(t)\}$ to be white noise.

We can, without loss of generality suppose that $E(\mathrm{e}^{Z(t)}) = 1$, so that $\mu_Z + \frac{1}{2}\sigma_Z^2 = 0$.

In the white-noise case there is thus just one extra parameter, namely σ_Z^2; if we allowed $\{Z(t)\}$ to have an exponential autocovariance function there would be two new parameters to be fitted.

Simple calculation with the properties of the log normal distribution shows that

$$\operatorname{var}\{\tilde{Y}(t)\} = \sigma_Y^2\, \mathrm{e}^{\sigma_Z^2} + \mu_Y^2(\mathrm{e}^{\sigma_Z^2} - 1), \tag{5.2}$$

$$c_{\tilde{Y}}(\tau) = c_Y(\tau)\exp\{c_Z(\tau) + \mu_Y^2[\exp\{c_Z(\tau)\} - 1]. \tag{5.3}$$

Note also that if σ_Z^2 is small we can write (5.2) and (5.3) in the form

$$\left.\begin{aligned} \operatorname{var}\{\tilde{Y}(t)\} &= \sigma_Y^2 + \sigma_Z^2(\mu_Y^2 + \sigma_Y^2) + O(\sigma_Z^4), \\ c_{\tilde{Y}}(\tau) &= c_Y(\tau) + c_Z(\tau)\{\mu_Y^2 + c_Y(\tau)\}. \end{aligned}\right\} \tag{5.4}$$

If $\{Z(t)\}$ is a white-noise process

$$c_{\tilde{Y}}(\tau) = c_Y(\tau) \quad (\tau \neq 0),$$

so that the effect of the jitter is to deflate the autocorrelations by a factor $\operatorname{var}\{Y(t)\}/\operatorname{var}\{\tilde{Y}(t)\}$.

The higher moments of $\tilde{Y}(t)$ can be evaluated directly from (5.1).

The effect on the aggregated process can be calculated from the general formula (2.14), numerical integration being needed in general. Note that if white-noise jitter is postulated, its effect disappears on aggregation, i.e. to get non-vacuous results we must start with non-degenerate $c_Z(\tau)$. Then, using the simplified form (5.4), we have that

$$\operatorname{var}(\tilde{Y}_i^{(h)}) = \operatorname{var}(Y_i^{(h)}) + 2\int_0^h (h-\tau)\, c_Z(\tau)\{c_Y(\tau)+\mu_Y^2\}\,\mathrm{d}\tau. \tag{5.5}$$

If $c_Z(\tau)$ decays to zero much more rapidly than $c_Y(\tau)$, we have approximately that

$$\operatorname{var}(\tilde{Y}_i^{(h)}) = \operatorname{var}(Y_i^{(h)}) + 2h(\mu_Y^2+\sigma_Y^2)\int_0^\infty c_Z(\tau)\,\mathrm{d}\tau. \tag{5.6}$$

Similarly

$$\operatorname{cov}(\tilde{Y}_i^{(h)}, \tilde{Y}_{i+k}^{(h)}) = \operatorname{cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) + \int_{-h}^h c_Z(kh+v)\{c_Y(kh+v)+\mu_Y^2\}(h-|v|)\,\mathrm{d}v \tag{5.7}$$

and under the assumptions leading to the approximation (5.5) the second term is negligible for $k > 1$, whereas for $k = 1$

$$\operatorname{cov}(\tilde{Y}_i^{(h)}, \tilde{Y}_{i+1}^{(h)}) = \operatorname{cov}(Y_i^{(h)}, Y_{i+1}^{(h)}) + h(\mu_Y^2+\sigma_Y^2)\int_0^\infty c_Z(\tau)\,\mathrm{d}\tau. \tag{5.8}$$

Thus in this situation the effect of jitter is to deflate all autocorrelations except the first by the same factor.

In the above discussion, the jitter has been applied to the total $\{Y(t)\}$. It is in some ways more natural to apply independent jitters to the component pulse depths X that make up the total. In this case (2.4) must be replaced by

$$X_t(u) = \begin{cases} X\exp\{Z_t(u)\} & \text{with probability } \mathscr{F}_L(u) \\ 0 & \text{with probability } 1-\mathscr{F}_L(u), \end{cases}$$

where $\{Z_t(u)\}$ is the jitter process applied to a cell with origin at t. The effect of this on the cluster-based models of §§3 and 4 is straightforward to obtain. In particular, for the covariance at lag τ the expressions (3.4) and (3.5) for the Neyman–Scott model and (4.12) and (4.13) for the Bartlett–Lewis model are modified by replacing $E(X^2)$ by

$$E(X^2)\,E(\mathrm{e}^{Z(u)+Z(u+\tau)}) = E(X^2)\,\mathrm{e}^{c_Z(\tau)};$$

as before we assume

$$E(\mathrm{e}^{Z(u)}) = \mathrm{e}^{\mu_Z+\frac{1}{2}\sigma_Z^2} = 1.$$

If σ_Z^2 is small then the inclusion of the factor $\mathrm{e}^{c_Z(\tau)}$ is equivalent, ignoring terms of $O(\sigma_Z^4)$, to adding the term

$$\rho\mu_Z\,E(X^2)\,\mathrm{e}^{-\eta\tau}\,c_Z(\tau)$$

to the covariances (3.4) and (4.12), and to adding $\rho\mu_C\,E(X^2)\,\sigma_Z^2$ to the variances (3.5) and (4.13).

If the $\{Z_t(\cdot)\}$ are all white-noise processes then the autocovariances (3.4) and (4.12) are unchanged, for $\tau \neq 0$, so that the autocorrelations for the cluster models with jitter will be smaller than those without jitter.

Note also that for the Bartlett–Lewis model the expression given in §4.3 for $E(\mathrm{e}^{-sY})$ still applies when the cell depths are subject to jitter if the function $M(\zeta)$ for the depth is taken to be $E\{\exp(-\zeta X\mathrm{e}^Z)\}$.

Properties for the aggregated process with either model can be deduced from the general equations (2.14) if the modifications to the autocovariance functions described above are made, and if a particular form for the autocovariance function $c_Z(\cdot)$ of the jitter process is assumed. If σ_Z^2 is small then the effect of the jitter with either model is to add

$$2\rho\mu_C E(X^2)\int_0^h (h-u)\,\mathrm{e}^{-\eta u}\,c_Z(u)\,\mathrm{d}u+O(\sigma_Z^4)$$

to the expressions for var $(Y_i^{(h)})$ given in (3.6) and (4.15), and to add

$$\rho\mu_C E(X^2)\int_{-h}^h (h-|u|)\,\mathrm{e}^{-\eta(kh+u)}\,c_Z(kh+u)\,\mathrm{d}u+O(\sigma_Z^4)$$

to the expressions in (3.7) and (4.15) for cov $(Y_i^{(h)}, Y_{i+k}^{(h)})$.

6. Concluding remarks

In this paper we have outlined some of the main properties of some relatively simple stochastic models of rainfall at a single site. The direction of further theoretical developments hinges in part on the outcome of more extensive analyses of empirical data. Some of the details in the present paper could certainly be elaborated, but more promising developments are probably the inclusion of seasonal effects and, very particularly, the incorporation of spatial aspects into the models.

We are grateful to B. Febres de Power for her computations connected with the empirical example. D.R.C. was supported by the Science and Engineering Research Council.

References

Cox, D. R. & Isham, V. 1980 *Point processes*. London: Chapman and Hall.
Cox, D. R. & Miller, H. D. 1965 *Theory of stochastic processes*. London: Chapman and Hall.
LeCam, L. 1961 In *Proc. 4th Berkeley Symp.* **3**, 165–186.
Rodriguez-Iturbe, I., Gupta, V. K. & Waymire, E. 1984 *Wat. Resour. Res.* **20**, 1611–1619.
Stern, R. D. & Coe, R. 1984 *Jl R. statist. Soc.* A**147**, 1–34.
Taylor, G. I. 1921 *Proc. Lond. math. Soc.* **20**, 196–211.
Vanmarcke, E. 1983 *Random fields: analysis and synthesis*. Cambridge: Massachusetts: MIT Press.
Waymire, E. & Gupta, V. K. 1981 *Wat. Resour. Res.* **17**, 1261–1272.