Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme

Improved methods for national spatial-temporal rainfall and evaporatoin modelling for BSM: Appendices

R&D Project Record FD2105/PR

Environment Agency

defra
Department for Environment Food and Rural Affairs

Appendices

Appendix A Moment-based inference for stochastic-mechanistic models

A.1 Introduction

Within this project, a substantial portion of the research is devoted to the development and application of models for rainfall based on point processes. These models are 'stochastic-mechanistic', in the sense that they attempt to provide a simplified stochastic representation of the mechanics of the rainfall process. They are parameterised in terms of physically interpretable quantities (e.g. storm arrival rate and mean cell intensity). However, estimation of the parameters is difficult, mainly because they are related rather indirectly to observable properties of rainfall sequences. Likelihood-based inference is generally infeasible, owing to the difficulty in formulating a likelihood function (this is a consequence of the complex dependencies induced by the model specification). Moreover, it has been argued by Rodriguez-Iturbe et al. (1988, Section 3) that likelihood-based inference is not necessarily appropriate in any case, because the models are necessarily over-simplified so that the joint distribution of an observed rainfall sequence differs substantially in some respects from that implied by the models. For example, the rectangular profile of rain cells in a single-site model leads to short-term deterministic features in model realisations; these are not present in real rainfall. This argument is to some extent supported by experience with the 'spectral likelihood' approach, which attempts to formulate an approximate Gaussian likelihood based on collections of sample Fourier coefficients. This approximate likelihood only involves the second-order moment properties of the data (mean, variance and autocorrelations); models fitted using this method are very good at reproducing these properties of observed rainfall sequences, but poor when it comes to other properties of interest such as lengths of dry intervals. Informally, the problem is that the likelihood method tries too hard to achieve a good match between model and data at very short timescales, whereas in practice this is not to be expected.

In the absence of a suitable likelihood-based approach, stochastic-mechanistic models are usually fitted using a generalised method of moments: select a set of properties of interest (e.g. mean, variance, autocorrelations and proportion of 'dry' intervals at various levels of aggregation) and choose parameter estimates that minimise some measure of discrepancy between model and data with respect to these properties. This measure is usually a (possibly weighted) sum of squared differences. A particular advantage of this approach is that the model parameterisation can be chosen to reproduce, as closely as possible, those properties that are deemed to be particularly important in any specific application. However, a major disadvantage (compared with, say, a likelihood-based approach) is that assessments of uncertainty are not readily available. This note is an attempt to summarise the available options for obtaining uncertainty estimates (e.g. confidence intervals) when model parameters are estimated using a generalised method of moments. The problem can be regarded as an application of the theory of estimating equations; the relevant aspects are summarised in Section A.3 below. Before this, however, we review some standard theory of likelihood-based inference, by way of illustrating the general concepts. In Section A.4, we present the moment-based estimation procedure within the estimating equation framework; and Section A.5 provides a concise summary along with some practical suggestions for implementation.

A.2 Inference in a likelihood-based setting

The relevant results from likelihood-based inference are most easily illustrated in the context of a problem in which a vector of observations $\mathbf{y} = (y_1 \ \ldots \ y_n)'$ has been generated from a joint probability distribution whose density has the form $f(\mathbf{y};\theta)$. The functional form of f is known but the exact value of θ is not. The LIKELIHOOD FUNCTION for θ given the data $\mathbf{y}$ is defined as

$$L(\theta|\mathbf{y}) = f(\mathbf{y};\theta) \ ,$$

and can be interpreted as the probability of obtaining the observed data for any given value of θ. The MAXIMUM LIKELIHOOD ESTIMATE (MLE) of θ is the value, $\hat{\theta}$ say, for which the likelihood function is

maximised (i.e. the value that allocates the highest probability to the observations). Equivalently, it is the value for which the log-likelihood

$$\ell(\theta|\mathbf{y}) = \ln L(\theta|\mathbf{y})$$

is maximised. In well-behaved problems, the MLE therefore satisfies the equation

$$U\left(\hat{\theta}|\mathbf{y}\right) = 0\,, \tag{A.1}$$

where $U(\theta|\mathbf{y}) = \partial\ell(\theta|\mathbf{y})/\partial\theta$ is the SCORE FUNCTION. We assume here that (A.1) has a unique solution. Note that

$$U(\theta|\mathbf{y}) = \frac{\partial \ln f(\mathbf{y};\theta)}{\partial\theta} = \frac{1}{f(\mathbf{y};\theta)}\frac{\partial f(\mathbf{y};\theta)}{\partial\theta}\,. \tag{A.2}$$

A.2.1 Properties of the score

Since $\mathbf{y}$ has been generated from a probability distribution, it can be regarded as the realised value of a vector $\mathbf{Y}$ of random variables. Hence the score function $U(\theta|\mathbf{y})$ is the realised value of a random variable $U_\theta = U(\theta|\mathbf{Y})$. The properties of this random variable depend on the true value of θ; call this θ_0. For example, we have

$$\mathsf{E}(U_\theta) = \int U(\theta|\mathbf{y})\, f(\mathbf{y};\theta_0)\, d\mathbf{y} = \int \frac{1}{f(\mathbf{y};\theta)}\frac{\partial f(\mathbf{y};\theta)}{\partial\theta} f(\mathbf{y};\theta_0)\, d\mathbf{y}\,,$$

the last step following from (A.2). This expression is valid for all θ. In particular, the expected score at the *true* parameter value is

$$\mathsf{E}\left(U_{\theta_0}\right) = \int \frac{1}{f(\mathbf{y};\theta_0)}\left.\frac{\partial f(\mathbf{y};\theta)}{\partial\theta}\right|_{\theta=\theta_0} f(\mathbf{y};\theta_0)\, d\mathbf{y} = \int \left.\frac{\partial f(\mathbf{y};\theta)}{\partial\theta}\right|_{\theta=\theta_0} d\mathbf{y}. \tag{A.3}$$

Now, since $f(\mathbf{y};\theta)$ is a probability density for all values of θ, we have

$$\int f(\mathbf{y};\theta)d\mathbf{y} = 1 \qquad \text{so that} \qquad \frac{\partial}{\partial\theta}\int f(\mathbf{y};\theta)d\mathbf{y} = 0\,.$$

In well-behaved problems we can interchange the order of differentiation and integration, to yield

$$\int \frac{\partial f(\mathbf{y};\theta)}{\partial\theta} d\mathbf{y} = 0\,.$$

This identity holds for all values of θ, and in particular for $\theta = \theta_0$. Hence, from (A.3), we have

$$\mathsf{E}\left(U_{\theta_0}\right) = 0\,. \tag{A.4}$$

We now turn to the variance of the score function. This can be related to the expected value of its derivative — or equivalently, of the second derivative of the log-likelihood. For, differentiating (A.2) with respect to θ, we obtain

$$\begin{aligned}\frac{\partial U(\theta|\mathbf{y})}{\partial\theta} = \frac{\partial^2\ell(\theta|\mathbf{y})}{\partial\theta^2} &= -\frac{1}{f^2(\mathbf{y};\theta)}\left(\frac{\partial f(\mathbf{y};\theta)}{\partial\theta}\right)^2 + \frac{1}{f(\mathbf{y};\theta)}\frac{\partial^2 f(\mathbf{y};\theta)}{\partial\theta^2} \\ &= -U^2(\theta|\mathbf{y}) + \frac{1}{f(\mathbf{y};\theta)}\frac{\partial^2 f(\mathbf{y};\theta)}{\partial\theta^2}\,.\end{aligned} \tag{A.5}$$

As before, all of these quantities are the realised values of random variables, so we can consider replacing $\mathbf{y}$ with $\mathbf{Y}$ and taking expectations. In particular, let $H_\theta = \partial^2\ell(\theta|\mathbf{Y})/\partial\theta^2$. Then we obtain

$$\mathsf{E}(H_\theta) = -\mathsf{E}\left[U_\theta^2\right] + \int \frac{1}{f(\mathbf{y};\theta)}\frac{\partial^2 f(\mathbf{y};\theta)}{\partial\theta^2} f(\mathbf{y};\theta_0)\, d\mathbf{y}\,.$$

Evaluated at $\theta = \theta_0$, the last term here is zero and we obtain

$$\mathsf{E}\left[U_{\theta_0}^2\right] = -\mathsf{E}\left(H_{\theta_0}\right) = I(\theta_0)\ \text{, say.}$$

But since $\mathsf{E}\left(U_{\theta_0}\right) = 0$, we must have $\mathsf{E}\left[U_{\theta_0}^2\right] = \mathsf{var}\left(U_{\theta_0}\right)$. Thus we have shown that

$$\mathsf{var}\left(U_{\theta_0}\right) = I(\theta_0)\,. \tag{A.6}$$

$I(\theta_0)$ is called the (FISHER) INFORMATION.

A.2.2 Large-sample properties of the MLE

The properties of the score function are fundamental to the development of asymptotic results for maximum likelihood estimators — in particular, to the construction of standard errors and confidence intervals for the parameters. The theory relies on the fact that, in well-behaved problems, as the sample size n tends to infinity the following two things happen:

1. The score function U_θ tends, when suitably normalised, to its expectation. For example, if the random variables in $\mathbf{Y}$ are independent and identically distributed (iid) then the log-likelihood is a sum of n independent contributions; it follows that the score function is also a sum of n independent contributions, and the law of large numbers dictates that $n^{-1}\left[U_\theta - \mathrm{E}\left(U_\theta\right)\right] \to 0$ as $n \to \infty$ in this case.

2. The distribution of $\left[U_\theta - \mathrm{E}\left(U_\theta\right)\right] / \sqrt{I(\theta)}$ tends to the standard normal distribution. Again, in the iid case this is easy to see: U_θ is a sum of independent terms, and the normality follows from the Central Limit Theorem.

Providing U_θ is continuous in θ, property 1 here implies that for large n, the score equation (A.1) has a solution in the neighbourhood of θ_0 (since, from (A.4), $\mathrm{E}\left(U_{\theta_0}\right) = 0$), and that this solution tends to θ_0 as $n \to \infty$. Hence, providing n is large enough, $\left|\hat{\theta} - \theta_0\right|$ will be small so that we can carry out a Taylor Series expansion for the score function in the neighbourhood of θ_0 and write

$$U_{\hat{\theta}} \approx U_{\theta_0} + \left(\hat{\theta} - \theta_0\right) H_{\theta_0} \tag{A.7}$$

(recall that H_θ is the second derivative of the log-likelihood at θ). But by definition, $U_{\hat{\theta}} = 0$, so that

$$\hat{\theta} - \theta_0 \approx -\frac{U_{\theta_0}}{H_{\theta_0}} = -\frac{U_{\theta_0}}{I(\theta_0)} \frac{I(\theta_0)}{H_{\theta_0}} \; .$$

As $n \to \infty$, H_θ tends to its expectation which is $-I(\theta)$, so that the second factor on the right-hand side here tends to -1. Hence we can approximate the estimation error $\hat{\theta} - \theta_0$ by $U_{\theta_0}/I(\theta_0)$. Strictly speaking, some care needs to be taken over the relative magnitudes of the various approximations here — for full details, see Cox and Hinkley (1974, Section 9.2).

Having expressed the estimation error in terms of the score, we can use property 2 above to deduce that for large samples, the estimation error has an approximate normal distribution. Specifically,

$$\sqrt{I(\theta_0)}\left(\hat{\theta} - \theta_0\right) \approx \frac{U_{\theta_0}}{\sqrt{I(\theta_0)}} = \frac{U_{\theta_0} - \mathrm{E}\left[U_{\theta_0}\right]}{\sqrt{I(\theta_0)}} \sim N(0,1) \; , \tag{A.8}$$

since $\mathrm{E}\left[U_{\theta_0}\right] = 0$. For practical purposes, an equivalent statement of this result is that for large n, the distribution of the MLE is approximately normal with mean θ_0 and variance $1/I(\theta_0)$. This can be used, for example, to construct approximate confidence intervals for θ_0: an approximate 95% interval is

$$\hat{\theta} \pm \frac{1.96}{\sqrt{I(\theta_0)}} \; . \tag{A.9}$$

Hypothesis tests based on (A.8) are referred to as WALD TESTS. As an alternative, inference could be based directly on the quantity $\left[U_\theta - \mathrm{E}\left(U_\theta\right)\right] / \sqrt{I(\theta)}$ at the right-hand side of (A.8), to yield a SCORE TEST. In general, the results from Wald and score tests will differ slightly due to the first approximation in (A.8).

A third possibility is to base inference on the log-likelihood function itself. A second-order Taylor expansion about the MLE yields, for some $\theta^\dagger$ between θ_0 and $\hat{\theta}$,

$$\begin{aligned} \ell(\theta_0|\mathbf{Y}) &= \ell\left(\hat{\theta}|\mathbf{Y}\right) + \left(\theta_0 - \hat{\theta}\right) \left.\frac{\partial \ell}{\partial \theta}\right|_{\theta=\hat{\theta}} + \frac{1}{2}\left(\theta_0 - \hat{\theta}\right)^2 \left.\frac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta^\dagger} \\ &= \ell\left(\hat{\theta}|\mathbf{Y}\right) + \frac{1}{2}\left(\theta_0 - \hat{\theta}\right)^2 H_{\theta^\dagger} \; , \end{aligned} \tag{A.10}$$

as the first derivative of the log-likelihood is zero at $\hat{\theta}$ by definition. Since $\theta^\dagger$ is between θ_0 and $\hat{\theta}$ we have $H_{\theta^\dagger} \approx H_{\hat{\theta}}$. Moreover, again using the fact that $H_\theta \approx -I(\theta)$ for large n, we find

$$2\left[\ell\left(\hat{\theta}|\mathbf{Y}\right) - \ell\left(\theta_0|\mathbf{Y}\right)\right] \approx \left(\theta_0 - \hat{\theta}\right)^2 I\left(\hat{\theta}\right) = \left[\left(\theta_0 - \hat{\theta}\right)\sqrt{I\left(\hat{\theta}\right)}\right]^2 . \tag{A.11}$$

But from (A.8), the right hand side here is just the square of a standard normal random variable, and therefore has a chi-squared distribution with 1 degree of freedom. The asymptotic approximation

$$2\left[\ell\left(\hat{\theta}|\mathbf{Y}\right) - \ell\left(\theta_0|\mathbf{Y}\right)\right] \sim \chi^2_1 \tag{A.12}$$

can therefore be used to test hypotheses and construct confidence intervals. For example, a 95% confidence interval for θ consists of all values for which $2\left[\ell\left(\hat{\theta}|\mathbf{Y}\right) - \ell\left(\theta|\mathbf{Y}\right)\right]$ is less than the upper 95% point of a χ^2_1 distribution (which is 3.84).

Hypothesis tests based on (A.12) may be referred to as LIKELIHOOD RATIO TESTS. The three test procedures (Wald, score and likelihood ratio) are asymptotically equivalent, in the sense that their results will be very similar for large enough sample sizes. However, since the likelihood ratio test is derived from approximation (A.11) rather than from (A.7), its results will usually differ slightly, in finite samples, from both the Wald and score tests. There are grounds for preferring $2\left[\ell\left(\hat{\theta}|\mathbf{Y}\right) - \ell\left(\theta|\mathbf{Y}\right)\right]$ as a test statistic, although the accuracy of the χ^2 approximation is not guaranteed in finite samples.

A.2.3 More than one parameter

The theory above carries over straightforwardly to the case when there is more than one parameter. Specifically, denote the unknown parameter vector by $\theta = (\theta_1 \ \ldots \ \theta_p)'$. Then the log-likelihood for θ can be defined as previously, and the MLE satisfies the system of score equations

$$U_j\left(\hat{\theta}|\mathbf{y}\right) = 0 \qquad (j = 1, \ldots, p)$$

where now $U_j(\theta|\mathbf{y}) = \partial\ell(\theta|\mathbf{y})/\partial\theta_j$. These p equations can be written in vector form as

$$\mathbf{U}\left(\hat{\theta}|\mathbf{y}\right) = \mathbf{0} . \tag{A.13}$$

$\mathbf{U}(\theta|\mathbf{y})$ is the SCORE VECTOR, and can be regarded as the realised value of a vector of random variables $\mathbf{U}_\theta$. Denoting the true parameter by θ_0, we can show that

$$\mathrm{E}\left(\mathbf{U}_{\theta_0}\right) = \mathbf{0} \qquad \text{and} \qquad \mathrm{var}\left(\mathbf{U}_{\theta_0}\right) = \mathbf{I}(\theta_0) = -\mathrm{E}\left(\mathbf{H}_{\theta_0}\right) , \tag{A.14}$$

where $\mathbf{H}_\theta$ is the Hessian matrix of second derivatives of the log-likelihood at θ. For large samples, $\mathbf{U}_\theta$ again approaches its expectation and has an approximate normal distribution (this time in p dimensions):

$$\mathbf{U}_{\theta_0} \sim MVN\left(\mathbf{0}, \mathbf{I}(\theta_0)\right) . \tag{A.15}$$

The Taylor expansion corresponding to (A.7) is now

$$\mathbf{U}_{\hat{\theta}} \approx \mathbf{U}_{\theta_0} + \mathbf{H}_{\theta_0}\left(\hat{\theta} - \theta_0\right) , \tag{A.16}$$

so that $\hat{\theta} - \theta_0 \approx -\mathbf{H}^{-1}{}_{\theta_0}\mathbf{U}_{\theta_0} \approx \mathbf{I}^{-1}(\theta_0)\mathbf{U}_{\theta_0}$. Hence $\mathrm{E}\left(\hat{\theta} - \theta_0\right) \approx 0$ and $\mathrm{var}\left(\hat{\theta} - \theta_0\right) \approx \mathbf{I}^{-1}(\theta_0)\,\mathrm{var}\left(\mathbf{U}_{\theta_0}\right)\mathbf{I}^{-1}(\theta_0) = \mathbf{I}^{-1}(\theta_0)$. For large samples we therefore have, approximately,

$$\hat{\theta} \sim MVN\left(\theta_0, \mathbf{I}^{-1}(\theta_0)\right) . \tag{A.17}$$

In the multiparameter case, the equivalent of (A.10) is

$$\ell(\theta_0|\mathbf{Y}) = \ell\left(\hat{\theta}|\mathbf{Y}\right) + \frac{1}{2}\left(\theta_0 - \hat{\theta}\right)'\mathbf{H}_{\theta^\dagger}\left(\theta_0 - \hat{\theta}\right)$$

for some $\theta^\dagger$ between θ_0 and $\hat{\theta}$. Now for large n, the elements of the matrix $\mathbf{H}_{\theta^\dagger} - \mathrm{E}\left[\mathbf{H}_{\theta^\dagger}\right]$ are order $n^{1/2}$ in probability. Also, since $\left(\theta_0 - \theta^\dagger\right)$ is order $n^{-1/2}$, the elements of $\mathrm{E}\left[\mathbf{H}_{\theta^\dagger}\right] - \mathrm{E}\left[\mathbf{H}_{\theta_0}\right]$ are themselves $\left(o_p\left(n^{1/2}\right)\right)$. Therefore we can write $\mathbf{H}_{\theta^\dagger} = -\mathbf{I}(\theta_0) + \mathbf{E}$, where the elements of $\mathbf{E}$ are $O_p\left(n^{1/2}\right)$. Hence

$$\begin{aligned} 2\left[\ell\left(\hat{\theta}|\mathbf{Y}\right) - \ell\left(\theta_0|\mathbf{Y}\right)\right] &= \left(\theta_0 - \hat{\theta}\right)'\mathbf{I}(\theta_0)\left(\theta_0 - \hat{\theta}\right) + O_p\left(n^{-1/2}\right) \\ &= \left[\mathbf{A}(\theta_0)\left(\theta_0 - \hat{\theta}\right)\right]'\mathbf{A}(\theta_0)\left(\theta_0 - \hat{\theta}\right) + O_p(n^{-1/2}) , \end{aligned} \tag{A.18}$$

where $\mathbf{A}(\theta_0)$ is a matrix such that $\mathbf{A}'(\theta_0)\mathbf{A}(\theta_0) = \mathbf{I}(\theta_0)$. $\mathbf{A}(\theta_0)$ is not uniquely defined but could be, for example, the Cholesky square root of $\mathbf{I}(\theta_0)$ (which is guaranteed to exist since $\mathbf{I}(\theta_0)$ is a covariance matrix and is therefore positive definite). Now, since $\hat{\theta} \sim MVN\left(\theta_0, \mathbf{I}^{-1}(\theta_0)\right)$, we must have

$$\mathbf{A}(\theta_0)\left(\theta_0 - \hat{\theta}\right) \sim MVN\left(\mathbf{0}, \mathbf{A}(\theta_0)\mathbf{I}^{-1}(\theta_0)\mathbf{A}'(\theta_0)\right)$$

approximately. Now, assuming the information matrix $\mathbf{I}(\theta_0)$ is nonsingular, we must have

$$\begin{aligned}\mathbf{A}(\theta_0)\mathbf{I}^{-1}(\theta_0)\mathbf{A}'(\theta_0) &= \mathbf{A}(\theta_0)\left[\mathbf{A}'(\theta_0)\mathbf{A}(\theta_0)\right]^{-1}\mathbf{A}(\theta_0)' \\ &= \mathbf{A}(\theta_0)\left[\mathbf{A}(\theta_0)\right]^{-1}\left[\mathbf{A}'(\theta_0)\right]^{-1}\mathbf{A}(\theta_0)' = \mathbf{1}_{p\times p}\,,\end{aligned}$$

the $p \times p$ identity matrix. Together with (A.18), this shows that $2\left[\ell\left(\hat{\theta}|\mathbf{Y}\right) - \ell(\theta_0|\mathbf{Y})\right]$ is approximately a sum of squares of p standard normal random variables. Asymptotically therefore,

$$2\left[\ell\left(\hat{\theta}|\mathbf{Y}\right) - \ell(\theta_0|\mathbf{Y})\right] \sim \chi^2_p\,. \tag{A.19}$$

This is the multiparameter equivalent of (A.12).

Either (A.15) or (A.17) can be used to construct confidence intervals for individual parameters, as well as confidence *regions* for subsets of the parameters. Moreover, (A.19) allows the construction of a confidence region for the entire parameter vector — for example, an approximate 95% confidence region for θ consists of all values such that $2\left[\ell\left(\hat{\theta}|\mathbf{Y}\right) - \ell(\theta|\mathbf{Y})\right]$ is less than the 95% point of χ^2_p. As it stands however, (A.19) does not allow the construction of confidence regions for subsets of the parameter vector. We now address this problem.

A.2.4 Profile likelihood

Suppose now that the parameter vector is partitioned into two subsets: $\theta = \left(\psi'\ \lambda'\right)'$, with target value $\theta_0 = \left(\psi_0'\ \lambda_0'\right)'$. Write $\ell(\theta) = \ell(\psi, \lambda)$ for the log-likelihood,

$$\mathbf{U}(\theta) = \begin{pmatrix} \mathbf{U}_\psi(\psi,\lambda) \\ \mathbf{U}_\lambda(\psi,\lambda) \end{pmatrix} \quad \text{for the score vector,}$$

$$\mathbf{I} = \begin{pmatrix} \mathbf{I}_{\psi\psi} & \mathbf{I}_{\psi\lambda} \\ \mathbf{I}_{\lambda\psi} & \mathbf{I}_{\lambda\lambda} \end{pmatrix} \text{ for } \mathrm{var}\left[\mathbf{U}(\theta_0)\right], \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_{\psi\psi} & \mathbf{H}_{\psi\lambda} \\ \mathbf{H}_{\lambda\psi} & \mathbf{H}_{\lambda\lambda} \end{pmatrix} \text{ for } \mathrm{E}\left[\partial^2\ell/\partial\theta^2_{|\theta=\theta_0}\right].$$

Suppose also that ψ is held fixed, and that the likelihood is maximised with respect to λ alone for this value of ψ. In general, the resulting estimate of λ will depend on ψ, so call it $\hat{\lambda}(\psi)$. The value of the resulting maximised likelihood, $\ell\left(\psi, \hat{\lambda}(\psi)\right)$ will also depend on ψ; this is called the PROFILE LIKELIHOOD for ψ.

Let $\hat{\psi}$ be the overall MLE for ψ; then the overall MLE for λ is $\hat{\lambda}(\hat{\psi})$. By definition, $\ell\left(\hat{\psi}, \hat{\lambda}(\hat{\psi})\right)$ cannot be less than the maximised log-likelihood at any other value of ψ. Therefore the likelihood ratio statistic

$$\Lambda(\psi) = 2\left[\ell\left(\hat{\psi}, \hat{\lambda}(\hat{\psi})\right) - \ell\left(\psi, \hat{\lambda}(\psi)\right)\right] \tag{A.20}$$

is always positive-valued, although we would expect $\Lambda(\psi_0)$ to be 'small' in general, if $\hat{\psi}$ is close to ψ_0. This suggests that when ψ is unknown, a confidence region could be determined as the set of values for which $\Lambda(\psi)$ is less than some threshold — or equivalently, as the set of values for which the profile likelihood exceeds a corresponding threshold. An appropriate threshold can be determined by considering the distribution of $\Lambda(\psi_0)$. We have

$$\Lambda(\psi_0) = 2\left\{\left[\ell\left(\hat{\psi}, \hat{\lambda}(\hat{\psi})\right) - \ell(\psi_0, \lambda_0)\right] - \left[\ell\left(\psi_0, \hat{\lambda}(\psi_0)\right) - \ell(\psi_0, \lambda_0)\right]\right\}\,. \tag{A.21}$$

Now, using essentially the same argument as that given in the previous section we find that the term

$2\left[\ell\left(\hat{\psi},\hat{\lambda}(\hat{\psi})\right)-\ell(\psi_0,\lambda_0)\right]$ can be written as

$$\begin{aligned}
& -(\hat{\theta}-\theta_0)'\mathbf{H}(\hat{\theta}-\theta_0)+O_p\left(n^{-1/2}\right) \\
= & -\left((\hat{\psi}-\psi_0)' \quad \left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)'\right)'\begin{pmatrix}\mathbf{H}_{\psi\psi} & \mathbf{H}_{\psi\lambda}\\ \mathbf{H}_{\lambda\psi} & \mathbf{H}_{\lambda\lambda}\end{pmatrix}\begin{pmatrix}\hat{\psi}-\psi_0\\ \hat{\lambda}(\hat{\psi})-\lambda_0\end{pmatrix}+O_p\left(n^{-1/2}\right) \\
= & -(\hat{\psi}-\psi_0)'\mathbf{H}_{\psi\psi}(\hat{\psi}-\psi_0)-\left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)'\mathbf{H}_{\lambda\psi}(\hat{\psi}-\psi_0) \\
& -(\hat{\psi}-\psi_0)'\mathbf{H}_{\psi\lambda}\left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)-\left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)'\mathbf{H}_{\lambda\lambda}\left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)+O_p\left(n^{-1/2}\right)\ .
\end{aligned} \tag{A.22}$$

For the second term in (A.21), the analysis can repeated as though ψ_0 is known and λ is the unknown parameter vector, to yield

$$2\left[\ell\left(\psi_0,\hat{\lambda}(\psi_0)\right)-\ell(\psi_0,\lambda_0)\right]=-\left(\hat{\lambda}(\psi_0)-\lambda_0\right)'\mathbf{H}_{\lambda\lambda}\left(\hat{\lambda}(\psi_0)-\lambda_0\right)+O_p\left(n^{-1/2}\right)\ . \tag{A.23}$$

We now substitute (A.22) and (A.23) into (A.21). This requires a relationship between $\hat{\lambda}(\hat{\psi})$ and $\hat{\lambda}(\psi_0)$. To find this we use (A.16), which we now write as

$$\begin{pmatrix}\mathbf{U}_\psi\left(\hat{\psi},\hat{\lambda}(\hat{\psi})\right)\\ \mathbf{U}_\lambda\left(\hat{\psi},\hat{\lambda}(\hat{\psi})\right)\end{pmatrix}=\begin{pmatrix}\mathbf{U}_\psi(\psi_0,\lambda_0)\\ \mathbf{U}_\lambda(\psi_0,\lambda_0)\end{pmatrix}+\begin{pmatrix}\mathbf{H}_{\psi\psi} & \mathbf{H}_{\psi\lambda}\\ \mathbf{H}_{\lambda\psi} & \mathbf{H}_{\lambda\lambda}\end{pmatrix}\begin{pmatrix}\hat{\psi}-\psi_0\\ \hat{\lambda}(\hat{\psi})-\lambda_0\end{pmatrix}+o_p\left(n^{1/2}\right)\ . \tag{A.24}$$

If ψ_0 is known so that only λ is being estimated, the corresponding expansion is

$$\mathbf{U}_\lambda\left(\psi_0,\hat{\lambda}(\psi_0)\right)=\mathbf{U}_\lambda(\psi_0,\lambda_0)+\mathbf{H}_{\lambda\lambda}\left(\hat{\lambda}(\psi_0)-\lambda_0\right)+o_p\left(n^{1/2}\right)\ . \tag{A.25}$$

Since the left hand sides of both (A.24) and (A.25) are zero by definition, we can equate (A.25) with the bottom row of (A.24) to obtain

$$\mathbf{H}_{\lambda\lambda}\left(\hat{\lambda}(\psi_0)-\lambda_0\right)=\mathbf{H}_{\lambda\psi}(\hat{\psi}-\psi_0)+\mathbf{H}_{\lambda\lambda}\left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)+o_p\left(n^{1/2}\right)\ ,$$

so that

$$\hat{\lambda}(\psi_0)-\lambda_0=\hat{\lambda}(\hat{\psi})-\lambda_0+\mathbf{H}_{\lambda\lambda}^{-1}\mathbf{H}_{\lambda\psi}(\hat{\psi}-\psi_0)+o_p\left(n^{1/2}\right)\ ,$$

the order of magnitude of the error following from the fact that the elements of $\mathbf{H}_{\lambda\lambda}$ are $O_p(n)$. The quadratic term in (A.23) can now be written as

$$\begin{aligned}
& -\left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)'\mathbf{H}_{\lambda\lambda}\left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)-\left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)'\mathbf{H}_{\lambda\psi}(\hat{\psi}-\psi_0) \\
& -(\hat{\psi}-\psi_0)'\mathbf{H}_{\psi\lambda}\left(\hat{\lambda}(\hat{\psi})-\lambda_0\right)-(\hat{\psi}-\psi_0)'\mathbf{H}_{\psi\lambda}\mathbf{H}_{\lambda\lambda}^{-1}\mathbf{H}_{\lambda\psi}(\hat{\psi}-\psi_0)+o_p(1)\ .
\end{aligned} \tag{A.26}$$

We now combine (A.21), (A.22), (A.23) and (A.26), and find

$$\begin{aligned}
\Lambda(\psi_0) &= -(\hat{\psi}-\psi_0)'\mathbf{H}_{\psi\psi}(\hat{\psi}-\psi_0)+(\hat{\psi}-\psi_0)'\mathbf{H}_{\psi\lambda}\mathbf{H}_{\lambda\lambda}^{-1}\mathbf{H}_{\lambda\psi}(\hat{\psi}-\psi_0)+o_p(1) \\
&= -(\hat{\psi}-\psi_0)'\left[\mathbf{H}_{\psi\psi}-\mathbf{H}_{\psi\lambda}\mathbf{H}_{\lambda\lambda}^{-1}\mathbf{H}_{\lambda\psi}\right](\hat{\psi}-\psi_0)+o_p(1)\ .
\end{aligned} \tag{A.27}$$

Next, observe that $\left[\mathbf{H}_{\psi\psi}-\mathbf{H}_{\psi\lambda}\mathbf{H}_{\lambda\lambda}^{-1}\mathbf{H}_{\lambda\psi}\right]^{-1}$ is the submatrix of $\mathbf{H}^{-1}$ corresponding to ψ (this is not immediately obvious, but is a standard result in matrix algebra — see, for example, Horn and Johnson 1985, page 18). For notational convenience therefore, if we write $\mathbf{H}^{-1}$ as

$$\mathbf{H}^{-1}=\begin{pmatrix}\mathbf{H}^{(\psi\psi)} & \mathbf{H}^{(\psi\lambda)}\\ \mathbf{H}^{(\lambda\psi)} & \mathbf{H}^{(\lambda\lambda)}\end{pmatrix}\ ,$$

the likelihood ratio statistic (A.27) can be written as

$$\Lambda(\psi_0)=-(\hat{\psi}-\psi_0)'\left[\mathbf{H}^{(\psi\psi)}\right]^{-1}(\hat{\psi}-\psi_0)+o_p(1)\ . \tag{A.28}$$

But from (A.17), the distribution of $\hat{\theta}$ is approximately $MVN\left(\theta_0, \mathbf{I}^{-1}\right) = MVN\left(\theta_0, -\mathbf{H}^{-1}\right)$, so that the distribution of $\hat{\psi} - \psi_0$ is approximately $MVN\left(\mathbf{0}, -\mathbf{H}^{(\psi\psi)}\right)$. The development from (A.18) to (A.19) can now be repeated, to conclude that in large samples

$$\Lambda(\psi_0) \sim \chi^2_q \tag{A.29}$$

approximately, where q is the dimension of ψ. Therefore, if ψ_0 is unknown, a confidence region can be determined as the set $\{\psi : \Lambda(\psi) < c\}$, where c is the appropriate percentile of the χ^2_q distribution.

A.3 Estimating equations

In the likelihood setting above, the asymptotic results depend on the following properties of the score function:

1. The expected score is zero at the true parameter value.
2. The variance of the score can be calculated at the true parameter value.
3. The score, when suitably normalised, tends to its expectation as $n \to \infty$.
4. The score is a continuous function of the parameter, in the neighbourhood of the true parameter value.
5. As $n \to \infty$, the distribution of the score vector tends to the multivariate normal.
6. The second derivative of the score is bounded in the neighbourhood of the true parameter value (this was not made explicit in the discussion above, but is necessary to control the magnitudes of the various approximations leading to (A.8) and (A.17)).

These observations suggest that, as an alternative to likelihood-based inference, we may consider obtaining parameters by solving the equation

$$\mathbf{g}\left(\hat{\theta}|\mathbf{y}\right) = \mathbf{0}\,, \tag{A.30}$$

where $\mathbf{g}$ is a function such that the associated vector of random variables $\mathbf{g}_\theta$ has properties 1–6 above. Any such equation is called an ESTIMATING EQUATION; we will call $\mathbf{g}_\theta$ an ESTIMATING FUNCTION. If $\hat{\theta}$ solves an estimating equation of the form (A.30), then the arguments of the previous section can be repeated. Let $\mathbf{H}(\theta)$ denote the expected value of the Hessian matrix $\mathbf{H}_\theta = \partial \mathbf{g}_\theta / \partial \theta$, and let $\mathbf{J}(\theta)$ be the covariance matrix of $\mathbf{g}_\theta$. Then for large n,

$$\hat{\theta} \sim MVN(\theta_0, \mathbf{V}(\theta_0)) \text{ approximately,} \tag{A.31}$$

where $\mathbf{V}(\theta) = [\mathbf{H}(\theta)]^{-1}\mathbf{J}(\theta)\mathbf{H}(\theta)^{-1}$.

Many numerical optimisation methods (for example, those based around Newton-Raphson iterative schemes) compute the Hessian $\mathbf{H}_{\hat{\theta}}$ as a by-product of the optimisation procedure. We may therefore use this to approximate $\mathbf{H}(\theta_0)$ in the calculation of $\mathbf{V}(\theta_0)$. Further simplification is possible if we can choose $\mathbf{g}$ in such a way that $\mathbf{H}(\theta) = -\mathbf{J}(\theta)$ (from (A.14), this is the case for score-based estimation), since in this case $\mathbf{V}(\theta) = -[\mathbf{H}(\theta)]^{-1}$.

As in the case of likelihood-based inference, tests of hypotheses can be based either on (A.31) (which is the equivalent of a Wald test) or on the multivariate normal distribution of the estimating function itself (the equivalent of a score test):

$$\mathbf{g}_{\theta_0} \sim MVN(\mathbf{0}, \mathbf{J}(\theta_0))\ . \tag{A.32}$$

Much of the literature on estimating equations takes (A.30) as its starting point. In this case, the resulting estimate is not necessarily the maximiser (or minimiser) of a function such as the log-likelihood, whence there is no obvious equivalent to the likelihood ratio test. However, a generalisation *is* possible if $\mathbf{g}_\theta$ is the gradient vector of some objective function. Specifically, suppose that the estimating equations arise from minimising a measure of discrepancy between data and model,

$S(\theta|\mathbf{y})$ say. In this case, a confidence region could be defined as the set of all values for which $S(\theta|\mathbf{y})$ is less than some threshold. We now establish what this threshold should be.

A second-order Taylor expansion of S about $\hat{\theta}$ yields

$$2\left[S(\theta_0|\mathbf{Y}) - S\left(\hat{\theta}|\mathbf{Y}\right)\right] \quad = \quad \left(\theta_0 - \hat{\theta}\right)' \mathbf{H}(\theta_0)\left(\theta_0 - \hat{\theta}\right) + o_p(1).$$

This is dominated by the first term, which is a quadratic form in normal random variables. Distributions of quadratic forms are difficult to compute exactly. However, it is common (e.g. Bowman and Azzalini 1997, p.88) to approximate their quantiles with those of a scaled and shifted χ^2 distribution. The shift, scale and degrees of freedom of the approximating distribution are chosen to match the first three moments (or equivalently, cumulants) of the quadratic form. The rth cumulant is given (Kuonen, 1999, Section 2) by

$$\kappa_r = 2^{r-1}\Gamma(r)\mathrm{tr}\left\{\left[\mathbf{V}(\theta_0)\mathbf{H}(\theta_0)\right]^r\right\}, \tag{A.33}$$

with $\mathrm{tr}()$ denoting the trace operator. The distribution of the quadratic form is then approximated by that of $aX + c$, where $X \sim \chi^2_b$ and

$$a = \frac{|\kappa_3|}{4\kappa_2} \qquad b = \frac{8\kappa_2^3}{\kappa_3^2} \qquad c = \kappa_1 - ab\,. \tag{A.34}$$

In practice, it is necessary to replace θ_0 in (A.33) with $\hat{\theta}$. Since $\mathbf{V}\left(\hat{\theta}\right) = \left[\mathbf{H}\left(\hat{\theta}\right)\right]^{-1}\mathbf{J}\left(\hat{\theta}\right)\left[\mathbf{H}\left(\hat{\theta}\right)\right]^{-1}$, we therefore compute

$$\kappa_r = 2^{r-1}\Gamma(r)\mathrm{tr}\left\{\left[\mathbf{H}^{-1}\left(\hat{\theta}\right)\mathbf{J}\left(\hat{\theta}\right)\right]^r\right\}, \tag{A.35}$$

This then yields a reasonably straightforward procedure for constructing confidence regions based on the values of the objective function. For example, a 95% region consists of the set of values for which

$$a^{-1}\left\{2\left[S(\theta_0|\mathbf{Y}) - S\left(\hat{\theta}|\mathbf{Y}\right)\right] - c\right\}$$

is less than the 95th percentile of the χ^2_b distribution.

Confidence regions for subsets of the parameters can be constructed in a similar way, following the theory outlined in Section A.2.4. Specifically, suppose a profile objective function is calculated for a subvector ψ, by holding this subvector fixed and maximising over the remaining parameters λ, say. With notation as in Section A.2.4, define a profile test statistic as

$$\Lambda(\psi) = 2\left[S\left(\psi, \hat{\lambda}(\psi)|\mathbf{Y}\right) - S\left(\hat{\psi}, \hat{\lambda}(\hat{\psi})|\mathbf{Y}\right)\right], \tag{A.36}$$

which is positive-valued by definition. Expansion (A.28) holds here, as in the likelihood setting; the difference is, once again, that for estimating equations the covariance matrix of $\hat{\psi}$ is not directly related to the Hessian. It should be clear, however, that the same procedure can be applied as in the case of the full parameter vector above. All that is required is to replace $\mathbf{H}\left(\hat{\theta}\right)$ with the estimated value of $\left[\mathbf{H}^{(\psi\psi)}\right]^{-1}$, and to extract the submatrix of $\mathbf{V}\left(\hat{\theta}\right)$ corresponding to ψ. In fact, once the full matrices $\mathbf{H}^{-1}$ and $\mathbf{V}$ have been estimated, the modification simply consists of extracting the elements corresponding to ψ from each of these matrices when calculating the κs in (A.33). Notice that, in general, different choices of ψ will lead to different thresholds. Notice also that the cancellation leading to (A.35) does not hold in general, when considering subsets of the parameter vector.

If ψ consists of a single parameter ψ, the procedure outlined above is particularly simple. In this case $\left[\mathbf{H}^{(\psi\psi)}\right]^{-1}$ is a scalar, as is the corresponding submatrix of $\mathbf{V}\left(\hat{\theta}\right)$. Denote these scalars by h and v respectively; then direct calculation shows that the constants defined in (A.34) are given by $a = hv$, $b = 1$ and $c = 0$. A confidence interval for ψ can therefore be defined as the set values for which the profile test statistic (A.36) is less than hv times the appropriate percentage point of a χ^2_1 distribution.

There is a substantial body of theory on the use of estimating equations. However, for current purposes there is no need to go beyond what has been presented above.

A.4 Estimating equations for the method of moments

We now return to the problem of parameter estimation for stochastic-mechanistic models, using a generalised method of moments. To formalise the problem, it is necessary to establish some notation. Specifically:

- Let $\mathbf{y}$ be a vector of observations as previously; this is regarded as the realised value of a vector $\mathbf{Y}$ of random variables.
- Let $\theta = (\theta_1 \ \ldots \ \theta_p)'$ be a vector of unknown parameters in the model.
- Let $\mathbf{T}(\mathbf{y}) = (T_1(\mathbf{y}) \ \ldots \ T_k(\mathbf{y}))'$ be a vector of summary statistics computed from the observations. $\mathbf{T}(\mathbf{y})$ is the realised value of a random vector $\mathbf{T} = (T_1 \ \ldots \ T_k)$ say. Denote the expected value of this random vector by $\mathsf{E}_\theta(\mathbf{T}) = \tau(\theta) = (\tau_1(\theta) \ \ldots \ \tau_k(\theta))'$.

The idea here is that $\mathbf{T}$ is a vector of data properties (means, variances, autocorrelations etc.) and that $\tau(\theta)$ is the corresponding set of theoretical properties derived from the model. The generalised method of moments seeks to minimise some measure of disagreement between $\mathbf{T}$ and $\tau(\theta)$. Following the notation above, denote this measure by $S(\theta|\mathbf{y})$. In practice, this is invariably a (possibly weighted) sum of squares:

$$S(\theta|\mathbf{y}) = \sum_{i=1}^{k} w_i(\theta)\left[T_i(\mathbf{y}) - \tau_i(\theta)\right]^2 \ . \tag{A.37}$$

for some collection of positive weights $\{w_i(\theta) : i = 1, \ldots, k\}$. For the moment, we allow the possibility that these may be parameter-dependent, although we will see below that this is actually a bad idea. In well-behaved problems, the minimiser of this function satisfies the vector equation $\mathbf{g}(\theta|\mathbf{y}) = \partial S/\partial\theta = \mathbf{0}$. We have

$$\mathbf{g}(\theta|\mathbf{y}) = \sum_{i=1}^{k}\left\{\frac{\partial w_i(\theta)}{\partial\theta}\left[T_i(\mathbf{y}) - \tau_i(\theta)\right]^2 - 2w_i(\theta)\frac{\partial\tau_i(\theta)}{\partial\theta}\left[T_i(\mathbf{y}) - \tau_i(\theta)\right]\right\} \ , \tag{A.38}$$

so that the parameter estimate $\hat{\theta}$ satisfies $\mathbf{g}\left(\hat{\theta}|\mathbf{y}\right) = \mathbf{0}$, as in (A.30).

To apply the theory of estimating equations here, we need to ensure that properties 1–6 in Section A.3 are satisfied by the random variables $\mathbf{g}_\theta$ whose values are given by (A.38). Properties 4 and 6 ($\mathbf{g}_\theta$ is continuous in θ with bounded second derivatives) are unlikely to cause problems. The remainder require some thought.

A.4.1 Zero mean

For $\mathbf{g}_\theta$ to have zero mean, we require

$$\sum_{i=1}^{k}\left\{\frac{\partial w_i(\theta)}{\partial\theta}\mathsf{E}_\theta\left[T_i - \tau_i(\theta)\right]^2 - 2w_i(\theta)\frac{\partial\tau_i(\theta)}{\partial\theta}\mathsf{E}_\theta\left[T_i - \tau_i(\theta)\right]\right\} \ = \ \mathbf{0}$$

at $\theta = \theta_0$. Since $E(\mathbf{T}) = \tau(\theta_0)$, this reduces to the requirement that

$$\sum_{i=1}^{k}\frac{\partial w_i(\theta)}{\partial\theta}\mathsf{var}_\theta(T_i) = \mathbf{0} \ ,$$

which is trivially true providing the weights are independent of θ. If the weights depend on θ, however, the requirement is not fulfilled in general. In particular, it is not fulfilled if $w_i(\theta)$ is set proportional to $1/\mathsf{var}_\theta(T_i)$ (which is a natural weighting scheme to consider, given the received wisdom that 'in least squares problems with unequal variances, observations should be weighted according to the inverse of their variances', and that in such problems, the weighted least squares estimates are known to be unbiased). To see this, consider any collection of weights satisfying

$$\sum_{i=1}^{k} w_i(\theta)\,\mathsf{var}_\theta(T_i) = \text{constant, independent of } \theta.$$

Differentiating both sides with respect to θ_j yields

$$\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} \text{var}(T_i) + \sum_{i=1}^{k} w_i(\theta) \frac{\partial \text{var}_\theta(T_i)}{\partial \theta_j} = 0$$

in which case, the first term can only be zero if the second is also. But since the ws are positive, the second term can only be zero if $\text{var}_\theta(T_i)$ is independent of θ_j for each i. Also, since each element of $\mathbf{g}_\theta$ must have zero expectation, the result must hold for all j: hence no collection of weights satisfying the constraint above will yield a valid estimating equation, unless $\text{var}_\theta(T_i)$ is independent of θ for each i.

At first sight, this appears to contradict the 'standard' theory of weighted least squares in regression problems. The resolution of the problem appears to lie in the fact that in regression problems, the weights do not depend on the regression parameters (which are the θs in the present context) — hence $\partial w_i/\partial\theta = 0$ in such problems. I suspect that the difficulty, when the weights depend on θ, is related to the known problems of bias in estimating equations when nuisance parameters are present (Liang and Zeger, 1995), although the current setting is slightly different.

The upshot of all this is that if we want to weight the fitting properties, the weights should not depend on θ; otherwise the resulting estimates will be biased (at least, for the kind of weighting scheme that may be considered in practice). In many situations, it is likely that the bias will tend to zero as the sample size (i.e. dimension of $\mathbf{Y}$) increases. However, as a first step in obtaining sampling distributions for moment-based estimators, it seems reasonable to restrict ourselves to estimators that are exactly unbiased. Hence the objective function (A.37) becomes

$$S(\theta|\mathbf{y}) = \sum_{i=1}^{k} w_i [T_i(\mathbf{y}) - \tau_i(\theta)]^2 \tag{A.39}$$

and the corresponding estimating equation becomes

$$\mathbf{g}(\theta|\mathbf{y}) = -2 \sum_{i=1}^{k} w_i \frac{\partial \tau_i(\theta)}{\partial \theta} [T_i(\mathbf{y}) - \tau_i(\theta)] = \mathbf{0}\,. \tag{A.40}$$

The factor of -2 is retained here to avoid confusion later on.

θ-dependent weights — a cunning plan

The problems above, regarding the use of weights depending on θ, can be resolved completely if we modify the objective function (A.37) slightly, to

$$\sum_{i=1}^{k} \left\{ w_i(\theta) [T_i(\mathbf{y}) - \tau_i(\theta)]^2 - \ln w_i(\theta) \right\}\,. \tag{A.41}$$

If we do this, the estimating function becomes

$$\mathbf{g}_\theta = \sum_{i=1}^{k} \left\{ \frac{\partial w_i(\theta)}{\partial \theta} \left[[T_i - \tau_i(\theta)]^2 - \frac{1}{w_i(\theta)} \right] - 2 w_i(\theta) \frac{\partial \tau_i(\theta)}{\partial \theta} [T_i - \tau_i(\theta)] \right\} \;=\; \mathbf{0}\,,$$

which clearly has zero expectation if we set $w_i(\theta) = 1/\text{var}_\theta(T_i)$. We do not pursue this any further here; however, it may be prove useful in the future.

A.4.2 Asymptotic normality

From (A.40), it is clear that the estimating function $\mathbf{g}_\theta$ will have an approximate normal distribution if either of the following conditions hold:

1. k is large, and the components of $\mathbf{T}$ are not too strongly dependent. For in this case, $\mathbf{g}_\theta$ is a sum of a large number of terms and the Central Limit Theorem applies.
2. $\mathbf{T}$ itself has an approximate multivariate normal distribution.

In practice, providing the elements of $\mathbf{T}$ are chosen appropriately, condition 2 is likely to be satisfied for large datasets, since most statistics of interest have an approximate normal distribution in the limit. Obviously, the closer this approximation, the better will be the normal approximation to the distribution of $\mathbf{g}_\theta$. This suggests that we should seek fitting properties with distributions that are 'as normal as possible', for example by transformation.

A.4.3 Consistency

The moment estimator will be consistent if $\mathrm{E}\left[\mathbf{g}_{\theta_0}\right] = \mathbf{0}$ and, when suitably normalised, $\mathbf{g}_\theta$ converges in probability to its expectation as $n \to \infty$. Again from (A.40), this convergence will occur if $\mathbf{T}$ converges to $\tau(\theta)$; and again, most statistics of interest *do* converge to their expectations in the required sense.

A.4.4 Variance calculation

To complete the estimating equation framework it is necessary to calculate, or at least estimate, $\mathbf{J}(\theta_0) = \mathrm{var}\left[\mathbf{g}_{\theta_0}\right]$, since this is required for the calculation of $\mathbf{V}(\theta_0)$ in (A.31) and (A.33). A number of options are available here:

1. Find an analytical expression for $\mathbf{J}(\theta)$, and use $\mathbf{J}\left(\hat{\theta}\right)$ as an estimate of $\mathbf{J}(\theta_0)$.
2. Obtain an empirical estimate of $\mathbf{J}\left(\hat{\theta}\right)$, and use this to estimate $\mathbf{J}(\theta_0)$.
3. If possible, set up the estimating equation in such a way that $\mathbf{J}(\theta_0) \propto \mathbf{H}(\theta_0)$. In this case, $\mathbf{V}(\theta_0) \propto [\mathbf{H}(\theta)]^{-1}$ and we can use the observed Hessian to estimate $\mathbf{V}(\theta_0)$ without ever needing to calculate $\mathbf{J}(\theta_0)$.

For the first two options, it may be useful to note that $\mathbf{g}_\theta$ can be written in matrix form as

$$\mathbf{g}_\theta = -2\left[\mathbf{W}(\theta)\right]'\left(\mathbf{T} - \tau(\theta)\right) ,$$

where $\mathbf{W}(\theta)$ is a $k \times p$ matrix whose (i,j)th element is $w_i \partial\tau_i(\theta)/\partial\theta_j$. Standard results for covariance matrices then give us

$$\mathbf{J}(\theta) = 4\mathrm{var}\left\{\left[\mathbf{W}(\theta)\right]'\left(\mathbf{T} - \tau(\theta)\right)\right\} = 4\left[\mathbf{W}(\theta)\right]'\mathrm{var}(\mathbf{T})\,\mathbf{W}(\theta) . \tag{A.42}$$

Hence $\mathbf{J}(\theta)$ can be calculated from the covariance matrix of $\mathbf{T}$. A specific suggestion for estimating this covariance matrix empirically is given in Section A.5 below. In practice, the derivatives of $\tau(\theta)$ appearing in $\mathbf{W}(\theta)$ can be evaluated numerically if necessary.

Variance calculation using the Hessian

In the third option above, the idea is to define the objective function in such a way that $\mathbf{J}(\theta_0) \propto \mathbf{H}(\theta_0) = \mathrm{E}[\partial\mathbf{g}_\theta/\partial\theta]$. We now investigate how to achieve this. The starting point is the zero-mean requirement for the estimating function, which implies that

$$\int \mathbf{g}(\theta|\mathbf{y})\, f(\mathbf{y};\theta)\, d\mathbf{y} = \mathbf{0} .$$

Differentiating both sides with respect to θ yields

$$\int \left[\frac{\partial\mathbf{g}(\theta|\mathbf{y})}{\partial\theta} f(\mathbf{y};\theta) + \mathbf{g}(\theta|\mathbf{y})\left(\frac{\partial f(\mathbf{y};\theta)}{\partial\theta}\right)'\right] d\mathbf{y} = \mathbf{0} ,$$

so that

$$\begin{aligned} \mathrm{E}\left[\frac{\partial\mathbf{g}_\theta}{\partial\theta}\right] &= -\int \mathbf{g}(\theta|\mathbf{y})\left(\frac{1}{f(\mathbf{y};\theta)}\frac{\partial f(\mathbf{y};\theta)}{\partial\theta}\right)' f(\mathbf{y};\theta)\, d\mathbf{y} \\ \text{i.e.} \quad \mathrm{E}\left[\mathbf{H}_\theta\right] &= -\mathrm{E}\left[\mathbf{g}_\theta\left(\frac{\partial \ln f(\mathbf{Y};\theta)}{\partial\theta}\right)'\right] . \end{aligned}$$

At θ_0, we require the right-hand side here to be proportional to $\mathbf{J}(\theta) = \text{var}[\mathbf{g}_\theta] = \text{E}[\mathbf{g}_\theta \mathbf{g}_\theta']$. It is not obvious that this is the case, unless $\mathbf{g}(\theta|\mathbf{y})$ is proportional to the score function

$$\mathbf{g}(\theta|\mathbf{y}) = \frac{\partial \ln f(\mathbf{y};\theta)}{\partial \theta} .$$

However, note that $\mathbf{g}_\theta$ is a function of $\mathbf{T}$, which in turn is a function of $\mathbf{Y}$. Hence we can obtain an equivalent development to the one above, by taking expectations with respect to $\mathbf{T}$; the only difference is that $f(\mathbf{y};\theta)$ will be replaced by the density of $\mathbf{T}$, $f_{\mathbf{T}}$ say, throughout.

This shows that $\mathbf{J}(\theta_0) \propto \mathbf{H}(\theta_0)$, as required, if $\mathbf{g}(\theta|\mathbf{y})$ is the θ-derivative of the log density for $\mathbf{T}$. But in Section A.4.2 above we argued that for large samples, $\mathbf{T}$ is likely to have an approximate multivariate normal distribution. Suppose, for the sake of argument, that we can choose $\mathbf{T}$ in such a way that (i) its elements are mutually uncorrelated (ii) $\text{var}[T_i]$ is independent of θ for each i. In this case the θ-derivative of the log density has jth element

$$\frac{\partial \ln f_{\mathbf{T}}}{\partial \theta_j} = \sum_{i=1}^{n} \frac{1}{\text{var}[T_i]} \frac{\partial \tau_i(\theta)}{\partial \theta} [T_i(\mathbf{y}) - \tau_i(\theta)]$$

which, if we take the weight $w_i = 1/\text{var}[T_i]$, is equal to $-\frac{1}{2}\mathbf{g}(\theta|\mathbf{y})$ in (A.40). In this case, therefore, $\text{var}\left[\partial \ln f_{\mathbf{T}}/\partial\theta\right] = \frac{1}{4}\text{var}[\mathbf{g}_\theta]$, and $\text{E}\left[\partial^2 \ln f_{\mathbf{T}}/\partial\theta^2\right] = -\frac{1}{2}\text{E}[\partial \mathbf{g}_\theta/\partial\theta]$. Since the left-hand sides here differ by a factor of -1, we must have $\mathbf{J}(\theta) = \text{var}[\mathbf{g}_\theta] = 2\text{E}[\partial \mathbf{g}_\theta/\partial\theta] = 2\mathbf{H}(\theta)$.

Of course, it is unrealistic to expect that the elements of $\mathbf{T}$ should be uncorrelated and that their variances should be independent of θ. The argument above does suggest, however, that if we choose $\mathbf{T}$ in such a way that as many components as possible have variances that are independent of θ; and to set the weights for the remaining components to a 'ballpark' figure that roughly reflects their uncertainty, $\mathbf{J}(\theta_0)$ should be approximated reasonably by $2\mathbf{H}(\theta_0)$ so that $\mathbf{V}(\theta_0)$ can be calculated as $2[\mathbf{H}(\theta_0)]^{-1}$.

In passing, it is also worth noting that θ-dependent weights can be accommodated within this framework, by changing the objective function to (A.41). In this case, the resulting estimating function is exactly the θ-derivative of a normal density for uncorrelated Ts.

A.5 Summary, and implications

The main points to emerge from the discussion above are the following:

1. Using a generalised method of moments, unbiased estimators can be obtained by minimising an expression of the form
$$S(\theta|\mathbf{Y}) = \sum_{i=1}^{k} w_i [T_i(\mathbf{Y}) - \tau_i(\theta)]^2$$
where the Ts are properties of the data and the τs are their expected values under the model.

2. The weights $\{w_i\}$ must *not* depend on the model parameters (or on the data!). If parameter-dependent weights are used, the objective function must be modified to that given in (A.41).

3. Under fairly general conditions, the estimator resulting from the above minimisation has a multivariate normal distribution. This can be used, for example, to construct approximate confidence intervals for the model parameters. The mean of the distribution is θ_0 (the true parameter vector), and its covariance matrix is $\mathbf{V}(\theta_0)$ where $\mathbf{V}(\theta) = [\mathbf{H}(\theta)]^{-1}\mathbf{J}(\theta)\mathbf{H}(\theta)^{-1}$. Here, $\mathbf{H}(\theta)$ is the expected second derivative of the objective function, which can be estimated from the Hessian output of a numerical minimisation routine. $\mathbf{J}(\theta)$ is the covariance matrix of the objective function derivatives.

4. An alternative way to construct confidence regions uses the objective function itself. Specifically, an approximate confidence region at a specified level consists of all points θ such that
$$a^{-1}\left\{2\left[S(\theta|\mathbf{Y}) - S\left(\hat{\theta}|\mathbf{Y}\right)\right] - c\right\}$$

is less than the appropriate percentile of a chi-squared distribution with b degrees of freedom. The constants a, b and c are given by

$$a = \frac{|\kappa_3|}{4\kappa_2} \qquad b = \frac{8\kappa_2^3}{\kappa_3^2} \qquad c = \kappa_1 - ab \ ,$$

with $\kappa_r = 2^{r-1}\Gamma(r)\text{tr}\left\{\left[\mathbf{V}\left(\hat{\theta}\right)\mathbf{H}\left(\hat{\theta}\right)\right]^r\right\} = 2^{r-1}\Gamma(r)\text{tr}\left\{\left[\mathbf{H}^{-1}\left(\hat{\theta}\right)\mathbf{J}\left(\hat{\theta}\right)\right]^r\right\}$.

Confidence regions for subsets of parameters can be constructed using profile objective functions, defined for a subset of parameters ψ as $S(\psi) = S\left(\psi, \hat{\lambda}(\psi)\,|\mathbf{Y}\right)$ where $\hat{\lambda}(\psi)$ minimises the objective function for a fixed value of ψ. The procedure is exactly the same as for the full parameter vector, except that the κs are calculated from the appropriate submatrices of $\mathbf{H}^{-1}$ and $\mathbf{V}$. In the case of a single parameter, let v be the appropriate diagonal element of $\mathbf{V}$, and h^{-1} the corresponding element of $\mathbf{H}^{-1}$; then $a = hv$, $b = 1$ and $c = 0$ in this case.

5. A final way to carry out tests uses the fact that at the true parameter value θ_0, the objective function gradient vector is distributed as $MVN(\mathbf{0}, \mathbf{J}(\theta_0))$. Any θ where the gradient is 'large' according to this distribution is therefore not supported by the data. This does not require calculation of the Hessian, which may be seen as a potential advantage.
6. The matrix $\mathbf{J}(\theta_0)$ can be estimated in any of three ways:
   (a) Find an analytical expression for the covariance matrix of the fitting properties under the model; then estimate $\mathbf{J}(\theta_0)$ as $4\left[\mathbf{W}\left(\hat{\theta}\right)\right]' \text{var}_{\hat{\theta}}(\mathbf{T})\,\mathbf{W}\left(\hat{\theta}\right)$, where $\mathbf{W}(\theta)$ is a $k \times p$ matrix whose (i,j)th element is $w_i \partial\tau_i(\theta)/\partial\theta_j$. If necessary, use numerical differentiation to evaluate $\partial\tau_i(\theta)$.
   (b) Calculate an empirical estimate of $\text{var}(\mathbf{T})$, and use $4\left[\mathbf{W}\left(\hat{\theta}\right)\right]' \widehat{\text{var}}(\mathbf{T})\,\mathbf{W}\left(\hat{\theta}\right)$ as an estimate of $\mathbf{J}(\theta_0)$. For example, if $n > 1$ years of data are available, fitting properties $\mathbf{T}_1, \ldots, \mathbf{T}_n$ can be computed separately for each year: $\mathbf{T}$ can then be taken as the mean over all years, and $\widehat{\text{var}}(\mathbf{T})$ as $n^{-2}\sum_{i=1}^{n}(\mathbf{T}_i - \mathbf{T})(\mathbf{T}_i - \mathbf{T})'$. This suggestion follows Rodriguez-Iturbe et al. (1988). Notice, however, that some components of each $\mathbf{T}_i$, in particular those relating to daily data, will be computed using relatively small samples. It is therefore important to use estimators that are, as far as possible, unbiased in small samples. This applies particularly to estimators of autocorrelation coefficients, for example — standard estimators can suffer from serious bias problems in small samples. Methods for correcting this are given by Kendall and Ord (1990, page 79), for example.
   (c) Choose fitting properties $\{T_i\}$ in such a way that (i) the chosen properties are approximately uncorrelated (ii) as many components as possible have variances that are independent of θ. Make an educated guess as to the variances of the remaining properties. Then, in the objective function, set $w_i = 1/\text{var}[T_i]$. Throw the result at a nonlinear minimisation routine that returns the Hessian as a by-product. Multiply this Hessian by 2, and take the result as an estimate of $\mathbf{J}(\theta_0)$; invert this to obtain an estimate of $\mathbf{V}(\theta_0)$ without any further matrix multiplication.

   For practical purposes, $\mathbf{J}(\theta_0)$ and $\mathbf{J}\left(\hat{\theta}\right)$ are interchangeable.

The guidelines in (6c) above, regarding choice of fitting properties, apply more generally — indeed, lack of correlation was one of the criteria given by Rodriguez-Iturbe et al. (1988) for choosing fitting properties. The theory outlined in the preceding sections also suggests the following considerations:

1. The chosen statistics should be unbiased for the corresponding theoretical properties i.e. $\text{E}[T_i] = \tau_i$.
2. The chosen statistics should have a normal distribution to a reasonable degree of approximation. This might involve, for example, taking logarithms of quantities that are essentially positive (another suggestion of Rodriguez-Iturbe et al. 1988), or applying a z-transformation to autocorrelations as in Wheater et al. (2000b, Section 2.8.5).
3. The chosen statistics should have variances that are as small as possible. This is intuitively obvious; in terms of the mathematics, it is easiest to see in the case of a single parameter, so that all matrices become scalars. In this case the variance of the parameter estimate is proportional to a weighted sum of variances of fitting properties.

4. The chosen statistics should vary rapidly with respect to the model parameters. Mathematically, this requirement corresponds to large values of the Hessian matrix (i.e. the matrix of derivatives of fitting properties with respect to parameters). For a single parameter, the variance is inversely proportional to this Hessian.

The theory outlined here represents an alternative to the approach suggested at the bottom of page 290 of Rodriguez-Iturbe et al. (1988). Instead of calculating the covariance matrix $\mathbf{V}(\theta_0)$, they suggested perturbing each of the fitting properties by a small amount, and re-estimating the model parameters at each of the perturbed configurations. This determines an approximate linear transformation from fitting properties to parameter estimates, which can be combined with an estimate of var$(\mathbf{T})$ to estimate the covariance matrix. The difference here is that we avoid refitting the model many times by transforming in the opposite direction (from θ to τ rather than from $\mathbf{T}$ to $\hat{\theta}$) and using an analytical (or numerical) linearisation of the transformation in the matrix $\mathbf{W}(\hat{\theta})$.

A further development is the ability to judge parameter sets on the basis of the objective function itself. This can be used, for example, to identify the region of the parameter space for which the objective function is 'almost' optimal.

Appendix B Mathematical expressions of generalised moments used in single-site rainfall models

B.1 Introduction

A range of single-site rainfall models are being compared for assessment in Chapter 2 for Work Package 1. Insofar as the selected fitting method is the generalised method of moments (Lekkas and Onof, 2003), mathematical expressions for each of these moments are required for input into the objective function.

Many such expressions are already available in the literature. However, as the result of a fairly high rate of typographical errors, they cannot be used without some form of checking, which may involve the full derivation of the formula.

Other expressions are not readily available. This is either because their derivations are unpublished, because the model in question has not yet been examined, or because the statistic in question has not yet been used.

This appendix seeks to bring all the relevant expressions together for use within the model comparison exercise. When useful or interesting, key elements of the derivations are given. To avoid unnecessary length, the appendix focusses upon models driven by a Bartlett-Lewis point process.

B.2 Background: the Bartlett-Lewis Rectangular Pulse Model

Since single-site models driven by a Bartlett-Lewis point process are all modifications of the Barlett-Lewis Rectangular Pulse model (BLRPM), the key expressions for this model are given here. First, we present the notation used throughout the appendix, as well as that specific for this model.

B.2.1 Notation and model specification

Three levels of description

Single site models represent the continuous-time rainfall $Y(t)$. They are calibrated and validated by examining properties of one or both the following processes:

- discrete time aggregated process at time-scale h: $Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(t)\,dt$
- continuous-time moving average process at time-scale h: $Y_{(h)}(t) = \frac{1}{h}\int_{t-h/2}^{t+h/2} Y(t)\,dt$

Data sets of observed data at time-scale h can be considered as samples of the aggregated process, but also as providing samples of the continuous-time moving average process.

Model description

The main Bartlett-Lewis point process is a cluster Poisson process characterised by the arrival of random clusters of points according to a Poisson process. In terms of the representation of rainfall, the clusters are *storms* and the points correspond to the arrivals of *cells*.

Within each cluster, points arrive throughout a period of storm activity which is a random variable.

To each cell arrival time is assigned a rainfall pulse of random duration and intensity.

Parameters

The following notation is used throughout:

- λ: Poisson cluster (storm) arrival rate
- β: Poisson point (cell) arrival rate activity

In the BLRPM, the storm and cell durations are standardly taken as exponentially distributed:

- γ: Exponential parameter of storm duration
- η: Exponential parameter of cell duration

The cell intensity distribution is characterised by three parameters:

- μ_x: Mean cell intensity
- μ_{x^2}: Mean of squares of cell intensities
- μ_{x^3}: Mean of cubes of cell intensities

Three distributions are considered for the intensity: the exponential, Gamma and general Pareto distributions. Details of the notation used and the main relevant properties of these distributions are given in B.7.

In the distributions considered here, one or two parameters are sufficient to fully characterise the distribution. The BLRPM can therefore be characterised by the following set of parameters:

$$\{\lambda, \mu_x, \mu_{x^2}, \eta, \beta, \gamma\}$$

These parameters do not, however, all have direct physical meaning. It is therefore useful to re-parameterize the model in terms of a set of *mechanistic parameters*:

$$\{\lambda, \mu_x, \sigma_x, \delta_c, \mu_c, \delta_s\}$$

defined by:

Storm arrival rate	λ		(hr^{-1})
Mean cell intensity	μ_x		$(mm.hr^{-1})$
Standard deviation of cell intensity	σ_x	$= \sqrt{\mu_{x^2} - \mu_x^2}$	$(mm.hr^{-1})$
Mean cell duration	δ_c	$= 1/\eta$	(hr)
Mean number of cells per storm	μ_c	$= 1 + \beta/\gamma$	
Mean duration of storm activity	δ_s	$= 1/\gamma$	(hr)

For the sake of simplicity, the equations are given in terms of the original parameter set. They can easily be re-expressed in terms of the mechanistic parameters using the following relations:

$$\begin{aligned} \mu_{x^2} &= \sigma_x^2 + \mu_x^2 \\ \eta &= 1/\delta_c \\ \beta &= (\mu_c - 1)/\delta_s \\ \gamma &= 1/\delta_s \end{aligned}$$

Properties of the aggregated process

The properties of the process $Y_i^{(h)}$ which are considered for model calibration and validation are functions of the following:

- $M(h)$: mean of the rainfall depth (in mm)
- $V(h)$: variance of the rainfall depth (in mm^2)
- $C(k,h)$: autocovariance lag-k of the depth (in mm^2)
- $A(k,h)$: autocorrelation lag-k of the depth (in mm^2)
- $M^p(h)$: non-centered moment of order p $(p>1)$ of the intensity (in mm^p)
- $P_d(h)$: proportion of dry periods
- $M_d(h)$: mean duration of a dry period (in hours)
- $M_w(h)$: mean duration of a wet period (in hours)

where all the properties are for time-scale h hours.

Properties of the moving average process

Note that the properties below could also be expressed as properties of the aggregated process, since $\text{var}\left[Y_{(h)}(t)\right] = \text{var}\left[\frac{Y_i^{(h)}}{h}\right]$. The main property of interest is the variance of the moving average process. This variance can be related to the variance of the underlying continuous-time process $Y(t)$ by defining a variance reduction factor called the *variance function* and denoted $\omega(h)$:

$$\text{var}\left[Y_{(h)}(t)\right] = \omega(h)\text{var}[Y(t)]$$

The variance function can easily be calculated (Vanmarcke, 1993)as:

$$\omega(h) = \frac{1}{h^2}\int_0^h\int_0^h \rho(t_1-t_2)\,dt_1\,dt_2 = \frac{2}{h}\int_0^h\left(1-\frac{\tau}{h}\right)\rho(\tau)\,d\tau \tag{B.1}$$

where $\rho(\tau)$ is the autocorrelation function at lag τ of process $Y(t)$.

Of particular interest is the behaviour of the variance function as the scale increases. If this is not a long memory process, we must have:

$$\lim_{h\to\infty}\omega(h) = 0$$

For many processes, the convergence to 0 is in $1/h$. Consequently, Vanmarcke (1993) defines the *scale of fluctuation* as:

$$\Theta = \lim_{h\to\infty} h\omega(h)$$

Since:

$$\omega(h) = \frac{1}{2h}\left[\int_0^h \rho(\tau)\,d\tau - \frac{1}{h}\int_0^h \tau\rho(\tau)\,d\tau\right]$$

and since the scale of fluctuation only exists if

$$\lim_{h\to\infty}\frac{1}{h}\int_0^h \tau\rho(\tau)\,d\tau = 0$$

we therefore obtain (Vanmarcke, 1993):

$$\Theta = \lim_{h\to\infty} h\omega(h) = 2\int_0^\infty \rho(\tau)\,d\tau$$

The scale of fluctuation is so called since, when h is large, $Y_{(h)}(t)$ has a variance which is approximately $\text{var}[Y(t)]\,\Theta/h$. $Y_{(h)}(t)$ is therefore equivalent to the mean of h/Θ independent observations from the continuous-time process.

To summarize, the following moving average process properties can be used in the calibration:

- $\omega(h)$: Variance function at time-scale h
- Θ: Scale of fluctuation

B.2.2 BLRPM Continuous-time properties

The two important properties for the calculation of the aggregated process properties are the mean and covariance function. From Rodriguez-Iturbe et al. (1987), we have:

$$E[Y(t)] = \lambda\mu_c\mu_x/\eta \tag{B.2}$$

where $\mu_c = 1 + \frac{\beta}{\gamma}$ is the mean number of cells per storm, and

$$c_Y(\tau) = \frac{\lambda\mu_c}{\eta}\left[\mu_{x^2} + \frac{\beta\gamma\mu_x^2}{\gamma^2-\eta^2}\right]e^{-\eta\tau} - \frac{\lambda\mu_c}{\eta}\frac{\beta\eta\mu_x^2}{\gamma^2-\eta^2}e^{-\gamma\tau} \tag{B.3}$$

B.2.3 BLRPM First- and second order aggregated depth moments

The first two moments of the marginal distribution of rainfall depths are obtained by integration:

$$M(h) = \quad E[Y_i^{(h)}] \quad = E[Y(t)] \tag{B.4}$$

$$V(h) = \quad \text{var}[Y_i^{(h)}] \quad = 2\int_0^h (h-u)c_Y(u)\,du \tag{B.5}$$

$$C(k,h) = \quad \text{cov}[Y_i^{(h)}, Y_{i+k}^{(h)}] \quad = \int_{-h}^h (h-|v|)c_Y(kh+v)\,dv \tag{B.6}$$

The moments are given by (Rodriguez-Iturbe et al., 1987):

$$M(h) = \frac{\lambda h\mu_x\mu_c}{\eta} \tag{B.7}$$

$$V(h) = \frac{2\lambda\mu_c}{\eta}\left[\frac{(\mu_{x^2}+\beta\mu_x^2/\gamma)h}{\eta} + \frac{\mu_x^2\beta\eta(1-e^{-\gamma h})}{\gamma^2(\gamma^2-\eta^2)} - \left(\mu_{x^2}+\frac{\beta\gamma\mu_x^2}{\gamma^2-\eta^2}\right)\frac{1-e^{-\eta h}}{\eta^2}\right] \tag{B.8}$$

The autocovariance of lag-k is given by:

$$C(k,h) = \frac{\lambda\mu_c}{\eta}\left[\left(\mu_{x^2}+\frac{\beta\gamma\mu_x^2}{\gamma^2-\eta^2}\right)\frac{(1-e^{-\eta h})^2e^{-\eta(k-1)h}}{\eta^2} - \frac{\mu_x^2\beta\eta(1-e^{-\gamma h})^2e^{-\gamma(k-1)h}}{\gamma^2(\gamma^2-\eta^2)}\right] \tag{B.9}$$

B.2.4 BLRPM Wet-dry discrete-time properties

The expression for the proportion dry is (Rodriguez-Iturbe et al., 1987):

$$P_d(h) = \exp\{-\lambda(h+\mu_T) + \lambda G_P^*(0,0)(\gamma+\beta e^{-(\beta+\gamma)h})/(\beta+\gamma)\} \tag{B.10}$$

where μ_T is the mean storm duration, given by Onof (1992):

$$\mu_T = \frac{1}{\gamma} + \frac{\gamma}{\eta^2}\int_0^1 v^{-1}\,dv\int_0^1 t^{\frac{\gamma}{\eta}-1}\left[1-(1-vt)e^{-\frac{\beta v(1-t)}{\eta}}\right]dt$$

and

$$G_P^*(z,s) = \eta^{-1}e^{-\frac{\beta(1-z)}{\eta}}\int_0^1 t^{\frac{\gamma+s}{\eta}-1}\left[1-(1-z)t\right]e^{\frac{\beta(1-z)t}{\eta}}\,dt$$

Since these expressions are not easy to compute, the following approximations can be used. They are valid if $\beta << \eta$ and $\gamma << \eta$ (i.e. if there is enough cell overlap and cell durations are much smaller than storm durations):

$$\mu_T \approx \frac{1}{\gamma}\left\{1 + \frac{\gamma(\beta+\gamma/2)}{\eta^2} - \frac{\gamma(5\gamma\beta+\beta^2+2\gamma^2)}{4\eta^3} + \frac{\gamma(4\beta^3+31\beta^2\gamma+99\beta\gamma^2+36\gamma^3)}{72\eta^4}\right\}$$

and

$$G_P^*(z,s) \approx \frac{1}{\gamma}\left\{1 - \frac{\beta+\gamma}{\eta} + \frac{3\beta\gamma+2\gamma^2+\beta^2}{2\eta^2}\right\}$$

as in Onof (1992).

However, if these requirements on small values of β and γ are not fulfilled, we can approximate these terms follows:

$$\begin{aligned}
\mu_T &\approx \eta^{-1}\left(1+\phi\sum_{j=1}^{M}\frac{(-\kappa)^{j-1}(\kappa-j^2-j)}{j(j+1)!}B(j+1,\phi)+\phi^{-1}\right) \\
&\text{and} \\
G_P^*(0,0) &\approx \eta^{1-}e^{-\kappa}\left(\sum_{j=0}^{M'}\frac{\kappa^j}{j!}B(j+\phi,2)+\frac{\delta_{M'}(\kappa)}{(M'+\phi+1)(M'+\phi+2)}\right) \\
&\text{with} \\
\delta_{M'}(\kappa) &= e^{\kappa}-\sum_{j=0}^{M'}\frac{\kappa^j}{j!}
\end{aligned} \tag{B.11}$$

where $\kappa = \beta/\eta$ and $\phi = \gamma/\eta$ and the values of M and M' are to be chosen large enough so as to reduce the error. For the LR model (discussed further), identical approximations are required. Upper bounds for the errors involved are estimated in B.11, together with numerical investigations into their values for different values of κ and ϕ. B.11 also presents the derivation of these approximations.

The mean duration dry is then a function of the proportion dry Onof et al. (1994):

$$M_d(h) = \frac{P_d(h)}{P_d(h)-P_d(2h)} \tag{B.12}$$

Note we can easily derive another useful statistic, namely the mean number of events at time-scale h in a period of duration $n(h)$ time-intervals of h hours. Since the probability of the arrival of an event at time-scale h hours is given by:

$$p_e(h) = Pr\{\text{event start in } [(n-1)h, nh)\} = Pr\{Y_n^{(h)} > 0 | Y_{n-1}^{(h)} = 0\} Pr\{Y_{n-1}^{(h)} = 0\}$$

this yields:

$$p_e(h) = \left(1 - \frac{P_d(2h)}{P_d(h)}\right) P_d(h) = P_d(h) - P_d(2h)$$

If the mean storm duration is very small compared to the duration of the period, i.e. $\left(\frac{\mu_T}{h}+1\right) p_e(h) << n(h)$ (the condition is, for instance, met if the period under consideration is the month and the time-scale less than 6 hours), then the mean number of events is approximately given by:

$$n_e(h) \approx (P_d(h) - P_d(2h))\, n(h) \tag{B.13}$$

since, as the event duration goes to zero, the distribution of the number of events is approximately binomially distributed $B(n(h), P_d(h) - P_d(2h))$.

B.2.5 BLRPM Third-order aggregated depth moment

The third-order moment yields information about the asymmetry of a distribution. Because of the need to obtain a good fit for extreme values, it is useful to include this moment in the fitting process. The main steps of the derivation are as follows.

General form of the integral

As a first step, we need to relate this moment to moments of the underlying continuous-time process. In general, we can write that, if

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(u)\,du$$

then,

$$E\left[Y_i^{(h)}, Y_{i+j}^{(h)}, Y_{i+k}^{(h)}\right] = E\left[\int_{(i-1)h}^{ih}\int_{(i+j-1)h}^{(i+j)h}\int_{(i+k-1)h}^{(i+k)h} Y(u)Y(v)Y(w)\,du\,dv\,dw\right]$$

The change of variables:

$$x = u; y = v - u - jh; z = w - u - kh$$

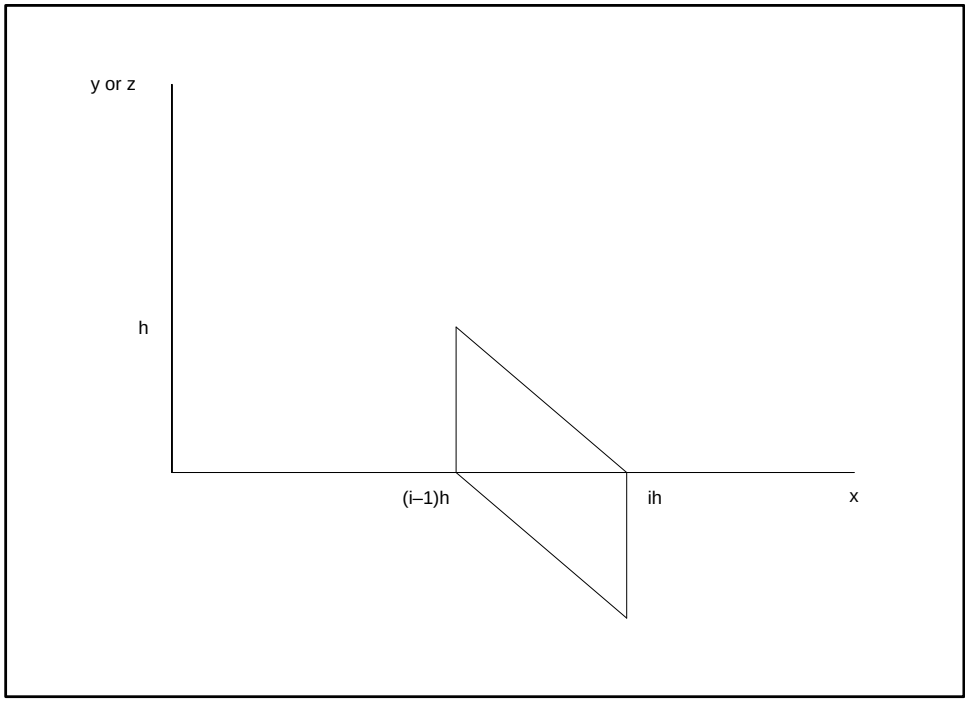

Figure B.21 Domain of integration in the (x,y) or (x,z) planes

yields the domain of integration shown in figure B.21.

The sum of six integrals must then be computed:

$$E\left[Y_i^{(h)}, Y_{i+j}^{(h)}, Y_{i+k}^{(h)}\right] = I_1 + I_2 + I_3 + I_4 + I_5 + I_6 \tag{B.14}$$

where:

$$\begin{aligned}
I_1 &= \int_{z=0}^{h}\int_{y=-h+z}^{0}\int_{x=(i-1)h-y}^{ih-z} E[Y(x)Y((x+y+jh)Y((x+z+kh)]\,dx\,dy\,dz \\
I_2 &= \int_{z=0}^{h}\int_{y=0}^{z}\int_{x=(i-1)h}^{ih-z} E[Y(x)Y((x+y+jh)Y((x+z+kh)]\,dx\,dy\,dz \\
I_3 &= \int_{z=0}^{h}\int_{y=z}^{h}\int_{x=(i-1)h}^{ih-y} E[Y(x)Y((x+y+jh)Y((x+z+kh)]\,dx\,dy\,dz \\
I_4 &= \int_{z=-h}^{0}\int_{y=-h}^{z}\int_{x=(i-1)h-y}^{ih} E[Y(x)Y((x+y+jh)Y((x+z+kh)]\,dx\,dy\,dz \\
I_5 &= \int_{z=-h}^{0}\int_{y=z}^{0}\int_{x=(i-1)h-z}^{ih} E[Y(x)Y((x+y+jh)Y((x+z+kh)]\,dx\,dy\,dz \\
I_6 &= \int_{z=-h}^{0}\int_{y=0}^{z+h}\int_{x=(i-1)h-z}^{ih-y} E[Y(x)Y((x+y+jh)Y((x+z+kh)]\,dx\,dy\,dz
\end{aligned}$$

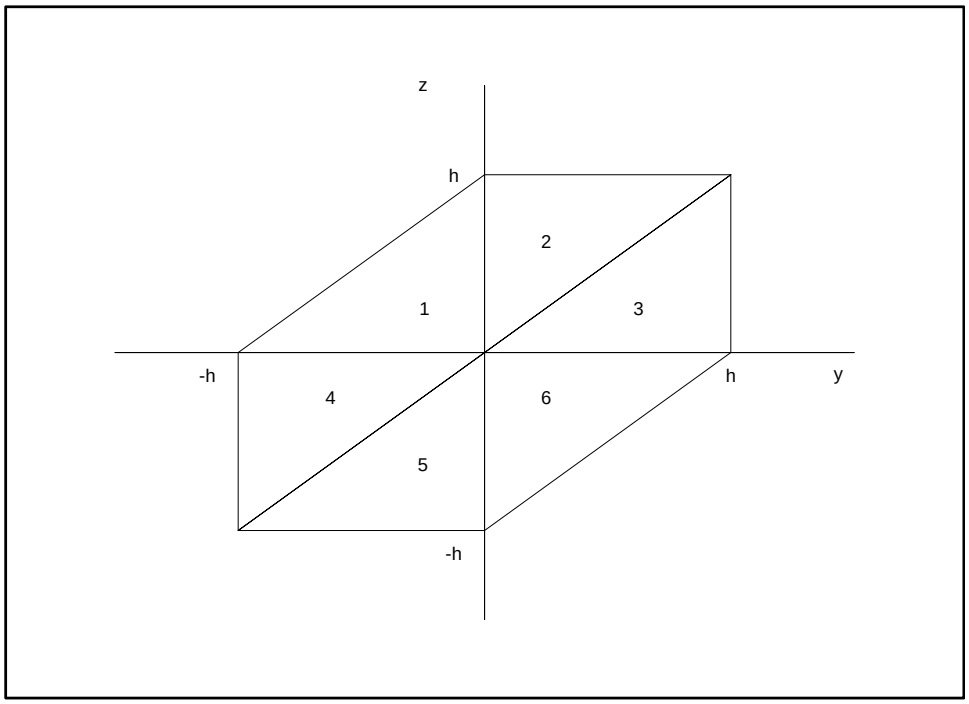

Figure B.22 Domain of integration in the (x,y) or (x,z) planes

This corresponds to the subdivision of the domain of integration according to figure B.22.

For the third-order moment, we have $j = k = 0$. The computation of the integrand, $E[Y(x)Y(x+y)Y(x+z)]$, requires that third-order properties of counts and of cell intensities be computed.

Third-order properties of counts

These are obtained for $x_1 < x_2 < x_3$ as follows:

$$\begin{aligned} E\left[\delta N(x_1)\delta N(x_2)\delta N(x_3)\right] &= (\lambda\mu_c)^3\, dx_1\, dx_2\, dx_3 \\ &\quad +\lambda\mu_c\beta^2 e^{-\gamma(x_3-x_1)}\, dx_1\, dx_2\, dx_3 \\ &\quad +(\lambda\mu_c)^2\beta\left[e^{-\gamma(x_3-x_2)}+e^{-\gamma(x_3-x_1)}+e^{-\gamma(x_2-x_1)}\right] dx_1\, dx_2\, dx_3 \end{aligned}$$

where the first term corresponds to 3 cells in different storms, the second to 3 cells in the same storm and the third to 1 cell in one storm and two in another. This yields:

$$\frac{E\left[\delta N(x_1)\delta N(x_2)\delta N(x_3)\right]}{dx_1\, dx_2\, dx_3} = \left\{(\lambda\mu_c)^3+\lambda\mu_c\beta^2 e^{-\gamma(x_3-x_1)}+(\lambda\mu_c)^2\beta\left[e^{-\gamma(x_3-x_2)}+e^{-\gamma(x_3-x_1)}+e^{-\gamma(x_2-x_1)}\right]\right\} \tag{B.15}$$

Third-order properties of cell intensities

The following expression is required in the computations:

$$\begin{aligned} E\left[X_{t_1-u_1}(u_1)X_{t_1+\tau_1-u_2}(u_1)X_{t_1+\tau_1+\tau_2-u_3}(u_3)\right] &= \mu_{x^3}e^{-\eta(u_1+\tau_1+\tau_2)} \\ &\quad +\mu_x mu_{x^2}e^{-\eta(u_1+u_2+\tau_2)} \\ &\quad +\mu_x mu_{x^2}e^{-\eta(u_1+u_2+\tau_1+\tau_2)} \\ &\quad +\mu_x mu_{x^2}e^{-\eta(u_1+u_3+\tau_1)} \\ &\quad +\mu_x^3 e^{-\eta(u_1+u_2+u_3)} \end{aligned} \tag{B.16}$$

where the terms correspond to the following respective 5 cases:

1. $t_1 - u_1 = t_1 + \tau_1 - u_2$ $t_1 - u_1 = t_1 + \tau_1 + \tau_2 - u_3$
2. $u_3 = u_2 + \tau_2$ $u_2 \neq u_1 + \tau_1$
3. $u_3 = u_1 + \tau_1 + \tau_2$ $u_2 \neq u_1 + \tau_1$
4. $u_2 = u_1 + \tau_1$ $u_3 \neq u_2 + \tau_2$
5. $u_2 = u_1 + \tau_1$ $u_3 \neq u_2 + \tau_2$ $u_3 \neq u_1 + \tau_1 + \tau_2$

Third-order moment of continuous-time process

The main term to compute is the expected value of a product of three rainfall depths of the continuous-time process. This is:

$$\begin{aligned} E[Y(x)Y(x+y)Y(x+z)] = \\ E[\int_0^\infty X_{x-u}(u)\,dN(x-u)\int_0^\infty X_{x+y-v}(v)\,dN(x+y-v) \\ \int_0^\infty X_{x+z-w}(w)\,dN(x+z-w)] \end{aligned} \tag{B.17}$$

This is evaluated as the following sum:

$$\begin{aligned} &E[Y(x)Y(x+y)Y(x+z)] = \\ &\int_{u=0}^\infty \int_{v=0,v\neq u+y}^\infty \int_{w=0,w\neq u+z,w\neq v+z-y}^\infty E[X_{x-u}(u)]E[X_{x+y-v}(v)]E[X_{x+z-w}(w)]\,E[dN(x-u)dN(x+y-v)dN(x+z-w)] \\ &+\int_{u=0}^\infty \int_{v=0,v\neq u+y}^\infty E[X_{x-u}(u)]E[X_{x+y-v}(v)X_{x+y-v}(v+z-y)]\,E[dN(x-u)dN(x+y-v)] \\ &+\int_{u=0}^\infty \int_{v=0,v\neq u+y}^\infty E[X_{x-u}(u)X_{x-u}(u+z)]E[X_{x+y-v}(v)]\,E[dN(x-u)dN(x+y-v)] \\ &+\int_{u=0}^\infty \int_{w=0,v\neq u+z}^\infty E[X_{x-u}(u)X_{x-u}(u+y)]E[X_{x+z-w}(w)]\,E[dN(x-u)dN(x+z-w)] \\ &+\int_{u=0}^\infty E[X_{x-u}(u)X_{x-u}(u+y)X_{x-u}(u+z)]\,E[dN(x-u)] \end{aligned}$$

The computation of this integral and the final form of its analytical expression are detailled in B.8.

Third-order moment of discrete-time process

The six integrals in B.14 can be rewritten as:

$$\begin{aligned} I_1 &= \int_{z=0}^h \int_{y=-h+z}^0 \int_{x=(i-1)h-y}^{ih-z} E[Y(x+y)Y((x+y)-y)Y((x+y)-y+z)]\,dxdydz \\ I_2 &= \int_{z=0}^h \int_{y=0}^z \int_{x=(i-1)h}^{ih-z} E[Y(x)Y(x+y)Y(x+z)]\,dxdydz \\ I_3 &= \int_{z=0}^h \int_{y=z}^h \int_{x=(i-1)h}^{ih-y} E[Y(x)Y(x+z)Y(x+y)]\,dxdydz \\ I_4 &= \int_{z=-h}^0 \int_{y=-h}^z \int_{x=(i-1)h-y}^{ih} E[Y(x+y)Y((x+y)+(z-y))Y((x+y)-y)]\,dxdydz \\ I_5 &= \int_{z=-h}^0 \int_{y=z}^0 \int_{x=(i-1)h-z}^{ih} E[Y(x+z)Y((x+z)+(y-z))Y((x+z)-z)]\,dxdydz \\ I_6 &= \int_{z=-h}^0 \int_{y=0}^{z+h} \int_{x=(i-1)h-z}^{ih-y} E[Y(x+z)Y((x+z)-z)Y((x+z)+(y-z))]\,dxdydz \end{aligned}$$

where the integrands have been written so as to contain products of $Y(r)Y(r+s)Y(r+t)$ with $0 \leq s \leq t$. This involves the following transformations:

For I_1: $r = x+y \quad s = -y \quad t = z-y$
For I_2: $r = x \quad s = y \quad t = z$
For I_3: $r = x \quad s = z \quad t = y$
For I_4: $r = x+y \quad s = z-y \quad t = -y$
For I_5: $r = x+z \quad s = y-z \quad t = z$
For I_6: $r = x+z \quad s = -z \quad t = y-z$

By introducing this change of variables, we find that all 6 integrals are identical to I so that:

$$M^3(h) = 6I = 6\int_{s=0}^{h}\int_{t=s}^{h}\int_{r=(i-1)h}^{ih-t} E[Y(r)Y(r+s)Y(r+t)]\,dr\,dt\,ds \tag{B.18}$$

The final expression for $M^3(h) = E[(Y_i^{(h)})^3]$ is given in B.9.

B.2.6 BLRPM moving average properties

Since, for the continuous-time process $Y(t)$, we have:

$$\text{var}[Y(t)] = \frac{\lambda\mu_c}{\eta}\left[\mu_{x^2} + \frac{\beta}{\gamma+\eta}\mu_x^2\right]$$

we have, for the variance function:

$$\omega(h) = \frac{2\left[\frac{(\mu_{x^2}+\beta\mu_x^2/\gamma)}{h\eta} + \frac{\mu_x^2\beta\eta(1-e^{-\gamma h})}{h^2\gamma^2(\gamma^2-\eta^2)} - \left(\mu_{x^2} + \frac{\beta\gamma\mu_x^2}{\gamma^2-\eta^2}\right)\frac{1-e^{-\eta h}}{h^2\eta^2}\right]}{\mu_{x^2} + \frac{\beta}{\gamma+\eta}\mu_x^2} \tag{B.19}$$

The scale of fluctuation is therefore given by:

$$\Theta = \frac{2}{\eta}\,\frac{\mu_{x^2} + \frac{\beta}{\gamma}\mu_x^2}{\mu_{x^2} + \frac{\beta}{\gamma+\eta}\mu_x^2} \tag{B.20}$$

Note that if $\gamma \to \infty$, we find the scale of fluctuation of the simpler Poisson Rectangular Pulse Model (Rodriguez-Iturbe et al., 1987), namely $\frac{2}{\eta}$.

B.3 The Dependent Depth-Duration Model

B.3.1 Model specification

One way in which the above model can be altered so as to improve its wet-dry properties is by introducing a dependence between cell intensity and cell duration distributions. This option, the DD model, has been examined by Kakou (1997).

The model is characterised by the same parameters $\lambda, \beta, \gamma, \eta$ for the storm and cell arrival rates, storm activity and cell duration. But the cell intensities X are now specified through the distribution of X conditional upon the cell duration L, i.e. $X|L$.

A first way of specifying the dependence is by choosing:

$$E[X|L=l] = fe^{-cl}$$

Kakou (1997) assumed an exponential distribution. More generally, we shall consider a second-order moment specified as:

$$E[X^2|L=l] = ge^{-dl}$$

Note that this entails the following first- and second-order unconditional moments:

$$\begin{aligned} E[X] &= \frac{f\eta}{c+\eta} \\ E[X^2] &= \frac{g\eta}{d+\eta} \end{aligned}$$

We shall refer to this as the DD1 model. Since the temporal structure is identical to that of the BLRPM, we need only examine its depth properties.

Parameters

The proposed DD1 model has 8 parameters:

$$\{\lambda, c, d, f, g, \eta, \beta, \gamma\}$$

Note that it is likely we may wish to simplify this and assume a relation between c and d for instance (in the exponential case considered by Kakou (1997), $d = 2c$ and $g = 2f^2$).

The *mechanistic parameters* for this model are:

$$\{\lambda, \mu_x, \sigma_x, \mu_{x|0}, \sigma_{x|0}, \delta_c, \mu_c, \delta_s\}$$

defined by:

Storm arrival rate	λ		(hr^{-1})
Mean cell intensity	μ_x	$= \frac{f\eta}{c+\eta}$	$(mm.hr^{-1})$
Std. deviation of cell intensity	σ_x	$= \frac{g\eta}{d+\eta} - \left[\frac{f\eta}{c+\eta}\right]^2$	$(mm.hr^{-1})$
Conditional mean cell intensity limit for 0 cell duration	$\mu_{x\|0}$	$= f$	$(mm.hr^{-1})$
Conditional std. deviation of cell intensity limit for 0 cell duration	$\sigma_{x\|0}$	$= \sqrt{g - f^2}$	$(mm.hr^{-1})$
Mean cell duration	δ_c	$= 1/\eta$	(hr)
Mean number of cells per storm	μ_c	$= 1 + \beta/\gamma$	
Mean duration of storm activity	δ_s	$= 1/\gamma$	(hr)

As before, the equations are given in terms of the original parameter set which can easily be re-expressed in terms of the mechanistic parameters using:

$$\begin{aligned} c &= \frac{1}{\delta_c}\left(\frac{\mu_{x|0}}{\mu_x} - 1\right) \\ d &= \frac{1}{\delta_c}\left(\frac{\sigma_{x|0}^2 + \mu_{x|0}^2}{\sigma_x^2 + \mu_x^2} - 1\right) \\ f &= \mu_{x|0} \\ g &= \sigma_{x|0}^2 + \mu_{x|0}^2 \\ \eta &= 1/\delta_c \\ \beta &= (\mu_c - 1)/\delta_s \\ \gamma &= 1/\delta_s \end{aligned}$$

B.3.2 Continuous-time depth properties

As in Kakou (1997), we find that for the DD model, the mean depth is:

$$E[Y(t)] = \lambda \mu_c E[XL] \tag{B.21}$$

For DD1, this yields:

$$E[Y(t)] = \lambda\mu_c \frac{\eta f}{(\eta + c)^2} \tag{B.22}$$

The other important property is the covariance of lag τ, which we derive as:

$$\begin{aligned} c_Y(\tau) &= \lambda\mu_c \int_\tau^\infty (l-\tau)\, E[X^2|l]\, f_L(l)\, dl + \frac{\lambda\beta\mu_c}{\gamma^2} \\ &\quad \{2\gamma D(0,\tau,l+\tau,0) + 2\gamma B(0,l+\tau,\infty,0) - 2\gamma\tau A(0,\tau,l+\tau,0) \\ &\quad -e^{-\gamma\tau}A(0,0,\infty,0) + e^{\gamma\tau}A(0,\tau,\infty,\gamma) - e^{\gamma\tau}A(-\gamma,l+\tau,\infty,\gamma) \\ &\quad +e^{-\gamma\tau}A(\gamma,0,\infty,0) - e^{-\gamma\tau}A(\gamma,0,l+\tau,-\gamma) + e^{-\gamma\tau}A(0,0,\tau,-\gamma)\} \end{aligned} \tag{B.23}$$

where:

$$\begin{aligned} A(\theta,a,b,\xi) &= \int_0^\infty dl \int_a^b dl'\, E[X|l]\, f_L(l)\, E[X'|l']\, f_L(l')\, e^{-\theta l} e^{-\xi l'} \\ B(\theta,a,b,\xi) &= \int_0^\infty dl \int_a^b dl'\, E[X|l]\, f_L(l)\, E[X'|l']\, f_L(l')\, l\, e^{-\theta l} e^{-\xi l'} \\ D(\theta,a,b,\xi) &= \int_0^\infty dl \int_a^b dl'\, E[X|l]\, f_L(l)\, E[X'|l']\, f_L(l')\, l'\, e^{-\theta l} e^{-\xi l'} \end{aligned}$$

For DD1, this becomes:

$$c_Y(\tau) = \frac{\lambda\mu_c g\eta e^{-(d+\eta)\tau}}{(d+\eta)^2} + \frac{\lambda\mu_c f^2\beta\eta^2\left[(c+\eta)e^{-\gamma\tau} - \gamma e^{-(c+\eta)\tau}\right]}{(c+\eta)^3\left[-\gamma^2 + (c+\eta)^2\right]} \tag{B.24}$$

B.3.3 Discrete-time depth properties

First- and second-order moments

The following relations (Rodriguez-Iturbe et al., 1987) are used to obtain first- and second-order properties of the aggregated process:

$$\begin{aligned} M(h) &= hE[Y(t)] \\ V(h) &= 2\int_0^h (h-\tau)\, c_Y(\tau)\, d\tau \\ C(k,h) &= \int_{-h}^{+h} (h-|\tau|)\, c_Y(kh+\tau)\, d\tau \end{aligned}$$

For model DD1, we trivially find for the mean depth at time-scale h:

$$M(h) = h\lambda\mu_c \frac{\eta f}{(c+\eta)^2} \tag{B.25}$$

The variance is derived as:

$$\begin{aligned} V(h) &= \frac{2\lambda\mu_c g\eta}{(d+\eta)^4}\left(e^{-(d+\eta)h} - 1 + (d+\eta)h\right) \\ &\quad + \frac{2\lambda\mu_c f^2\beta\eta^2}{(c+\eta)^3\left[(c+\eta)^2 - \gamma^2\right]}\left[\frac{c+\eta}{\gamma^2}(e^{-\gamma h} + \gamma h - 1) - \frac{\gamma}{(c+\eta)^2}\left(e^{-(c+\eta)h} + (c+\eta)h - 1\right)\right] \end{aligned} \tag{B.26}$$

and the covariance as:

$$\begin{aligned} C(k,h) &= \frac{\lambda\mu_c g\eta}{(d+\eta)^4} e^{-(d+\eta)(k-1)h}\left(1 - e^{-(d+\eta)h}\right)^2 \\ &\quad + \frac{\lambda\mu_c f^2\beta\eta^2}{(c+\eta)^3\left[(c+\eta)^2 - \gamma^2\right]}\left[\frac{c+\eta}{\gamma^2}(1 - e^{-\gamma h})^2 e^{-\gamma(k-1)h}\right. \\ &\quad \left. - \frac{\gamma}{(c+\eta)^2}\left(1 - e^{-(c+\eta)h}\right)^2 e^{-(c+\eta)(k-1)h}\right] \end{aligned} \tag{B.27}$$

Wet-Dry properties

These properties are the same as for the BLRPM model, so that equations (B.10) and (B.12) can be used, as well as the approximations (B.11), where, as above, $\kappa = \beta/\eta$ and $\phi = \gamma/\eta$.

Third-order moment

As with the BLRPM model, the calculation of $M^3(h)$ first requires the evaluation of

$$E[Y(x)Y(x+y)Y(x+z)] \text{ with } y > 0 z > y.$$

This integral involves more extensive calculations than for the BLRPM. It is the sum of 14 terms which are analytically derivable, but are not presented here because they are too cumbersome. To illustrate this, the first term is shown in B.10. The expressions for all these terms are available in Maple.

The computation of the third-order moment of the discrete-time process involves a triple integral of the sum of these 14 terms, as in the following expression in equation

$$M^3(h) = 6I = 6\int_{y=0}^{h}\int_{z=y}^{h}\int_{r=(i-1)h}^{ih-z} E[Y(x)Y(x+y)Y(x+z)]\,dx\,dz\,dy \tag{B.28}$$

which, since (as a result of stationarity) the integrand is not a function of x, reduces to:

$$M^3(h) = 6\int_{y=0}^{h}\int_{z=y}^{h} E[Y(x)Y(x+y)Y(x+z)](h-z)\,dx\,dz\,dy \tag{B.29}$$

These integrals can be computed analytically for the DD1 model, and the results are available in Maple.

B.3.4 Moving average properties

Since, for the continuous-time process $Y(t)$, we have, for the DD1 model:

$$\text{var}[Y(t)] = \frac{\lambda\mu_c g\eta}{(d+\eta)^2} + \frac{\lambda\mu_c f^2\beta\eta^2}{(c+\eta)^3(c+\eta+\gamma)}$$

we have, for the variance function:

$$\begin{aligned}\omega(h) \quad = \quad & \frac{\frac{2\lambda\mu_c g\eta}{h^2(d+\eta)^4}\left(e^{-(d+\eta)h} - 1 + (d+\eta)h\right)}{\frac{\lambda\mu_c g\eta}{(d+\eta)^2} + \frac{\lambda\mu_c f^2\beta\eta^2}{(c+\eta)^3(c+\eta+\gamma)}} \\ & + \frac{\frac{2\lambda\mu_c f^2\beta\eta^2}{h^2(c+\eta)^3\left[(c+\eta)^2-\gamma^2\right]}\left[\frac{c+\eta}{\gamma^2}(e^{-\gamma h} + \gamma h - 1) - \frac{\gamma}{(c+\eta)^2}\left(e^{-(c+\eta)h} + (c+\eta)h - 1\right)\right]}{\frac{\lambda\mu_c g\eta}{(d+\eta)^2} + \frac{\lambda\mu_c f^2\beta\eta^2}{(c+\eta)^3(c+\eta+\gamma)}}\end{aligned} \tag{B.30}$$

The scale of fluctuation is therefore given by:

$$\Theta = 2\,\frac{\frac{g}{(d+\eta)^3} + \frac{f^2\eta\beta}{(c+\eta)^4\gamma}}{\frac{g}{(d+\eta)^2} + \frac{f^2\eta\beta}{(c+\eta)^3(c+\eta+\gamma)}} \tag{B.31}$$

B.4 The N-Cell Model

B.4.1 Model specification

Since empirical observations confirm that rainfall produced by convective and frontal mechanisms have different features, and that many climates tend to experience both types, the model can be transformed to generate n types of cells. These are characterised by:

- n random variables for the intensity distributions $\{X_i, i = 1,..,n\}$, with means $\{\mu_{x_i}, i = 1,..,n\}$ and mean square intensities $\{\mu_{x_i^2}, i = 1,..,n\}$
- n duration distributions, with exponential parameters $\{\eta_i, i = 1,..,n\}$
- n probabilities $\{\psi_i, i = 1,..,n\}$ for each cell type

This is a model defined by $4n+3$ parameters. Because of the constraint $\sum_{i=1}^{n} \psi_i = 1$, there are in effect $4n+2$ parameters. In practice, it is likely that $n = 2$ will be used.

The parameters are:

$$\{\lambda, \mu_{x_1}, ..., \mu_{x_n}, \mu_{x_1^2}, ..., \mu_{x_n^2}, \eta_1, ...\eta_n, \psi_1, ..., \psi_n, \beta, \gamma\}$$

For this model, the following *mechanistic parameters* can be used:

$$\{\lambda, \mu_{x_1}, ...\mu_{x_n}, \sigma_{x_1}, ..., \sigma_{x_n}, \delta_{c_1}, ..., \delta_{c_n}, \psi_1, ..., \psi_n, \mu_c, \delta_s\}$$

defined by:

Storm arrival rate	λ		(hr^{-1})
Mean cell intensities	μ_{x_i}	for $i = 1,...,n$	$(mm.hr^{-1})$
Standard deviations of cell intensities	σ_{x_i}	$= \sqrt{\mu_{x_i^2} - \mu_{x_i}^2}$ for $i = 1,...,n$	$(mm.hr^{-1})$
Mean cell durations	δ_{c_i}	$= 1/\eta_i$ for $i = 1,...,n$	(hr)
Proportion of each rainfall type	ψ_i	for $i = 1,...,n$	
Mean number of cells per storm	μ_c	$= 1 + \beta/\gamma$	
Mean duration of storm activity	δ_s	$= 1/\gamma$	(hr)

The original parameters, in terms of which the equations are written, are the following functions of these *mechanistic parameters*:

$$\begin{aligned} \mu_{x_i^2} &= \sigma_{x_i}^2 + \mu_{x_i}^2 \\ \eta_i &= 1/\delta_{c_i} \\ \beta &= (\mu_c - 1)/\delta_s \\ \gamma &= 1/\delta_s \end{aligned}$$

B.4.2 Continuous-time depth properties

We derive the following expressions for the mean and covariance of the continuous-time process:

$$E[Y(t)] = \lambda \mu_c \sum_{i=1}^{n} \frac{\psi_i \mu_{x_i}}{\eta_i} \tag{B.32}$$

$$\begin{aligned} c_Y(\tau) &= \lambda \mu_c \sum_{i=1}^{n} \frac{\psi_i \mu_{x_i^2} e^{-\eta_i \tau}}{\eta_i} \\ &\quad + \lambda \mu_c \beta e^{-\gamma\tau} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\psi_i \psi_j \mu_{x_i} \mu_{x_j}}{(\eta_i - \gamma)(\eta_j + \gamma)} \\ &\quad + 2\lambda \mu_c \beta \gamma \sum_{j=1}^{n} \frac{\psi_j \mu_{x_j} e^{-\eta_j \tau}}{\gamma^2 - \eta_j^2} \sum_{i=1}^{n} \frac{\psi_i \mu_{x_i}}{\eta_i + \eta_j} \end{aligned} \tag{B.33}$$

This can be re-expressed as:

$$c_Y(\tau) = \sum_{i=1}^{n} C_i e^{-\eta_i \tau} + D e^{-\gamma\tau} \tag{B.34}$$

using the following notation:

$$\begin{aligned} C_i &= \lambda\mu_c\left[\frac{\psi_i\mu_{x_i^2}}{\eta_i}+2\frac{\beta\gamma\psi_i\mu_{x_i}}{\gamma^2-\eta_i^2}\sum_{j=1}^{n}\frac{\psi_i\mu_{x_i}}{\eta_j+\eta_i}\right] \\ D &= \lambda\mu_c\beta\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\psi_i\psi_j\mu_{x_i}\mu_{x_j}}{(\eta_i-\gamma)(\eta_j+\gamma)} \end{aligned}$$

B.4.3 Discrete-time properties

First- and second-order depth moments

The following first- and second-order properties are obtained by integration:

$$M(h) = h\lambda\mu_c\sum_{i=1}^{n}\frac{\psi_i\mu_{x_i}}{\eta_i} \tag{B.35}$$

$$V(h) = \sum_{i}^{n}\frac{2C_i}{\eta_i^2}(h\eta_i+e^{-\eta_i h}-1)+\frac{2D}{\gamma^2}(\gamma h+e^{-\gamma h}-1) \tag{B.36}$$

$$C(k,h) = \sum_{i}^{n}\frac{C_i e^{-\eta_i(k-1)h}}{\eta_i^2}(1-e^{-\eta_i h})^2+\frac{De^{-\gamma(k-1)h}}{\gamma^2}(1-e^{-\gamma h})^2 \tag{B.37}$$

Wet-dry properties

The proportion of dry periods can easily be derived on the basis of the derivation of the same property for BLRPM in Rodriguez-Iturbe et al. (1987). The two terms which depend upon the cell duration parameter η in the exact expression of $P_d(h)$ (see equation (B.10))are the mean duration of a storm μ_T and the term $G_P^*(0,0)$. Both these terms are functions of the probabilities $p_r(t)$ and $q_r(t)$ defined as:

$$\begin{aligned} p_r(t) &= \Pr\{\text{Storm live and r cells active at time t}\} \\ q_r(t) &= \Pr\{\text{Storm terminated and r cells active at time t}\} \end{aligned}$$

which satisfy the following differential equations:

$$\begin{aligned} dp_r(t)/dt &= -(\beta+\gamma+r\sum_{i=1}^{n}\psi_i\eta_i)\,p_r(t)+(r+1)(\sum_{i=1}^{n}\psi_i\eta_i)\,p_{r+1}(t)+\beta\,p_{r-1}(t) \\ dq_r(t)/dt &= -(r\sum_{i=1}^{n}\psi_i\eta_i)\,q_r(t)+\gamma p_r(t)+(r+1)(\sum_{i=1}^{n}\psi_i\eta_i)\,q_{r+1}(t) \end{aligned}$$

so that, if we define $\eta=\sum_{i=1}^{n}\psi_i\eta_i$, we have the same differential system as for the RBLPM (see Rodriguez-Iturbe et al., 1987, section 4.2).

We therefore have:

$$P_d(h) = \exp\{-\lambda(h+\mu_T)+\lambda G_P^*(0,0)(\gamma+\beta e^{-(\beta+\gamma)h})/(\beta+\gamma)\} \tag{B.38}$$

with:

$$\mu_T = \frac{1}{\gamma}+\frac{\gamma}{\eta^2}\int_0^1 v^{-1}\,dv\int_0^1 t^{\frac{\gamma}{\eta}-1}\left[1-(1-vt)e^{-\frac{\beta v(1-t)}{\eta}}\right]dt$$

and

$$G_P^*(z,s) = \eta^{-1}e^{-\frac{\beta(1-z)}{\eta}}\int_0^1 t^{\frac{\gamma+s}{\eta}-1}\,[1-(1-z)t]e^{\frac{\beta(1-z)t}{\eta}}\,dt$$

Since these expressions are not easy to compute, the following approximations can be used. They are valid if $\beta << \eta$ and $\gamma << \eta$ (i.e. if there is enough cell overlap and cell durations are much smaller than storm durations):

$$\mu_T \approx \frac{1}{\gamma}\left\{1+\frac{\gamma(\beta+\gamma/2)}{\eta^2}-\frac{\gamma(5\gamma\beta+\beta^2+2\gamma^2)}{4\eta^3}+\frac{\gamma(4\beta^3+31\beta^2\gamma+99\beta\gamma^2+36\gamma^3)}{72\eta^4}\right\}$$

and

$$G_P^*(z,s) \approx \frac{1}{\gamma}\left\{1-\frac{\beta+\gamma}{\eta}+\frac{3\beta\gamma+2\gamma^2+\beta^2}{2\eta^2}\right\}$$

as in Onof (1992).

As with previous models, we also require approximations when these conditions upon β and γ are not fulfilled. Defining, as previously, $\kappa = \beta/\eta$ and $\phi = \gamma/\eta$, μ_T and $G_P^*(0,0)$ can be approximated as in equation (B.11).

Third-order moment

The evaluation of this moment involves very lengthy analytical developments. It will therefore not be computed for the purpose of this project.

B.4.4 Moving-average properties

The variance function is easily obtained as:

$$\omega(h) = \frac{\sum_i^n \frac{2C_i}{\eta_i^2 h^2}(h\eta_i + e^{-\eta_i h} - 1) + \frac{2D}{\gamma^2 h^2}(\gamma h + e^{-\gamma h} - 1)}{\sum_i^n C_i + D} \tag{B.39}$$

The scale of fluctuation is therefore given by:

$$\Theta = 2\,\frac{\sum_i^n C_i/\eta_i + D/\gamma}{\sum_i^n C_i + D} \tag{B.40}$$

B.5 The Linear Random Parameter Model

B.5.1 Model specification

An important modification of the original BLRPM was proposed by Rodriguez-Iturbe et al. (1988). The observation that this model does not provide a satisfactory reproduction of the proportion of dry periods suggested introducing a greater diversity of the internal wet-dry structure of storms.

A first way in which this could be done would be to consider introducing a range of m types of storms, such that each storm is characterised by one of η_i, β_i and γ_i, with $i = 1,...,M$, each type appearing with probability ε_1. To preserve the overall structure of storms, β_1 and γ_i would be chosen proportional to η_i according to:

$$\begin{aligned}\beta_i &= \kappa\eta_i \\ \gamma_i &= \phi\eta_i\end{aligned}$$

so that the model parameters would be:

$$\{\lambda, \mu_x, \mu_{x^2}, \eta_1, ...\eta_m, \varepsilon_1, ..., \varepsilon_m, \kappa, \phi\}$$

thus yielding a $2m+4$ parameter model (since one ε_i can be calculated from the knowledge of the others to satisfy the condition that these parameters add up to 1).

The observation that the above approach amounts to randomising parameter η by assigning it a discrete distribution characterised by the m probabilities $\varepsilon_i, /, i = 1,...,m$, suggests the second approach which is adopted here. This consists in using a continuous distribution to randomise

parameter η. A flexible candidate is the Gamma distribution. Thus, η is now sampled for each storm from the distribution $\Gamma(\alpha,\nu)$, while β and γ are proportional according to the relations:

$$\begin{aligned}\beta &= \kappa\eta \\ \gamma &= \phi\eta\end{aligned}$$

As a consequence, we have a 7 parameter model characterised by the following set of parameters:

$$\{\lambda, \mu_x, \mu_{x^2}, \alpha, \nu, \kappa, \phi\}$$

The following *mechanistic parameters* can be used:

$$\{\lambda, \mu_x, \sigma_x, \delta_c, \sigma_c, \mu_c, \delta_s\}$$

defined by:

Storm arrival rate	λ		(hr^{-1})
Mean cell intensity	μ_x		$(mm.hr^{-1})$
Standard deviation of cell intensity	σ_x	$=\sqrt{\mu_{x^2}-\mu_x^2}$	$(mm.hr^{-1})$
Mean cell duration	δ_c	$=\frac{\nu}{\alpha-1}$	(hr)
Inter-Storm standard deviation of cell duration	ε_c	$=\frac{\nu}{\sqrt{(\alpha-1)^2(\alpha-2)}}$	(hr)
Mean n^o of cells/storm	μ_c	$=1+\frac{\kappa}{\phi}$	
Mean duration of storm activity	δ_s	$=\frac{\nu}{(\alpha-1)\phi}$	(hr)

The original parameters, which are used in the equations below, are expressed in terms of the mechanistic parameters as follows:

$$\begin{aligned}\mu_{x^2} &= \sigma_x^2+\mu_x^2 \\ \alpha &= 2+\frac{\delta_c^2}{\varepsilon_c^2} \\ \nu &= \delta_c\left(1+\frac{\delta_c^2}{\varepsilon_c^2}\right) \\ \kappa &= \frac{\delta_c}{\delta_s}(\mu_c-1) \\ \phi &= \frac{\delta_c}{\delta_s}\end{aligned}$$

B.5.2 Continuous-time properties

The expressions below are obtained by derivation as for the BLRPM (see Rodriguez-Iturbe et al., 1988). Note that they can also be obtained by integrating the equivalent expressions for the BLRPM over the parameter η.

$$E[Y(t)] = \lambda\mu_c\mu_x\frac{\nu}{\alpha-1} \tag{B.41}$$

where $\mu_c = 1+\frac{\kappa}{\phi}$ is the mean number of cells per storm, and

$$c_Y(\tau) = \frac{\lambda\mu_c\nu}{\alpha-1}\left[\left\{\mu_{x^2}+\frac{\kappa\phi}{\phi^2-1}\mu_x^2\right\}\left(\frac{\nu}{\nu+\tau}\right)^{\alpha-1} - \frac{\kappa\mu_x^2}{\phi^2-1}\left(\frac{\nu}{\nu+\phi\tau}\right)^{\alpha-1}\right] \tag{B.42}$$

Rodriguez-Iturbe et al. (1988) note that for $1<\alpha<2$, the integral

$$\int_0^\infty c_Y(\tau)\,d\tau$$

diverges, indicating asymptotic self-similarity over that range.

B.5.3 Discrete-time depth properties

First- and second-order moments

The mean depth is (Onof, 1992):

$$M(h) = \lambda h \mu_c \mu_x \frac{\nu}{\alpha - 1} \tag{B.43}$$

and the variance and co-variance are (Rodriguez-Iturbe et al., 1988):

$$V(h) = 2A_1 \left\{ (\alpha-3)h\nu^{2-\alpha} - \nu^{3-\alpha} + (\nu+h)^{3-\alpha} \right\} - 2A_2 \left\{ \phi(\alpha-3)h\nu^{2-\alpha} - \nu^{3-\alpha} + (\nu+\phi h)^{3-\alpha} \right\} \tag{B.44}$$

and

$$\begin{aligned} C(k,h) &= A_1 \left\{ [\nu+(k+1)h]^{3-\alpha} - 2[\nu+kh]^{3-\alpha} + [\nu+(k-1)h]^{3-\alpha} \right\} \\ &\quad -A_2 \left\{ [\nu+(k+1)\phi h]^{3-\alpha} - 2[\nu+k\phi h]^{3-\alpha} + [\nu+(k-1)\phi h]^{3-\alpha} \right\} \end{aligned} \tag{B.45}$$

where:

$$\begin{aligned} A_1 &= \frac{\lambda \mu_c \nu^\alpha}{(\alpha-1)(\alpha-2)(\alpha-3)} \left\{ \mu_{x^2} + \frac{\kappa\phi\mu_x^2}{\phi^2-1} \right\} \\ A_2 &= \frac{\lambda \mu_c \kappa \mu_x^2 \nu^\alpha}{\phi^2(\phi^2-1)(\alpha-1)(\alpha-2)(\alpha-3)} \end{aligned}$$

Wet-dry properties

As shown by Rodriguez-Iturbe et al. (1988), the proportion dry for the Linear Randomised model is obtained by starting with the expression for this property in the BLRPM and taking expectations over the term which is exponentiated. This leads to the following for the proportion of dry periods (Onof, 1992):

$$P_d(h) = \exp\left[-\lambda(h+\mu_T) + \frac{\lambda\nu e^{-\kappa}}{\alpha-1} \times \frac{\phi + \kappa\left(\frac{\nu}{\nu+(\kappa+\phi)h}\right)^{\alpha-1}}{\phi+\kappa} \int_0^1 dt\, t^{\phi-1}(1-t)e^{\kappa t} \right] \tag{B.46}$$

If $\kappa << 1$ and $\phi << 1$, which means that there is enough cell overlap and cell durations are much smaller than storm durations, then the following approximation can be used Onof (1992):

$$\begin{aligned} &P_d(h) \approx \\ &\exp\left\{ -\lambda h - \frac{\lambda\nu}{\phi(\alpha-1)} \left[1 + \phi\left(\kappa + \frac{\phi}{2}\right) - \frac{1}{4}\phi\left(5\phi\kappa + \kappa^2 + 2\phi^2\right) + \frac{1}{72}\phi\left(4\kappa^3 + 31\kappa^2\phi + 99\kappa\phi^2 + 36\phi^3\right) \right] \right. \\ &\left. + \frac{\lambda\nu}{\phi(\alpha-1)} \left(1 - \kappa - \phi + \frac{3}{2}\kappa\phi + \phi^2 + \frac{1}{2}\kappa^2 \right) \left[\frac{\phi}{\phi+\kappa} + \frac{\kappa}{\phi+\kappa} \left(\frac{\nu}{\nu+(\kappa+\phi)h} \right)^{\alpha-1} \right] \right\} \end{aligned} \tag{B.47}$$

However, if these requirements on small values of κ and ϕ are not fulfilled, we can approximate the proportion dry as follows:

$$P_d(h) \approx \exp\left[-\lambda(h+\hat{\mu_{T}}_M) + \frac{\lambda\nu e^{-\kappa}}{\alpha-1} \times \frac{\phi + \kappa\left(\frac{\nu}{\nu+(\kappa+\phi)h}\right)^{\alpha-1}}{\phi+\kappa} \hat{I}_{M'} \right] \tag{B.48}$$

where

$$\hat{\mu_{T}}_M = \frac{\nu}{\alpha-1} \left(1 + \phi \sum_{j=1}^{M} \frac{(-\kappa)^{j-1}(\kappa - j^2 - j)}{j(j+1)!} B(j+1,\phi) + \phi^{-1} \right)$$

and

$$\hat{I}_{M'} = \sum_{j=0}^{M'} \frac{\kappa^j}{j!} B(j+\phi,2) + \frac{\delta_{M'}(\kappa)}{(M'+\phi+1)(M'+\phi+2)}$$

with

$$\delta_{M'}(\kappa) = e^{\kappa} - \sum_{j=0}^{M'} \frac{\kappa^j}{j!}$$

where M and M' are to be chosen large enough for a good approximation. Upper bounds for the errors involved and numerical investigations into their values are presented, together with the derivations for the approximations are found in B.11.

Third-order moment

The third-order moment is best obtained by integration of the corresponding expression for the BLRPM multiplied by the density function of the gamma distribution $\Gamma(\alpha,\nu)$. The resulting expression can however not be integrated in a closed form. As a consequence, a numerical integration is required.

A note about this numerical integration is useful. For the integral:

$$E\left[\left(Y_i^{(h)}\right)^3\right] = \int_0^{\infty} f(\eta)\, d\eta \tag{B.49}$$

has the particularity that $f(\eta) \propto \frac{1}{\eta^4}$ in the neighbourhood of 0. $f(\eta)$ is therefore not integrable at 0. However, practically, values of η close to 0 are not physically representative (and their probability is very small). This would correspond to a storm with very long cells only, which is hardly appropriate for the representation of fine-scale rainfall. It is therefore realistic to neglect small values of η. Calculations with a lower bound of 10^{-7} were found to give results in line with the simulations. The upper bound does not have to be chosen as particularly large since the integrand $f(\eta)$ decreases very quickly. Thus we can approximate the integral as follows:

$$E\left[\left(Y_i^{(h)}\right)^3\right] = \int_{10^{-7}}^{100} f(\eta)\, d\eta \tag{B.50}$$

and calculate it using Simpson's rule.

B.5.4 Moving-average properties

The variance function is easily obtained as:

$$\omega(h) = \frac{2A_1\left\{(\alpha-3)h\nu^{2-\alpha} - \nu^{3-\alpha} + (\nu+h)^{3-\alpha}\right\} - 2A_2\left\{\phi(\alpha-3)h\nu^{2-\alpha} - \nu^{3-\alpha} + (\nu+\phi h)^{3-\alpha}\right\}}{\frac{h^2(\alpha-2)(\alpha-3)}{\nu^{\alpha-1}}(A_1 - \phi^2 A_2)} \tag{B.51}$$

The scale of fluctuation will then depend upon the value of α. If $\alpha < 2$, the scale is infinite, in line with the observation of asymptotic self-similarity. Else, if $\alpha > 2$ we find:

$$\Theta = \frac{2\nu}{\alpha-2} \times \frac{A_1 - A_2\phi}{A_1 - A_2\phi^2} \tag{B.52}$$

B.6 The Quadratic Random Parameter Model

B.6.1 Model specification

The observation that the LR model is liable to underestimate extreme rainfall depths suggests an alteration of the mechanism by which cells are produced in a storm. Rather than have the random cell

arrival rate β depend linearly upon the cell duration parameter η, this dependence could include a power function. So as to facilitate the computations, let us assume a quadratic dependence. We therefore assume that:

$$\beta = \kappa_1 \eta + \kappa_2 \eta^2 \tag{B.53}$$

The model therefore has 8 parameters:

$$\{\lambda, \mu_x, \mu_{x^2}, \alpha, \nu, \kappa_1, \kappa_2, \phi\}$$

The following *mechanistic parameters* can be used:

$$\{\lambda, \mu_x, \sigma_x, \delta_c, \varepsilon_c, \mu_c, \sigma_c, \delta_s\}$$

defined by:

Storm arrival rate	λ		(hr^{-1})
Mean cell intensity	μ_x		$(mm.hr^{-1})$
Std. deviation of cell intensity	σ_x	$= \sqrt{\mu_{x^2} - \mu_x^2}$	$(mm.hr^{-1})$
Mean cell duration	δ_c	$= \frac{\nu}{\alpha - 1}$	(hr)
I.S. std. deviation of cell duration	ε_c	$= \frac{\nu}{\sqrt{(\alpha-1)^2(\alpha-2)}}$	(hr)
Mean n^o of cells/storm	μ_c	$= 1 + \frac{\kappa_1}{\phi} + \frac{\kappa_2 \alpha}{\phi \nu}$	
Std. deviation of n^o of cells/storm	σ_c	$= \sqrt{\frac{2}{\phi^2}\left[(\kappa_1^2 + \kappa_1\phi) + (2\kappa_1\kappa_2 + \kappa_2\phi)\frac{\alpha}{\nu} + \kappa_2^2 \frac{\alpha(\alpha+1)}{\nu^2}\right]}$	
Mean duration of storm activity	δ_s	$= \frac{\nu}{(\alpha-1)\phi}$	(hr)

where 'I.S.' stands for 'Inter-Storm' and the calculation of the standard deviation of the number of cells per storm is shown in B.12.

The equations below are given in terms of the original parameters. These can be re-expressed in terms of the mechanistic parameters using:

$$\begin{aligned}
\mu_{x^2} &= \sigma_x^2 + \mu_x^2 \\
\alpha &= 2 + \frac{\delta_c^2}{\varepsilon_c^2} \\
\nu &= \delta_c \left(1 + \frac{\delta_c^2}{\varepsilon_c^2}\right) \\
\kappa_1 &= \phi \left(\mu_c - 1 - \sqrt{\frac{(1-\mu_c)^3(\alpha+1)}{\alpha} + \frac{\sigma_c^2 \alpha}{2}}\right) \\
\kappa_2 &= \frac{\nu\phi}{\alpha} \sqrt{\frac{(1-\mu_c)^3(\alpha+1)}{\alpha} + \frac{\sigma_c^2 \alpha}{2}} \\
\phi &= \frac{\delta_c}{\delta_s}
\end{aligned}$$

B.6.2 Continuous-time properties

The expressions below are obtained by integrating the equivalent expressions for the BLRPM over the parameter η. The same integrals of functions of η are used which were computed for the LR model. This yields:

$$E[Y(t)] = \lambda \mu_x \left[\frac{\nu}{\alpha - 1}\left(1 + \frac{\kappa_1}{\phi}\right) + \frac{\kappa_2}{\phi}\right] \tag{B.54}$$

and

$$
\begin{aligned}
c_Y(\tau) &= \frac{\lambda\left(1+\frac{\kappa_1}{\phi}\right)\nu}{\alpha-1}\left[\mu_{x^2}+\frac{\kappa_1\phi\mu_x^2}{\phi^2-1}\right]\left(\frac{\nu}{\nu+\tau}\right)^{\alpha-1} \\
&+\lambda\left[\frac{\left(1+\frac{\kappa_1}{\phi}\right)\kappa_2\phi\mu_x^2}{\phi^2-1}+\frac{\kappa_2}{\phi}\left(\mu_{x^2}+\frac{\kappa_1\phi\mu_x^2}{\phi^2-1}\right)\right]\left(\frac{\nu}{\nu+\tau}\right)^{\alpha} \\
&+\frac{\lambda\kappa_2^2\mu_x^2\alpha}{(\phi_2^2-1)\nu}\left(\frac{\nu}{\nu+\tau}\right)^{\alpha+1}-\frac{\lambda\mu_x^2}{\phi^2-1}\left[\left(1+\frac{\kappa_1}{\phi}\right)\left(\frac{\kappa_1\nu}{\alpha-1}+\kappa_2\right)+\frac{\kappa_2}{\phi}\left(\kappa_1+\kappa_2\frac{\alpha}{\nu}\right)\right]e^{-\gamma\tau}
\end{aligned}
\tag{B.55}
$$

We note that for $-1<\alpha<2$, the integral

$$\int_0^\infty c_Y(\tau)\,d\tau$$

diverges, indicating asymptotic self-similarity when $\alpha<2$.

B.6.3 Discrete-time properties

These properties can be obtained by using the general relations presented in equations (B.4) applied to the equations (B.54) and (B.55). But since the integration over the values of η can be performed last, they can also be obtained by directly integrating the properties of the BLRPM.

First- and second-order moments of depths

For these properties, the derivation proceeds by integrating the expressions in equations (B.7), (B.8) and (B.9). This yields:

$$M(h)=\lambda h\mu_x\left[\frac{\nu}{\alpha-1}\left(1+\frac{\kappa_1}{\phi}\right)+\frac{\kappa_2}{\phi}\right] \tag{B.56}$$

and

$$
\begin{aligned}
V(h) &= \frac{2\lambda h\kappa_2^2\mu_x^2}{\phi^2}+\frac{2\lambda\nu\kappa_2}{(\alpha-1)\phi}\left[h\mu_{x^2}+h\mu_x^2\left(1+\frac{2\kappa_1}{\phi}\right)+\frac{\mu_x^2\kappa_2}{\phi^2(\phi^2-1)}\xi(1,\phi)-\frac{\kappa_2\mu_x^2\phi}{\phi^2-1}\xi(1,1)\right] \\
&+\frac{2\lambda\nu^2}{(\alpha-1)(\alpha-2)}\left[\left(1+\frac{\kappa_1}{\phi}\right)h\left(\mu_{x^2}+\mu_x^2\frac{\kappa_1}{\phi}\right)+\frac{\kappa_2\mu_x^2}{\phi^2(\phi^2-1)}\left(1+\frac{2\kappa_1}{\phi}\right)\xi(2,\phi)\right. \\
&\left.-\kappa_2\xi(2,1)\left(\frac{\mu_{x^2}}{\phi}+\frac{\mu_x^2\phi}{\phi^2-1}\left(1+\frac{2\kappa_1}{\phi}\right)\right)\right] \\
&+\frac{2\lambda\nu^3\left(1+\frac{\kappa_1}{\phi}\right)}{(\alpha-1)(\alpha-2)(\alpha-3)}\left[\frac{\mu_x^2\kappa_1}{\phi^2(\phi^2-1)}\xi(3,\phi)-\left(\mu_{x^2}+\frac{\mu_x^2\phi}{\phi^2-1}\kappa_1\right)\xi(3,1)\right]
\end{aligned}
\tag{B.57}
$$

where:

$$\xi(k,l)=1-\left(\frac{\nu}{\nu+lh}\right)^{\alpha-k}$$

For the covariance, let us define:

$$
\begin{aligned}
A_1(x) &= \frac{\nu}{\alpha-1}\left[\left(\frac{\nu}{\nu+x(k-1)h}\right)^{\alpha-1}-2\left(\frac{\nu}{\nu+xkh}\right)^{\alpha-1}+\left(\frac{\nu}{\nu+x(k+1)h}\right)^{\alpha-1}\right] \\
A_2(x) &= \frac{\nu^2}{(\alpha-1)(\alpha-2)}\left[\left(\frac{\nu}{\nu+x(k-1)h}\right)^{\alpha-2}-2\left(\frac{\nu}{\nu+xkh}\right)^{\alpha-2}+\left(\frac{\nu}{\nu+x(k+1)h}\right)^{\alpha-2}\right] \\
A_3(x) &= \frac{\nu^3}{(\alpha-1)(\alpha-2)(\alpha-3)}\left[\left(\frac{\nu}{\nu+x(k-1)h}\right)^{\alpha-3}-2\left(\frac{\nu}{\nu+xkh}\right)^{\alpha-3}+\left(\frac{\nu}{\nu+x(k+1)h}\right)^{\alpha-3}\right]
\end{aligned}
$$

In terms of these expressions, the covariance is derived as:

$$\begin{aligned} C(k,h) &= \lambda\left(1+\frac{\kappa_1}{\phi}\right)\left[\mu_{x^2}+\frac{\kappa_1\phi\mu_x^2}{\phi^2-1}\right]A_3(1) \\ &\quad +\lambda\kappa_2\left[\frac{\mu_{x^2}}{\phi}+\frac{\phi\left(1+\frac{2\kappa_1}{\phi}\right)}{\phi^2-1}\mu_x^2\right]A_2(1)+\frac{\lambda\kappa_2^2\mu_x^2}{\phi^2-1}A_1(1) \\ &\quad -\frac{\lambda\mu_x^2}{\phi^2(\phi^2-1)}\left[\left(1+\frac{\kappa_1}{\phi}\right)\kappa_1 A_3(\phi)+\left(1+2\frac{\kappa_1}{\phi}\right)\kappa_2 A_2(\phi)+\frac{\kappa_2^2}{\phi}A_1(\phi)\right] \end{aligned} \tag{B.58}$$

Wet-dry properties

For the proportion dry, a closed expression cannot be obtained since the following integrals need to be evaluated:

$$\begin{aligned} I_1(t) &= \int_0^\infty \frac{\exp\left[-(\kappa_1+\kappa_2 u)(1-t)-\nu u\right]u^{\alpha-2}\nu^\alpha}{(\kappa_1+\kappa_2 u+\phi)\Gamma(\alpha)}\,du \\ I_2(t) &= \int_0^\infty \frac{\exp\left[-(\kappa_1+\kappa_2 u)(1-t)-(\kappa_1 u+\kappa_2 u^2+\phi u)h-\nu u\right]u^{\alpha-2}\nu^\alpha}{(\kappa_1+\kappa_2 u+\phi)\Gamma(\alpha)}\,du \\ I_3(t) &= \int_0^\infty \frac{\exp\left[-(\kappa_1+\kappa_2 u)(1-t)-(\kappa_1 u+\kappa_2 u^2+\phi u)h-\nu u\right]u^{\alpha-1}\nu^\alpha}{(\kappa_1+\kappa_2 u+\phi)\Gamma(\alpha)}\,du \end{aligned}$$

In terms of these integrals, and with the following expression for the mean storm duration:

$$\mu_T = \phi\int_0^1 du\int_0^1 dt\,u^{-1}t^{\phi-1}\frac{\nu}{\alpha-1}\left[1-(1-ut)e^{-\kappa_1 u(1-t)}\left(\frac{\nu}{\nu+\kappa_2 u(1-t)}\right)^{\alpha-1}\right]+\frac{\phi\nu}{\alpha-1}$$

the proportion dry is:

$$P_d(h) = \exp\left[-\lambda(h+\mu_T)+\int_0^1 dt\,t^{\phi-1}(1-t)\lambda\left(\phi I_1(t)+\kappa_1 I_2(t)+\kappa_2 I_3(t)\right)\right] \tag{B.59}$$

Third-order moment of depths

For the third-order moment of the Quadratic Randomised model, a similar numerical computation is required as in the case of the Linear Randomised model.In other words, the expression for the BLRPM multiplied by the density function of the gamma distribution $\Gamma(\alpha,\nu)$. The final expression is then integrated numerically using Simpson's rule. Because of the non-convergence of the integral in the neighbourhood of 0, the domain of integration is not $[0,\infty)$, but can, in practice, for instance be taken as $[10^{-7},100]$.

B.6.4 Moving-average properties

The variance function is given by:

$$\omega(h) = \frac{V(h)}{c_Y(0)h^2} \tag{B.60}$$

where the numerator and denominator are given in equations (B.57) and (B.55).

This leads to the following expression for the scale of fluctuation. If $\alpha<2$, it is infinite, which reflects the asymptotic self-similarity. If $\alpha>2$ we find::

$$\Theta = 2\frac{\frac{\lambda\kappa_2^2\mu_x^2}{\phi^2}+\frac{\lambda\nu\kappa_2}{(\alpha-1)\phi}\left[\mu_{x^2}+\mu_x^2\left(1+\frac{2\kappa_1}{\phi}\right)\right]+\frac{\lambda\nu^2}{(\alpha-1)(\alpha-2)}\left[\left(1+\frac{\kappa_1}{\phi}\right)\left(\mu_{x^2}+\mu_x^2\frac{\kappa_1}{\phi}\right)\right]}{\lambda\mu_x^2\left(1+\frac{\kappa_1}{\phi}\right)\frac{\kappa_1\nu+\kappa_2(\alpha-1)}{(\alpha-1)(\phi+1)}+\frac{\lambda\kappa_2\mu_x^2}{\phi(\phi+1)}\left(\frac{\kappa_2\alpha}{\nu}+\kappa_1\right)+\lambda\mu_{x^2}\left[\frac{\nu}{\alpha-1}\left(1+\frac{\kappa_1}{\phi}\right)+\frac{\kappa_2}{\phi}\right]} \tag{B.61}$$

B.7 Cell intensity distributions

This appendix includes density and cumulative distribution functions for the three distributions considered for the cell intensity. Also included are the third-order moments.

B.7.1 Exponential distribution

The exponential distribution is a one-parameter distribution. The density function is defined by:

$$f_X(x) = \frac{1}{\mu_x} e^{-x/\mu_x} \text{ for } x > 0 \tag{B.62}$$

and the cumulative distribution function (cdf) by:

$$F_X(x) = 1 - e^{-x/\mu_x} \text{ for } x > 0 \tag{B.63}$$

The following relations hold for the exponential distribution:

$$\mu_{x^2} = 2\mu_x^2 \text{ and } \mu_{x^3} = 3\mu_x^3 \tag{B.64}$$

B.7.2 Gamma distribution

The Gamma distribution is a two-parameter distribution denoted $\Gamma(\psi, \sigma)$, where ψ is the shape parameter and σ the scale parameter. The density function is given by:

$$f_X(x) = \frac{\sigma^\psi x^{\psi-1} e^{-\sigma x}}{\Gamma(\psi)} \text{ for } x > 0 \tag{B.65}$$

and the cdf is not available in a closed form:

$$F_X(x) = \int_0^x \frac{\sigma^\psi x^{\psi-1} e^{-\sigma t}}{\Gamma(\psi)} dt \text{ for } x > 0 \tag{B.66}$$

where:

$$\psi = \frac{\mu_x^2}{\mu_{x^2} - \mu_x^2} \text{ and } \sigma = \frac{\mu_x}{\mu_{x^2} - \mu_x^2} \tag{B.67}$$

The following relation holds for the Gamma distribution:

$$\mu_{x^3} = \frac{(2\mu_{x^2} - \mu_x^2)\mu_{x^2}}{\mu_x} \tag{B.68}$$

B.7.3 General Pareto distribution

The Pareto distribution is also a two parameter distribution, denoted $P(\psi, \sigma)$, where again ψ is the shape parameter and σ the scale parameter. The density function is given by:

$$f_X(x) = \frac{\psi \sigma^\psi}{x^{\psi+1}} \text{ for } x > \sigma \tag{B.69}$$

and the cdf by:

$$F_X(x) = 1 - (\frac{\sigma}{x})^\psi \text{ for } x > \sigma \tag{B.70}$$

where:

$$\psi = 1 + \sqrt{\frac{\mu_{x^2}}{\mu_{x^2} - \mu_x^2}} \text{ and } \sigma = \frac{\mu_x(\psi - 1)}{\psi} \tag{B.71}$$

The third-order moment is given by:

$$\mu_{x^3} = \frac{(\psi - 1)(\psi - 2)}{\psi(\psi - 3)} \mu_{x^2} \mu_x \tag{B.72}$$

Note that, in the above, the moment of order p is only defined for $p < \alpha$.

B.8 3rd-order continuous-time BLRPM properties

The computation of the integrals described in the text require that the following expected values of differential products be computed:

$$
\begin{aligned}
&E\left[dN(x-u)\,dN(x+y-v)\,dN(x+z-w)\right] = \\
&\quad \Big\{ (\lambda\mu_c)^3 + \lambda\mu_c\beta^2 e^{-\gamma(M-m)} \\
&\quad + (\lambda\mu_c)^2\beta\left[e^{-\gamma|z-w-y+v|} + e^{-\gamma|z-w+u|} + e^{-\gamma|y-v+u|}\right]\Big\}\,du\,dv\,dw
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[dN(x-u)\,dN(x+y-v)\right] &= cov\left[dN(x-u)\,dN(x+y-v)\right] + E\left[dN(x-u)\right]E\left[dN(x+y-v)\right] \\
&= \lambda\mu_c\left\{\delta(y-v+u) + h(y-v+u) - \lambda\mu_c\right\}du\,dv + (\lambda\mu_c)^2\,du\,dv \\
&= \lambda\mu_c h(y-v+u)\,du\,dv \\
&= \lambda\mu_c\left(\lambda\mu_c + \beta e^{-\gamma|y-v+u|}\right)du\,dv
\end{aligned}
$$

where: $m = \min(x-u, x+y-v, x+z-w)$ and $M = \max(x-u, x+y-v, x+z-w)$.

The computation of the integral then involves examining $M-m$ which is:

$$
M - m = \max(y-v+u, z-w+u, z-w-y+v, -y+v-u, -z+w-u, -z+w+y-v)
$$

for the different intervals of integration.

We find the following:

$$
\begin{array}{lll}
v \in [0, y+u] & \text{and } w \in [0, z-y+v] & \Rightarrow M-m = z-w+u \\
v \in [0, y+u] & \text{and } w \in [z-y+v, z+u] & \Rightarrow M-m = y-v+u \\
v \in [0, y+u] & \text{and } w \in [z+u, \infty) & \Rightarrow M-m = -z+w+y-v \\
v \in [y+u, \infty] & \text{and } w \in [0, z+u] & \Rightarrow M-m = z-w-y+v \\
v \in [y+u, \infty] & \text{and } w \in [z+u, z-y+v] & \Rightarrow M-m = -y+v-u \\
v \in [y+u, \infty] & \text{and } w \in [z-y+v, \infty) & \Rightarrow M-m = -z+w-u
\end{array}
$$

Lengthy but standard computations of integrals of exponential functions then lead to the following expression for $y \le 0$ and $z \le y$:

$$
\begin{aligned}
E[Y(x)Y(x+y)Y(x+z)] &= {\mu_x}^3\lambda\mu_c\left(\frac{{\mu_c}^2\lambda^2}{\eta^3} + 2\,\frac{e^{-\eta z}e^{-\eta y}\beta^2\gamma^2}{4\eta^5 - 5\eta^3\gamma^2 + \eta\gamma^4} - \frac{e^{-\gamma z}e^{-\eta y}\beta^2\gamma}{\eta\,(\eta-\gamma)\,(\eta+\gamma)\,(2\eta+\gamma)}\right. \\
&\quad + \frac{e^{-\eta y}\beta\gamma\lambda\mu_c}{-\eta^4+\eta^2\gamma^2} + \frac{e^{-\eta z}e^{\eta y}\beta\gamma\lambda\mu_c}{-\eta^4+\eta^2\gamma^2} + \frac{e^{-\eta z}e^{(\eta-\gamma)y}\beta^2\gamma}{-2\eta^4+\eta^3\gamma+2\eta^2\gamma^2-\eta\gamma^3} \\
&\quad - \frac{e^{-\eta z}e^{-\gamma y}\beta^2\gamma}{\eta\,(\eta-\gamma)\,(\eta+\gamma)\,(2\eta+\gamma)} - \frac{e^{-\gamma z}e^{\gamma y}\lambda\mu_c\beta}{(-\eta+\gamma)\,(\eta+\gamma)\,\eta} \\
&\quad \left. + \frac{e^{-\gamma z}\beta\,(\beta+\lambda\mu_c)}{\eta^3-\eta\gamma^2} + \frac{e^{-\eta z}\beta\gamma\lambda\mu_c}{-\eta^4+\eta^2\gamma^2} - \frac{e^{-\gamma y}\lambda\mu_c\beta}{(-\eta+\gamma)\,(\eta+\gamma)\,\eta}\right) \\
&\quad + \mu_x{\mu_x}^2\lambda\mu_c\left(-2\,\frac{e^{-\eta z}e^{-\eta y}\beta\gamma}{\eta\,(\eta-\gamma)\,(\eta+\gamma)} + \frac{e^{-\gamma z}e^{-\eta y}\beta}{(\eta-\gamma)\,(\eta+\gamma)} + \frac{e^{-\eta y}\lambda\mu_c}{\eta^2}\right. \\
&\quad + \frac{e^{-\eta z}e^{\eta y}\lambda\mu_c}{\eta^2} + \frac{e^{-\eta z}e^{(\eta-\gamma)y}\beta}{(\eta-\gamma)\,(\eta+\gamma)} + \frac{e^{-\eta z}e^{-\gamma y}\beta}{(\eta-\gamma)\,(\eta+\gamma)} \\
&\quad \left. + \frac{e^{-\eta z}\left(\beta\eta\gamma + \left(\gamma^2-\eta^2\right)\lambda\mu_c\right)}{-\eta^4+\eta^2\gamma^2}\right) + \frac{{\mu_x}^3\lambda\mu_c e^{-\eta z}}{\eta}
\end{aligned}
\tag{B.73}
$$

B.9 3rd-order moment of BLRPM aggregated process

This is the final expression of the third-order moment of rainfall depths at a time-scale h hours. If we define the following:

$$T1 = 12\gamma^7{\mu_x}^3\beta^2e^{h(\eta+\gamma)} - 48{\mu_x}^3e^{2h\eta}\eta^7\beta^2 + 72\gamma^7e^{h(2\eta+\gamma)}{\mu_{x3}}\eta^2 + 48\gamma\mu_x\mu_{x2}e^{h(2\eta+\gamma)}\beta\eta^7 + 24\gamma h{\mu_x}^3e^{h(2\eta+\gamma)}\beta\lambda\eta^7\mu_c - 36\mu_x\mu_{x2}\gamma^7h^2e^{h(2\eta+\gamma)}\lambda\mu_c\eta^3 - 24\gamma h{\mu_x}^3e^{h(2\eta+\gamma)}\eta^7\beta^2 + 24\mu_x\mu_{x2}\gamma^4he^{h(\eta+\gamma)}\beta\eta^5 + 24\mu_x\mu_{x2}\gamma^2e^{h(2\eta+\gamma)}\beta\eta^6 - 36\mu_x\mu_{x2}\gamma^3e^{h(2\eta+\gamma)}\beta\eta^5 - 6\gamma^8h{\mu_x}^3e^{h(2\eta+\gamma)}\beta\lambda\mu_c - 30\gamma^3h{\mu_x}^3e^{h(2\eta+\gamma)}\beta\lambda\eta^5\mu_c - 72\mu_x\mu_{x2}\gamma^6he^{h(2\eta+\gamma)}\beta\eta^3 + 6\gamma^5h{\mu_x}^3e^{h(2\eta+\gamma)}\beta\lambda\mu_c\eta^3 - 54\mu_x\mu_{x2}\gamma^5he^{h(2\eta+\gamma)}\lambda\mu_c\eta^4 - 84\gamma^2{\mu_x}^3e^{h(2\eta+\gamma)}\eta^5\beta^2 + 30\gamma^6h{\mu_x}^3e^{h(2\eta+\gamma)}\beta\lambda\mu_c\eta^2 - 36\gamma^5h{\mu_x}^3e^{h(2\eta+\gamma)}\beta^2\eta^3 + 24\mu_x\mu_{x2}\gamma^3he^{h(2\eta+\gamma)}\lambda\eta^6\mu_c + 54\gamma^3h{\mu_x}^3e^{h(2\eta+\gamma)}\eta^5\beta^2 + 36\mu_x\mu_{x2}\gamma^7he^{h(2\eta+\gamma)}\lambda\mu_c\eta^2 + 6\mu_x\mu_{x2}\gamma^5e^{h(2\eta+\gamma)}\beta\eta^3 + 6\gamma^7h{\mu_x}^3e^{h(2\eta+\gamma)}\eta\beta^2 - 24\mu_x\mu_{x2}\gamma^2e^{h(\eta+\gamma)}\beta\eta^6 + 117\mu_x\mu_{x2}\gamma^6e^{h(2\eta+\gamma)}\beta\eta^2 - 18\gamma^4{\mu_x}^3e^{h(\eta+\gamma)}\beta^2\eta^3 - 30\gamma^6h{\mu_x}^3e^{h(\eta+\gamma)}\beta\lambda\mu_c\eta^2 + 54\gamma^5e^{h(2\eta+\gamma)}h{\mu_{x3}}\eta^5 + 39\gamma^5{\mu_x}^3e^{h(2\eta+\gamma)}\beta^2\eta^2 - 36\gamma^7e^{h(2\eta+\gamma)}h{\mu_{x3}}\eta^3 - 24\gamma^3e^{h(2\eta+\gamma)}h{\mu_{x3}}\eta^7 - 12\gamma^9e^{h(2\eta+\gamma)}{\mu_{x3}} + 6\eta\gamma^9h{\mu_{x3}}e^{h(\eta+\gamma)}$$

$$T2 = -24\gamma^4h{\mu_x}^3e^{h(2\eta+\gamma)}\beta\lambda\mu_c\eta^4 + 6\mu_x\mu_{x2}\gamma^4e^{2h\eta}\beta\eta^4 - 30\mu_x\mu_{x2}\gamma^6he^{h(\eta+\gamma)}\beta\eta^3 - 48\mu_x\mu_{x2}\gamma^2he^{h(2\eta+\gamma)}\beta\eta^7 - 48\gamma\mu_x\mu_{x2}e^{2h\eta}\beta\eta^7 - 24\gamma h{\mu_x}^3e^{2h\eta}\eta^7\beta^2 + 30\gamma^3h{\mu_x}^3e^{2h\eta}\beta\lambda\eta^5\mu_c + 54\gamma^4{\mu_x}^3h^2e^{h(2\eta+\gamma)}\beta\lambda\eta^5\mu_c + 6\gamma^5{\mu_x}^3e^{2h\eta}\beta^2\eta^2 + 6\mu_x\mu_{x2}\gamma^8he^{h(\eta+\gamma)}\beta\eta - 36\mu_x\mu_{x2}\gamma^7he^{h(\eta+\gamma)}\lambda\mu_c\eta^2 - 138\mu_x\mu_{x2}\gamma^4e^{h(2\eta+\gamma)}\beta\eta^4 + 6\mu_x\mu_{x2}\gamma^9he^{h(\eta+\gamma)}\lambda\mu_c + 48{\mu_x}^3e^{h(2\eta+\gamma)}\eta^7\beta^2 + 30\gamma^3h{\mu_x}^3e^{2h\eta}\eta^5\beta^2 + 54\mu_x\mu_{x2}\gamma^5h^2e^{h(2\eta+\gamma)}\lambda\eta^5\mu_c - 24\mu_x\mu_{x2}\gamma^2e^{2h\eta}\beta\eta^6 + 9\gamma^5{\mu_x}^3h^3e^{h(2\eta+\gamma)}\lambda^2\eta^5{\mu_c}^2 + 36\mu_x\mu_{x2}\gamma^3e^{2h\eta}\beta\eta^5 + 24\mu_x\mu_{x2}\gamma^3e^{h(\eta+\gamma)}\beta\eta^5 + 6\mu_x\mu_{x2}\gamma^9h^2e^{h(2\eta+\gamma)}\lambda\eta\mu_c + 24\gamma^4h{\mu_x}^3e^{h(\eta+\gamma)}\beta\lambda\mu_c\eta^4 - 24\mu_x\mu_{x2}\gamma^3he^{h(\eta+\gamma)}\lambda\eta^6\mu_c - 132\mu_x\mu_{x2}\gamma^6e^{h(\eta+\gamma)}\beta\eta^2 - 6\mu_x\mu_{x2}\gamma^5e^{2h\eta}\beta\eta^3 - 6\gamma^5h{\mu_x}^3e^{2h\eta}\beta\lambda\mu_c\eta^3 + 54\mu_x\mu_{x2}\gamma^5he^{h(\eta+\gamma)}\lambda\mu_c\eta^4 - 24\gamma h{\mu_x}^3e^{2h\eta}\beta\lambda\eta^7\mu_c + 150\mu_x\mu_{x2}\gamma^4e^{h(\eta+\gamma)}\beta\eta^4 - 42\gamma^5{\mu_x}^3e^{h(\eta+\gamma)}\beta^2\eta^2 - 6\gamma^7{\mu_x}^3h^3e^{h(2\eta+\gamma)}\lambda^2{\mu_c}^2\eta^3 + \gamma^9{\mu_x}^3h^3e^{h(2\eta+\gamma)}\lambda^2{\mu_c}^2\eta + 6\gamma^8{\mu_x}^3h^2e^{h(2\eta+\gamma)}\beta\lambda\eta\mu_c - 6\gamma^5h{\mu_x}^3e^{2h\eta}\beta^2\eta^3 + 12\mu_x\mu_{x2}\gamma^8he^{h(2\eta+\gamma)}\beta\eta - 6\mu_x\mu_{x2}\gamma^9he^{h(2\eta+\gamma)}\lambda\mu_c - 6\mu_x\mu_{x2}\gamma^5e^{h(\eta+\gamma)}\beta\eta^3 - 24\eta^5\mu_{x2}\mu_x\beta\gamma^3e^{h\eta} - 12\eta^4\mu_{x2}\mu_x\gamma^4\beta e^{h\gamma} - 6\eta^4\mu_{x2}\mu_x\gamma^4\beta e^{h\eta} + 6\eta^3\gamma^5\mu_{x2}\mu_x\beta e^{h\eta} - 3\mu_{x2}\mu_x\gamma^8\beta e^{h\gamma} + 24\eta^6\mu_{x2}\mu_x\beta\gamma^2e^{h\eta} + 15\eta^2\mu_{x2}\mu_x\gamma^6\beta e^{h\gamma} - 3\gamma^7{\mu_x}^3\beta^2e^{h\gamma}$$

$$T3 = 18\eta^3\gamma^4{\mu_x}^3\beta^2e^{h\eta} - 12\eta^4\gamma^3{\mu_x}^3\beta^2e^{h\eta} - 6\eta^2\gamma^5{\mu_x}^3\beta^2e^{h\eta} + 3\eta^2\gamma^5{\mu_x}^3\beta^2e^{h\gamma} - 9\gamma^7e^{h(2\eta+\gamma)}{\mu_x}^3\beta^2 + 108\eta^4\gamma^5\mu_{x3}e^{h(\eta+\gamma)} + 48\gamma^3e^{h(2\eta+\gamma)}{\mu_{x3}}\eta^6 - 72\eta^2\gamma^7\mu_{x3}e^{h(\eta+\gamma)} - 48\eta^6\mu_{x3}\gamma^3e^{h(\eta+\gamma)} + 84\gamma^2{\mu_x}^3e^{2h\eta}\eta^5\beta^2 + 18\gamma^4{\mu_x}^3e^{h(2\eta+\gamma)}\beta^2\eta^3 + 24\mu_{x2}\mu_x\gamma^8\beta e^{h(\eta+\gamma)} + 54\eta^5\gamma^5h{\mu_{x3}}e^{h(\eta+\gamma)} - 24\eta^7h{\mu_{x3}}\gamma^3e^{h(\eta+\gamma)} - 36\eta^3\gamma^7h{\mu_{x3}}e^{h(\eta+\gamma)} - 21\gamma^8e^{h(2\eta+\gamma)}\mu_{x2}\mu_x\beta + 6\gamma^9e^{h(2\eta+\gamma)}h{\mu_{x3}}\eta + 12\gamma^3{\mu_x}^3e^{h(\eta+\gamma)}\beta^2\eta^4 + 12\gamma^3{\mu_x}^3e^{2h\eta}\beta^2\eta^4 - 18\gamma^4{\mu_x}^3e^{2h\eta}\beta^2\eta^3 - 24\gamma^2{\mu_x}^3h^2e^{h(2\eta+\gamma)}\beta\lambda\eta^7\mu_c - 12\gamma^3{\mu_x}^3e^{h(2\eta+\gamma)}\beta^2\eta^4 - 108\gamma^5e^{h(2\eta+\gamma)}{\mu_{x3}}\eta^4 + 6\gamma^8h{\mu_x}^3e^{h(\eta+\gamma)}\beta\lambda\mu_c - 4\gamma^3{\mu_x}^3h^3e^{h(2\eta+\gamma)}\lambda^2\eta^7{\mu_c}^2 + 108\mu_x\mu_{x2}\gamma^4he^{h(2\eta+\gamma)}\beta\eta^5 + 12\gamma^9{\mu_{x3}}e^{h(\eta+\gamma)} - 24\mu_x\mu_{x2}\gamma^3h^2e^{h(2\eta+\gamma)}\lambda\eta^7\mu_c - 36\gamma^6{\mu_x}^3h^2e^{h(2\eta+\gamma)}\beta\lambda\mu_c\eta^3$$

and

$$\chi = \frac{\lambda\mu_c e^{-h(2\eta+\gamma)}}{(\eta^2+2\gamma\eta+\gamma^2)(\gamma^4-2\eta\gamma^3-3\eta^2\gamma^2+8\eta^3\gamma-4\eta^4)\gamma^3\eta^4}$$

the third order moment is:

$$E\left[\left(Y_i^{(h)}\right)^3\right] = \chi(T1+T2+T3)$$

B.10 3rd-order continuous-time DD1 properties

We define:

$$
\begin{aligned}
N_1 \;=\;& 1/6\mu\lambda F^3\eta^3\Big(-18\eta\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\gamma C^2\beta^2 e^{\gamma z}\\
&+96\eta C^3\beta^2 e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}-48\mu^2\lambda^2\eta^3\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2 Ce^{\gamma y}e^{\gamma z}\\
&-12C^4\beta^2 e^{Cz}e^{\eta z}e^{\gamma y}+16\eta C^3\beta^2 e^{\gamma y}e^{\gamma z}\\
&-12\mu^2\lambda^2\eta^4\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2 e^{\gamma y}e^{\gamma z}-12\eta^4\left(e^{Cy}\right)^2\\
&\left(e^{\eta y}\right)^2\beta^2 e^{\gamma z}-24\mu\lambda\eta^4\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\beta e^{\gamma z}-12\mu^2\lambda^2\eta^4 e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}\\
&-5\mu^2\lambda^2\gamma^2C^2e^{\gamma y}e^{\gamma z}+18\eta^2\gamma C\beta^2 e^{\gamma y}e^{\gamma z}+6\gamma^2C^2\beta^2 e^{Cz}e^{\eta z}e^{\gamma y}\\
&-24\mu\lambda\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\beta C^4e^{\gamma z}+16\mu^2\lambda^2\eta^3Ce^{\gamma y}e^{\gamma z}\\
&-96\mu\lambda\eta\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\beta C^3e^{\gamma z}\\
&-48\mu^2\lambda^2\eta^3Ce^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}-12\mu\lambda\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\beta C^4e^{\gamma y}e^{\gamma z}\\
&-12\mu\lambda\eta^4\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\beta e^{\gamma y}e^{\gamma z}-3\mu^2\lambda^2\left(e^{Cy}\right)^2\\
&\left(e^{\eta y}\right)^2\gamma^4e^{\gamma y}e^{\gamma z}-48\mu\lambda\eta\beta C^3e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}+144\eta^2\beta^2C^2e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}\\
&+96\mu\lambda\eta^3\beta Ce^{Cz}e^{\eta z}e^{Cy}e^{\eta y}e^{\gamma z}+96\mu\lambda\eta\beta C^3e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}e^{\gamma z}\\
&-48\mu\lambda\eta\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\beta C^3e^{\gamma y}e^{\gamma z}-6\mu\lambda\gamma^2\beta C^2e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}e^{\gamma z}\\
&-36\mu\lambda\eta^2\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\gamma\beta Ce^{\gamma y}e^{\gamma z}+24\mu\lambda\left(e^{\gamma y}\right)^2\beta C^4e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}\\
&-48\mu\lambda\eta^3\beta Ce^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}-12\mu\lambda\gamma C^3\beta e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}\\
&+24\mu\lambda\beta C^4e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}e^{\gamma z}+96\mu\lambda\eta\beta C^3e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}\\
&-72\mu\lambda\eta^2\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\beta C^2e^{\gamma y}e^{\gamma z}+2\gamma^2C^2\beta^2e^{\gamma y}e^{\gamma z}\\
&-12\mu\lambda\eta^3\gamma\beta e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}-12\mu\lambda\eta^3\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\gamma\beta e^{\gamma y}e^{\gamma z}\\
&-12\eta\gamma^2C\beta^2e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}+24\mu\lambda\eta^4\left(e^{\gamma y}\right)^2\beta e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}\\
&-12\mu\lambda\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2\gamma C^3\beta e^{\gamma y}e^{\gamma z}\\
&+24\mu\lambda\eta^4\beta e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}e^{\gamma z}\\
&+30\mu^2\lambda^2\eta\gamma^2Ce^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}-6\eta^2\gamma^2\beta^2e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}\\
&+96\mu\lambda\eta\left(e^{\gamma y}\right)^2\beta C^3e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}-72\mu^2\lambda^2\eta^2\left(e^{Cy}\right)^2\left(e^{\eta y}\right)^2C^2e^{\gamma y}e^{\gamma z}\\
&-36\mu\lambda\eta^2\gamma\beta Ce^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}+3\mu\lambda\eta^2\gamma^2\beta e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}\\
&+24\eta^4\beta^2e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}+144\mu\lambda\eta^2\beta C^2e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}\\
&+144\mu\lambda\eta^2\left(e^{\gamma y}\right)^2\beta C^2e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}+12\mu\lambda\eta\gamma^2\beta Ce^{Cz}e^{\eta z}e^{\gamma y}\\
&-6\mu\lambda\gamma^2\beta C^2e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}-12\mu\lambda\eta\gamma^2\beta Ce^{Cz}e^{\eta z}e^{Cy}e^{\eta y}e^{\gamma z}
\end{aligned}
$$

$$
\begin{aligned}
&-6\mu\lambda\eta\gamma^2\beta Ce^{\gamma y}e^{\gamma z}+3\mu\lambda\eta\gamma^3\beta e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}+3\mu\lambda\gamma^3\beta Ce^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}\\
&+3\mu\lambda\eta^2\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma^2\beta e^{\gamma y}e^{\gamma z}+3\mu\lambda\gamma^2\beta C^2e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}\\
&-12\mu\lambda\eta\,(e^{\gamma y})^2\gamma^2\beta Ce^{Cz}e^{\eta z}e^{Cy}e^{\eta y}-96\mu\lambda\eta\beta C^3e^{Cz}e^{\eta z}e^{\gamma y}\\
&+6\mu\lambda\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma^2\beta C^2e^{\gamma z}+36\mu\lambda\eta^2\gamma\beta Ce^{\gamma y}e^{\gamma z}\\
&-96\mu\lambda\eta^3\left(e^{Cy}\right)^2(e^{\eta y})^2\beta Ce^{\gamma z}-12\left(e^{Cy}\right)^2(e^{\eta y})^2C^4\beta^2e^{\gamma z}\\
&-6\eta^3\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma\beta^2e^{\gamma z}+36\mu\lambda\eta\gamma\beta C^2e^{\gamma y}e^{\gamma z}\\
&+15\mu^2\lambda^2\eta^2\gamma^2e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}+6\mu\lambda\eta^2\gamma^2\beta e^{Cz}e^{\eta z}e^{\gamma y}\\
&+6\mu^2\lambda^2\gamma^4e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}e^{\gamma z}-6\mu\lambda\eta^2\gamma^2\beta e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}\\
&-72\eta^2\left(e^{Cy}\right)^2(e^{\eta y})^2\beta^2C^2e^{\gamma z}+15\mu^2\lambda^2\gamma^2C^2e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}\\
&+16\eta^3C\beta^2e^{\gamma y}e^{\gamma z}-6\mu\lambda\eta^2\gamma^2\beta e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}e^{\gamma z}\\
&+3\mu\lambda\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma^2\beta C^2e^{\gamma y}e^{\gamma z}\\
&+6\mu\lambda\eta\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma^2\beta Ce^{\gamma y}e^{\gamma z}\\
&+6\mu\lambda\eta\gamma^2\beta Ce^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}+4\mu^2\lambda^2C^4e^{\gamma y}e^{\gamma z}\\
&-6\mu\lambda\,(e^{\gamma y})^2\gamma^2\beta C^2e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}+30\mu^2\lambda^2\eta\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma^2Ce^{\gamma y}e^{\gamma z}\\
&-6\gamma^2C^2\beta^2e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}\\
&+12\mu\lambda\eta\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma^2\beta Ce^{\gamma z}-12\mu\lambda\eta\gamma^2\beta Ce^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}\\
&-36\mu\lambda\eta\gamma\beta C^2e^{Cz}e^{\eta z}e^{\gamma y}e^{\gamma z}\\
&-48\eta^3\left(e^{Cy}\right)^2(e^{\eta y})^2C\beta^2e^{\gamma z}+24\mu\lambda\beta C^4e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}\\
&+24\mu^2\lambda^2\eta^4e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}e^{\gamma z}\\
&+24\mu\lambda\eta^4\beta e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}+15\mu^2\lambda^2\eta^2\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma^2e^{\gamma y}e^{\gamma z}\\
&-30\mu^2\lambda^2\eta^2\gamma^2e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}e^{\gamma z}-12\eta^4\beta^2e^{Cz}e^{\eta z}e^{\gamma y}\\
&-30\mu^2\lambda^2\gamma^2C^2e^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}e^{\gamma z}\\
&-60\mu^2\lambda^2\eta\gamma^2Ce^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}e^{\gamma z}\\
&-36\mu\lambda\eta\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma\beta C^2e^{\gamma y}e^{\gamma z}\\
&+6\eta^2\gamma^2\beta^2e^{Cz}e^{\eta z}e^{\gamma y}+96\mu\lambda\eta^3\beta Ce^{Cz}e^{\eta z}e^{\gamma y}e^{Cy}e^{\eta y}\\
&-48\eta C^3\beta^2e^{Cz}e^{\eta z}e^{\gamma y}-72\eta^2\beta^2C^2e^{Cz}e^{\eta z}e^{\gamma y}\\
&-6\mu\lambda\eta^2\,(e^{\gamma y})^2\gamma^2\beta e^{Cz}e^{\eta z}e^{Cy}e^{\eta y}\\
&+15\mu^2\lambda^2\left(e^{Cy}\right)^2(e^{\eta y})^2\gamma^2C^2e^{\gamma y}e^{\gamma z}\\
&-48\mu\lambda\eta^3\left(e^{Cy}\right)^2(e^{\eta y})^2\beta Ce^{\gamma y}e^{\gamma z}-6\eta^3\gamma\beta^2e^{Cz}e^{\eta z}e^{\gamma y}-6\gamma C^3\beta^2e^{Cz}e^{\eta z}e^{\gamma y}\\
&+18\eta\gamma C^2\beta^2e^{\gamma y}e^{\gamma z}+48\mu\lambda\eta\beta C^3e^{\gamma y}e^{\gamma z}
\end{aligned}
$$

$$
\begin{aligned}
&-12\mu^2\lambda^2C^4e^{C_z}e^{\eta_z}e^{\gamma_y}e^{\gamma_z}+6\mu\lambda\gamma^2\beta C^2e^{C_z}e^{\eta_z}e^{\gamma_y}\\
&+96\mu\lambda\eta^3\left(e^{\gamma_y}\right)^2\beta Ce^{C_z}e^{\eta_z}e^{C_y}e^{\eta_y}-12\mu\lambda\eta^4\beta e^{C_z}e^{\eta_z}e^{\gamma_y}e^{\gamma_z}\\
&+24C^4\beta^2e^{C_z}e^{\eta_z}e^{\gamma_y}e^{C_y}e^{\eta_y}-6\left(e^{C_y}\right)^2\left(e^{\eta_y}\right)^2\gamma C^3\beta^2e^{\gamma_z}\\
&-18\eta^2\gamma C\beta^2e^{C_z}e^{\eta_z}e^{\gamma_y}+12\eta\gamma^2C\beta^2e^{C_z}e^{\eta_z}e^{\gamma_y}\\
&+12\mu\lambda\beta C^4e^{\gamma_y}e^{\gamma_z}+6\mu\lambda\eta^2\left(e^{C_y}\right)^2\left(e^{\eta_y}\right)^2\gamma^2\beta e^{\gamma_z}\\
&+96\mu^2\lambda^2\eta^3Ce^{C_z}e^{\eta_z}e^{\gamma_y}e^{C_y}e^{\eta_y}e^{\gamma_z}\\
&-48\eta^3C\beta^2e^{C_z}e^{\eta_z}e^{\gamma_y}+3\mu\lambda\eta\left(e^{C_y}\right)^2\left(e^{\eta_y}\right)^2\gamma^3\beta e^{\gamma_y}e^{\gamma_z}\\
&+96\mu^2\lambda^2\eta C^3e^{C_z}e^{\eta_z}e^{\gamma_y}e^{C_y}e^{\eta_y}e^{\gamma_z}+144\mu^2\lambda^2\eta^2C^2e^{C_z}e^{\eta_z}e^{\gamma_y}e^{C_y}e^{\eta_y}e^{\gamma_z}\\
&+24\mu^2\lambda^2\eta^2C^2e^{\gamma_y}e^{\gamma_z}+24\mu^2\lambda^2C^4e^{C_z}e^{\eta_z}e^{\gamma_y}e^{C_y}e^{\eta_y}e^{\gamma_z}\\
&+144\mu\lambda\eta^2\beta C^2e^{C_z}e^{\eta_z}e^{C_y}e^{\eta_y}e^{\gamma_z}-48\eta\left(e^{C_y}\right)^2\left(e^{\eta_y}\right)^2C^3\beta^2e^{\gamma_z}\\
&-3\mu\lambda\gamma^3\beta Ce^{\gamma_y}e^{\gamma_z}-48\mu^2\lambda^2\eta C^3e^{C_z}e^{\eta_z}e^{\gamma_y}e^{\gamma_z}\\
&-144\mu\lambda\eta^2\beta C^2e^{C_z}e^{\eta_z}e^{\gamma_y}-12\mu^2\lambda^2\left(e^{C_y}\right)^2\left(e^{\eta_y}\right)^2C^4e^{\gamma_y}e^{\gamma_z}\\
&+24\eta^2\beta^2C^2e^{\gamma_y}e^{\gamma_z}-3\mu^2\lambda^2\gamma^4e^{C_z}e^{\eta_z}e^{\gamma_y}e^{\gamma_z}\\
&-24\mu\lambda\beta C^4e^{C_z}e^{\eta_z}e^{\gamma_y}+4\eta\gamma^2C\beta^2e^{\gamma_y}e^{\gamma_z}\\
&-72\mu^2\lambda^2\eta^2C^2e^{C_z}e^{\eta_z}e^{\gamma_y}e^{\gamma_z}-10\mu^2\lambda^2\eta\gamma^2Ce^{\gamma_y}e^{\gamma_z}\\
&+96\eta^3C\beta^2e^{C_z}e^{\eta_z}e^{\gamma_y}e^{C_y}e^{\eta_y}-12\mu\lambda\beta C^4e^{C_z}e^{\eta_z}e^{\gamma_y}e^{\gamma_z}\\
&+3\mu\lambda\left(e^{C_y}\right)^2\left(e^{\eta_y}\right)^2\gamma^3\beta Ce^{\gamma_y}e^{\gamma_z}+16\mu^2\lambda^2\eta C^3e^{\gamma_y}e^{\gamma_z}\\
&-18\eta^2\left(e^{C_y}\right)^2\left(e^{\eta_y}\right)^2\gamma C\beta^2e^{\gamma_z}-18\eta\gamma C^2\beta^2e^{C_z}e^{\eta_z}e^{\gamma_y}\\
&+48\mu\lambda\eta^3\beta Ce^{\gamma_y}e^{\gamma_z}+6\gamma C^3\beta^2e^{\gamma_y}e^{\gamma_z}+12\mu\lambda\gamma C^3\beta e^{\gamma_y}e^{\gamma_z}\\
&-3\mu\lambda\gamma^2\beta C^2e^{\gamma_y}e^{\gamma_z}-24\mu\lambda\eta^4\beta e^{C_z}e^{\eta_z}e^{\gamma_y}+72\mu\lambda\eta^2\beta C^2e^{\gamma_y}e^{\gamma_z}\\
&-48\mu^2\lambda^2\eta\left(e^{C_y}\right)^2\left(e^{\eta_y}\right)^2C^3e^{\gamma_y}e^{\gamma_z}-144\mu\lambda\eta^2\left(e^{C_y}\right)^2\left(e^{\eta_y}\right)^2\beta C^2e^{\gamma_z}\\
&-72\mu\lambda\eta^2\beta C^2e^{C_z}e^{\eta_z}e^{\gamma_y}e^{\gamma_z}-96\mu\lambda\eta^3\beta Ce^{C_z}e^{\eta_z}e^{\gamma_y}+4C^4\beta^2e^{\gamma_y}e^{\gamma_z}\\
&+4\eta^4\beta^2e^{\gamma_y}e^{\gamma_z}+6\eta^3\gamma\beta^2e^{\gamma_y}e^{\gamma_z}+2\eta^2\gamma^2\beta^2e^{\gamma_y}e^{\gamma_z}\\
&+4\mu^2\lambda^2\eta^4e^{\gamma_y}e^{\gamma_z}-5\mu^2\lambda^2\eta^2\gamma^2e^{\gamma_y}e^{\gamma_z}+\mu^2\lambda^2\gamma^4e^{\gamma_y}e^{\gamma_z}+12\mu\lambda\eta^4\beta e^{\gamma_y}e^{\gamma_z}\\
&-3\mu\lambda\eta^2\gamma^2\beta e^{\gamma_y}e^{\gamma_z}-3\mu\lambda\eta\gamma^3\beta e^{\gamma_y}e^{\gamma_z}+12\mu\lambda\eta^3\gamma\beta e^{\gamma_y}e^{\gamma_z}\Big)
\end{aligned}
$$

and

$$
\begin{aligned}
D_1 \;=\;& e^{C_y}e^{\eta_y}e^{C_z}e^{\eta_z}e^{\gamma_z}e^{\gamma_y}\Big(-280C^5\eta^3\gamma^2+6\gamma^4C^5\eta\\
&-40\eta^7\gamma^2C-40\gamma^2C^7\eta+15\gamma^4C^4\eta^2+6\gamma^4C\eta^5-350C^4\eta^4\gamma^2\\
&+15\gamma^4C^2\eta^4-280C^3\eta^5\gamma^2-140C^2\eta^6\gamma^2-140\gamma^2C^6\eta^2\\
&+40\eta^9C+180C^2\eta^8-5C^8\gamma^2+\gamma^4C^6+840C^6\eta^4+40C^9\eta+480C^7\eta^3+4C^{10}+480C^3\eta^7\\
&+1008C^5\eta^5+20\gamma^4C^3\eta^3+840C^4\eta^6+180C^8\eta^2+\gamma^4\eta^6-5\eta^8\gamma^2+4\eta^{10}\Big)
\end{aligned}
$$

which yields the first of the 14 integrals as:

$$
I_1 = N_1/D_1 \tag{B.74}
$$

B.11 Probability dry approximation for LR model

B.11.1 Integral term

For the random-parameter Bartlett-Lewis model, the exact probability that an arbitrary interval is dry depends on an integral of the form

$$I(\phi,\kappa) = \int_0^1 t^{\phi-1}(1-t)\,e^{\kappa t}\,dt \tag{B.75}$$

(see equation B.46). This cannot be evaluated analytically, although it is (almost) a 'standard' integral, in the sense that it has a name, since $I(\phi,\kappa)/B(\phi,2)$ is a confluent hypergeometric function — $M(\phi,2+\phi,\kappa)$ in the notation of Abramowitz and Stegun (1965, equation 13.2.1). $B(a,b)$ here is the beta function. There appear not to be any nice ways of evaluating $I(\phi,\kappa)$ or relating it to other special functions that can be calculated easily — I've checked everything in Abramowitz and Stegun (1965) and in Gradshteyn and Ryzhik (1980).

Rodriguez-Iturbe et al. (1987) approximated the integral using a third-order series expansion that is valid when ϕ and κ are both small (i.e. substantially less than 1). This approach runs through all subsequent developments of the Bartlett-Lewis model, and is still used in our fitting programs. However, the requirement for κ and ϕ to be small seems to have gone largely unnoticed (or been forgotten) in our fitting work. Now that we're thinking about fitting models to lots of datasets, it may be worth examining. For example, some recent problems with the SCE fitting code 'blowing up' for some datasets appear to be caused solely by the failure of this approximation in a region of the parameter space that the algorithm was exploring. Moreover, the exact magnitude of the approximation error for any given κ and ϕ is not known, which makes me feel a bit uncomfortable ...

In view of this, it may be worth exploring alternative means of evaluating the integral. A possible solution is to use standard quadrature methods; however, since the integrand becomes infinite at $t=0$ for $\phi<1$, this may be delicate. Instead, consider expanding the $e^{\kappa t}$ term, to get

$$\begin{aligned} I(\phi,\kappa) = \int_0^1 t^{\phi-1}(1-t)\sum_{j=0}^{\infty}\frac{(\kappa t)^j}{j!}dt &= \sum_{j=0}^{\infty}\frac{\kappa^j}{j!}\int_0^1 t^{j+\phi-1}(1-t)\,dt \\ &= \sum_{j=0}^{\infty}\frac{\kappa^j}{j!}B(j+\phi,2)\ . \end{aligned}$$

This suggests truncating the infinite sum at a suitably large value, say M, and approximating the integral by

$$\tilde{I}_M(\phi,\kappa) = \sum_{j=0}^{M}\frac{\kappa^j}{j!}B(j+\phi,2)\ . \tag{B.76}$$

The point about this is that standard algorithms exist for calculating the Beta function to a high degree of accuracy (it can be expressed in terms of gamma functions — $B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$), and these are readily available in both R and FORTRAN (there's a FORTRAN routine to evaluate gamma functions in file `OURPROGS/rec_math.f` on argos).

The error in approximating I with $\tilde{I}_M$ is

$$I(\phi,\kappa)-\tilde{I}_M(\phi,\kappa) = \int_0^1 t^{\phi-1}(1-t)\left[e^{\kappa t}-\sum_{j=0}^{M}\frac{(\kappa t)^j}{j!}\right]dt = \int_0^1 t^{\phi-1}(1-t)\sum_{j=M+1}^{\infty}\frac{(\kappa t)^j}{j!}dt\ . \tag{B.77}$$

Each term in the infinite sum is non-negative and increasing in t. Therefore, over the range of the integral, it takes its maximum value at $t=1$. The maximum value of the sum is

$$\delta_M(\kappa) = \sum_{j=M+1}^{\infty}\frac{\kappa^j}{j!} = e^{\kappa}-\sum_{j=0}^{M}\frac{\kappa^j}{j!} \tag{B.78}$$

which is, again, easily evaluated providing M is not too large. We therefore have

$$0 < I(\phi,\kappa)-\tilde{I}_M(\phi,\kappa) < \delta_M(\kappa)\int_0^1 t^{\phi-1}(1-t)\,dt\ = \delta_M(\kappa)B(\phi,2)\ . \tag{B.79}$$

Therefore, for any value of M we can calculate an upper bound on the approximation error. This enables us to find a value of M that will approximate the integral to any desired accuracy. A pragmatic criterion, for example, may be to choose M such that $\delta_M(\kappa)B(\phi,2) < 0.01 \times \tilde{I}_M(\kappa,\phi)$.

From equation (B.79), it is clear that $\tilde{I}_M$ will always underestimate I. It is natural to ask whether a correction can be made for this, to improve the approximation. From equation (B.77) we have

$$I(\phi,\kappa) - \tilde{I}_M(\phi,\kappa) = \int_0^1 \sum_{j=M+1}^{\infty} \frac{\kappa^j}{j!} t^{j+\phi-1} dt \; - \int_0^1 \sum_{j=M+1}^{\infty} \frac{\kappa^j}{j!} t^{j+\phi} dt \; .$$

Each of the integrands here increases monotonically from 0 to $\delta_M(\kappa)$ as t ranges from 0 to 1. Moreover, since $t < 1$ throughout the range of integration, the $j = M+1$ term is the largest in each sum. This suggests approximating the error by taking just the $M+1$ term from each sum and scaling it to match the correct value at each end of the range of integration. This yields the approximation

$$\delta_M(\kappa) \int_0^1 \left(t^{M+\phi} - t^{M+\phi+1}\right) dt = \frac{\delta_M(\kappa)}{(M+\phi+1)(M+\phi+2)}$$

which, in turn, suggests that

$$\hat{I}_M(\phi,\kappa) = \tilde{I}_M(\phi,\kappa) + \frac{\delta_M(\kappa)}{(M+\phi+1)(M+\phi+2)} \tag{B.80}$$

will be an improved estimate of $I(\phi,\kappa)$. Note that the improvement is obtained almost 'free of charge' — it depends only on M (which is known) and upon $\delta_M(\kappa)$ (which has already been calculated to determine the accuracy of $\tilde{I}_M$).

Numerical investigation

To assess the adequacy of these approximations, some numerical experiments have been carried out for several values of κ and ϕ. It is of particular interest to determine how large M needs to be to achieve a specified degree of accuracy. Define $\tilde{M}_\alpha$ and $\hat{M}_\alpha$ to be the values required to obtain a relative error of less than $100\alpha\%$ for particular values of κ and ϕ, using $\tilde{I}_M$ and $\hat{I}_M$ respectively. Table B.2 shows the values of $\tilde{M}_{0.01}$, $\hat{M}_{0.01}$, $\tilde{M}_{10^{-6}}$ and $\hat{M}_{10^{-6}}$ for values of ϕ and κ between 0 and 10. In all cases, the 'exact' expression was calculated as $\tilde{I}_{100}(\phi,\kappa)$. As a check on the adequacy of this (and on the overall accuracy of the theory and programming!), the results for $\phi = 0.1, \kappa = 0.1$ and for $\phi = 0.1, \kappa = 1$ have also been evaluated manually using Table 13.1 of Abramowitz and Stegun (1965).

Table B.1 shows that for the values of ϕ and κ likely to be encountered in rainfall modelling applications, a small value of M yields very high accuracy. In such applications it would be unusual to find values in excess of 1) For example, using $\hat{I}_M$, when $\phi = \kappa = 1$ a relative error of less than 1% can be achieved with $M = 2$. Indeed, $M = 6$ is sufficient to ensure a relative error of less than 10^{-6} in this case. As expected, $\hat{I}_M$ is more accurate than $\tilde{I}_M$ and hence is preferable (since it is no more expensive to compute). The magnitude of this improvement can be illustrated by comparing $\tilde{I}_0(1,1) = 0.6667$ and $\hat{I}_0(1,1) = 0.7265$ (not shown in Table B.2) with the actual value of $I(1,1)$, which is 0.7183. In this case, $\hat{I}$ improves considerably over $\tilde{I}$. The magnitude of the error here suggests that taking M as small as zero may adequate for some applications, if using $\hat{I}$.

B.11.2 Mean duration of a storm

Derivation of the approximation

The theoretical expression for the mean storm duration which is required for the estimation of the proportion of dry periods μ_T is (Onof, 1992):

$$\mu_T = \frac{\phi\nu}{\alpha - 1} \int_0^1 dv \int_0^1 dt \, v^{-1} t^{\phi-1} [1 - (1-vt)e^{-\kappa v(1-t)}] + \frac{\phi^{-1}\nu}{\alpha - 1} \tag{B.81}$$

Table B.1 Exact values of $I(\phi,\kappa)$, together with values of M required to achieve relative errors of less than 10^{-2} and 10^{-6} respectively.

ϕ	κ	$I(\phi,\kappa)$	$\tilde{M}_{0.01}$	$\hat{M}_{0.01}$	$\tilde{M}_{10^{-6}}$	$\hat{M}_{10^{-6}}$
0.01	0	99.0099	0	0	0	0
0.10	0	9.0909	0	0	0	0
1	0	0.5000	0	0	0	0
10	0	0.0091	0	0	0	0
0.01	0.01	99.0148	0	0	1	0
0.10	0.01	9.0952	0	0	1	1
1	0.01	0.5017	0	0	2	1
10	0.01	0.0092	0	0	2	1
0.01	0.10	99.0600	0	0	2	1
0.10	0.10	9.1350	0	0	3	2
1	0.10	0.5171	1	0	3	2
10	0.10	0.0099	1	0	4	2
0.01	1	99.6013	0	0	5	4
0.10	1	9.6160	1	1	6	5
1	1	0.7183	3	2	7	6
10	1	0.0210	4	2	8	6
0.01	10	385.9201	15	13	26	23
0.10	10	288.2351	16	13	26	23
1	10	220.1547	16	13	26	23
10	10	56.3963	17	13	27	24

Since approximating this integral involves expanding the exponential term as the sum of a series, a change of variable in the integral in t would be preferable (the new variable is $1-t$). This yields:

$$\mu_T = \frac{\phi\nu}{\alpha-1}\int_0^1 dv \int_0^1 dt\, v^{-1}(1-t)^{\phi-1}[1-(1-v(1-t))e^{-\kappa\nu t}] + \frac{\phi^{-1}\nu}{\alpha-1} \tag{B.82}$$

By using the Taylor expansion of the exponential function, the double integral inside this expression can be written as:

$$\begin{aligned} I &= \int_0^1 dv \int_0^1 dt\, v^{-1}(1-t)^{\phi-1}\left[-\frac{\sum_{j=1}^{\infty}(-\kappa\nu t)^j}{j!} + v\frac{\sum_{j=0}^{\infty}(-\kappa\nu t)^j}{j!} - vt\frac{\sum_{j=0}^{\infty}(-\kappa\nu t)^j}{j!}\right] \\ &= \int_0^1 dv \int_0^1 dt\,(1-t)^{\phi-1}\left[-\frac{\sum_{j=1}^{\infty}(-\kappa\nu t)^j}{\nu j!} + \frac{\sum_{j=0}^{\infty}(-\kappa\nu t)^j}{j!} - t\frac{\sum_{j=0}^{\infty}(-\kappa\nu t)^j}{j!}\right] \end{aligned}$$

which is the sum of three terms:

$$\begin{aligned} I_1 &= -\sum_{j=1}^{\infty}\frac{(-\kappa)^j}{j!}\int_0^1 v^{j-1}dv\int_0^1 (1-t)^{\phi-1}t^j\,dt \\ I_2 &= \sum_{j=0}^{\infty}\frac{(-\kappa)^j}{j!}\int_0^1 v^j dv\int_0^1 (1-t)^{\phi-1}t^j\,dt \\ I_3 &= -\sum_{j=0}^{\infty}\frac{(-\kappa)^j}{j!}\int_0^1 v^j dv\int_0^1 (1-t)^{\phi-1}t^{j+1}\,dt \end{aligned}$$

Having thus separated the variables of integration, the expressions simplify in terms of beta functions:

$$I_1 = -\sum_{j=1}^{\infty} \frac{(-\kappa)^j}{jj!} B(j+1,\phi) \tag{B.83}$$

$$I_2 = \sum_{j=0}^{\infty} \frac{(-\kappa)^j}{(j+1)!} B(j+1,\phi) \tag{B.84}$$

$$I_3 = -\sum_{j=0}^{\infty} \frac{(-\kappa)^j}{(j+1)!} B(j+2,\phi) \tag{B.85}$$

The sum $I = I_1 + I_2 + I_3$ can therefore be approximated by $I_M = I_{1_M} + I_{2_M} + I_{3_M}$ where:

$$I_{1_M} = -\sum_{j=1}^{M} \frac{(-\kappa)^j}{jj!} B(j+1,\phi) \tag{B.86}$$

$$I_{2_M} = \sum_{j=0}^{M} \frac{(-\kappa)^j}{(j+1)!} B(j+1,\phi) \tag{B.87}$$

$$I_{3_M} = -\sum_{j=0}^{M} \frac{(-\kappa)^j}{(j+1)!} B(j+2,\phi) \tag{B.88}$$

Note that, with computational efficiency in mind, I_M can be rewritten so as to minimise the calls a program has to make to the Beta function. Thus:

$$\begin{aligned} I_{1_M} &= -\sum_{j=1}^{M} \frac{(-\kappa)^j}{jj!} B(j+1,\phi) \\ I_{2_M} &= \sum_{j=0}^{M} \frac{(-\kappa)^j}{(j+1)!} B(j+1,\phi) \\ I_{3_M} &= -\sum_{j=1}^{M+1} \frac{(-\kappa)^{j-1}}{j!} B(j+1,\phi) \end{aligned}$$

this suggests we take the following as approximation:

$$I'_M = -\sum_{j=1}^{M} \frac{(-\kappa)^j}{jj!} B(j+1,\phi) + \sum_{j=0}^{M} \frac{(-\kappa)^j}{(j+1)!} B(j+1,\phi) - \sum_{j=1}^{M} \frac{(-\kappa)^{j-1}}{j!} B(j+1,\phi)$$

which yields:

$$I'_M = B(1,\phi) + \sum_{j=1}^{M} (-\kappa)^{j-1} \left(\frac{\kappa}{jj!} - \frac{\kappa}{(j+1)!} - \frac{1}{j!} \right) B(j+1,\phi)$$

or:

$$I'_M = \frac{1}{\phi} + \sum_{j=1}^{M} \frac{(-\kappa)^{j-1}(\kappa - j^2 - j)}{j(j+1)!} B(j+1,\phi) \tag{B.89}$$

where I'_M is related to I_M by:

$$I_M = I'_M - \frac{(-\kappa)^M}{(M+1)!} B(M+2,\phi) \tag{B.90}$$

Error estimation

To analyse the error on the approximation of I, let us return to the expression in terms of three integrals. The general term of the sequence which is being summed to compute I_M is:

$$u_j = -\frac{(-\kappa)^j}{j(j+1)!} (B(j+1,\phi) + jB(j+2,\phi))$$

which is alternatively negative (for j even) and positive. The sequence has the property that it is strictly decreasing in absolute value, therefore:

$$u_{2m+1}+u_{2m+2}>0 \quad \Rightarrow \quad \sum_{j=2m+1}^{n} u_j > 0 \text{ for any } n \geq 2m+1$$
$$u_{2m}+u_{2m+1}<0 \quad \Rightarrow \quad \sum_{j=2m}^{n} u_j < 0 \text{ for any } n \geq 2m$$

and therefore, the limit of the series (taking the sum from 0) is between $\sum_{j=0}^{M} u_j$ and $\sum_{j=0}^{M+1} u_j$ and upper bounds for the error made in approximating I with I_M are:

$$|I-I_M|<|u_{M+1}|=|\frac{(-\kappa)^M}{M(M+1)!}(B(M+1,\phi)+MB(M+2,\phi))|<\frac{(\kappa)^M}{MM!}B(M+1,\phi) \tag{B.91}$$

since $B(M+2,\phi)<B(M+1,\phi)$.

Consequently, the error on the computationally more efficient approximation I'_M can be bounded as follows:

$$\begin{aligned} |I-I'_M| &< |I-I_M|+|I_M-I'_M| \\ &< \frac{(\kappa)^M}{MM!}B(M+1,\phi)+\frac{(\kappa)^M}{(M+1)!}B(M+2,\phi) \\ &< 2\frac{(\kappa)^M}{MM!}B(M+1,\phi) \end{aligned} \tag{B.92}$$

from equation (B.90).

This is a coarse upper bound and in fact, more can be said about the errors involved in approximating with I_M or I'_M. In particular, we have the result:

Lemma 1 *I'_M is a better approximation of I than I_M for $\kappa > 1$.*

This can easily be seen by observing that, assuming M even, we have:

$$\begin{aligned} I'_M &= I_M+\frac{\kappa^M}{(M+1)!}B(M+2,\phi) \\ \text{thus:} \\ I'_M &> I_M \\ \text{and} \\ I_{M+1}=I_M+u_M &= I_M+\frac{\kappa^{M+1}}{M(M+1)!}\left(B(M=1,\phi)+MB(M+2,\phi)\right) \\ &= I_M+\frac{\kappa^M}{(M+1)!}\frac{\kappa\left(B(M=1,\phi)+MB(M+2,\phi)\right)}{M} \\ &> I_M+\kappa\left(1+\frac{1}{M}\right)B(M+2,\phi) \\ \text{so that, if } \kappa>1, \\ I_{M+1} &> I'_M \end{aligned}$$

Therefore, assuming $\kappa > 1$, we obtain:

$$\begin{aligned} &I'_M \in (I_M, I_{M+1}) \quad \text{for M even,} \\ &\text{and similarly} \\ &I'_M \in (I_{M+1}, I_M) \quad \text{for M odd.} \end{aligned} \tag{B.93}$$

This entails that

$$|I'_{M+1}-I'_M|<|I_{M+1}-I_M|$$

and, more generally:

$$\begin{aligned} |I'_{M+p} - I'_M| &< |I_{M+p} - I_M| \text{ for p odd} \\ |I'_{M+p} - I'_M| &< |I_{M+p+1} - I_M| \text{ for p even} \end{aligned}$$

so that, taking limits as $p \to \infty$, we have:

$$\forall M \, |I - I'_M| < |I - I_M| \tag{B.94}$$

q.e.d.

On the contrary, for small values of κ, the series which converges fastest is I_n if:

$$\kappa \left(\frac{n+\phi+1}{(n+2)n} + 1 \right) < 1$$

which is true for n large enough (for given values of κ and ϕ). However, as the numerical experiments below show, small values of κ in any case lead to fast convergence of I'_n.

Numerical investigation

Using Maple which calculates sums of terms with Beta functions in terms of the Generalised Hypergeometric Function, we can however evaluate the exact relative error $\Delta I_M = |I - I'_M|/I$ for a range of values of M, for given values of parameters κ and ϕ.

The results are shown in table B.2 below.

Table B.2 Relative errors on the estimation of μ_T

ϕ	κ	I	ΔI_M: $M=3$	$M=5$	$M=7$	$M=10$	$M=15$	$M=20$	$M=30$
0.01	0.01	1.98	$4.2\,10^{-7}$	$1.3\,10^{-12}$	$2.5\,10^{-18}$	$2.5\,10^{-27}$			
0.10	0.01	1.00	$3.8\,10^{-7}$	$1.2\,10^{-12}$	$2.1\,10^{-18}$	$2.1\,10^{-27}$			
1	0.01	0.50	$1.7\,10^{-7}$	$4.0\,10^{-13}$	$5.6\,10^{-19}$	$4.2\,10^{-28}$			
10	0.01	0.09	$4.6\,10^{-9}$	$1.9\,10^{-15}$	$6.3\,10^{-22}$	$7.9\,10^{-32}$			
0.01	0.1	10.63	$4.5\,10^{-5}$	$1.5\,10^{-8}$	$2.7\,10^{-11}$	$2.7\,10^{-17}$			
0.10	0.1	1.78	$4.0\,10^{-5}$	$1.3\,10^{-8}$	$2.3\,10^{-11}$	$2.2\,10^{-17}$			
1	0.1	0.54	$1.8\,10^{-5}$	$4.3\,10^{-9}$	$5.9\,10^{-13}$	$4.5\,10^{-18}$			
10	0.1	0.09	$4.8\,10^{-7}$	$2.0\,10^{-10}$	$6.6\,10^{-15}$	$8.3\,10^{-23}$			
0.01	1	79.75	$1.6\,10^{-1}$	$5.7\,10^{-3}$	$1.0\,10^{-4}$	$1.1\,10^{-6}$			
0.10	1	8.05	$1.4\,10^{-1}$	$4.8\,10^{-3}$	$8.7\,10^{-5}$	$8.8\,10^{-7}$			
1	1	0.86	$4.7\,10^{-2}$	$1.2\,10^{-3}$	$1.7\,10^{-5}$	$1.3\,10^{-7}$			
10	1	0.10	$9.3\,10^{-4}$	$4.0\,10^{-5}$	$1.3\,10^{-7}$	$1.7\,10^{-12}$			
0.01	5	217.81	1.4	1.8	0.9	0.1	$9.0\,10^{-4}$	$1.2\,10^{-6}$	$7.9\,10^{-13}$
0.10	5	20.98	1.3	1.5	0.8	0.1	$7.2\,10^{-4}$	$9.6\,10^{-7}$	$6.0\,10^{-14}$
1	5	1.66	0.6	0.5	0.2	$2.1\,10^{-2}$	$1.0\,10^{-4}$	$1.0\,10^{-7}$	$4.6\,10^{-15}$
10	5	.11	0.0	$2.7\,10^{-3}$	$2.4\,10^{-4}$	$4.0\,10^{-6}$	$1.7\,10^{-9}$	$2.6\,10^{-13}$	$6.6\,10^{-22}$
0.01	10	286.74	2.1	20.3	50.0	64.0	14.9	0.7	$4.7\,10^{-5}$
0.10	10	27.64	1.9	17.5	42.0	52.3	11.7	0.5	$3.4\,10^{-5}$
1	10	2.18	0.8	5.2	9.6	9.1	1.5	$5.1\,10^{-2}$	$2.5\,10^{-5}$
10	10	0.13	$3\,10^{-2}$	$2\,10^{-2}$	$1\,10^{-1}$	$1.4\,10^{-3}$	$2.0\,10^{-5}$	$1.1\,10^{-6}$	0.

Conclusion

In conclusion, we note that although I_M (computed for instance as $I_M = I'_M - \frac{(-\kappa)^M}{(M+1)!} B(M+2,\phi)$) could be used as approximation,it is preferable to approximate I with I'_M, so that μ_T can be approximated as:

$$\mu_T \approx \frac{\nu}{\alpha - 1}\left(\phi I'_M + \phi^{-1}\right) \tag{B.95}$$

B.12 Std. deviation of number of cells/storm for QR model

Conditional upon η, the mean number of cells N_c in a storm is geometrically distributed with mean $E[N_c|\eta] = 1 + \frac{\kappa_1 + \kappa_2 \eta}{\phi}$. Let $a = 1 - E[N_c|\eta]^{-1}$. The distribution is thus given by:

$$Pr\{N_c = n|\eta\} = (1-a)a^{n-1} \text{ for } n > 0$$

The variance of N_c is:

$$\text{var}[N_c] = E[\text{var}[N_c|\eta]] \tag{B.96}$$

Let us first calculate the conditional variance:

$$\text{var}[N_c|\eta] = E[N_c^2|\eta] - E[N_c|\eta]^2$$

and

$$E[N_c^2|\eta] = (1-a)\sum_{n=1}^{\infty} n^2 a^{n-1}$$

The following sums are useful:

$$\begin{aligned}
\sum_{n=1}^{\infty} n a^{n-1} &= \frac{\partial}{\partial a}\sum_{n=0}^{\infty} a^n \\
&= \frac{1}{(1-a)^2} \\
\sum_{n=1}^{\infty} n(n-1)a^{n-2} &= \frac{\partial^2}{\partial a^2}\sum_{n=0}^{\infty} a^n \\
&= \frac{2}{(1-a)^3}
\end{aligned}$$

Hence:

$$\sum_{n=1}^{\infty} n^2 a^{n-1} = \frac{a+1}{(1-a)^3}$$

thus,

$$E[N_c^2|\eta] = \frac{a+1}{(1-a)^2}$$

and:

$$\begin{aligned}
\text{var}[N_c|\eta] &= \frac{2a}{(1-a)^2} \\
&= 2\left[\left(\frac{1}{1-a}\right)^2 - \frac{1}{1-a}\right]
\end{aligned}$$

Therefore:

$$\text{var}[N_c|\eta] = 2\left[\left(1 + \frac{\kappa_1 + \kappa_2\eta}{\phi}\right)^2 - 1 - \frac{\kappa_1 + \kappa_2\eta}{\phi}\right]$$

or,

$$\text{var}[N_c|\eta] = 2\left[\left(\frac{\kappa_1 + \kappa_2\eta}{\phi}\right)^2 + \frac{\kappa_1 + \kappa_2\eta}{\phi}\right] \tag{B.97}$$

From equation (B.96), the unconditional variance is:

$$\text{var}[N_c] = \frac{2}{\phi^2}\left[(\kappa_1^2 + \kappa_1\phi) + (2\kappa_1\kappa_2 + \kappa_2\phi)E[\eta] + \kappa_2^2 E[\eta^2]\right]$$

Since:

$$\begin{aligned} E[\eta] &= \alpha/\nu \\ E[\eta^2] &= (\alpha+1)\alpha/\nu^2 \end{aligned}$$

we finally obtain:

$$\text{var}[N_c] = \frac{2}{\phi^2}\left[(\kappa_1^2+\kappa_1\phi)+(2\kappa_1\kappa_2+\kappa_2\phi)\frac{\alpha}{\nu}+\kappa_2^2\frac{\alpha(\alpha+1)}{\nu^2}]\right] \tag{B.98}$$

Appendix C Further details of the single-site rainfall model

This appendix provides further details of the single-site rainfall model presented within Chapter 2, along with greater detail on simulation results.

C.1 Simulation program details

Since the distributions used in our stochastic models have no upper bound, they are liable to give rise to extreme values of certain random variables which are not considered to be realistic. Without introducing any significant bias into the simulation, some of the sampling procedures are set limits. This means that, when these limits are exceeded, the variable in question must be re-sampled, and this can happen a number of times before the program stops. The chosen limits have been selected on the basis of previous experience and general observations about the structure of storms in observed data.

C.1.1 Duration of storm activity

An upper limit of 1600 hours has been set for the duration over which a storm can produce cells. If this limit is exceeded, another storm is generated. This process can be repeated 5 times. If these are unsuccessful, the simulation stops and another simulation is attempted. If this fails again, the program stops.

C.1.2 Number of storms per year

An upper limit of 2000 storms has been set for the year. If this limit is exceeded, another year is simulated. This process can be repeated 3 times. If all these attempts fail, the simulation terminates and a new simulation is attempted. If this fails again, the program stops.

C.1.3 Number of cells per storm

An upper limit of 2000 has been set for the number of cells per storm. If this limit is exceeded, another storm is generated. This process can be repeated 5 times. If all 5 attempts are unsuccessful, the simulation terminates and another simulation is carried out. If this in turn fails, the program stops.

C.1.4 Choice of values of η for each storm

The Random Parameter Bartlett-Lewis model is characterised by the sampling of a value of η for each storm. Should this value be too small, the parameters selected to govern the cell structure for that small will be unrealistic. Very long cells would arrive at a very low rate over a very long period of time. To avoid this situation, a lower threshold for η is selected by the user. It is generally recommended to choose a value of 0.1 or 0.2.

If 40 consecutive samplings from the Gamma distribution for η yield values that lie below the threshold, this will be deemed inappropriate and the program stops. The user then has to restart the simulation(s) with a lower threshold for η.

C.2 Profile objective functions

The plots in this section show profile objective functions for various models fitted to data from Heathrow, Elmdon and Plynlimon, and supplement the text in Section 2.4.3. Table 2.6 (page 17) summarises the models.

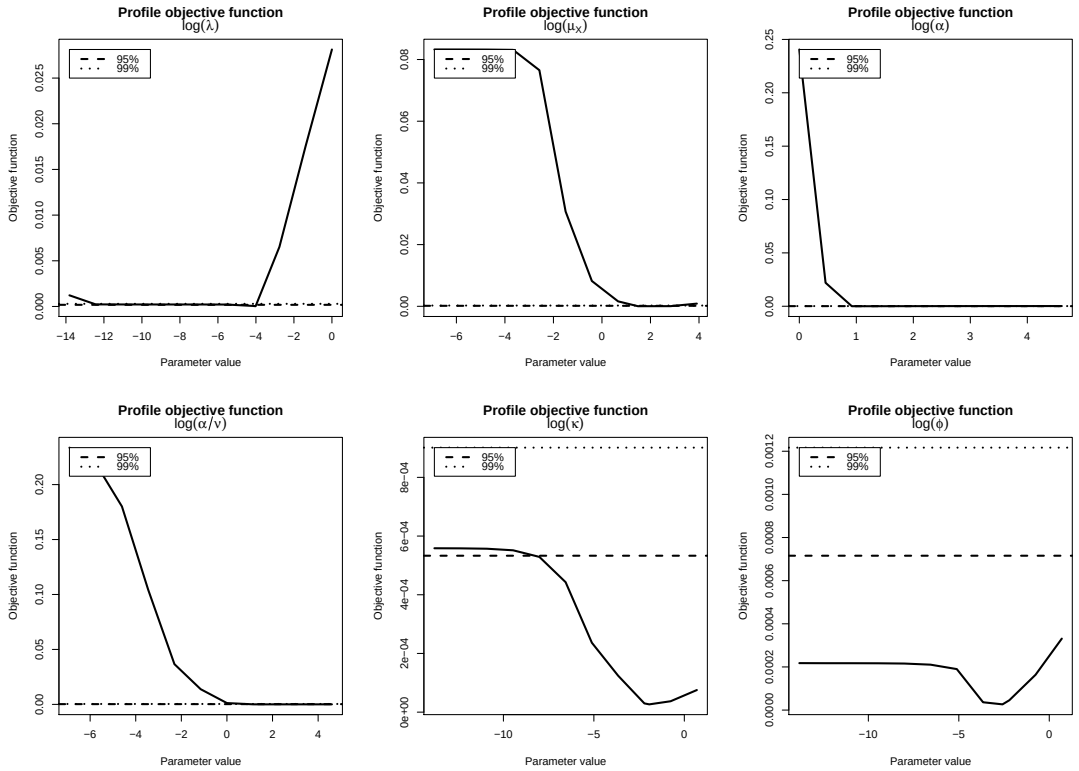

Figure C.21 Profile objective functions for model 4 fitted to July data from Heathrow, using objective function 1.

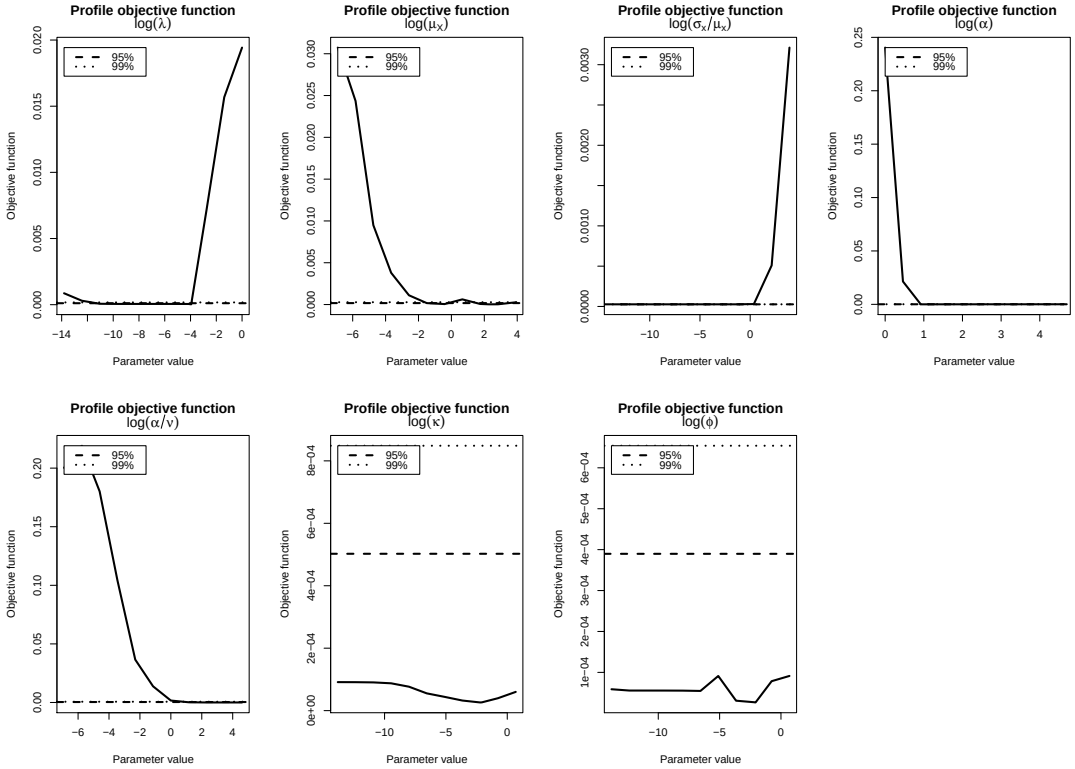

Figure C.22 Profile objective functions for model 5 fitted to July data from Heathrow, using objective function 1.

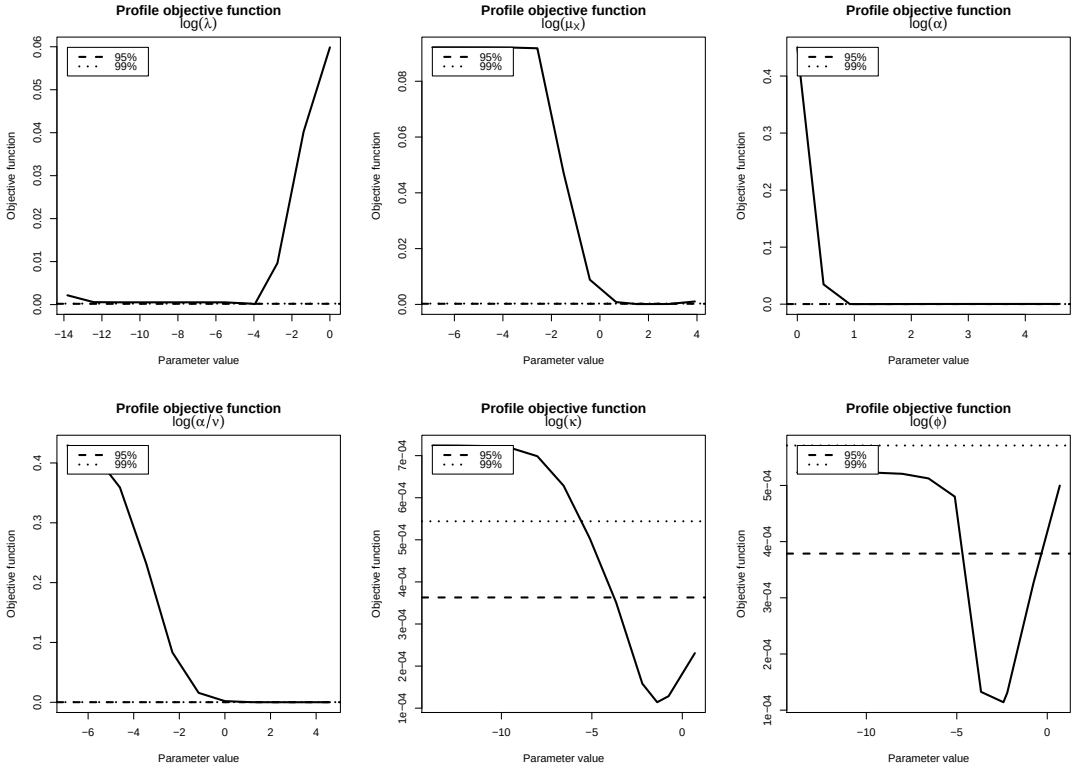

Figure C.23 Profile objective functions for model 4 fitted to July data from Heathrow, using objective function 2.

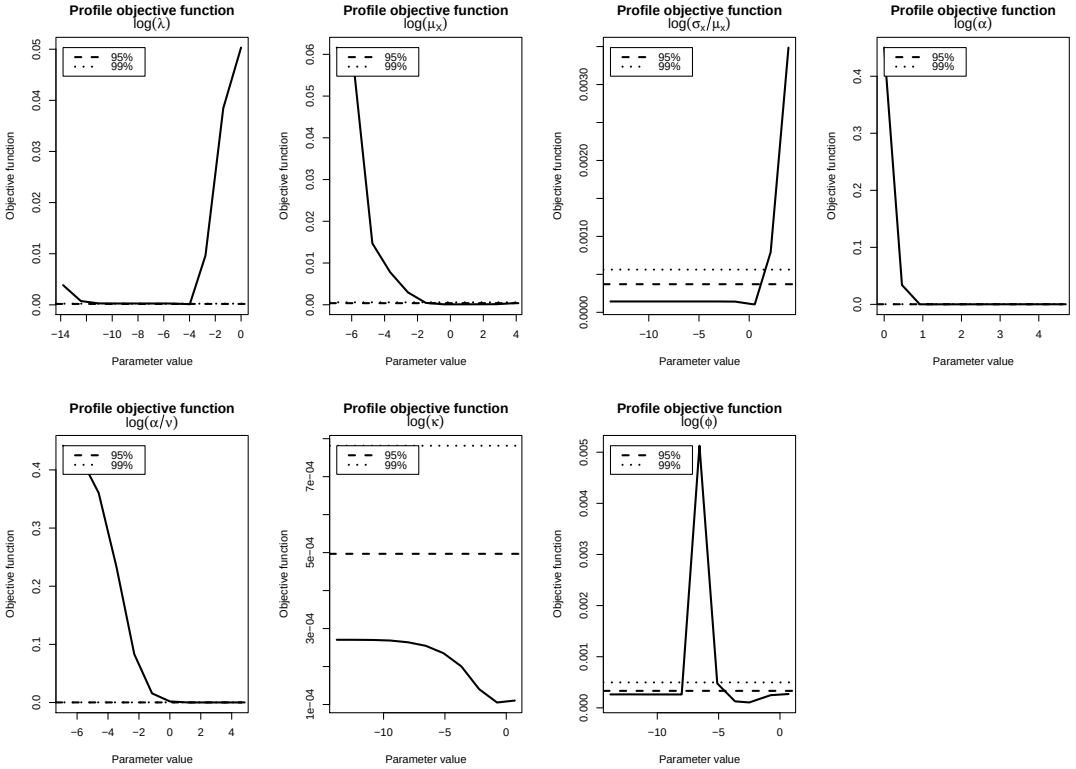

Figure C.24 Profile objective functions for model 5 fitted to July data from Heathrow, using objective function 2.

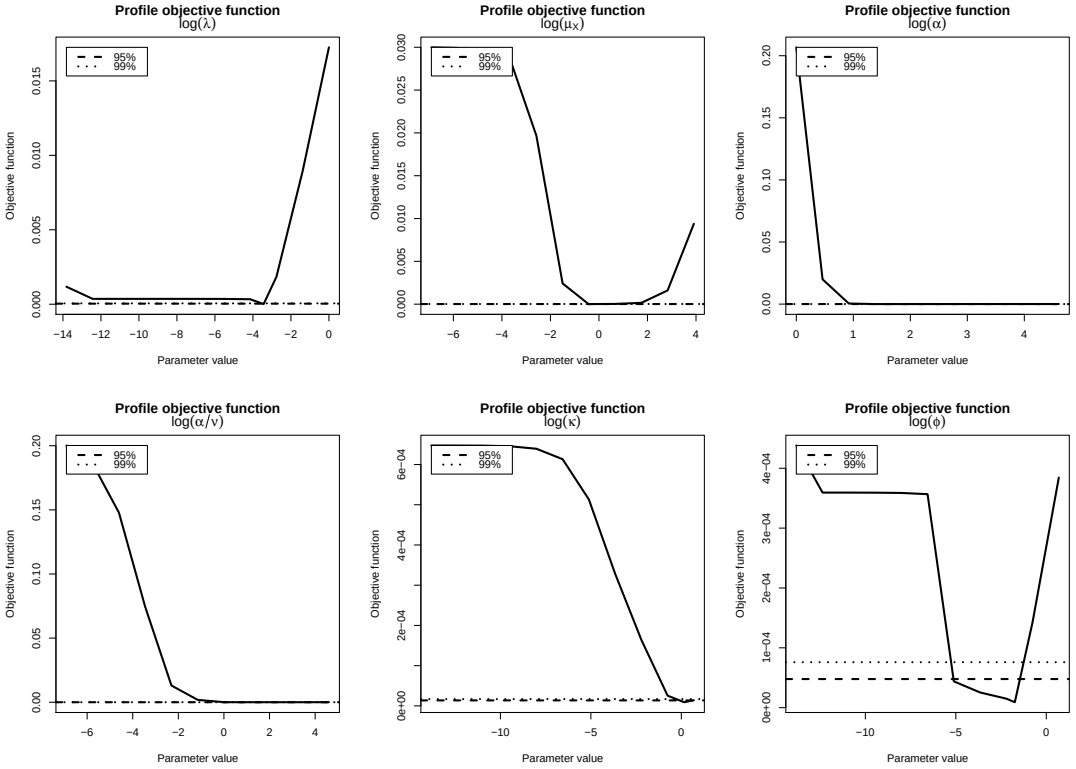

Figure C.25 Profile objective functions for model 4 fitted to January data from Elmdon, using objective function 1.

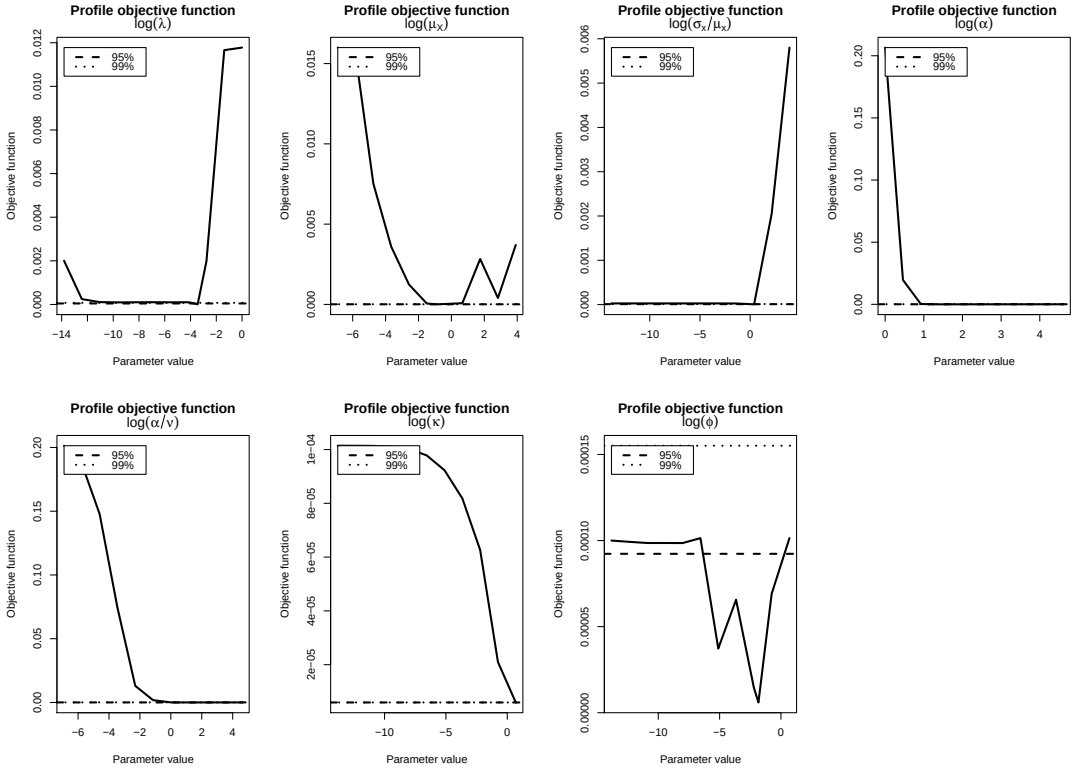

Figure C.26 Profile objective functions for model 5 fitted to January data from Elmdon, using objective function 1.

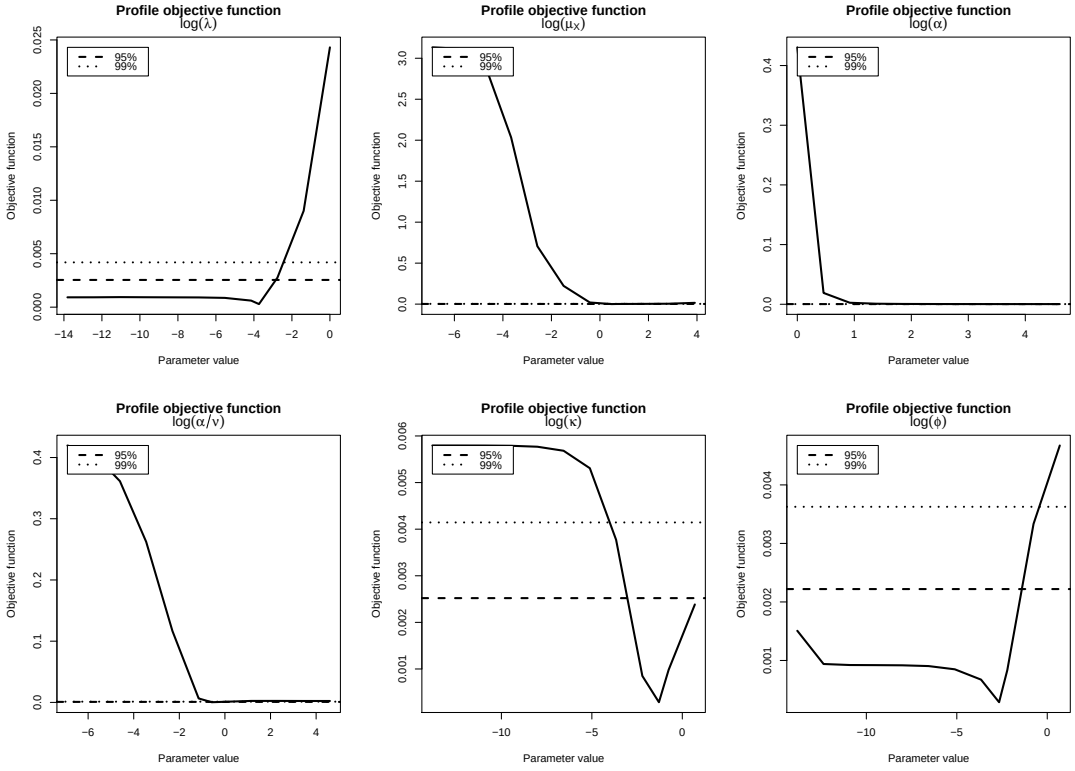

Figure C.27 Profile objective functions for model 4 fitted to January data from Plynlimon, using objective function 1.

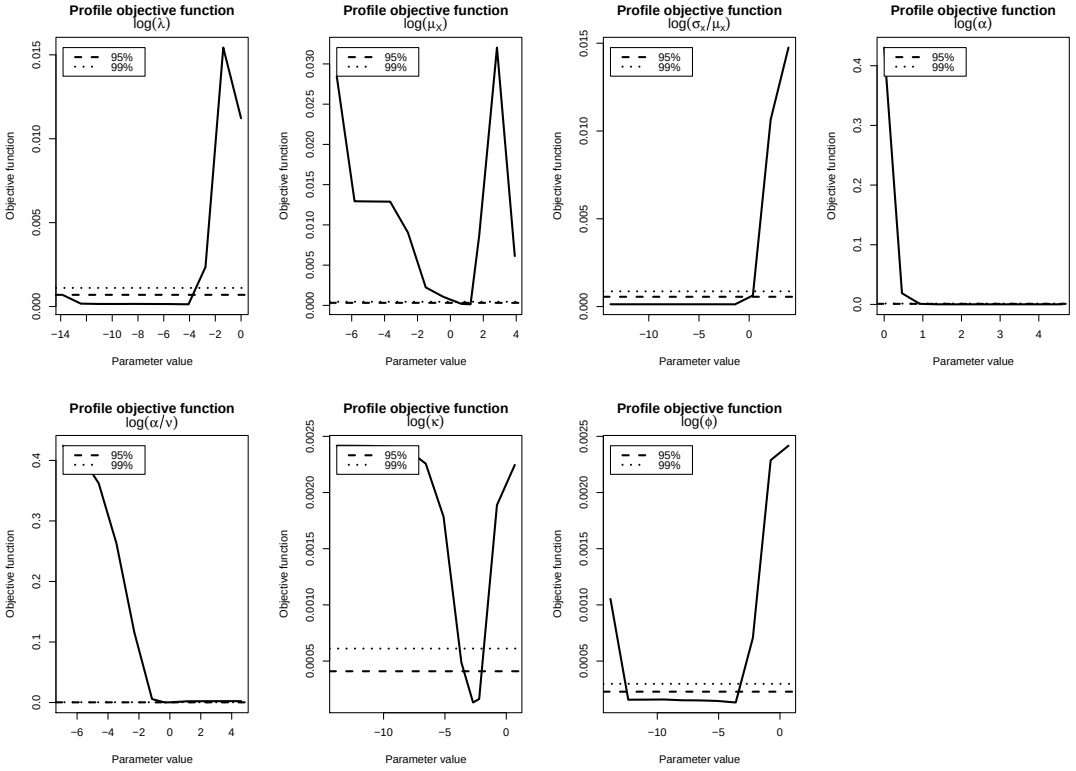

Figure C.28 Profile objective functions for model 5 fitted to January data from Plynlimon, using objective function 1.

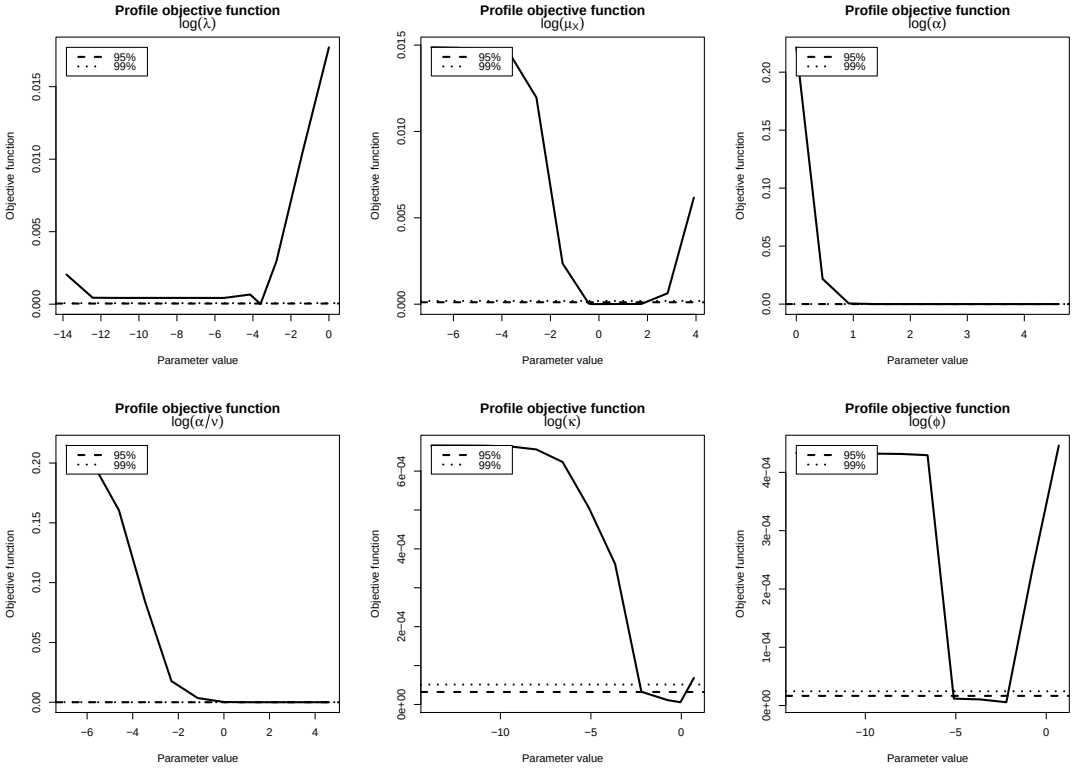

Figure C.29 Profile objective functions for model 4 fitted to January data from 1949 to 1973 at Heathrow, using objective function 1.

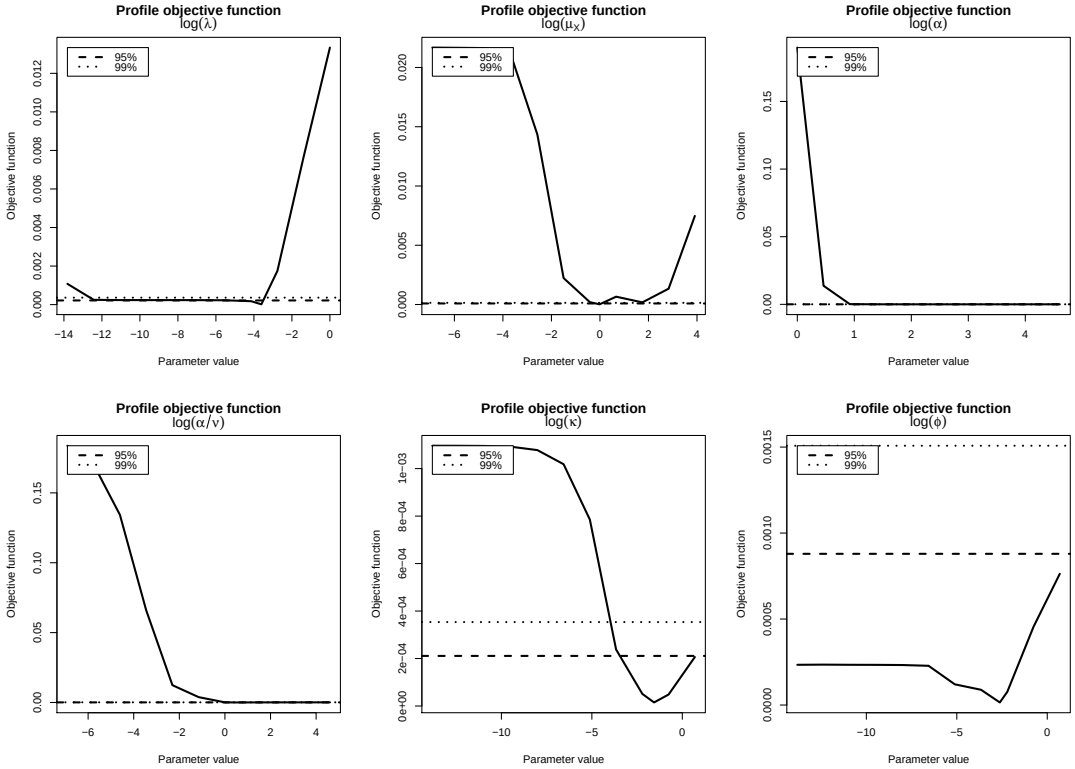

Figure C.210 Profile objective functions for model 4 fitted to January data from 1977 to 2001 at Heathrow, using objective function 1.

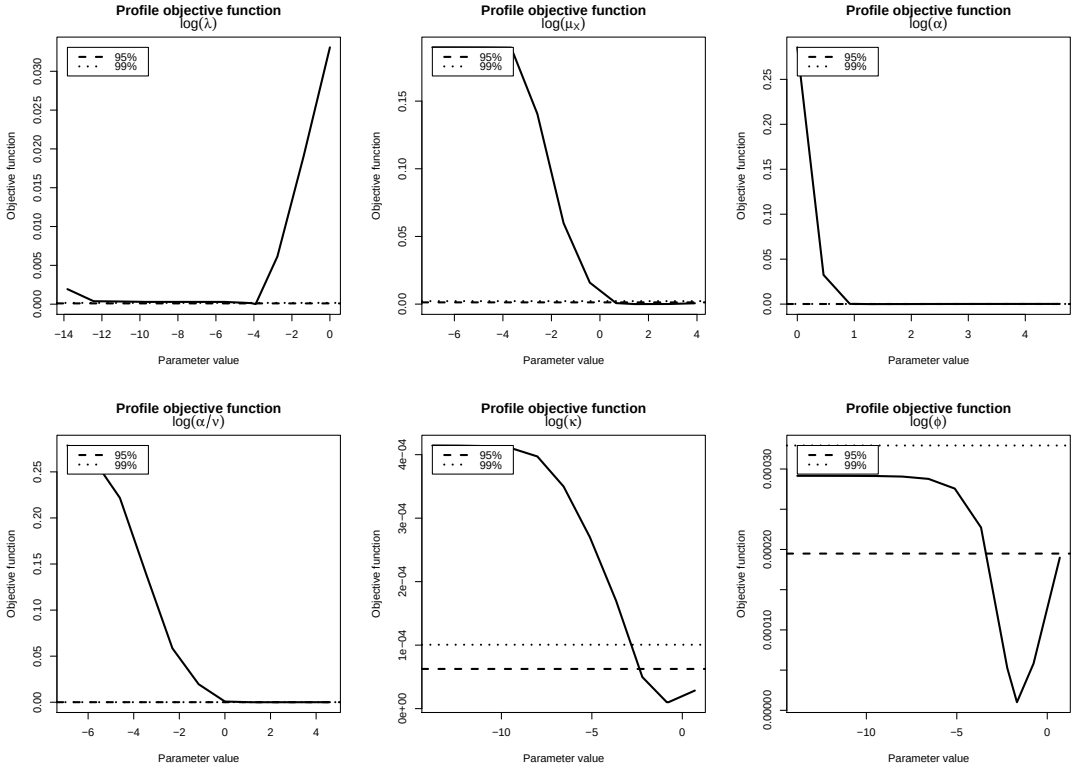

Figure C.211 Profile objective functions for model 4 fitted to July data from 1949 to 1973 at Heathrow, using objective function 1.

C.3 Monthly extreme value performance plots

Figures C.313 and C.314 show the performance of, respectively, the gamma and Pareto versions of model 5 in Table 2.6, for July hourly extremes at Elmdon.

Figures C.315 and C.316 show the performance of model 4 for January and July hourly extremes at Plynlimon. Figure C.317 shows the performance for July daily extremes. Figures C.318 and C.319 show the performance of model 5 (gamma version) for January and July hourly extremes. Figures C.320 and C.321 are the corresponding plots for the Pareto version of the model.

Figures C.322 and C.323 show the hourly performance of model 4 at Heathrow in January and July. Figures C.324 and C.325 are the corresponding plots for daily performance. Figure C.326 shows the performance of model 5 (gamma version) for January hourly extremes.

C.4 Annual extreme value performance plots

Figure C.427 shows the annual performance of the best parameters at Plynlimon at the daily time-scale.

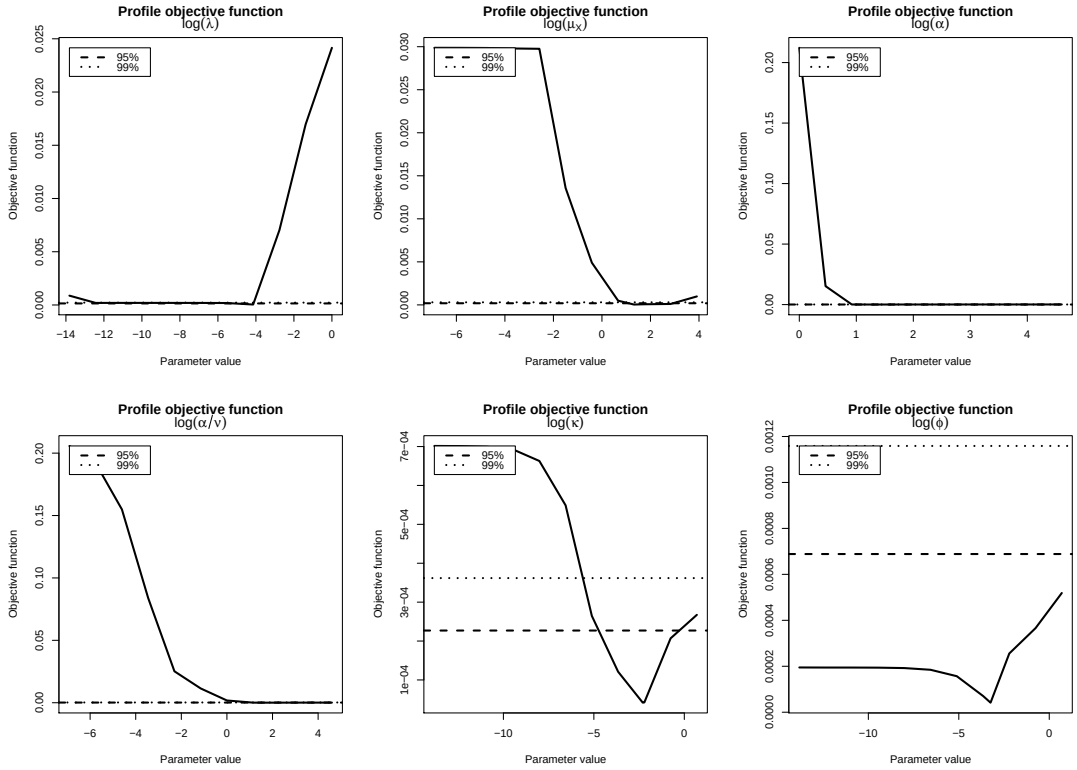

Figure C.212 Profile objective functions for model 4 fitted to July data from 1977 to 2001 at Heathrow, using objective function 1.

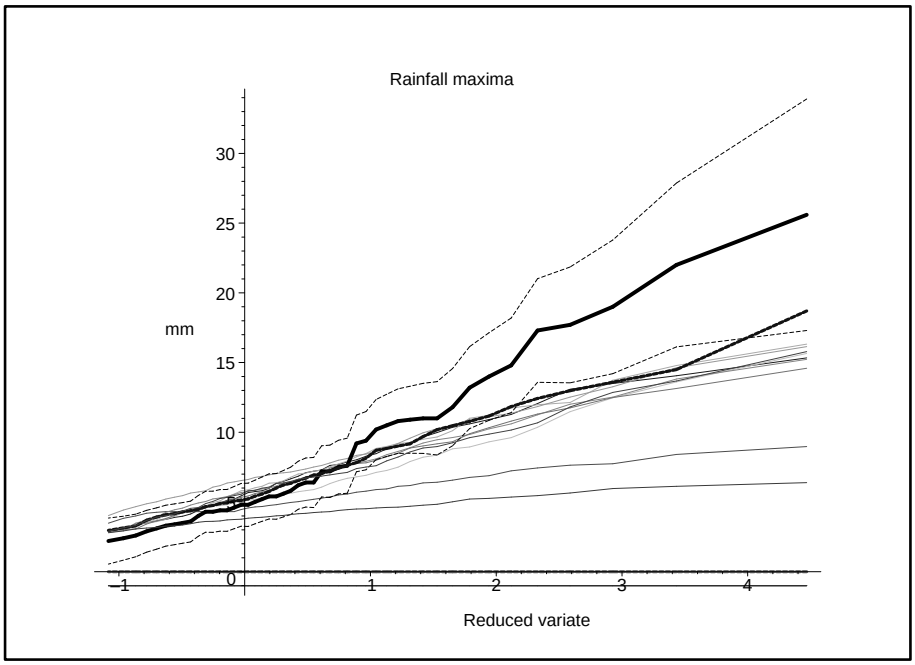

Figure C.313 Model 5 (gamma version) at Elmdon — July hourly extremes (see details in Section 2.5.2).

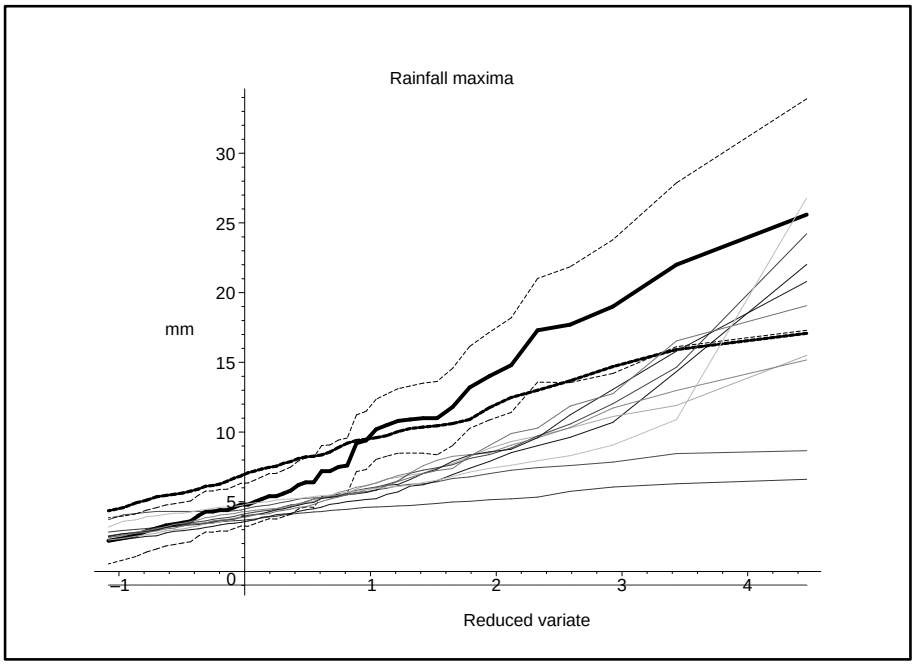

Figure C.314 Model 5 (Pareto version) at Elmdon — July hourly extremes (see details in Section 2.5.2).

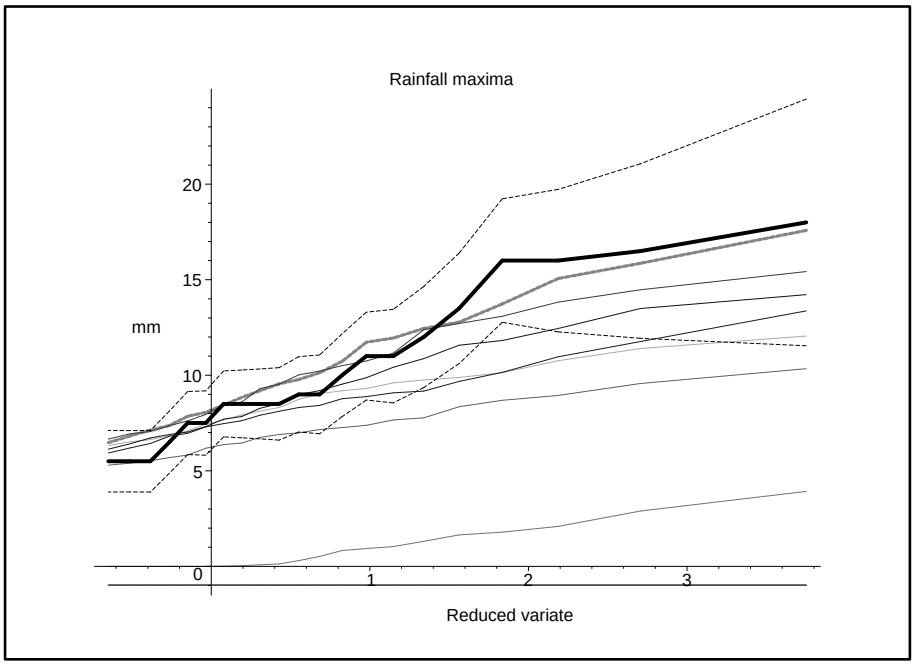

Figure C.315 Model 4 at Plynlimon — Jan hourly extremes (see details in Section 2.5.2)

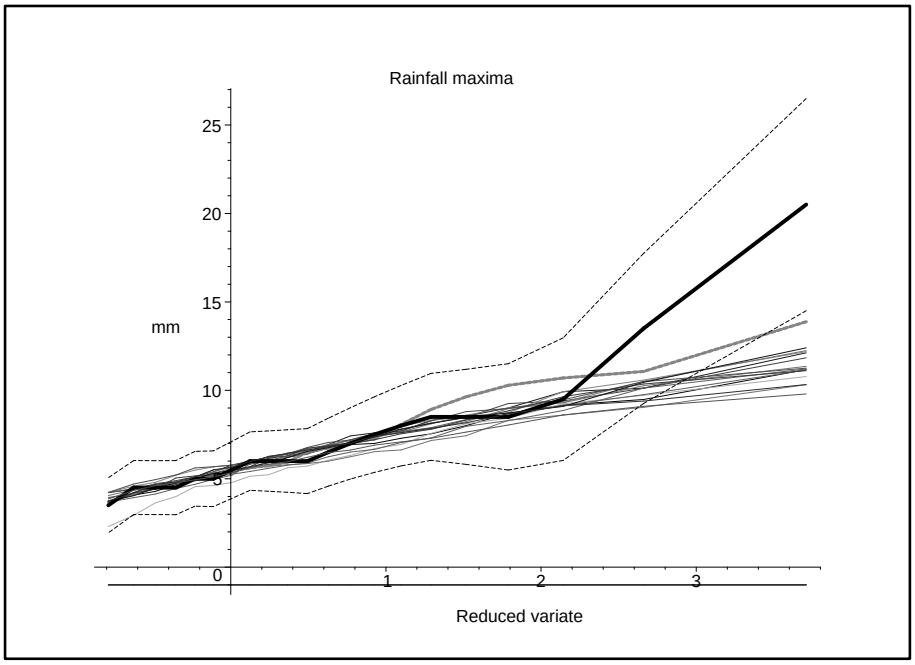

Figure C.316 Model 4 at Plynlimon — July hourly extremes (see details in Section 2.5.2)

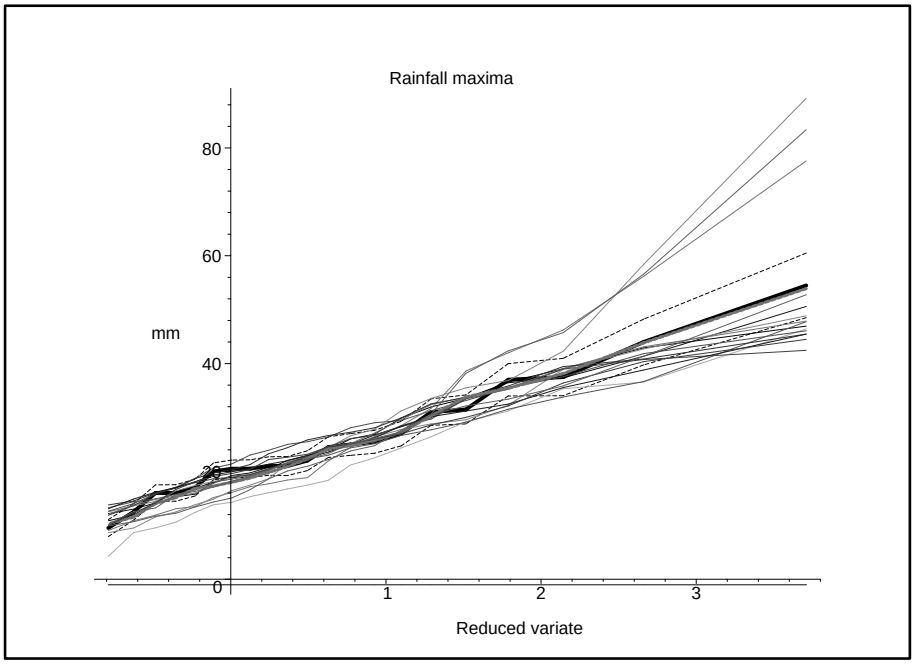

Figure C.317 Model 4 at Plynlimon — July daily extremes (see details in Section 2.5.2)

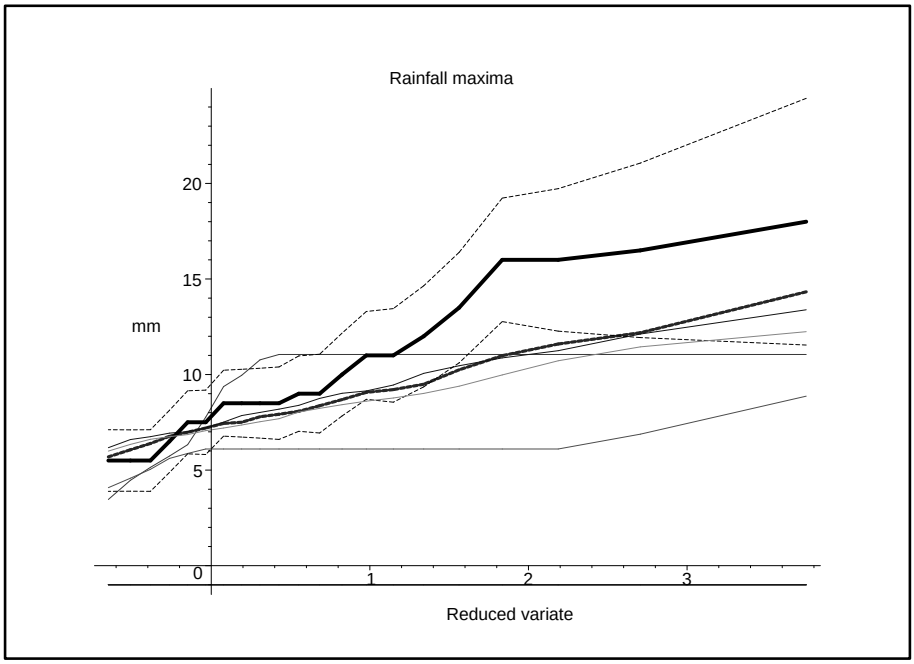

Figure C.318 Model 5 (gamma version) at Plynlimon — January hourly extremes (see details in Section 2.5.2).

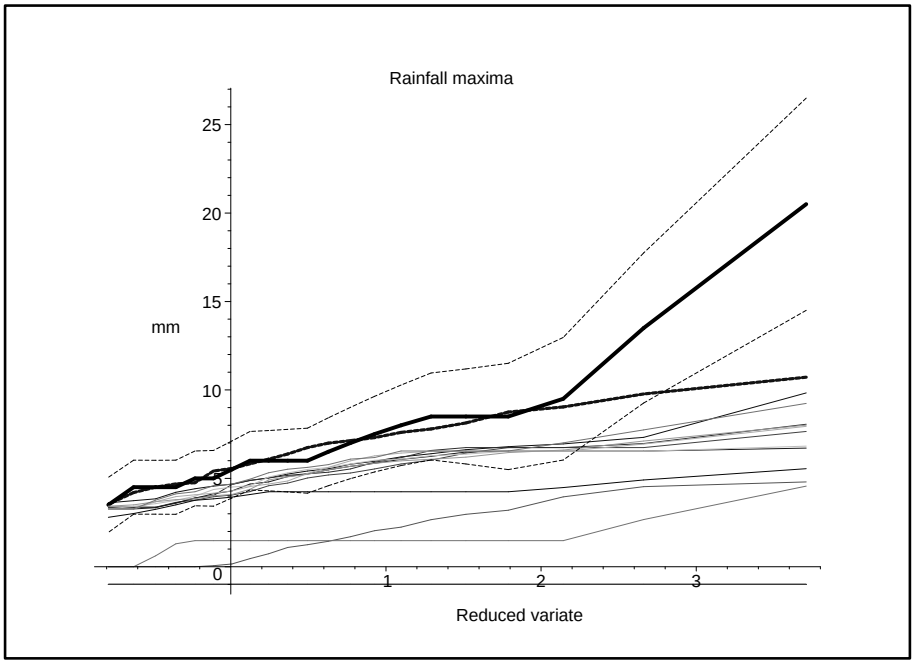

Figure C.319 Model 5 (gamma version) at Plynlimon — July hourly extremes (see details in Section 2.5.2).

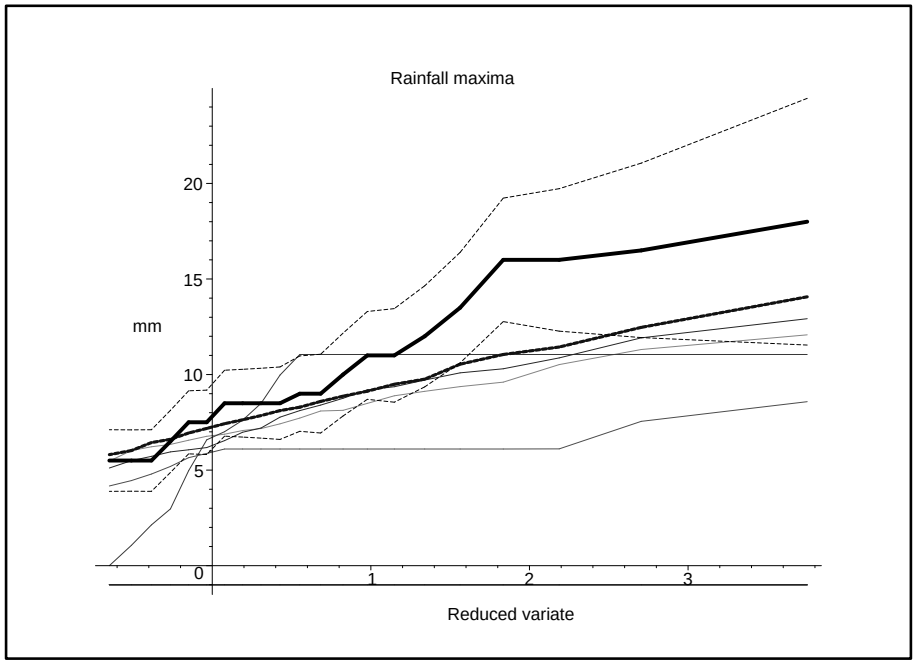

Figure C.320 Model 5 (Pareto version) at Plynlimon — Jan hourly extremes (see details in Section 2.5.2).

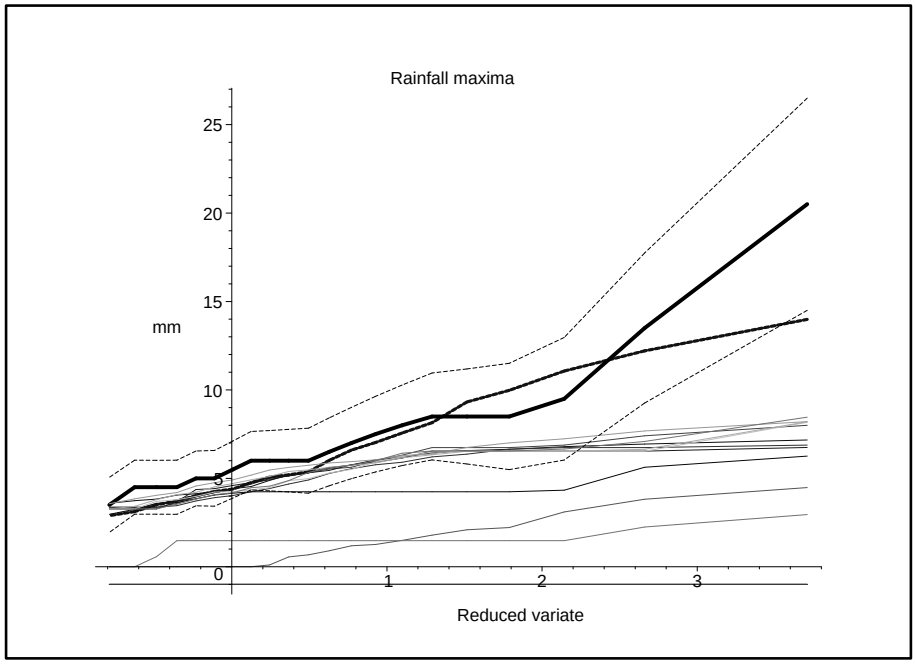

Figure C.321 Model 5 (Pareto version) at Plynlimon — July hourly extremes (see details in Section 2.5.2).

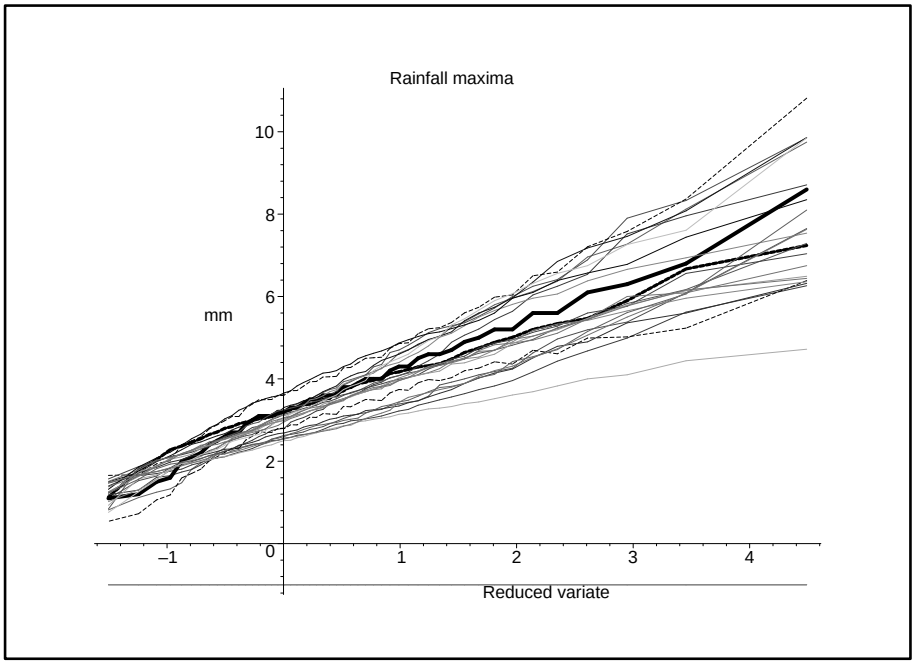

Figure C.322 Model 4 at Heathrow — January hourly extremes (see details in Section 2.5.2).

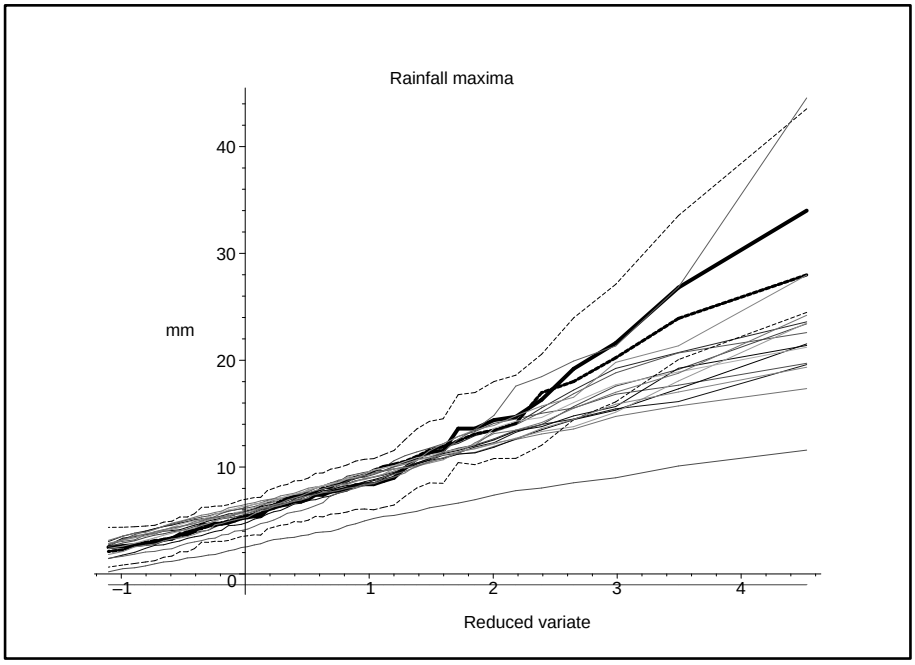

Figure C.323 Model 4 at Heathrow — July hourly extremes (see details in Section 2.5.2).

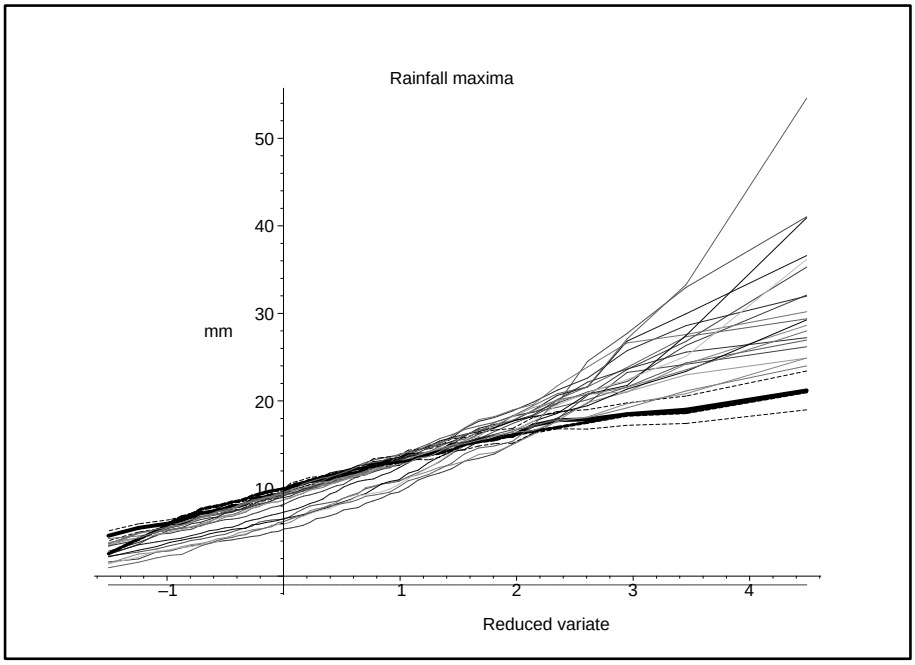

Figure C.324 Model 4 at Heathrow — January daily extremes (see details in Section 2.5.2).

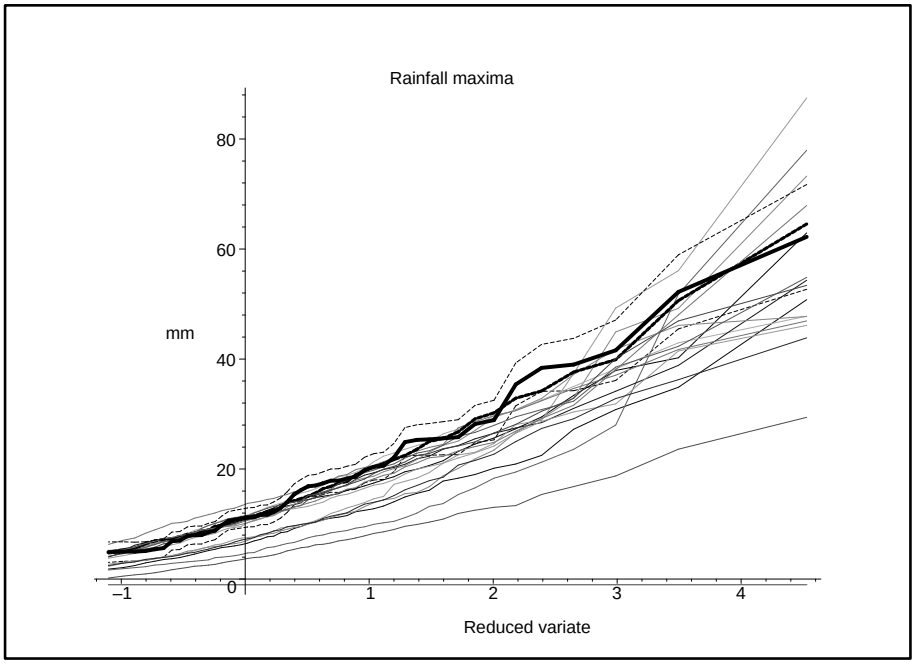

Figure C.325 Model 4 at Heathrow — July daily extremes (see details in Section 2.5.2).

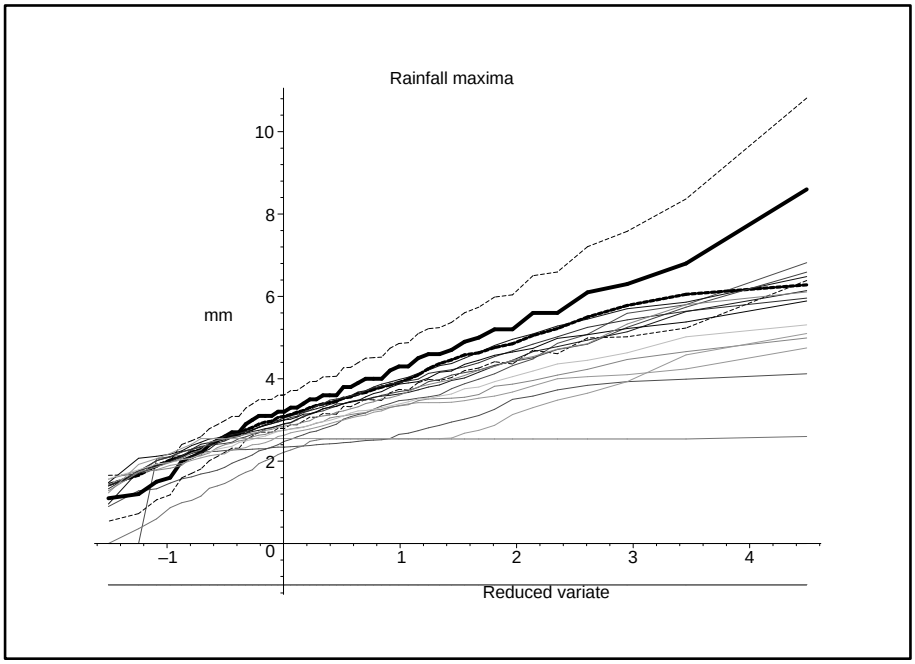

Figure C.326 Model 5 (gamma version) at Heathrow — Jan hourly extremes (see details in Section 2.5.2).

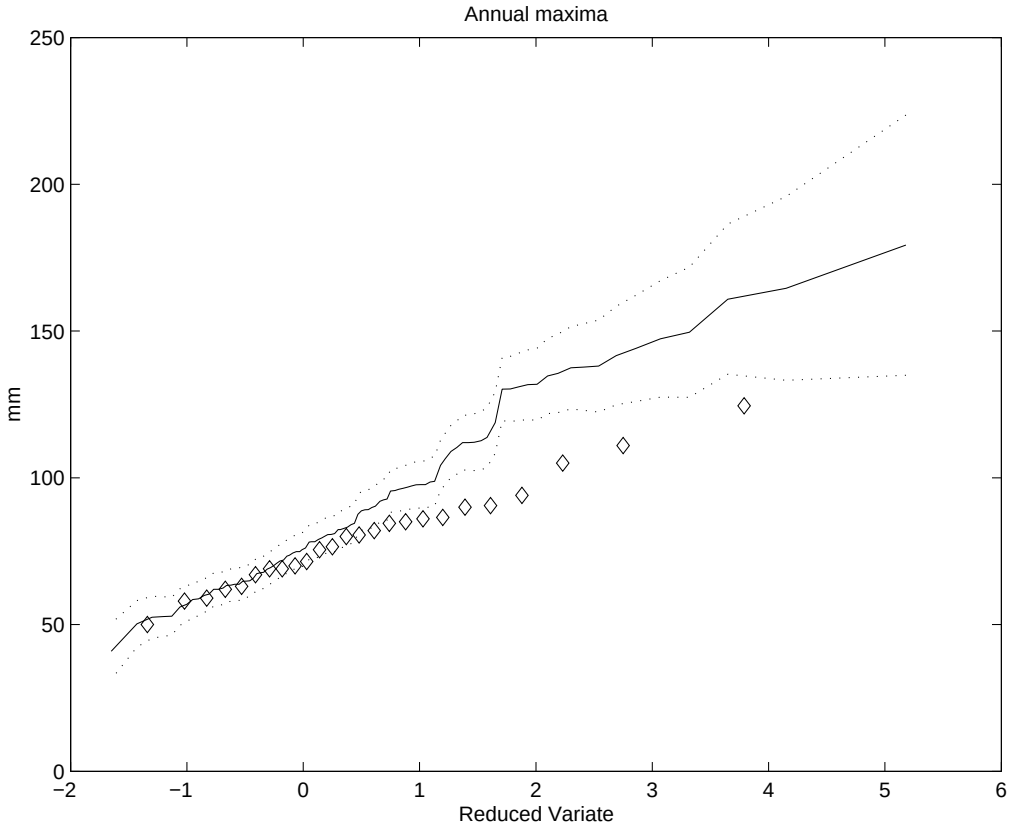

Figure C.427 Model 4 at Plynlimon — Annual daily extremes. See Figure 2.16 (page 47) for legend.

Appendix D Derivation of daily evaporation model properties

In this appendix we derive expressions required for application of the evaporation model presented within Chapter 3.

D.1 Mean and covariance of the daily evaporation model

The two expression derived are (3.18) and (3.19) for the mean and covariance structure of the selected model for daily PE sequences (model 3 in Table 3.4). The model can be written as

$$Y_t = \beta_0 + \sum_{j=1}^{2} \beta_j^* \cos\left[2\pi\left(\omega_j t - \phi_j\right)\right] + \beta_3 Y_{t-1} + \varepsilon_t \ , \tag{D.1}$$

where ε_t is distributed as $N\left(0, \sigma_t^2\right)$ independently (but not identically) for each t, and σ_t is given by (3.17). The parameterisation in (D.1) is the same as that in (3.16).

The model is essentially a first-order autoregression, superimposed onto a deterministic seasonal cycle and with the added complication of a deterministic cycle in the variance structure. Although the model is nonstationary, it has some similarities to a stationary process with a mixed spectrum (Priestley, 1981, Chapter 8). It would be surprising if the time domain properties of such a process had not been studied previously; however, we have not found any relevant references in the modern time series literature.

To derive the properties of model (D.1) it is useful to introduce the backshift operator, B (see Priestley, 1981, p.123, for example). The effect of the operator is to shift time backwards one unit: $B\left(Y_t\right) = Y_{t-1}, B^2\left(Y_t\right) = Y_{t-2}$ and so on. An appealing property is that B follows algebraic rules so that series expansions can be carried out. In particular, it is legitimate to write $\left(1 - aB\right)^{-1} = \sum_{r=0}^{\infty} \left(aB\right)^r$ if $|a| < 1$.

In backshift notation, model (D.1) can be written as

$$\begin{aligned} Y_t - \beta_3 Y_{t-1} = \left(1 - \beta_3 B\right) Y_t &= \beta_0 + \sum_{j=1}^{2} \beta_j^* \cos\left[2\pi\left(\omega_j t - \phi_j\right)\right] + \varepsilon_t \\ &= \beta_0 + \Re\left\{\sum_{j=1}^{2} \beta_j^* \exp\left[2\pi i\left(\omega_j t - \phi_j\right)\right]\right\} + \varepsilon_t \ , \end{aligned}$$

where $\Re(z)$ denotes the real part of a complex number z and $i = \sqrt{-1}$. Multiplying both sides by by $\left(1 - \beta_3 B\right)^{-1}$ and applying the series expansion given above (which is legitimate since, in Table 3.4, the estimate of β_3 is less than 1 in magnitude), we obtain

$$\begin{aligned} Y_t &= \left(1 - \beta_3 B\right)^{-1} \left[\beta_0 + \Re\left\{\sum_{j=1}^{2} \beta_j^* \exp\left[2\pi i\left(\omega_j t - \phi_j\right)\right]\right\} + \varepsilon_t\right] \\ &= \sum_{r=0}^{\infty} \beta_3^r B^r \left[\beta_0 + \Re\left\{\sum_{j=1}^{2} \beta_j^* \exp\left[2\pi i\left(\omega_j t - \phi_j\right)\right]\right\} + \varepsilon_t\right] \\ &= \beta_0 \sum_{r=0}^{\infty} \beta_3^r + \Re\left\{\sum_{j=1}^{2} \beta_j^* \sum_{r=0}^{\infty} \beta_3^r \exp\left[2\pi i\left(\omega_j (t-r) - \phi_j\right)\right]\right\} + \sum_{r=0}^{\infty} \beta_3^r \varepsilon_{t-r} \\ &= \beta_0 \sum_{r=0}^{\infty} \beta_3^r + \Re\left\{\sum_{j=1}^{2} \beta_j^* \exp\left[2\pi i\left(\omega_j t - \phi_j\right)\right] \sum_{r=0}^{\infty} \beta_3^r \exp\left[-2\pi i \omega_j r\right]\right\} + \sum_{r=0}^{\infty} \beta_3^r \varepsilon_{t-r} \end{aligned}$$

The first two terms involve infinite sums of geometric series, so that the equation can be rewritten as

$$Y_t = \frac{\beta_0}{1-\beta_3} + \Re\left\{\sum_{j=1}^{2} \beta_j^* \exp\left[2\pi i\left(\omega_j t - \phi_j\right)\right]\left(1-\beta_3\exp\left[-2\pi i\omega_j\right]\right)^{-1}\right\} + \sum_{r=0}^{\infty}\beta_3^r\varepsilon_{t-r}\ .$$

Simplifying the central expression, and taking the real part, yields

$$Y_t = \frac{\beta_0}{1-\beta_3} + \sum_{j=1}^{2}\beta_j^*\left\{\frac{\cos\left[2\pi\left(\omega_j t-\phi_j\right)\right]-\beta_3\cos\left[2\pi\left(\omega_j(t+1)-\phi_j\right)\right]}{1-2\beta_3\cos\left(2\pi\omega_j\right)+\beta_3^2}\right\} + \sum_{r=0}^{\infty}\beta_3^r\varepsilon_{t-r}\ . \qquad \text{(D.2)}$$

Here, Y_t is represented as the sum of a deterministic function and a sequence of uncorrelated terms (the εs). Its properties can now be derived straightforwardly. For example, taking expectations of both sides of (D.2) yields (3.18) directly, since the εs all have zero mean.

The covariance structure of the model is slightly more complicated. Since the only random variables in (D.2) are the εs, we have

$$\text{Cov}\left(Y_t, Y_{t-k}\right) = \text{Cov}\left[\sum_{r=0}^{\infty}\beta_3^r\varepsilon_{t-r}, \sum_{r=0}^{\infty}\beta_3^r\varepsilon_{t-k-r}\right] = \sum_{r=k}^{\infty}\beta_3^{2r-k}\sigma_{t-r}^2, \qquad \text{(D.3)}$$

since $\text{Cov}\left(\varepsilon_t,\varepsilon_s\right) = \sigma_t^2$ if $t=s$, and zero otherwise.

Using parameterisation (3.17) for σ_t, we have

$$\sigma_t^2 = \left(\gamma_0 + \sum_{j=1}^{2}\gamma_j\cos\left[2\pi\left(\omega_j t - \psi_j\right)\right]\right)^2\ .$$

Expanding this, and simplifying using standard trigonometric identities, yields

$$\sigma_t^2 = \gamma_0^* + \sum_{p=1}^{6}\gamma_p^*\cos\left[2\pi\left(\omega_p^* t - \psi_p^*\right)\right]\ ,$$

with notation defined in (3.20). Moving back to complex numbers, and substituting into (D.3), we obtain

$$\begin{aligned}\text{Cov}\left(Y_t, Y_{t-k}\right) &= \Re\left\{\sum_{r=k}^{\infty}\beta_3^{2r-k}\left[\gamma_0^* + \sum_{p=1}^{6}\gamma_p^*\exp\left[2\pi i\left(\omega_p^*(t-r)-\psi_p^*\right)\right]\right]\right\} \\ &= \gamma_0^*\sum_{r=k}^{\infty}\beta_3^{2r-k} + \Re\left\{\sum_{p=1}^{6}\gamma_p^*\exp\left[2\pi i\left(\omega_p^* t-\psi_p^*\right)\right]\sum_{r=k}^{\infty}\beta_3^{2r-k}\exp\left[-2\pi i\omega_p^* r\right]\right\}\ .\end{aligned}$$

As before, the infinite sums here relate to geometric series; the resulting expressions can be simplified, and the real part taken to yield the reult (3.19).

The derivation here illustrates the reason for excluding rainfall from the daily evaporation model (despite the fact that in Section 3.4.2 it was found to be a significant predictor). Specifically, if rainfall was included in the model then it would have to be treated as a deterministic component of (D.1); the subsequent manipulation would then lead to an infinite sum of past rainfall values as well as past εs in (D.2), which may lead to practical difficulties in calculation.

Appendix E Review of some multifractal models for rainfall

One of the aims of Work Package 2 (detailed within Part III) was to investigate fractal based methods as a method of spatial-temporal generation of rainfall series. This appendix details a review of such methods available.

E.1 Introduction

The effort to represent rainfall processes over a large range of scales motivates the development of scaling models for precipitation. The appeal of scaling models is due to their parsimony in describing the variability of rainfall at a large range of scales in space and time. Earlier work in this direction has focused on simple scaling, which assumes self-similarity of the process. The main idea is that rainfall is scale invariant, so that its statistical distribution at different levels of magnification (or scales) is the same up to multiplicative factors that only involve the scale ratios. The form of these factors is λ^{θ}, where λ is the scale ratio and θ is a constant scaling exponent.

Since simple scaling has not been widely successful in reproducing observed properties of precipitation (e.g. Schertzer and Lovejoy 1987, Gupta and Waymire 1990, Kedem and Chiu 1987), more attention in the last 15 years has focused on multi-scaling models, mainly based on multifractal theory. The basic idea behind this direction of research is that moments of different order scale with different scaling exponents.

Several multifractal models have been developed to represent spatial or temporal rainfall and, more recently, spatio-temporal rainfall. Multifractal fields are mostly generated through multiplicative cascade processes, which will be described in Section E.2.2. Based on the discrete or continuous nature of the generating mechanism, one may identify two main groups of models. Some authors (e.g. Gupta and Waymire) focus on discrete cascade models, which propagate an initial mass through a *discrete* range of smaller and smaller scales. Others (e.g. Lovejoy and Schertzer) prefer cascades that develop over a *continuous* range of scales. Sections E.3 and E.4 contain descriptions of the two approaches and underline their advantages and disadvantages.

Not all multifractal models for rainfall are covered here. The purpose of this review is to provide some indication of whether readily usable multifractal approaches to rainfall modelling and simulation exist. In particular, the interest is in models that can generate continuous *space-time* simulations and do not require a lot of additional effort in terms of model development and validation. Therefore, only multifractal models for which a space-time extension already exists are considered. The reader that is mainly interested in the conclusions, may focus on Sections E.3.4, E.4.1, E.4.2 and E.5.

Besides different drawbacks specific to each model, there is of a lack of evidence that multifractals can reproduce various properties of observed rainfall. While the scaling properties of the simulated process are thoroughly verified against data, very little or no attention is concentrated on other characteristics of rainfall.

E.2 Background

Before describing specific multifractal models for rainfall, it is useful to review the general concepts of fractal sets and multifractal measures. Section E.2.1 outlines the characteristics usually associated with fractals and the idea of dimension of such sets, while Section E.2.2 contains an overview of multifractal measures. The connection between fractal sets and multifractal measures, which explains the term “multifractal”, is sketched, together with the cascade process that leads to multifractals and the self-similarity properties that characterize these measures.

E.2.1 Fractal Sets

Fractals are very irregular sets, with a structure that is too complicated to be explained by the rules of classical geometry. Typical examples are snowflakes and irregular coastlines. No accepted definition of a fractal set exists, but one can identify several properties that are *usually* verified for such sets (Falconer, 1990). A property that is particularly important in the context of scaling models for rainfall is self-similarity. The parts, at any scale, of a self-similar set have the same structure as the whole, apart from a scale factor. That is, at any scale, the set can be decomposed into subsets that are scaled down versions of itself. Self-similarity can also be approximate (i.e., distortion is allowed), or statistical. The parts, at any scale, of a statistically self-similar set have the same distribution as the whole, again apart from a scale factor.

The properties of fine structure and local irregularity are related to self-similarity, although they can exist in a set that is not self-similar. Fractals have fine structure in the sense that increasing the level of magnification at which the set is examined, reveals more and more detail. Also, fractals are locally, as well as globally, irregular. That is, the irregularity of the whole set is propagated to all scales.

Another property that is commonly found in fractal sets refers to their dimension. Specifically, the dimension of a fractal is non-integral (i.e. *fractional*) and larger than its classical dimension. For example (Harte, 2001), an irregular coastline may be thought to have dimension between 1 (the dimension of a simple line) and 2 (the dimension of an area). Originally, Mandelbrot (1975) defined fractals by this property. However, it has been found that such definition excludes some sets that should nonetheless be considered as fractals. It remains intuitive to think of a fractal as a very irregular set that has a fractional dimension (Harte, 2001).

As for the concept of fractals, no generally accepted definition of fractal dimension is available. For a strictly self-similar object, one may intuitively think of the fractal dimension as related to the number of scaled down copies of the set that form the set itself. In particular, an object that is formed by m copies of itself, scaled by $1/n$, has similarity dimension $\log m / \log n$. Since self-similarity need not be exact, several more general definitions have been proposed, still based on the idea of covering the fractal set. The shape of the covers, which could be boxes or spheres, their fixed versus variable size and their disjoint or overlapping nature are among the elements that distinguish the various definitions. A detailed account of the different concepts of fractal dimension is outside the scope of this review. The *box counting* and *Hausdorff* dimensions are among the most commonly adopted definitions and the latter will be used in the subsequent description of multifractal measures and multifractal models for rainfall. For completeness, specific definitions are given below for these two dimension concepts only, although the mathematical details are not crucial to the rest of the exposition.

The box counting dimension is the most intuitive and it is relatively easy to compute.

Definition 1 *Let A be a non-empty and bounded subset of $\mathbf{R}^n$ and let $N_\delta(A)$ denote the number of boxes of the δ-grid of $\mathbf{R}^n$ that intersect A. Then the box counting dimension of the set A is defined as:*

$$d_B = -\lim_{\delta \to 0} \frac{\log N_\delta(A)}{\log \delta^{-1}}$$

when the limit exists.

Before giving a definition of the Hausdorff dimension, one needs to define the Hausdorff measure:

Definition 2 *Let X be a metric space, A be a subset of X, and d a non negative number (not necessarily an integer). The d-dimensional Hausdorff measure of A, $H^d(A)$, is the infimum of positive numbers y such that for every $r > 0$, A can be covered by a countable family of closed sets, each of diameter less than r, such that the sum of the d-th powers of their diameters is less than y. (From Weisstein 1992).*

Now, the Hausdorff dimension of a set can be defined as follows:

Definition 3 *Let A be a subset of a metric space X. Then the Hausdorff dimension D(A) of A is the infimum of $d \geq 0$ such that the d-dimensional Hausdorff measure of A is 0. (From Weisstein 1992).*

In less rigorous terms, the Hausdorff dimension of a set A can described by letting $N(\varepsilon)$ denote the smallest number of balls of diameter $\varepsilon > 0$ required to cover A. The Hausdorff dimension, $D(A)$ is the constant that satisfies $N(\varepsilon) \approx \varepsilon^{-D(A)}$ as $\varepsilon \to 0$ (in the many circumstances when $N(\varepsilon)$ follows such a power law). The Hausdorff dimension is less computationally convenient than the box dimension and its practical meaning is not as immediately apparent. However, it has other advantages (Harte, 2001) and is useful in understanding the characteristics of a multifractal measure.

E.2.2 Multifractal Measures

Connections with Fractals

While fractals are sets, the term "multifractal" refers to measures. Multifractal measures, like fractal sets, are extremely irregular. They are "singular" measures, that are not discrete but do not have a local density. However, the fractal appearance of multifractal measures is not at the origin of their name, nor does the term "multifractal" strictly refer to the support of the measure, which may or may not be a fractal set. Rather, the name of these measures is due to the fact that their support may be partitioned into multiple fractal sets, such that all points in each set exhibit the same order of *singularity* (i.e. irregularity).

In general, a singularity (or singular point) is a point at which an equation, surface, etc., blows up or becomes degenerate (Weisstein, 1992). For multifractals, the presence of singularities is better understood when describing their generating mechanism and will be clarified in this context (see Section E.2.2). For the moment, it suffices to recognize that the order of singularity, or irregularity, of a measure at a point x (in 1-D, for ease of notation) may be summarized by the Hölder exponent $\alpha(x)$. Letting μ be the multifractal measure, and $\{\Delta_n\}$ denote a sequence of decreasing intervals containing x, the Hölder exponent $\alpha(x)$ is defined as:

$$\alpha(x) \equiv \lim_{n\to\infty} \alpha_n(x), \tag{E.1}$$

when the limit exists and where $\alpha_n(x)$ is given by

$$\alpha_n(x) \equiv \frac{\log \mu(\Delta_n)}{\log |\Delta_n|}. \tag{E.2}$$

In order to clarify the interpretation of the Hölder exponent, define $M(x) \equiv \mu([0,x])$, and take $x' = x + \delta$, so that $|M(x') - M(x)| = \mu(\Delta)$, with $\Delta \equiv [x, x+\delta]$. For small δ, $\mu(\Delta)$ is approximately equal to $|\Delta|^{\alpha(x)}$, from (E.1) and (E.2). Then

$$\left|M(x') - M(x)\right| \approx \left|x' - x\right|^{\alpha(x)},$$

and the exponent $\alpha(x)$ may be interpreted as a generalization of the local degree of differentiability of the measure at x (Riedi, 1997).

In summary, for a multifractal measure, the set of points x in its support for which $\alpha(x)$ takes on any specific value α, form a fractal set. In the case of monofractal measures, all singularities are of the same order, i.e. there is only one valid Hölder exponent α_0.

Multifractals are usually constructed through multiplicative cascades, where the mass is iteratively redistributed into smaller and smaller subintervals. The generating mechanism is described in Section E.2.2. Here, it is useful to introduce the concept of a coarse Hölder exponent by successively subdividing the support into b^n equal intervals, and calculating $\alpha_n(x) = \frac{\log \mu(\Delta_n)}{\log b^{-n}}$. In order to study the statistical properties of the coarse Hölder exponent, Calvet et al. (1997) describe the $\alpha_n(x)$'s, corresponding to the different b-adic intervals, as drawn from a random variable. When the measure μ is deterministic, they consider the mass of a random cell $\Delta_n^i \equiv [x_i, x_i + \Delta_n]$. The corresponding $\alpha_n(x_i)$, may be viewed as a draw of the random variable α_n, which represents the coarse Hölder exponent of the random interval. Multifractal models for rainfall involve random, rather than deterministic, measures. For the case of random measures, the intervals may be kept fixed. Since the mass on each fixed cell is random, so is the corresponding coarse Hölder exponent. Because of the cascade construction outlined in Section E.2.2, the mass is also identically distributed across intervals, thus the coarse Hölder

exponents at different values of x may be viewed as draws of the same random variable α_n (see Calvet et al., 1997).

Several asymptotic results for α_n are described in Calvet et al. (1997). By the strong law of large numbers, $\alpha_n \xrightarrow{p} \alpha_0$ as $n \to \infty$. That is, as the support is subdivided into smaller and smaller sub-intervals, and the mass is redistributed, almost all Hölder exponents concentrate in a small neighborhood of α_0. In other words, for large n, almost all intervals have Hölder exponent in a neighborhood of α_0. However, the mass of these intervals goes to zero as n goes to infinity, since, typically, for multifractal measures the mass concentrates on cells with Hölder exponent bounded away from α_0. Thus, it is important to consider also the other values of the Hölder exponent. The central limit theorem provides another asymptotic result, i.e. $\sqrt{n}(\alpha_n - \alpha_0) \xrightarrow{d} N(0, \sigma^2)$. Also, large deviation theory provides an asymptotic result for the tail properties of the Hölder exponent (see Calvet et al., 1997).

The strengths of the singularities (i.e. the Hölder exponents) associated with a multifractal measure constitute the support of a multifractal spectrum $f(\alpha)$. Several multifractal spectra have been defined, which in turn may be derived in different ways and are subject to various interpretations. One of the possible definitions involves a renormalized histogram of coarse Hölder exponents. The idea is to partition the range of α's into small intervals. At stage n in the subdivision of the mass support, let $N_n(\alpha)$ denote the number of coarse Hölder exponents in the interval $(\alpha, \alpha + \Delta\alpha]$ and consider the renormalized histogram given by $\frac{N_n(\alpha)}{b^n}$. Its limit, as n tends to infinity, is taken to be a multifractal spectrum. By the central limit theorem result mentioned above, this $f(\alpha)$ is locally quadratic around α_0. Alternatively, the multifractal spectrum is defined as the Hausdorff dimension of the set $\mathcal{A}(\alpha)$ of points with Hölder exponent equal to α, $\alpha \geq 0$. Thus $N(\alpha, \Delta_n) \approx (\Delta_n)^{-f(\alpha)}$, where $N(\alpha, \Delta_n)$ denotes the number of intervals $[x, x + \Delta_n]$ required to cover $\mathcal{A}(\alpha)$. A third definition of the multifractal spectrum is based on large deviation theory and takes $f(\alpha)$ to represent the limit of $n^{-1} \log_b P(\alpha_n > \alpha) + 1$. Although the three definitions are not equivalent in general, they agree for many multifractals. Calvet et al. (1997) contains a more detailed description of the three versions of the multifractal spectrum and their properties and derivation.

In the following, $f(\alpha)$ will be called a spectrum of singularities or spectrum of scaling exponents, and only one of the above versions will be considered, because of its connections to multifractal models for rainfall. In Section E.2.2, the spectrum of singularities is defined through a Legendre transform and is equivalent to the multifractal spectrum derived from large deviation theory. According to this definition, the spectrum of singularities may be negative, which is not the case for the other two versions of the multifractal spectrum. However, when it is non-negative, $f(\alpha)$ may be interpreted as the Hausdorff dimension of the set of points with Hölder exponent equal to α. Since monofractal measures only have one Hölder exponent, their spectrum of singularities is defined exclusively at a single point α_0. Thus, only one fractal dimension is associated with a monofractal, while multifractal measures are characterized by a whole range of dimensions (an infinite hierarchy of dimensions in the Lovejoy and Schertzer terminology).

Multiplicative Cascades

Multifractal rainfall fields are usually generated via the multiplicative random cascade theory. To describe a multiplicative cascade, consider an iterative procedure with an infinite number of steps. A fixed mass, initially uniformly distributed over its support, is iteratively redistributed over finer and finer partitions of the support by multiplication with random weights. Schematically, the iterative process consists of the following steps:

Step 0. Start with an initial mass density W_0, which is uniformly distributed over its support. To simplify notation, assume the support is the unit square and proceed with a 2-D example.

Step 1. Divide the support into $b = \lambda^2$ sub-squares of side length $l_1 = L/\lambda$, where L is the largest scale of interest ($L = 1$ in this example) and λ is an integer. Generate b random weights $W_1(1), W_1(2), \ldots, W_1(b)$ which are independent and identically distributed as some W, with $E(W) = 1$. Now redistribute the mass density on the b sub-squares by multiplying W_0 by the weights, resulting in the densities $W_0W_1(1), W_0W_1(2), \ldots, W_0W_1(b)$.

$\vdots$

Iteratively repeat the procedure on each of the sub-squares

$\vdots$

Step n. Divide each of the sub-squares defined by steps up to $n-1$ into b sub-squares of side length $l_n = L/\lambda^n$. The resulting number of sub-squares at step n is b^n. Also, for each of the b^{n-1} sub-squares from steps up to $n-1$, generate b random weights. The b^n weights generated at step n, the W_n's, are iid as W and are independent of the W_1's, W_2's etc. Obtain $W_0 W_1(\Delta_n) \cdots W_n(\Delta_n)$, where $W_j(\Delta_n)$ $(j = 1, \ldots, n)$ is the weight assigned at step j to the subsquare on the path to Δ_n, and Δ_n is one of the b^n subsquares at step n.

$\vdots$

The random variable W is called the cascade *generator*, and the number b is referred to as *branching number*. The choice of distribution for the cascade generator is among the elements that distinguish different multiplicative cascade models for rainfall. In all cases, the condition $E(W) = 1$ ensures conservation of the average mass density W_0.

The steps outlined above give rise to a *discrete* cascade, in the sense that the cascade develops over a discrete range of scales. In fact, the scale ratio for each step is the constant integer λ. The choice of explaining the multiplicative process that leads to multifractal measures via discrete cascades is due to the resulting intuitive description of the generating mechanism. The *continuous* version of multiplicative cascades is not as illustrative. Thus, the details of the construction of a multifractal field via continuous cascades are postponed to Section E.3, in the context of specific models for rainfall that use the continuous version.

Multiplicative cascade processes result in multifractal fields where the mass is increasingly concentrated on sparser and sparser fractal sets (Lovejoy and Schertzer, 1990). As the multiplicative process continues to smaller and smaller scales, i.e. as the scale ratio between the largest and smallest scale goes to infinity, the field at most points goes to zero, while it goes to infinity at some points. The latter points are called *singularities*, the former *regularities* (Schertzer and Lovejoy, 1997).

Scaling Properties

The scaling properties of multifractals motivate their use in rainfall modelling. Like fractal sets, multifractal measures exhibit some form of scale invariance. Monofractals are consistent with simple scaling, while multifractals lead to multi-scaling models.

To illustrate the difference between simple and multi-scaling and the connection with multifractals, let φ_l denote the random measure φ on a box of side length l. Simple scaling of φ corresponds to equality of the joint probability distributions of φ at scales l and $l\lambda$, (for $\lambda > 0$) up to a multiplicative correction factor that only depends on λ, i.e.

$$\{\varphi_{\lambda l}\} \stackrel{d}{=} \{g(\lambda)\varphi_l\}, \tag{E.3}$$

where $g(\lambda) > 0$. The scale function $g(\lambda)$ is of the form

$$g(\lambda) = \lambda^{\theta}, \tag{E.4}$$

with θ denoting a constant scaling exponent. From (E.3) and (E.4), the moments of φ follow the scaling relationship

$$E(\varphi_{\lambda l}^q) = \lambda^{\theta q} E(\varphi_l). \tag{E.5}$$

Since θ is the same for all q, the log-log linear relationship between moments of order q and scale has a slope that is linear in q.

Simple scaling models for rainfall have not provided an adequate representation of the process (see e.g. Gupta and Waymire 1990). The shortcomings of this approach have motivated changes in the scaling literature in two different directions. While most authors focused on the development of multi-scaling models, Kumar and Foufoula-Georgiou (1993) argued that rainfall fluctuations, rather than rainfall

intensities, exhibit simple scaling behaviour. Only a brief summary of the latter approach is described here, while the rest of the exposition concentrates on multi-scaling models. In the spatial setting, Perica and Foufoula-Georgiou (1996), modelled standardized rainfall fluctuations as Gaussian fields that follow simple scaling. The standardized fluctuations at pixel size l are defined as $\zeta_{l,i} = \frac{R'_{l,i}}{\overline{R}_l}$, where $\overline{R}_l$ is the field of average rainfall intensities at pixel size l, and $R'_{l,i}$ represents the corresponding fluctuations in direction i ($i = 1, \ldots, 3$, i.e. horizontal, vertical or diagonal direction). That is, for each pixel j, $R'_{l,i}$ gives the difference between the value at the adjacent pixel in direction i and the value at j. According to the normality assumption, $\zeta_{l,i}$ has a Gaussian distribution with mean zero and standard deviation σ_{ζ_l} which, because of the simple scaling hypothesis, is related to the standard deviation at different scales by the relationship

$$\frac{\sigma_{\zeta_{l\lambda}}}{\sigma_{\zeta_l}} = \lambda^{\theta}. \tag{E.6}$$

Perica and Foufoula-Georgiou (1996) found the scaling exponent θ to depend on the convective available potential energy (CAPE). In their model, the Haar wavelet (see e.g. Percival and Walden 2000) decomposes the average field at pixel size l into the average field at the next larger dyadic pixel size and the three rainfall fluctuations at the same larger pixel size. This formulation can be used to disaggregate the output of mesoscale climate models. Given an estimate of CAPE, one obtains a value for θ and, through (E.6), the value of σ_ζ at several scales given its value at a reference scale (which, in Perica and Foufoula-Georgiou 1996, also depends on CAPE). Then, the standardized rainfall fluctuations at the current pixel size may be generated from a Gaussian distribution with the appropriate standard deviation. The inverse wavelet transform then leads from the average and three fluctuation fields to the average field at the next smaller pixel size. The simulation and filtering procedure may be repeated until the rainfall field at smallest scale of interest is reconstructed. The intermittency typical of precipitation fields is preserved by thresholding the reconstructed average field.

The space-time extension of Venugopal et al. (1999) is based on *dynamic scaling*. Their analysis is centered on the rate of evolution of rainfall, defined as $\Delta \ln R_{l,\tau} \equiv \ln R_l(t+\tau) - \ln R_l(t)$. Here τ represents the time lag and, as above, l is the pixel size. Dynamic scaling requires the distribution of $\Delta \ln R$ to remain the same through spatial and temporal scales, provided that $\frac{\tau_1}{\tau_2} = \left(\frac{l_1}{l_2}\right)^z$, where z is called the dynamic scaling exponent. The downscaling procedure of Venugopal et al. (1999) first disaggregates the large scale spatial precipitation field (e.g. the output of a regional climate model) at one time instant. That is, one goes from spatial scale l_1 to the smaller spatial scale of interest l_2. Then, the distribution of $\Delta \ln R_{l_1,\tau_1}$ is evaluated from the large scale field at time points separated by lag τ_1. According to dynamic scaling, $\Delta \ln R_{l_2,\tau_2}$ has the same distribution, provided that $\tau_2 = \tau_1 \left(\frac{l_1}{l_2}\right)^z$. Thus, given the initial large scale field at time t, one can obtain the field at fine scale l_2 at time points $t+\tau_2, t+2\tau_2$, etc.

The literature on simple-scaling applied to rainfall fluctuations, both in its spatial and space-time versions, is one of the few examples where precipitation data are used to validate the performance of the scaling model with respect to a range of characteristics of rainfall, besides scaling itself. For example, the percentage of rain-covered area, temporal and spatial correlations are reproduced adequately, although, as the initial field is evolved, some deviations from the observed spatial structure appear. The extreme intensities seem more problematic and are often underestimated or overestimated. The spatial approach of Perica and Foufoula-Georgiou (1996) was tested on a single event from the Oklahoma-Kansas Preliminary Regional Experiment, while the space-time extension was applied to rainy season convective storms in Darwin, Australia. Extensive investigation of more storms, under different climates, is necessary to assess the performance of the model, and may provide further indications on some of its aspects, such as the possible dependence of the dynamic scaling exponent z on physical quantities.

Rather than studying rainfall fluctuations, most authors, following the poor performance of simple scaling models for rainfall intensity, turned to the development of multi-scaling models, where moments of different order scale with different scaling exponents. The relationship between moments of order q and scale is still log-log linear, but its slope is not linear in q. Multifractals are central to multi-scaling approaches to rainfall modelling. In fact, multifractals are characterized by their scaling properties, which can be expressed equivalently in terms of moments or exceedance probabilities. Consider a field φ that has constant mean across scales ($E(\varphi_\lambda) = \text{constant}$), i.e. a *conserved* field. Multifractality of φ is equivalent to the property

$$E(\varphi_\lambda^q) \approx \lambda^{K(q)}, \tag{E.7}$$

with $\lambda > 1$ and $\approx$ denoting equality up to proportionality constants and logarithmic corrections. The function $K(q)$ is the multiple-scaling exponent and in general it is not a linear function of the moment order. The linear case reduces to the moments scaling with constant exponent, and thus to simple scaling. The linearity (in q) of $K(q)$ also corresponds to monofractality, as is more apparent after describing the following characterization of multifractals – which is equivalent to (E.7) – in terms of exceedance probabilities:

$$P(\varphi_\lambda \geq \lambda^\gamma) \approx \lambda^{-c(\gamma)}, \tag{E.8}$$

for a range of values of γ, where the scaling functions $K(q)$ and $c(\gamma)$ are related to each other through the Legendre transform, that is

$$K(q) = \max_\gamma[q\gamma - c(\gamma)], \quad c(\gamma) = \max_q[q\gamma - K(q)]. \tag{E.9}$$

The scaling exponent for the probability distributions, $c(\gamma)$, is also related to the spectrum of singularities mentioned in Section E.2.2. In fact, the spectrum of singularities is defined as $f(\alpha) \equiv D - c(D-\alpha)$, where D is the dimension of the underlying space and α is any possible value of the Hölder exponent. When $D \geq c(D-\alpha)$, i.e. when $f(\alpha)$ is non-negative, $c(\gamma)$ can then be interpreted as the codimension of the set with singularities of order γ. If φ is monofractal, there is only one Hölder exponent $\alpha = \theta$, since the singularities are all of the same order. Then $f(\alpha)$ and $c(\gamma)$ are only defined at a point, and from the first equality in (E.9) it follows that $K(q)$ is linear in q. On the other hand, non-linearity of $K(q)$ (i.e., multi-scaling) implies that $c(\gamma)$ and $f(\alpha)$ are defined on a range of values, which in turn is equivalent to multifractality.

The characterizations of multifractals given in this section follow Lovejoy and Schertzer (1990) and Tessier et al. (1993). The same properties are derived by Gupta and Waymire (1993) with different notation. The latter authors also establish the notation changes required to obtain one formulation from the other.

E.3 Continuous Cascade Models

A lot of research in the past two decades has focused on multifractal rainfall models where multiplicative cascades develop over a continuous range of scales (e.g. Schertzer and Lovejoy 1997, Lovejoy and Schertzer 1990, Tessier et al. 1993 and Marsan and Schertzer 1996). Earlier work has developed methods for spatial or temporal applications, while Marsan and Schertzer (1996) provided a space-time extension by introducing anisotropy and causality. Sections E.3.1 and E.3.2 describe the general framework of some continuous cascade models and the methods of estimation and simulation applicable to spatial or temporal rainfall. Then a space-time extension is given in Section E.3.3. Finally, some comments on this direction of research are provided in Section E.3.4.

E.3.1 Universal Multifractals

The cascade process results in a conserved field φ (that is, a field that has constant mean across scales) that is characterized by the scaling properties in (E.7) and (E.8). The current knowledge of the principles that govern precipitation does not allow for a clear identification of the physical field that corresponds to φ. Tessier et al. (1993) assume the relationship

$$\Delta R_\lambda \approx \varphi_\lambda \lambda^{-H}, \tag{E.10}$$

where ΔR_λ denotes the rain fluctuations at scale λ. Since $E(\varphi_\lambda)$ is independent of scale, then $E(|\Delta R_\lambda|) \approx \lambda^{-H}$, i.e. H is the order of fractional integration that transforms the field φ into the rainfall field R. In other words, H summarizes the degree of departure of the rainfall field from the conserved field.

Inferences on R are obtained through φ. Since φ is characterized by its scaling properties, one needs to estimate the scaling functions $K(q)$ and $c(\gamma)$. The corresponding scaling functions for the non-conserved field ΔR_λ are obtained by the transformations $\gamma \mapsto \gamma - H$ and $K(q) \mapsto K(q) - Hq$. The first mapping corresponds to the translation of the order of singularity by $-H$ in the exponent for the exceedance probabilities. Its motivation lies in the fact that, from (E.10), $\varphi_\lambda = \lambda^\gamma$ implies $\Delta R_\lambda = \lambda^{\gamma - H}$. The mapping of the moment scaling function is also explained by the relationship in (E.10), from which it follows that $E(|\Delta R_\lambda|^q) = \lambda^{K(q)} \lambda^{-Hq}$.

In general, estimation of the scaling exponents requires estimation of the entire $K(q)$ and $c(\gamma)$ functions. However a simplification of the problem is provided in the context of *universal* multifractals. Only a brief explanation of universality and how it is achieved is given here. For a full mathematical account one can refer to Schertzer and Lovejoy (1997). The main idea is that nonlinear mixing and scale densification in physical processes lead to universal multifractals. Nonlinear mixing refers to the multiplication of independent and identically distributed processes over the same range of scales $\Lambda = L/l$, where L is the largest scale and l the smallest resolved scale. Scale densification, on the other hand, involves introducing more and more intermediate scales over a fixed range of scales Λ. In the multiplicative cascade process that generates multifractals, before proceeding to smaller and smaller scales (i.e. taking the limit as $\Lambda \to \infty$), one can include nonlinear mixing and/or scale densification by multiplying the process by a larger and larger number of independent replicas and/or introducing more and more intermediate scales. Now, both renormalized nonlinear mixing over a finite range of scales of iid cascade processes, and scale densification over a finite range of scales of a cascade process, converge to a multifractal that admits a Lévy generator (see Schertzer and Lovejoy 1997). For this reason, multifractals with Lévy generator are called universal. A Lévy generator W is defined as $W = e^{-X}$, where X is a Lévy-stable random variable (i.e. the renormalized sum of any n variables that are iid as X, is still distributed as X, e.g. Feller 1996).

The concept of universality is not generally accepted. Gupta and Waymire (1993; 1997) argue that, while Lévy distributions represent a broad class of generators, there are other admissible choices for the distribution of W. Their criteria for selecting the distribution of the generator will be outlined in Section E.4.

The scale functions of universal multifractals are completely described by the three parameters H, C_1 and a. Using the notation of Tessier et al. (1993)

$$c(\gamma - H) = \begin{cases} C_1 \left(\frac{\gamma}{C_1 a'} + \frac{1}{a} \right)^{a'}, & a \neq 1 \\ C_1 \exp\left(\frac{\gamma}{C_1} - 1 \right), & a = 1 \end{cases} \tag{E.11}$$

and

$$K(q) + qH = \begin{cases} \frac{C_1}{a-1}(q^a - q), & a \neq 1 \\ C_1 q \log(q), & a = 1 \end{cases} \tag{E.12}$$

for $0 \leq a \leq 2$, $q \geq 0$, and with $\frac{1}{a'} + \frac{1}{a} = 1$. As mentioned above, H represents the degree to which the observed field deviates from the conserved field φ. The sparseness of the conserved field is represented by C_1, which is the order of singularity associated with the mean of the process. The conservation property, combined with (E.7) and (E.9), imply that $E(\varphi_\lambda) \approx \lambda^{K(1)} = \lambda^{\max_\gamma[\gamma - c(\gamma)]} =$ constant. Since the exponent must be equal to zero for the last equality to hold, it follows that $c(\gamma) \geq \gamma \, \forall \gamma$ and there exists a $\gamma = C_1$ (the singularity associated with the mean of the process) such that $c(C_1) = C_1$. Recalling the interpretation of $c(\gamma)$ as the codimension of the set with singularities of order γ, $C_1 \leq D$ is required for the mean not to be so sparse that the field is equal to zero almost everywhere. In general, features with $c(\gamma) > D$ are too sparse to be observed in a single D-dimensional picture (e.g. a satellite photograph in a 2-D context). Finally, a denotes the degree of multifractality, which is given by the rate of change in the slope of $c(\gamma)$ at C_1, i.e. it represents the rate at which the degree of sparseness of the field varies as we move away from the mean. The monofractal case corresponds to $a = 0$, since the singularities are all of the same order C_1. The upper bound is $a = 2$, which corresponds to the maximum degree of multifractality.

The multifractality parameter a also corresponds to the Lévy index and, according to its value, the universal multifractal belongs to one of five qualitatively different classes. The cases $a = 2$, $1 < a < 2$ and $a = 1$ correspond to the lognormal, log-Lévy (with unbounded singularities) and log-Cauchy multifractals respectively. All these are unconditionally hard multifractals (Tessier et al., 1993) in the

sense that, for any D, there is a finite q_D such that moments of order q_D and higher diverge. The case $0 < a < 1$ corresponds to the (log)Lévy multifractal with bounded singularities, which displays conditionally hard behavior, i.e. for large values of D, the critical order for divergence (q_D) is not finite. The divergence of high order moments is related to the distinction between "bare" and "dressed" cascades (Lovejoy and Schertzer, 1990; Tessier et al., 1993). The properties of multifractals described earlier apply to the bare quantities, which result from the construction of a cascade over a finite range of scales. In practice, bare quantities are not experimentally available, since measuring devices can only resolve scales that are much larger than the scale at which the process becomes homogeneous. Atmospheric process can usually be regarded as homogeneous on intervals (squares) of $1mm$ or less (see e.g. Tessier et al. 1993), while, for example, rainfall radars have a resolution of the order of kilometers. Thus, the cascade is completed, i.e. developed to its small scale limit, which is mathematically singular (Lovejoy and Schertzer, 1990), and then integrated over a finite scale. The resulting quantities are called dressed and their high order moments generally diverge (except for the case when $a < 1$), i.e. they follow the hard behaviour mentioned above.

E.3.2 Estimation and Simulation Methods

Several techniques have been proposed for the estimation of the scaling functions $c(\gamma)$ and $K(q)$. Some authors estimate $K(q)$ and then obtain $c(\gamma)$ via the Legendre transform (see Halsey et al., 1986). Lovejoy and Schertzer (1990) propose a Probability Distribution/Multiple Scaling (PDMS) technique to estimate $c(\gamma)$ directly (see Lavallée et al. 1990 for another version of the PDMS method). The basic idea is to successively degrade the resolution of the field, form probability distributions at the different scales and determine $c(\gamma)$ from (E.8). This technique is related to functional box counting (Lovejoy et al., 1987), where the field is first transformed into exceedance sets and the resolution of the exceedance sets (rather than the measure) is degraded. Details of these methods are not described here, since Tessier et al. (1993) later applied the Double Trace Moment (DTM) technique (Lavallée, 1991), which determines the scaling functions by estimating the universal parameters C_1, a and H directly. This approach seems preferable to the estimation of the entire scaling functions.

To outline the DTM method, consider the two dimensional example of a satellite picture of side length $L = 1$. Denote by λ' the ratio of the picture scale to the scale of homogeneity for the conserved field φ (note that, since φ is conserved, we assume $H = 0$). For ease of exposition, assume that λ' is equal to the resolution of the measuring device, i.e. the pixel resolution. Then cover the picture with λ^2 disjoint sub-squares $\Delta^i_\lambda, i = 1, \ldots, \lambda^2$, of resolution $\lambda < \lambda'$. That is, the resolution is degraded by considering squares, B, larger than the pixels. Integration of the field $\varphi_{\lambda'}$ over each sub-square yields the dressed fluxes $\Phi_{\lambda'}(\Delta^i_\lambda)$. Now, the q-th order trace moments are defined as

$$Tr(\varphi_{\lambda'})^q \equiv E\left(\sum_i \left(\Phi_{\lambda'}(\Delta^i_\lambda)\right)^q\right) \approx \lambda^{K(q)-(q-1)D}, \tag{E.13}$$

where $D = 2$ in the present example. Note that dividing the Φ's by $(\lambda'/\lambda)^2$ (in general $(\lambda'/\lambda)^D$), i.e. the ratio of the sizes of the squares at scales λ and λ', ensures the conservation property is respected and leads to the (dressed) conserved field at scale λ. The q-th moments of these quantities are responsible for the $\lambda^{K(q)}$ term. Since the summation is over λ^2 disjoint sub-squares, the term λ^2 (in general λ^D) appears. The λ^{-qD} term is due to the fact that, having divided by $(\lambda'/\lambda)^{qD}$ to obtain the conserved field at scale λ, it is also necessary to multiply by the same quantity. The approximate equality in (E.13) is valid only for $q < q_D$, with q_D denoting the critical order of divergence. For practical applications, (E.13) also requires $q < q_s$, where q_s is the maximum-order moment that can be estimated with a finite sample. See Tessier et al. (1993) for formulas that express q_D and q_s in terms of the dimension of the underlying space and the universal multifractal parameters C_1 and a. The formula for q_s involves the sampling dimension $D_s = \log N_s / \log \lambda$, where N_s in this case is the number of available satellite pictures. With N_s photographs one can explore structures for which $\gamma < \gamma_s$ with $c(\gamma_s) = D + D_s$. Larger numbers of photographs translate into accessibility of features with higher orders of singularity.

One may obtain $K(q)$ through (E.13) and then, via (E.12), estimate C_1 and a as the parameters of a non-linear regression of $K(q)$ on q. However, the high correlation between a and C_1 results in an ill-conditioned regression. To overcome this problem, the DTM technique introduces a second exponent η by the transformation $\varphi_{\lambda'} \mapsto \varphi^\eta_{\lambda'}$. Integration of $\varphi^\eta_{\lambda'}$ over the sub-squares Δ^i_λ gives the "η fluxes"

$\Phi_{\lambda'}^{(\eta)}(\Delta_\lambda^i)$. The *double* trace moment is then defined as

$$Tr(\varphi_{\lambda'}^{\eta})^q \equiv E\left(\sum_i \left(\Phi_{\lambda'}^{(\eta)}(\Delta_\lambda^i)\right)^q\right) \approx \lambda^{K(q,\eta)-(q-1)D}, \tag{E.14}$$

where $K(q,\eta)$ is a new exponent that obeys $K(q,1) = K(q)$. The transformation to double trace moments (by using φ^η rather than φ) induces the mappings $\gamma \mapsto \gamma^*$, $c(\gamma) \mapsto c^*(\gamma^*)$, $q \mapsto q^*$, $K(q) \mapsto K^*(q^*) = K(q,\eta)$ (see Tessier et al. 1993 for the specific formulas). The main point of introducing the second exponent η is the transformation obtained for C_1 (Lavallée, 1991), i.e. $C_1 \mapsto C_1^* = C_1\eta^a$. In fact, since $C_1 = \left.\frac{dK}{dq}\right|_{q=1}$ and $C_1^* = \left.\frac{dK^*}{dq^*}\right|_{q^*=1}$, the following relationship holds

$$K(q,\eta) = \eta^a K(q). \tag{E.15}$$

Thus, a can be estimated as the slope parameter in the simple linear regression of $\log K(q,\eta)$ on $\log K(q)$, for fixed q. From the intercept one can obtain C_1 (through (E.12) with $H = 0$). This formulation only holds for $\max(q\eta, q) < \min(q_s, q_D)$.

Up to this point, the DTM method allows the estimation of a and C_1 for a conserved process ($H = 0$). From (E.10), the non-conserved process ΔR_λ must be fractionally differentiated by order $-H$ to obtain a conserved quantity, i.e. its power spectrum (the Fourier transform of the autocovariance function) must be power-law filtered by k^H. However, Tessier et al. (1993) show by simulation that exact knowledge of H is not required to estimate a and C_1. It is sufficient to ensure that the spectrum of the analysed field is less steep than the spectrum of the underlying conserved process, which in turn is of the form $k^{-\beta}$, with $\beta = 1 - K(2)$ ($K(1) = 0$ because of the conservation property). If this is not the case, a can still be estimated accurately, but the estimate of C_1 is further and further from the true value as the spectral slope for the analysed field (β') is above the conserved case analogue (β) by larger and larger values (i.e. the simulated conserved process is fractionally integrated by progressively larger orders). In practice, one can power law filter the spectra of the observed process until the estimates of a and especially C_1 stabilize. To avoid the use of Fourier space, Tessier et al. (1993) propose the alternative of analysing the gradient or the Laplacian of the observed field rather than the field itself. This approach corresponds to much larger orders of differentiation than necessary, but it gives the desired results. Now, the estimates of a and C_1 result in a value for $K(2)$, and finally H can be obtained as

$$H = \frac{\beta' - 1 + K(2)}{2} = \frac{\beta' - 1}{2} + \frac{C_1(2^a - 2)}{2(a-1)}, \tag{E.16}$$

where β' is the spectral slope of the observed process, i.e. the slope of the log-log linear relationship between the power spectrum $S(k)$ and k.

The whole DTM procedure may be summarized into a few schematic steps:

- Compute the spectral slope β' of the observed process, by fitting a line through the log-log plot of estimated spectrum versus frequency.
- Take the modulus of the gradient (or the Laplacian) of the observed field, e.g. replace the satellite picture by the modulus of its gradient, to obtain the conserved field $\varphi_{\lambda'}$. Since in practice one works with discrete pixels, the gradient is simply derived by taking differences.
- Plot $\log Tr(\varphi_{\lambda'}^{\eta})^q$ versus $\log\lambda$ for several low values of η and with q equal to an arbitrary small value. Obtain the slope of the relationship, i.e. $K(q,\eta)$.
- Plot $\log K(q,\eta)$ versus $\log\eta$ for some small value(s) of q. Estimate a as the slope and obtain C_1 from the intercept.
- Combine the estimates of β', C_1 and a to compute the order of integration H required to go from the conserved process φ to the non-conserved quantity ΔR_λ.

The estimated universal parameters can now be used for simulation purposes. For continuous cascades, it is more convenient to substitute the multiplicative process described in Section E.2.2 with the equivalent additive process for the logarithm of the field. Note that, given the form of the Lévy generators, this approach results in sums of iid Lévy-stable random variables. In the notation of Tessier et al. (1993), the generator is the limit of an infinite sum of iid Lévy-stable random variables and, at

resolution $\lambda = L/l$, it is denoted by Γ_λ. Such a generator is related to the conserved multifractal field by $\varphi_\lambda = e^{\Gamma_\lambda}$. Generation of Γ_λ starts with a subgenerator G_λ, which is a white noise consisting of iid random variables that have a Lévy-stable distribution with characteristic exponent a and amplitude determined by C_1. To obtain Γ_λ, the subgenerator must be fractionally integrated (on the range $[1/L, \lambda/L]$) to give a spectrum $S(k) \approx k^{-1}$. This guarantees multiple scaling of the moments of φ_λ and thus its multifractality. The conserved field φ_λ then results from exponentiating Γ_λ. Since the goal is to obtain a non-conserved process with spectral slope β, the final step is a fractional integration of order H of φ_λ, which is obtained by multiplying its Fourier transform by k^{-H}.

E.3.3 Space-Time Extension

The methods described in the previous section are valid for a generic dimension D of the underlying space. Thus, in principle, they may be applied to space-time problems by considering time as equivalent to an additional spatial dimension. However, this approach ignores the scaling *anisotropy* between space and time and the *causality* that characterizes the time axis. To account for anisotropy and causality, Marsan and Schertzer (1996) propose an extension to the continuous cascade processes that generate universal multifractals. The starting point is the scenario of Section E.3.2, i.e. the construction of a spatial isotropic multifractal field resulting from a continuous cascade process from scale L down to scale l. The procedure of Section E.3.2 – where a white noise Lévy subgenerator G_λ is fractionally integrated on the range $[1/L, \lambda/l]$ to obtain the generator Γ_λ – may be formulated as

$$\Gamma_\lambda(\underline{x}) = g_\lambda(\underline{x}) \star G_\lambda(\underline{x}), \tag{E.17}$$

where $\star$ denotes convolution and the spectrum $S(k)$ of the scaling filter $g_\lambda(\underline{x}) \approx |\underline{x}|^h$ (where h must be equal to D/a; see Marsan and Schertzer 1996) is non-zero only for $|k| \in [1/L, \lambda/l]$.

The first step of the extension consists in introducing anisotropy in a spatial setting by allowing the scale changing operator to act differently in different directions. In particular, anisotropy is considered in its simplest form, i.e. self-affinity. For a 2-D spatial example as in Marsan and Schertzer (1996), self-affinity corresponds to statistical invariance with respect to the scale transformation given by $x_1 \mapsto x_1/\lambda$ and $x_2 \mapsto x_2/\lambda^{1-Q}$. Equivalently, self-affinity may be expressed as statistical invariance under the action of the scale changing operator T_λ given by

$$T_\lambda = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \lambda^{-\mathcal{G}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \tag{E.18}$$

with

$$\mathcal{G} = \begin{pmatrix} 1 & 0 \\ 0 & 1-Q \end{pmatrix}. \tag{E.19}$$

When the anisotropy exponent Q is equal to zero, self-affinity reduces to isotropy. There is no unique form of the scaling filter g_λ that is required to generate a self-affine field. It suffices to ensure that

$$g_\lambda(T_\Lambda[\underline{x}]) \approx \Lambda^{-h} g_\lambda(\underline{x}). \tag{E.20}$$

Consider a scale function $\|\cdot\|$ that satisfies

$$\|T_\lambda[\underline{x}]\| = \lambda^{-1}\|\underline{x}\|, \tag{E.21}$$

then (E.20) holds for the scaling filter

$$g_\lambda(\underline{x}) \approx \|\underline{x}\|^{-h}; \tag{E.22}$$

see Marsan and Schertzer (1996) for specific scale functions that satisfy property (E.21).

When considering a space-time process, in addition to anisotropy, causality becomes important. The idea of Marsan and Schertzer (1996) is to perform a causal filtering of the subgenerator G_λ by implementing the convolution of (E.17) only on the part of the subgenerator that corresponds to the past. This is equivalent to defining a "retarded" filter $g_\lambda^{\rightarrow}(|\underline{x}|, t)$

$$g_\lambda^{\rightarrow}(|\underline{x}|, t) = g_\lambda(|\underline{x}|, t)\Theta_t, \tag{E.23}$$

where $g_\lambda(|\underline{x}|,t)$ is an anisotropic filter and Θ_t is the heaviside function

$$\Theta_t(t') = \begin{cases} 1, & t' \leq t \\ 0, & t' > t \end{cases}$$

Equivalently, in Fourier space

$$\tilde{g}_\lambda^{\rightarrow}(|\underline{k}|,\omega) = \tilde{g}_\lambda(|\underline{k}|,\omega) \star \tilde{\Theta}_t. \tag{E.24}$$

See Marsan and Schertzer (1996) for the form of $\tilde{\Theta}_t$.

The space-time version of the anisotropic filter g_λ of (E.22) is of the form

$$g_\lambda(\underline{x},t) \approx \|(\underline{x},t)\|^{-h} \tag{E.25}$$

and, in Fourier domain

$$\tilde{g}_\lambda(\underline{k},\omega) \approx \|(\underline{k},\omega)\|^{-D_{el}+h}, \tag{E.26}$$

where the elliptic dimension $D_{el} = D+1-Q$ is the effective dimension of the space-time domain. The arbitrariness of the scale function $\|\cdot\|$ may be exploited to obtain a causal filter, i.e. a retarded g_λ function. Marsan and Schertzer (1996) start the construction of the causal filter from the scale function defined by

$$\|(\underline{x},t)\| = \left(|\underline{x}|^{D_{el}-h} + |t|^{\frac{D_{el}-h}{1-Q}}\right)^{\frac{1}{D_{el}-Q}}, \tag{E.27}$$

where $\underline{x}$ and t have been non-dimensionalized by dividing by the largest spatial and temporal scales considered (here $L = T = 1$ to simplify the notation) and with the constraint $D_{el} > h$. The scale function of (E.27) satisfies the space-time equivalent of property (E.21), that is

$$\|T_\lambda[(\underline{x},t))]\| = \|\lambda^{-1}\underline{x}, \lambda^{1-Q}t)\| = \lambda^{-1}\|(\underline{x},t)\|. \tag{E.28}$$

The corresponding anisotropic scaling filter in the Fourier domain is given by

$$\tilde{g}_\lambda(\underline{k},\omega) \approx \frac{1}{|\underline{k}|^{\frac{D+1}{\nu}} + |\omega|^{\frac{D+1}{\nu(1-Q)}}}, \tag{E.29}$$

where a is the usual Lévy index and ν is such that $1/a + 1/\nu = D_{el}/(D+1)$. In the isotropic case, the effective space-time dimension is equal to $D+1$ and ν becomes the a' of Section E.3.2. Finally, a causal function $g_\lambda^{\rightarrow}$ is obtained by modifying (E.29) as

$$\tilde{g}_\lambda^{\rightarrow}(\underline{k},\omega) \approx \frac{1}{|\underline{k}|^{\frac{D+1}{\nu}} + (i\omega)^{\frac{D+1}{\nu(1-Q)}}}. \tag{E.30}$$

In order to apply these methods to practical problems, one needs to estimate the anisotropy parameter Q. Marsan and Schertzer (1996) propose a rough estimate of Q based on spectral slopes. The context is a space-time example with spatial dimension $D = 2$. For the 2-D sections $x_1 - t$ and $x_2 - t$, they estimate the slopes of the spectra $S(k)$ for $\omega = 0$ and $S(\omega)$ for $k = 0$, and then derive Q from these.

Marsan and Schertzer (1996) describe a method for forecasting space-time rainfall by using their anisotropic and causal extension to the generating mechanism of multifractals. The multifractal field φ_λ is generated by applying a transformation to a subgenerator G_λ. Thus, given a field φ_λ up to time t, the idea is to invert the transform and obtain its subgenerator G_λ, which is then used to generate future values of φ. Because of the causality of the filtering described above, this approach ensures that the predicted field results from a causal process, i.e. the past values of φ_λ are not modified by the predictor. Marsan and Schertzer (1996) also point out the importance of respecting the limits of predictability. That is, past the predictability time τ_l associated to structures at a given scale l – which they claim follows the relation $\tau_l \approx l^{1-Q}$ – these structures are erased from the predictor.

E.3.4 Comments

There are several issues that should be addressed when modelling rainfall via continuous cascade models. For example, continuous cascades do not produce real zeros, i.e. they do not naturally

incorporate dry areas in rainfall fields. Whether or not this is a problem is not clear. On one hand, appropriate thresholding may adequately deal with the presence of zeros. On the other hand, the wet-dry properties of the generated (universal) multifractal fields have not been checked against observed data. In general, research on the universal multifractal models of the previous sections has focused on validating their (multi)scaling properties, which seem to reproduce fairly adequately the scaling of observed rainfall over the considered range of scales. The performance of these models in reproducing other properties of rainfall has not been explored.

Gupta and Waymire (1993) raise another issue, related to the unboundedness of the Lévy generators. The problem is described in non-rigorous terms here and more appropriately in Section E.4.1, after introducing some notation for the approach of Gupta and Waymire (1993). The main point is that the estimation of multifractal parameters from a single realization requires convergence of the spatial sample moments to the corresponding ensemble moments as the resolution goes to infinity. However, since the masses of a cascade are highly correlated in space, the strong law of large numbers fails and the convergence of the sample moments does not hold in general. Boundedness conditions, which are not satisfied by Lévy generators, ensure that the sample moments scale with the same slope as ensemble moments. Thus, in general the universal multifractal parameters must be estimated across realizations, which may not be feasible due to non-stationarity problems.

Parameter uncertainty and its consequences on the properties of generated rainfall fields are issues that also deserve further investigation. The parameters a and C_1 are estimated by the DTM through a linear regression of $\log K(q,\eta)$ on $\log \eta$. Since a and C_1 do not depend on q, the value of q for which the regression is carried out is arbitrary. In practice, different values of q will lead to different estimates of universal parameters. Tessier et al. (1993) suggest using several values of q to improve the statistical accuracy of the estimates. However, both in a spatial (horizontal scaling of radar reflectivities) and a temporal analysis of rainfall they only consider two values, $q = 0.5$ and $q = 2$. For the temporal case, the authors report an accuracy of ± 0.2 for both a and C_1 (± 0.1 in the spatial case). Now, considering that $0 \leq a \leq 2$, and within this range there are five qualitatively different cases, an interval of width 0.4 around the estimate seems fairly large. The consequences of varying the value of a on a range of properties of simulated rainfall have not been explored.

The choice of an estimator for the power spectrum of the non conserved process is another non trivial problem. Although the estimation method is not explained in details in Tessier et al. (1993), the authors seem to resort to a periodogram (see e.g. Percival and Walden 2000). As an estimator of the power spectrum, the periodogram has an asymptotic relative bias (Hurvich and Beltrao, 1993). Therefore, fitting a line through the log-log plot of periodogram versus frequency may give misleading information on the spectral slope β' of the rainfall process. Since β' is then used to infer the amount of fractional integration that transforms a simulated conserved process into simulated rainfall, this issue should be considered carefully.

When choosing a method for the simulation of rainfall fields, it is also important to consider the required computing resources. The simulation techniques described in the earlier sections involve two fractional integrations. For the isotropic case, Tessier et al. (1993) report that 2-D simulations can be easily obtained on a personal computer, while the 3-D case – e.g. space-time simulations – is much more computationally intensive. In fact, the authors produced 3-D multifractal fields on a Cray 2. Although computer technology has improved since 1993, the required time and resources for space-time simulations are still issues that need to be investigated.

The space-time case also raises some additional modelling and estimation issues. For example, Tessier et al. (1993) had already sketched an anisotropic extension of their model, which was then developed by Marsan and Schertzer (1996). However, Tessier et al. envisage difficulties due to the fact that their temporal and spatial rainfall applications correspond to estimated values of a belonging to two qualitatively different classes. That is, the multifractal behaviour of rainfall in time and space is qualitatively different. It is not clear whether or not the approach of Marsan and Schertzer (1996) addresses this issue.

The space-time model of Marsan and Schertzer (1996) also requires estimation of the anisotropy parameter Q. The data analysed in their paper leads to an estimate of Q that departs both from the value previously obtained for raindrops by Lovejoy and Schertzer (1991), and from the turbulent value $Q = 1/3$. Thus, Marsan and Schertzer advocate the need for more extensive data analysis in order to obtain an accurate estimate of the scaling anisotropy parameter. In general, they set a complete

analysis of rainfall data as the next step of their work. However, an account of such data analysis has not been found in the literature.

E.4 Discrete Cascade Models

Multifractal measures may be generated through discrete cascades, as outlined in Section E.2.2. This section describes multifractal models for rainfall that use discrete random cascades in two different ways. The approach of Gupta and Waymire (1993) and Over and Gupta (1994) assumes that the volume of rainfall over a subsquare results from a discrete cascade, and Over and Gupta (1996) provide a space-time extension. Section E.4.2 outlines a different model, where the random coefficients of what the author calls a wavelet expansion of the rainfall field are generated via a discrete cascade (Deidda, 1999, 2000).

E.4.1 Gupta, Waymire and Over's Approach

Spatial approach

Gupta and Waymire (1993) establish an approach to spatial rainfall modelling based on discrete cascades and describe the connections with the multifractal literature of the previous sections. With the appropriate notation changes, the multiscaling properties that characterize multifractals are expressed as in the continuous cascade literature, i.e. as in (E.7) and (E.8). The generating mechanism, however, is based on a cascade that develops over a discrete range of scales. The basic construction is outlined in Section E.2.2. Here some more notation is introduced to describe the specific approach of Gupta and Waymire (1993).

Since Gupta and Waymire (1993) describe a spatial model, the dimension of the underlying space is taken to be $D = 2$. Now, at step n of the cascade, denote the i-th subsquare by Δ_n^i, $i = 1, \ldots, b^n$, and let $\mu_n(\Delta_n^i)$ be the total mass assigned to Δ_n^i. The mass $\mu_n(\Delta_n^i)$ is obtained by multiplying the the initial mass W_0 by all the n weights on the path to the i-th subsquare, and normalizing by the size of the subsquare b^{-n} (Lb^{-n} with the initial square of side length L=1), i.e.

$$\mu_n(\Delta_n^i) = W_0^{(i)} W_1^{(i)} \cdots W_n^i b^{-n}. \tag{E.31}$$

As the resolution increases, that is, as $n \to \infty$, the mass distribution μ_n tends to a limit μ_∞. Cases where μ_∞ is degenerate – i.e. μ_∞ is zero on the whole support with probability 1 – are not of interest. [Note that μ_∞ and μ_n correspond to the dressed and bare cascade measures of Tessier et al. (1993) respectively]. Let Z_∞ denote the total limit mass on the whole support (the unit square in this case), that is $Z_\infty = \mu_\infty([0,1]^2)$. When μ_∞ is non-degenerate, the mean conservation condition corresponds to $E(Z_\infty) = 1$. The limit mass in the i-th subsquare $\mu_\infty(\Delta_n^i)$ satisfies the recursive relation

$$\mu_\infty(\Delta_n^i) = \mu_n(\Delta_n^i) Z_\infty(i),\ i = 1, 2, \ldots, b^n, \tag{E.32}$$

where $Z_\infty(i)$ are iid as Z_∞ and are independent of the $\mu_n(\Delta_n^i)$. Equation (E.32) decomposes the total limit mass on a subsquare at step n into a large scale component, represented by $\mu_n(\Delta_n^i)$, and a small scale component $Z_\infty(i)$ that reflects subsquare variability.

In the context of spatial rainfall, Gupta and Waymire (1993) take $\mu_\infty(\Delta_\lambda)$ to represent the total volume of water that falls on a $l \times l$ square in the unit time, with $\lambda = 1/l$. The rainfall intensity

$$R_\lambda \stackrel{d}{=} \frac{\mu_\infty(\Delta_\lambda)}{\lambda^{-2}} \tag{E.33}$$

then directly corresponds to the conserved field resulting from the cascade process. That is, R_λ is the equivalent of field φ_λ in Section E.3. Multiscaling properties are then obtained directly for R_λ, resulting in equivalent formulations of (E.7) and (E.8). The moment scaling function is derived to be $K(q) = 2\kappa(q)$, with

$$\kappa(q) = \log_b E(W^q) = \chi_b(q) + (q-1) \tag{E.34}$$

(see Gupta and Waymire 1993 for the details of the derivation). The function $\chi_b(q) = \log_b E(W^q) - (q-1)$ is called the modified cumulant generating function, or MKP function (from Mandelbrot, Kahane and Peyriere, who first recognized its importance for random cascade theory), and the condition $-2\chi_b'(1) > 0$ is necessary and sufficient for non-degeneracy of μ_∞. Note that the multiscaling properties for R_λ are derived in terms of the dressed quantities μ_∞. Thus, the scaling of the moments of R_λ holds only for moment orders that are smaller than the critical order of divergence.

In order to estimate the slope of the log-log linear relationship of moments (or exceedance probabilities) versus scale from a single spatial realization, it is necessary that the sample moments (or proportion of pixels at scale λ that have mass at least λ^γ) get close in some sense to the theoretical equivalents as the resolution increases, i.e. as $n \to \infty$. Define the sample moments $M_n(q)$ for μ_∞ as

$$M_n(q) \equiv \sum_{i=1}^{b^n} \mu_\infty^q(\Delta_n^i). \tag{E.35}$$

The ensemble quantities are given by

$$E(M_n(q)) = W_0^q (b^{1-q} E(W^q))^n E(Z_\infty^q). \tag{E.36}$$

From (E.36), the slope of the log-log linear relationship between $E(M_n)$ and scale is obtained as $2\chi_b(q)$ (see Over and Gupta 1996). In rigorous terms one cannot speak of convergence of $M_n(q)$ to $E(M_n(q))$, since the ensemble moments diverge to infinity or go to zero with n. Instead, consider the ratio of sample to ensemble moments, i.e. the sequence

$$\{Y_n(q)\} = \left\{ \frac{M_n(q)}{E(M_n(q))} \right\}, \; n = 1, 2, \ldots, \tag{E.37}$$

and study the scaling of sample moments via

$$\log_b M_n(q) = log_b \frac{M_n(q)}{E(M_n(q))} + log_b E(M_n(q)). \tag{E.38}$$

If $\{Y_n(q)\}$ converges to some limit $Y(q)$ as $n \to \infty$, then from (E.38) it follows that, for large n, the sample moments scale with the same slope as the ensemble moments. That is, the slope of the log-log linear relationship of sample moments versus scale tends to the corresponding slope for the ensemble moments. The law of large numbers – which would imply $\{Y_n(q)\} \to 1$ as $n \to \infty$ – is not applicable because the masses of the cascade are highly correlated in space. However, Holley and Waymire (1992) showed that, if the generator W is strongly bounded, $\{Y_n(q)\}$ converges to a random variable $Y(q)$ with probability one. Then

$$\tau(q) \equiv \lim_{\lambda_n \to \infty} \frac{\log M_n(q)}{\log \lambda_n} = 2\chi_b(q). \tag{E.39}$$

See Gupta and Waymire (1993) for the original boundedness conditions of Holley and Waymire (1992). Slightly generalized sufficient conditions for the range of values of q where (E.39) holds are (Over and Gupta, 1996)

$$E(W^{2q})/E^2(W^q) < b \quad \text{and} \quad E(Z_\infty^{2q}) < \infty. \tag{E.40}$$

Similar arguments hold for the exceedance probabilities. In particular, Gupta and Waymire (1993) obtain expressions for scaling slopes of the "fractional wetted area", i.e the proportion $\overline{p}(\lambda)$ of rainy pixels in a single spatial scan, and the corresponding ensemble quantity $p(\lambda) = P(R_\lambda > 0)$. Under the same boundedness conditions on W mentioned above, as $\lambda \to \infty$ the scaling slope for $\overline{p}(\lambda)$ converges to the scaling slope of $p(\lambda)$, which is $2\log_b(1-r)$, with $r = P(W = 0) > 0$.

The results on the convergence of the scaling slopes affect the choice of a generator W for the random cascade. In fact, only bounded W's allow estimation of the scaling slopes from single realizations. When W is unbounded (as in the case of Lévy generators), the estimation must rely on multiple spatial scans, which in turns requires stationarity. Gupta and Waymire (1993) argue in favour of single scan analysis, in order to avoid assuming stationarity a priori.

In Over and Gupta (1994), the spatial analysis of single realizations from the GATE dataset over a 10 day period, relates the parameters of the generator with large scale forcing, represented by the large

scale rain rate $\overline{R}$. The chosen generator is composite, i.e. $W = BY$, where B partitions the domain into dry and wet areas, and Y is a strictly positive generator. Specifically, B is the β-model, which is described by

$$P(B=0) = 1 - b^{-\beta} \quad \text{and} \quad P(B = b^{\beta}) = b^{-\beta}. \tag{E.41}$$

Y is taken to be lognormal, i.e. $Y = b^{\nu+\sigma X}$, where X is a standard Normal random variable. Appropriate restrictions on the parameters ensure validity of the conditions in (E.40) – at least for $q \in [0,1]$ –, and non-degeneracy of the limit mass (see Over and Gupta 1996 for a summary). The MKP functions for the β-model, the lognormal generator, and the composite generator are denoted by $\chi_{b,B}(q)$, $\chi_{b,Y}(q)$ and $\chi_{b,W}(q)$ respectively. Fixing $b = 4$, the estimates of the MKP functions lead to estimates of β and σ (ν is determined from σ and the condition $E(Y) = 1$) for each spatial scan. The parameter β is found to vary with $\overline{R}$, while the values of σ do not show strong dependence on the large scale forcing.

Space-Time Extension

Over and Gupta (1996) propose a space-time extension of the discrete cascade models of Section E.4.1. Their main goal is to obtain a model that satisfies the properties of consistency, causality and contingency. The requirement of consistency means that the space-time cascade reduces to a spatial cascade at any fixed time point. Contingency, instead, requires the evolution of the cascade to respond to time-varying forcing.

The space-time model is a natural extension of the spatial version, where the cascade weights are iid as a stochastic process $\{W_t\}$ rather than a random variable W. For any fixed t, $P(W_t \geq 0) = 1$, and the mean conservation property requires $E(W_t) = 1$. This space-time cascade reduces to the spatial equivalent at any fixed time point, thus satisfying the consistency property. Causality here is imposed directly on W_t, for example by assuming Markovian dependence. Finally, contingency on time-varying forcing can be built in the specific form of W_t.

At any fixed points in time, e.g. t_1 and t_2, the marginal distributions of the spatial generators W_{t_1} and W_{t_2} may be estimated via the scaling properties of the spatial moments, as in the previous section. In order to obtain information on the joint distributions of W_t, Over and Gupta (1996) study the scaling properties of the Lagrangian temporal cross moments and obtain analogue results with respect to the spatial case. The sample Lagrangian temporal cross moments are given by

$$M_n(q;t_1,t_2) = \sum_{i=1}^{b^n} \mu_\infty^q(\Delta_n^i;t_1)\mu_\infty^q(\Delta_n^i;t_2). \tag{E.42}$$

The same subsquare Δ_n^i is considered at times t_1 and t_2, i.e. the subsquare is tracked while the field advects (i.e. moves horizontally). The expected values of the sample Lagrangian moments are the ensemble moments (see Over and Gupta 1996 for the ensemble moment formula), which scale with slope $\chi_b(q;t_1,t_2) = \log_b E[W_{t_1}^q W_{t_2}^q] - (2q-1)$. The slope $\chi_b(q;t_1,t_2)$ is called the generalized MKP function. Note that W_t is not a priori required to be stationary. If W_t is stationary, then the scaling slope of the ensemble Lagrangian cross moments only depends on the lag $\Delta t = t_2 - t_1$.When W_t is not stationary, the joint distribution of the generator must be investigated via single space-time realizations. Estimation of $\chi_b(q;t_1,t_2)$ via a single realization is possible under the same conditions as in the spatial case. Following the steps of the spatial derivation, define the sequence $\{Y_n(q;t_1,t_2)\}$ as

$$\{Y_n(q;t_1,t_2)\} = \left\{ \frac{M_n(q;t_1,t_2)}{E(M_n(q;t_1,t_2))} \right\}, \;\; n = 1,2,\ldots \tag{E.43}$$

Under the equivalent of (E.40), $Y_n(q;t_1,t_2)$ converges to the random limit $Y(q;t_1,t_2)$ as $n \to \infty$. Therefore, the sample temporal cross moments $\tau(q;t_1,t_2)$ from a single space-time realization scale with a slope that converges to the generalized MKP function, as the spatial resolution goes to infinity.

Comments

With respect to the continuous cascades of Section E.3, discrete cascades are more intuitive generating mechanisms and allow direct inclusion of dry areas. However, a disadvantage of the discrete version is

the artificial rectangular structure that results from iteratively subdividing a rectangular region into smaller and smaller rectangles.

Perhaps the main issue that arises when applying the methods of Gupta, Waymire and Over to the analysis of rainfall, is the absence of techniques that allow for the estimation of both the branching number b and the parameters of the distribution of W. The scaling slope for the sample moments provides an estimate for the MKP function, from which the parameters of the distribution of the generator are inferred, provided a value for b is given. For example, in the GATE application of Over and Gupta (1994), the branching number is fixed at $b = 4$. While the idea of simulating rainfall fields conditionally on large scale forcing may be appealing, it is important to consider the problem of whether the assumed value for b is appropriate. When the main concern is to study the scaling properties of rainfall, a solution consists of restricting W to belong to the class of log-infinitely divisible generators. In fact, if the distribution of $\log W$ is infinitely divisible, the branching number and the parameters of the distribution of W may be varied together leaving the MKP function unchanged (Over and Gupta, 1996). However, the distribution of the limit masses μ_∞ – and thus the distribution of R_λ – still depends on b. For this reason, when scaling is not the only property of interest, the choice of b is still an issue to be taken into account.

The consequences of different choices for the value of b, in terms of the model ability to reproduce many important properties of rainfall, have not been explored. In general, attention is focused exclusively on investigating the scaling properties of rainfall. While Gupta and Waymire (1993) describe cascade models as an ideal setting for the analysis of intermittency and extremes, the model performance in terms of such properties appears not to have been assessed.

In the context of the space-time extension of discrete cascade models for rainfall, Over and Gupta (1996) list several open problems. For example, the specific parametric form of the space-time cascade generators is an issue that needs further investigation. In order to identify an adequate structure for the generators, Over and Gupta suggest model validation against data with respect to different "predictions" of the theory. In conclusion, it seems that Gupta, Waymire and Over's approach is not readily applicable, but rather requires considerable model development.

E.4.2 Expansions with Random Coefficients

Spatial and Space-Time Approaches

Deidda (1999) proposed a discrete cascade model that differs from the approach of Section E.4.1. In his spatial formulation, the multifractal field φ_λ on the unit square is obtained as a wavelet expansion with coefficients from a discrete random cascade,i.e.

$$\varphi(x,y) = \sum_{j=0}^{N} \sum_{k_x=0}^{2^j-1} \sum_{k_y=0}^{2^j-1} a_{j,k_x,k_y} \Psi_{j,k_x,k_y}(x,y), \tag{E.44}$$

where j denotes the level in the cascade and k_x and k_y indicate the spatial position. The two-dimensional wavelet $\psi_{j,k_x,k_y}(x,y)$ is obtained as a product of two one-dimensional wavelet basis functions $\Psi(z)$, i.e.

$$\psi_{j,k_x,k_y}(x,y) = 2^{2j}\Psi(2^j x - k_x)\Psi(2^j y - k_y). \tag{E.45}$$

A multiplicative discrete cascade generates the coefficients a in (E.44). Each a_{j,k_x,k_y} is derived from multiplying its parent at level $j-1$ by a random weight W_{j,k_x,k_y}. That is

$$a_{j,k_x,k_y} = a_{j-1,\frac{k_x}{2},\frac{k_y}{2}} W_{j,k_x,k_y} \tag{E.46}$$

All the W_{j,k_x,k_y}'s are iid from a generator W. Specifically, Deidda (1999) chose a log-Poisson distribution for the generator, i.e. $W = e^{\xi}\beta^Y$, where ξ and β are parameters and Y is a Poisson distributed random variable with parameter ϑ. The moment scaling function of the resulting multifractal field φ is given by (Deidda, 1999)

$$K(q) = \log_2 EW^q - q(2 + \log_2 EW) = -2q - \vartheta\frac{q(\beta-1)-(\beta^q-1)}{\ln 2}. \tag{E.47}$$

The function $K(q)$ depends only on the parameters β and ϑ, which may then be estimated by using sample moment scaling slopes.

The space-time extension of Deidda (2000) rescales the time dimension by a velocity V, and then describes space-time rainfall intensity as a 3-D process, where two coordinates correspond to space and the third to rescaled time (Vt). In most of the data analysis, V is kept constant with scale, although the case $V \approx \lambda^{-Q}$ is also considered (recall that $\lambda = 1/l$, where l is the side length of the subsquares at the considered spatial resolution). In the expansion of (E.44), ψ is now a three-dimensional function defined by the product of three one-dimensional functions. Also, a fourth index k_t is introduced to denote the time position. As in Deidda (1999), the generator is taken to be log-Poisson and the moment scaling function is given by

$$K(q) = \log_2 EW^q - q(3 + \log_2 EW) = -3q - \vartheta \frac{q(\beta - 1) - (\beta^q - 1)}{\ln 2}. \tag{E.48}$$

Denoting the sample moment scaling slopes by $\tau(q)$, the parameters ϑ and β are derived from the minimization of $\sum_q \left[\frac{\tau(q) - K(q)}{q-1} \right]^2$.

The spatial and space-time models of Deidda (1999; 2000) are used to downscale the large-scale rainfall intensity, which enters in the first term a_0 of the cascade, to the rainfall intensities at the desired scale in space and time.

Comments

There are several issues with Deidda's formulations. In Section E.4.2, the term "wavelet" was used inappropriately, following Deidda (1999; 2000), to describe the expansion in (E.44). In fact, the basis function $\Psi(z)$ adopted by Deidda is a Gaussian density with mean $1/2$ and standard deviation 0.15, truncated to the interval $[0, 1]$. Thus $\Psi(z)$ does not integrate to zero (in fact, it integrates to almost one) and consequently it is not a wavelet in the usual sense (see e.g. Chui 1992 or Percival and Walden 2000). Although the expansion provided by the author may be a valid approach, neither the spatial nor space-time formulation is a wavelet-based model. Also, there are additional cases of misuse of statistical terminology.

Some mathematical details of the approach need to be checked further, for example the formula for the initial coefficient a_0 at the top of the cascade in terms of the integral of the process φ over the spatial or space-time support. In general, the imprecisions in the description of basic concepts, although they may be due simply to careless exposition, cast some doubt on the validity of the whole procedure.

Another issue is the estimation of ξ in the log-Poisson distribution. It is not clear from Deidda's exposition whether or how ξ is estimated. When downscaling the large scale rainfall intensity, in order to create the coefficients of the expansion, random weights are generated from a log-Poisson with parameters ξ, β and ϑ. Thus, a value for ξ is needed.

Despite these problems, it is interesting that Deidda verifies the performance of his model in terms of several properties in addition to scaling. The results of the data analysis in Deidda (1999) are not representative, since the data are selected by requiring a very strong log-log linear relationship between moments and scale; that is, by requiring that scaling holds. However, the data selection criteria of Deidda (2000) are different and more objective. The ability of the model to reproduce several observed properties of rainfall is investigated by simulation. The coefficients of variation, skewness and kurtosis seem to be reproduced adequately for a range of values of the large scale field. Cumulative distribution functions of rainfall rates are also shown, with 90% confidence intervals constructed by simulation. Although these results seem fairly adequate, it is difficult to judge the performance at small values of the rainfall rate on the log-scale plots.

Further study would be required in order to use the model of Deidda (2000) for long (e.g. several decades) continuous simulations of rainfall. In fact, either one assumes that time scaling continues to hold for extended time periods (with respect to the 16 hours of Deidda's case study), or the simulations are done in shorter time chunks that need to be linked into a coherent rainfall sequence. It also seems that implementing the methods described above, especially when generating long space-time rainfall simulations, would require substantial computing resources.

E.5 Discussion

Multifractal models for rainfall are in general mathematically complex. While the basic idea of (multi)scaling may be intuitive, the same cannot be said of the concept of multifractal measures. The end methods applied to observed data usually involve simple estimations of slopes of log-log linear relationships. However, the derivations involve complex measure theoretical arguments that are bound to resurface whenever the models need to be modified for specific applications. An example that arises from the previous sections is the choice of a cascade generator. In fact, the distribution of the generator has important consequences on properties of the cascade and viable estimation methods, e.g. the non-degeneracy of the limit masses and the feasibility of single scan analysis. This is an important issue, especially when the goal is to provide a space-time model that may then be used for national application.

If multifractal models are considered as an alternative to existing point process approaches to space-time rainfall modelling, it should also be clear from the literature that multifractals lead to an improved ability to reproduce a range of observed properties of rainfall. However, with very few exceptions, only scaling properties of rainfall have been investigated in the context of multifractal models. For example, it has been argued that cascade theory accommodates the behaviour of extremes, but the claim has not been verified against data. The performance of multifractal models with respect to this and other properties of rainfall still remains to be assessed and compared with the performance of point processes.

Unpublished work by Mouhous (2003) identified difficulties and shortcomings of the random cascade approach to rainfall modelling. First, the use of the structure function to estimate the cascade parameters is limited by the bias of the estimator. Corrections to this bias only lead to an increase in the estimator variance. Second, the inability to reproduce dry period statistics that is inherent to the model structure has not been addressed by researchers in this field. Third, Mouhous argues that simulations with random cascades overestimate rainfall extremes, basing her conclusion on a comparison of Gumbel fits to simulated and observed data. The results obtained by Mouhous however rely heavily upon the claim that the log-Poisson cascade is more appropriate. Schertzer (2003) argues that the choice of the type of cascade has to be guided by more than the examination of statistics such as the structure function, but also by physical considerations. Additionally, the Gumbel assumption made in the extreme value validation is not germane to a multifractal framework (Chaouche, 2001) and not clearly grounded statistically. Mouhous' results are however interesting in that they have brought to light problematic issues that need to be addressed by multifractal modellers and will thus (hopefully) provide an impetus for a real debate between them and the hydrological community.

Specific issues that arise in the context of different multifractal models for rainfall have been described in previous sections and are briefly summarized here. The continuous cascades of Schertzer, Lovejoy, Tessier and Marsan are the least intuitive version of multifractal models and their generating mechanism seems fairly computationally intensive. The appeal of this class of models is the parsimony that derives from universality. In fact, only three parameters entirely characterize the scaling functions for the moments and exceedance probabilities. Despite this advantage, some estimation and uncertainty assessment issues are still open. In the space-time context the problem is exacerbated by the fact that the multifractal behaviour of rainfall in space and time may be qualitatively different, hence more difficulties in estimating the Lévy index a arise. Finally, appropriate thresholding is required to incorporate dry areas in the context of continuous cascades. Details of a thresholding method and its consequences on the wet-dry properties of simulated rainfall have not been described.

The discrete cascade approach of Gupta, Waymire and Over allows direct inclusion of non-rainy areas. However, simulations from this type of models have an artificial rectangular structure that derives from the construction of the cascade. Also, the choice of a branching number b is still an open problem. In fact, b cannot be estimated together with the parameters of the distribution of the generator, but needs to be fixed at an arbitrary value. Since the distribution of the limit masses of the cascade depends on b, this may have practical consequences on the properties of simulated rainfall fields. Although a space-time extension of this approach has been outlined, several aspects of the space-time model, including the parametric form of the generator, need further development.

Finally, Deidda's approach is based on an expansion with random coefficients from a discrete cascade. Besides investigating the scaling properties of precipitation fields, Deidda's research does also focus on

evaluating the model performance in terms of several other observed properties of rainfall. The relevance of the results of the spatial model is limited by the data selection criteria. That is, the data were chosen so that the model would fit well. When validating the space-time model, different more appropriate criteria were used to select parts of the GATE dataset. Simulations from the space-time model seem to reproduce several observed properties of rainfall fairly adequately. However the log-scale in the plots of the cumulative distribution function tends to obscure the behaviour at small (or zero) precipitation values. More extensive validation is still needed, as well as further checking of some mathematical details.

In conclusion, a readily applicable multifractal approach to space-time rainfall modelling and simulation is not available. Further model development, and especially extensive validation against data in terms of several properties, are required for all the multifractal approaches described in the previous sections.

Appendix F Radar data issues

Work Package 2 (detailed within Part III) involved modelling of spatial-temporal rainfall data, with radar rainfall fields used as input to these models. There were several issues encountered in the case studies relating to the data used - this appendix details those issues.

F.1 Introduction

Earlier work, under FD0426, made use of a very limited set of data from the Wardon Hill radar. For the current project, a much more extensive set of data from the Chenies radar, north west of London, has been available. This appendix describes these radar data, identifies some problems that have arisen with their use, and discusses ways in which these have been resolved. The Chenies data set comprises radar data from 1990 to 2002. Images at a 5×5km^2 pixel resolution are available at 15-minute intervals within a disc of 210km radius from the radar, while images at a higher, 2×2km^2, resolution are available at 5-minute intervals up to a radius of 76km. It takes one minute for the radar beam to make a complete revolution and thus to produce a single image, which will be treated as representing an instantaneous snapshot of the spatial rainfall field. The left hand panel of Figure 3.1 shows the radar location, as well as the extent of the 2km and 5km data. There are frequent missing data in the Chenies record: breaks in the record may have durations ranging from a single image to several days. The recorded intensities are discretised so that the smallest non-zero value is 0.03mm/hr.

Radar data need to be calibrated against rain gauge measurements if they are to represent 'ground truth'. There are a number of reasons for this. First, in radar measurement, empirical relationships are used to transform the reflected energy within the radar beam to a rainfall intensity (see Collier (2000)), and these may introduce inaccuracies. Also, the radar beam measures reflected energy at a height in the atmosphere that increases with distance from the radar and may pass over the precipitation occurring at ground level. Obstacles may intercept the radar beam causing systematic anomalies, while background noise can cause 'clutter' in radar images. Further, the radar images represent spatial averages whereas the gauges are point values; Austin (2001) gives more detail on such issues. An example of a radar image at the 2×2km^2 resolution is shown in the right hand panel of Figure 3.1). This image has not been calibrated, and illustrates spurious radial features caused by systematic anomalies and, at the north western edge, a failure to detect rainfall because the radar beam is above the cloud level.

The first step in calibrating the Chenies radar data is to remove systematic anomalies. These can be identified by averaging all of the radar images over a long time period (e.g. a year) and examining the resulting spatial image, which should be fairly smooth. Any pixels with anomalous behaviour are marked and are treated as missing data. Clutter is removed from each radar image by setting to zero any isolated pixels that appear to be 'wet'. The images are then calibrated against 15-minute data from a network of 122 tipping bucket rain gauges (with a resolution of 0.2mm/hr), with the aim of producing radar images that are in reasonable agreement with the gauge data. In the UK, the standard procedure is that described by Moore et al. (1994). Section F.2 discusses issues that became evident in our calibration of the Chenies data and ways in which these were addressed. For purposes of flood risk assessment, it is important to reproduce correctly the distributions of rainfall intensity at various spatial and temporal scales, particularly in the upper tails. In general, we have found that uncalibrated radar data overestimate the larger hourly rainfalls, and that the calibration corrects this. There is a tendency for calibrated radar to underestimate annual totals, but the relative differences are generally small (a few percent) and acceptable for our purposes.

Once the radar data have been suitably calibrated, they can be used to fit the models described in Chapter 9. In Section F.4, we discuss how to select those events in the data set that are suitable for fitting the model for event interiors, and consider possible biases that may have been introduced in the events selected by changes in radar recording practice.

F.2 Calibration of Chenies Data: Problems and Changes

F.2.1 Method Overview

As noted above, in the UK, the standard procedure for radar calibration is that described by Moore et al. (1994). This Section summarizes the problems encountered during the practical application of the calibration procedure to the Chenies radar, and the modifications to the procedures that were adopted to address them. We consider only the calibration of the 2km resolution data, that cover a circular area of about 76 km radius around Chenies and are needed for fitting the stochastic spatial-temporal model.

We start with a brief summary of the calibration method. After removing anomalies and decluttering the image as described in Section F.1, the first step is to compute calibration factors at the gauge locations. The calibration factor corresponding to the j-th gauge, at any specific time point, is

$$z_j = \frac{R_g^j + \varepsilon_g}{R_r^i + \varepsilon_r}, \tag{F.1}$$

where R_g^j denotes the hourly intensity recorded at gauge j, i.e. the 15 minute total multiplied by 4, and R_r^i is the average of the 3 hourly intensities recorded by the radar at the pixel $i = i(j)$ that contains gauge j, during the 15 minute interval ending at the time of the gauge record.

A calibration surface is then computed as

$$\underline{s} = G\underline{a} + \underline{a_0}, \tag{F.2}$$

where $\underline{s}$ is the vector of scaling factors to be applied to the radar data. The elements of the matrix G represent the Euclidean distance between the M pixel centers and the N gauges:

$$G_{ij} = d_{ij}, \quad i = 1,\ldots,M, \quad j = 1,\ldots,N. \tag{F.3}$$

Setting $s_i = z_j$ whenever gauge j is located at the exact centre of pixel i, gives

$$\underline{a} = G^{*-1}(\underline{z} - \underline{a_0}), \tag{F.4}$$

where G^* is the $N \times N$ matrix of inter-gauge distances. Rather than force the calibration surface to reproduce exactly the calibration factors at the N gauges, the method adopted by Moore et al. (1994) is to modify the matrix G^* in (F.4) by replacing the zeros along its diagonal by a parameter K. That is, G^* is defined by

$$G^*_{ij} = \begin{cases} d_{ij} & i \neq j, \;\; i,j = 1,\ldots,N \\ K \leq 0 & i = j \end{cases} \tag{F.5}$$

while G remains as in (F.3). Effectively, a non-zero value for K implies that the correlation between the calibration factors at pairs of gauges does not tend to 1 as the distance goes to zero.

The form of $\underline{a_0}$ is determined by the requirement that the surface is flat at large distances from the raingauges. The corresponding constraint is

$$\underline{a}^T \underline{1} = 0, \tag{F.6}$$

which results in the following expression for $\underline{a_0}$:

$$\underline{a_0} = a_0 \underline{1} = \frac{\underline{1}^T G^{*-1} \underline{z}}{\underline{1}^T G^{*-1} \underline{1}} \; \underline{1}. \tag{F.7}$$

The constraint in (F.6) does not always achieve its goal of avoiding problematic values of the calibration surface. The reasons are detailed in Section F.2.2. Briefly, equation (F.6) only guarantees that the slope of the surface tends to zero as the distance from the gauges goes to infinity. It does not impose any restrictions on the slope or value of the surface immediately outside the area spanned by the gauges.

The recommended values (Moore et al., 1994) for K, ε_g and ε_r, are based on the examination of a few events rather than the entire sequence, and vary in time as follows:

- Before 1990 – $K = -15$, $\varepsilon_g = 3$, $\varepsilon_r = 5$.
- Between 1990 and August 1994 – $K = -30$, $\varepsilon_g = 2$, $\varepsilon_r = 6$.
- After August 1994 – $K = -55$, $\varepsilon_g = 1$, $\varepsilon_r = 1$.

The first set of values was obtained for an area of 60 km $\times$ 60 km, using 30 gauges, while the other two were computed for the whole Thames basin, with 98 supporting gauges. We will refer to these values as the MMJB recommended values. However, in our experience, the calibration results obtained are very sensitive to the values chosen for K, ε_g and ε_r, and in the light of problems to be summarized in the next sections some of these values have been changed. It has not been within the scope of this project to explore extensively the calibration issue. Our aim has rather been to find a fairly simple rule that gives realistic rainfall images. Thus, our choice of parameter values is not optimal in any formal way and there is still room for further improvement in terms of the final gauge-radar comparison. Nevertheless, the procedure suggested here produces reasonably good agreement between the distributions of calibrated radar and gauge rainfall. In this comparison, particular attention has been paid to the extreme values, as they are central to this project. Thus Section F.2.2 describes a preliminary change to the calibration procedure, while the later sections deal with the choice of values for the three parameters and their effect on the radar-pixel comparison for Chenies radar data.

F.2.2 Calibration of regions Outside the Raingauge Range

To avoid problematic (e.g. negative) values, the calibration surface should be flat away from the area spanned by the raingauges. That is, the surface should have slope equal to 0 outside the range of the gauge network. Equation (F.6) was introduced by Bob Moore to ensure flatness at *large* distances from the raingauges.

In order to describe the effect of constraint (F.6) on the calibration surface, consider first a one-dimensional case. The value of the surface at an arbitrary point located at position x on the line is

$$\begin{aligned} s &= \sum_j |x-u_j|a_j + a_0 \\ &= \sum_{j:u_j<x} (x-u_j)a_j + a_0 + \sum_{j:u_j \geq x} (u_j - x)a_j + a_0, \end{aligned}$$

where u_j is the position of raingauge j. Taking the derivative with respect to the position x at which the calibration surface is to be evaluated, we obtain

$$\frac{\partial s}{\partial x} = \sum_{j:u_j<x} a_j - \sum_{j:u_j \geq x} a_j. \tag{F.8}$$

Outside the interval spanned by the raingauges, either $x < u_j \ \forall j$, or $x > u_j \ \forall j$. Therefore, the derivative in (F.8) is equal to $-\sum_{j=1}^{N} a_j$ (to the left of the gauges) or $\sum_{j=1}^{N} a_j$ (to the right of the gauges). The constraint in (F.6) then ensures that the calibration curve is flat everywhere outside the range of the gauges. Furthermore, the curve flattens out exactly at the left- and right-most gauge positions.

A synthetic example is shown in Figure F.21. To obtain the calibration curve (the line in Figure F.21), ten gauge and radar values were randomly generated and truncated at the respective cutoffs to produce zeros. These values result in ten calibration factors (the circles in Figure F.21), with $\varepsilon_g = \varepsilon_r = 0.1$. Setting $K = -55$ and following the procedure of Section F.2.1 we obtain the calibration curve of Figure F.21. The curve is flat outside the interval that defines the gauge positions, and its value at locations to the left (right) of the gauges is equal to its value at the left-most (right-most) raingauge.

For radar calibration, the two-dimensional case is of interest. The value of the calibration surface at an arbitrary pixel located at (x, y) is

$$s = \sum_j \left[(x-u_j)^2 + (y-v_j)^2\right]^{1/2} a_j + a_0, \tag{F.9}$$

where (u_j, v_j) denotes the position of raingauge j. The derivative with respect to the easting of the pixel

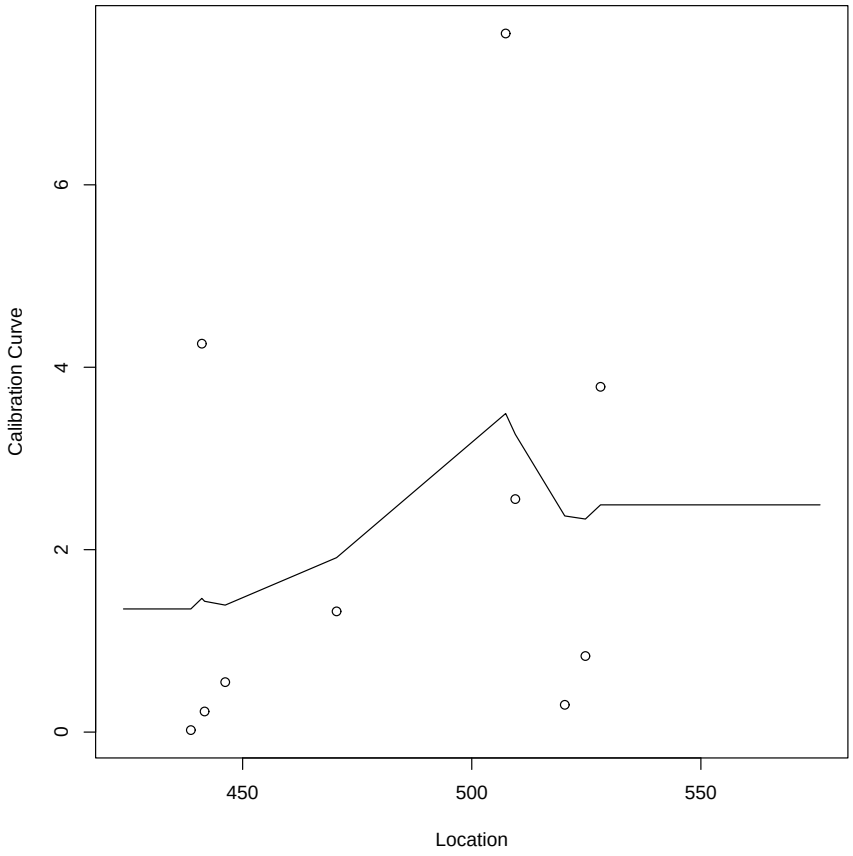

Figure F.21 Synthetic example of one-dimensional calibration curve

is

$$\begin{aligned}\frac{\partial s}{\partial x} &= \sum_{j=1}^{N} \frac{x-u_j}{[(x-u_j)^2+(y-v_j)^2]^{1/2}}\, a_j \\ &= \sum_{j=1}^{N} \text{sgn}(x-u_j)\left[1+\frac{(y-v_j)^2}{(x-u_j)^2}\right]^{1/2} a_j, \text{ provided } (x-u_j)\neq 0,\ \forall j. \end{aligned} \tag{F.10}$$

Similarly, the derivative with respect to the northing of the pixel is

$$\frac{\partial s}{\partial y} = \sum_{j=1}^{N} \text{sgn}(y-v_j)\left[\frac{(x-u_j)^2}{(y-v_j)^2}+1\right]^{1/2} a_j, \text{ provided } (y-v_j)\neq 0,\ \forall j. \tag{F.11}$$

Constraint (F.6), i.e. $\sum_{j=1}^{N} a_j = 0$, does not guarantee that either derivative is equal to zero outside the region spanned by the gauges. However, as the distance of the pixel to the gauges goes to infinity, in *both* the easting and northing components, the surface becomes flat. In fact, as both x and y go to infinity along a straight line, the pixel they define will eventually be in the region where its eastings and northings are outside the range defined by the raingauges. That is, $\text{sgn}(x-u_j)$ is the same for all j's, and similarly for $\text{sgn}(y-v_j)$. These terms can then be taken out of the sums in (F.10) and (F.11). Also, the ratio $(y-v_j)^2/(x-u_j)^2$ (and equivalently $(x-u_j)^2/(y-v_j)^2$) will tend to the same value for all j. These arguments result in

$$\frac{\partial s}{\partial x} \to c_1 \sum_{j=1}^{N} a_j, \quad \frac{\partial s}{\partial y} \to c_2 \sum_{j=1}^{N} a_j, \tag{F.12}$$

where c_1 (c_2) includes the sign of the difference in easting (northing) between the pixel and all gauges, and the common limiting value of the ratio of square difference in northing (easting) to square difference in easting (northing). Constraint (F.6) ensures that both limits in (F.12) are equal to zero.

Note, however, that if x goes to infinity while y is kept fixed in the range of the gauge northings, the partial derivative with respect to x tends to zero, but the partial derivative with respect to y does not. As $x \to \infty$ (or $x \to -\infty$), the sign of $(x-u_j)$ will be the same for all j's, and it can be taken outside the summation in (F.10). Also, since y is kept fixed, the ratio of the square difference in northings to the square difference in eastings will tend to zero for all gauges. Therefore $\partial s/\partial x \to \text{sgn}(x-u_1)\sum_{j=1}^{N} a_j$, and constraint (F.6) ensures that this limit is equal to zero. On the other hand, since y is in the range of the gauge northings, $\text{sgn}(y-v_j)$ is not the same for all j's and cannot be taken out of the summation in (F.11), and the ratio $(x-u_j)^2/(y-v_j)^2$ will tend to infinity for all gauges. Then, constraint (F.6) does not ensure that $\partial s/\partial y \to 0$ as $x \to \pm\infty$. Similarly, if y tends to infinity while x is kept fixed in the range of the gauge eastings, only the partial derivative with respect to y tends to zero. In summary, at large distances from the gauges, the surface becomes flat in the region where both eastings and northings are outside the range of the gauge eastings and northings. In the region where either the eastings or northings are in the range defined by the gauges, even at large distances, the calibration surface becomes flat only in one direction.

Since the requirement in (F.6) does not impose any restriction on the value at which the calibration surface flattens, it is possible that the surface becomes negative before it levels off. Figure F.22 shows an example calibration surface for Chenies data where the North-West corner, drawn in white for ease of visualization, is negative. The crosses mark the positions of the gauges, none of which is located inside the problematic area.

Negative values of the surface are not acceptable, since they would produce negative calibrated rainfall intensities. The easiest way (and that adopted here) to fix this is to set any values of the calibration surface that fall outside the interval $[\min z, \max z]$ either to the minimum or to the maximum calibration factor, depending on which side they fall.

F.2.3 Calibration Results with MMJB Parameter Values

The calibration method described in Section F.2.1, with the change of Section F.2.2, was applied to the Chenies radar data, for the period following 1990, and excluding 1992 (data from the years preceding 1990, and one month in 1992, had been crudely calibrated at site by the Met Office; these data cannot be properly calibrated using the current method until the effects of this precalibration have been removed). Initially, the values of K, ε_g and ε_r were taken to be the MMJB recommended values given in Section F.2.1.

If the calibration procedure is successful, the calibrated radar values at pixels that contain gauges should be fairly close to the raingauge data themselves. Three gauges – at close, mid- and large distances from the radar centre – were selected to verify the adequacy of the calibration method. Figures F.23, F.24, and F.25 show the cumulative sums of monthly rainfall from the three selected gauges and radar pixels for a selection of years. The bold dashed lines represent the gauge values, the solid lines correspond to the uncalibrated radar data, and the third line type denotes the calibrated radar data, obtained with the MMJB recommended parameter values. Figure F.23 is typical of the period between 1990 and 1994. While the uncalibrated radar data can be more or less close in either direction to the gauge values throughout the period, the common feature is that calibration always considerably lowers the radar rainfall and makes the radar versus gauge comparison even worse with respect to the original data. This is due to the chosen values of the increments ε_g and ε_r (both their size and ratio are important). Since the increments are quite large with respect to most of the recorded rainfall values, the corresponding calibration factors are often very close to $1/3$ (See Section F.2.4 for further discussion of this issue). After 1994, when the increment values are both set to 1, calibration improves the radar versus gauge comparison, although not drastically. Figure F.24 is an example of a situation where the uncalibrated radar data is quite far from the gauge equivalents. Calibration reduces the gap between radar and gauge, but leaves a considerable difference in annual totals. Figure F.25 shows a year when the gauge versus uncalibrated radar comparison is relatively good, and calibration using the MMJB recommended parameter values improves it further, at least for some gauges. However, the correction is not substantial, the solid and lighter dashed curves being fairly close to each other.

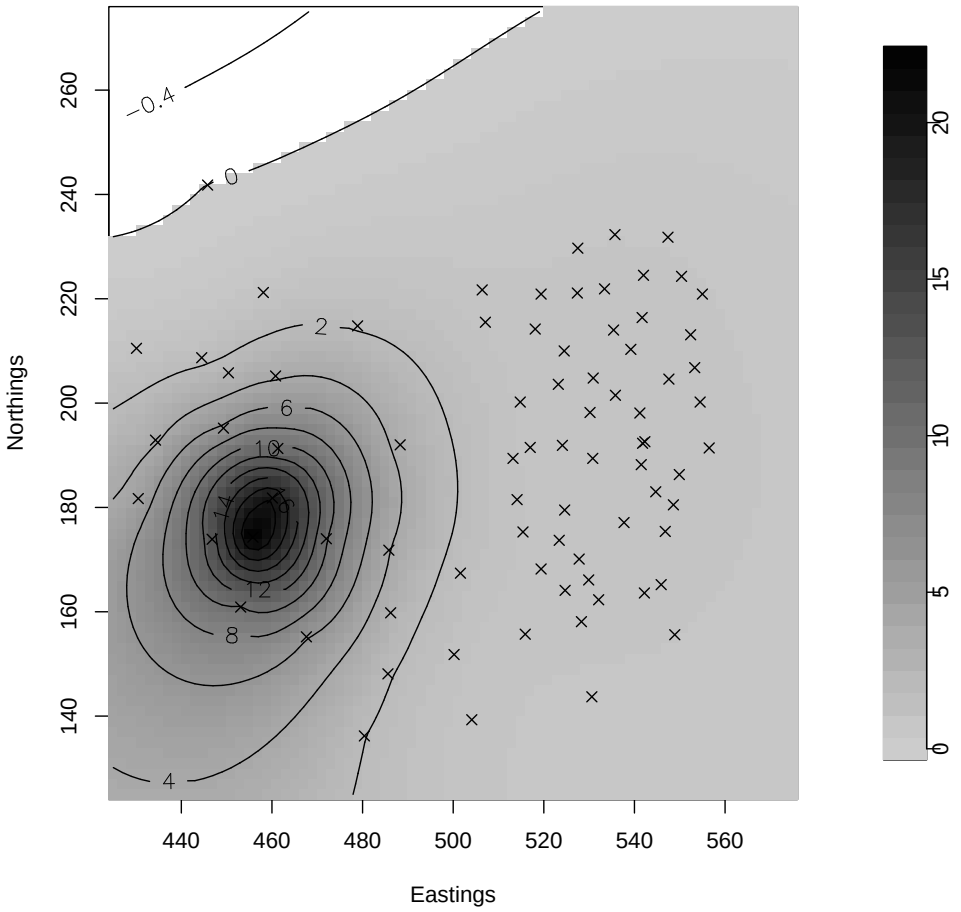

Figure F.22 Example of calibration surface (crosses show gauge positions)

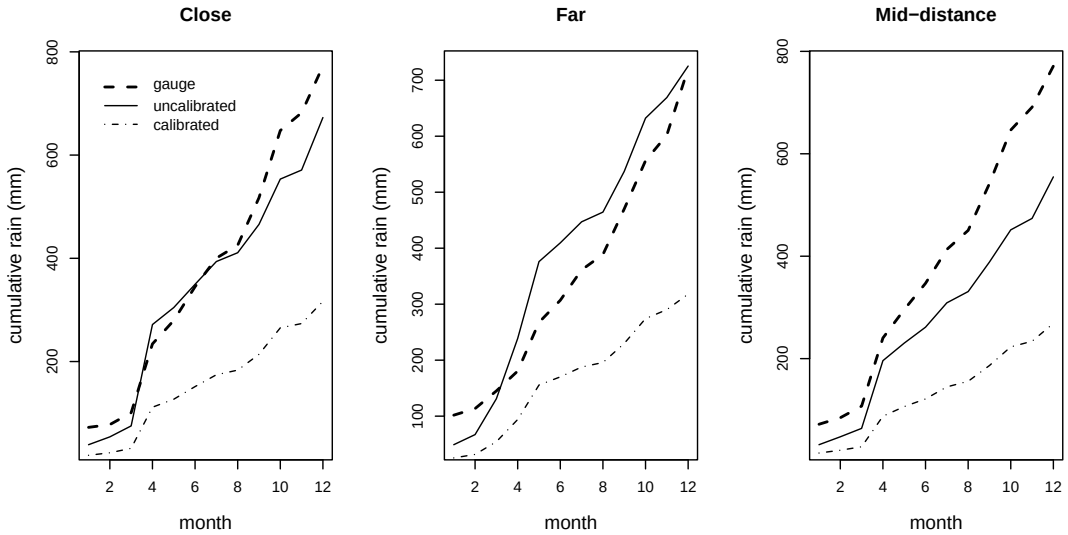

Figure F.23 1993: cumulative sums of gauge and radar monthly rainfall

In order to gain some insight on the calibration results for the whole network, it is also useful to compare the spatial mean of the gauges to the mean of the corresponding pixels, as in Figure F.26. The bold

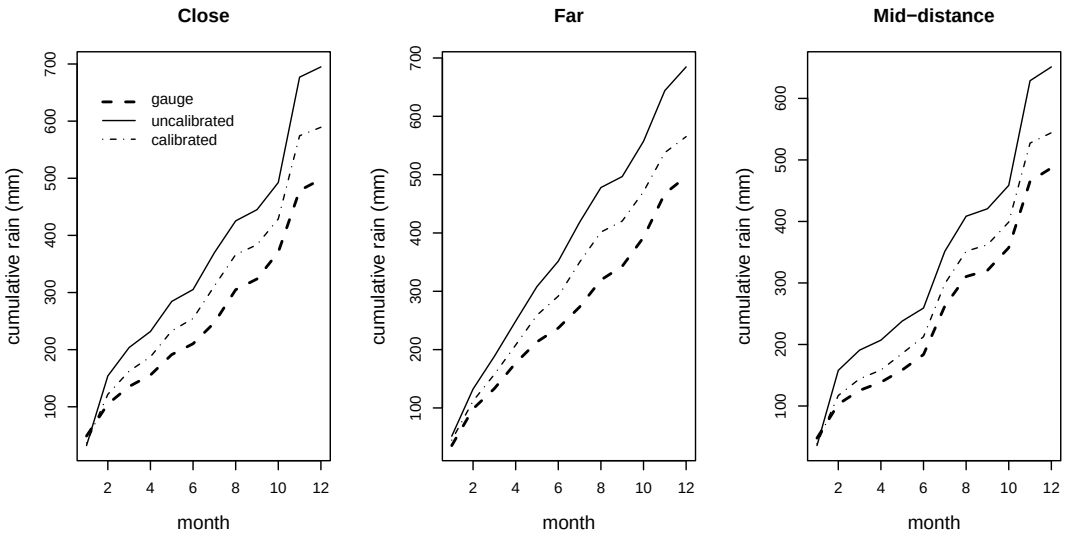

Figure F.24 1996: cumulative sums of gauge and radar monthly rainfall

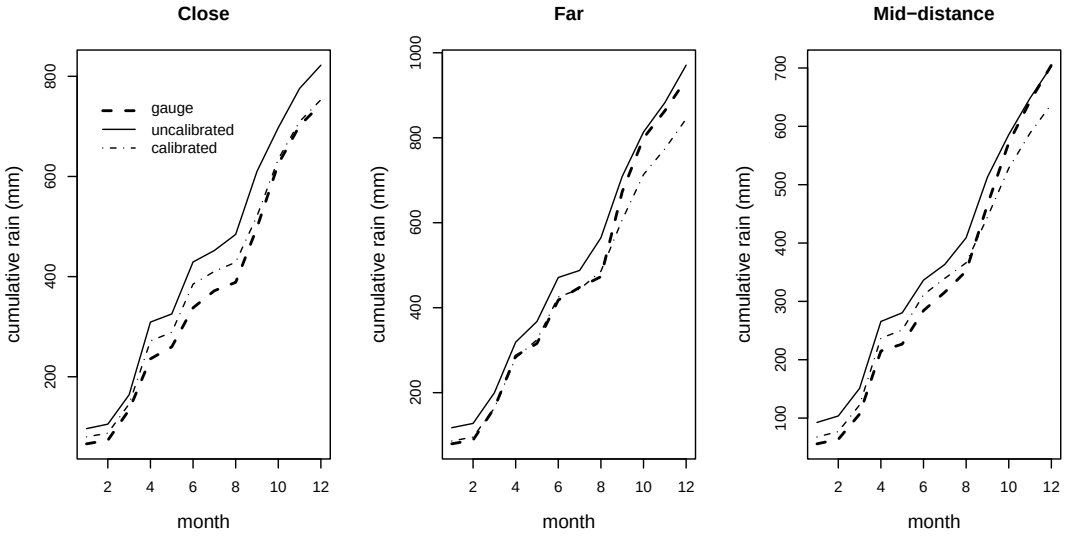

Figure F.25 1998: cumulative sums of gauge and radar monthly rainfall

dashed line is obtained by averaging all available gauge values at each time point and then computing the monthly totals for this network mean. The remaining lines are the analogous quantities from the uncalibrated and calibrated radar data, respectively. In Figure F.26, the comparison of spatial averages is shown for the same years as in Figures F.23, F.24, and F.25. The calibration results are similar to the individual gauge versus pixel comparisons above. Since radar values equal to zero are not changed in the calibration procedure, it is necessary to check whether the differences between calibrated radar data and gauge rainfall are mainly due to those values. Plots (not shown here) similar to Figure F.26, but where the spatial average was calculated excluding all pixels (and corresponding gauges) with a 0 radar value, show that this is not case. In fact, the calibrated radar data are even further away from the gauge equivalent when the zero radar values (and corresponding gauge values) are excluded from the average.

F.2.4 The Effect of the increments ε_g and ε_r

The increments applied to the radar and gauge data in (F.1) ensure that the calibration factors are defined also when the radar values are equal to zero. The effect of the increments on the calibration

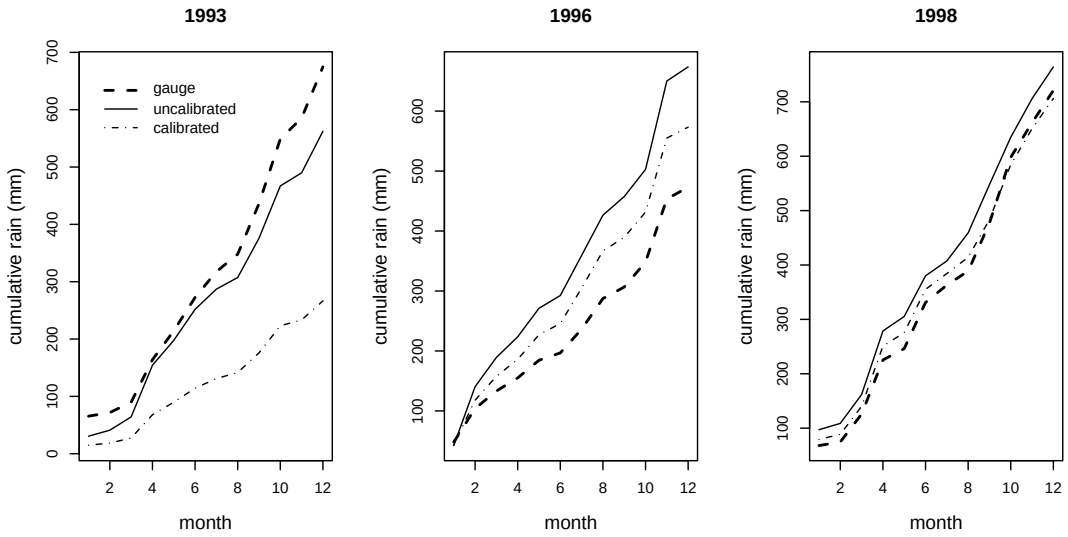

Figure F.26 Cumulative sums of gauge and radar spatially averaged rainfall

surface is important, since different sets of values for ε_g and ε_r result in very different calibrated images. In general, the values of these increments affect the calibration factors, while K determines how steeply the calibration surface departs from them (see next sections for details on the effect of K).

Figure F.27 gives a one-dimensional synthetic example of the change in the calibration curve due to varying the increment values. The circles show the calibration factors for each gauge with $\varepsilon_g = \varepsilon_r = 0.1$; the solid curve is the resulting 'surface'. In contrast, when $\varepsilon_g = \varepsilon_r = 1$, the individual calibration factors are shown as triangles, and the resulting surface as a dashed line. Since the value of K is set to -55 in both cases, the change is due to the difference in calibration factors. The dashed curve is much flatter than the solid one, since the individual triangular calibration factors are much less variable. Setting the increments to 0.001 or 0.0001, on the other hand, results in much larger peaks. Although in theory one would like the increments to be very small, to compare the gauge and radar data alone, in practice the presence of zeros complicates the issue. For example, if the radar value is zero but the gauge records positive rainfall, even its smallest possible value of 0.8 mm/hr, the corresponding calibration factor can be made arbitrarily large by choosing the order of magnitude of the increments. When $\varepsilon_g = \varepsilon_r = 0.1$ the calibration factor will be equal to 8, while when both increments are 0.0001 its value is 8000. If the gauges were located exactly at the pixel centres, the smaller the increment, the closer would the calibrated radar value be to the rainfall recorded by the corresponding gauge. However, when the gauges are away from pixel centres, the calibrated value of the pixel that contains gauge j will be affected not only by the corresponding calibration factor z_j, but also by the arbitrarily large z's close by.

Another issue, which arises from the calibration results of the previous section, is the effect of having $\varepsilon_g = \varepsilon_r$ or $\varepsilon_g \neq \varepsilon_r$. Assume for now that the gauges are located at pixel centres and the calibration surface goes through the calibration factors. Then the calibrated radar value R_r^{*i} at pixel i, which contains gauge j, is $R_r^{*i} = R_r^i \times z_j$. Therefore, if the increments in (F.1) are equal, $R_r^{*i} < R_g^j$ only if $R_r^i < R_g^j$. In other words, the calibrated radar values can be closer to the corresponding gauge values than the uncalibrated equivalents, but there is no overcompensation. If the radar values are thought to be generally larger than the gauge values, calibration will not eliminate this feature but only diminish its extent. On the other hand, having $\varepsilon_r = 2$ and $\varepsilon_g = 6$ as recommended by Moore et al. (1994) (or in general $\varepsilon_g/\varepsilon_r = 1/3$), would result in $R_r^{*i} < R_j^g$ whenever $R_r^i < 3R_g^j$. Since the gauges are not located exactly at pixel centres in the Chenies case, and the calibration surface is not forced through the z's, the above reasoning does not hold exactly. However, in the case of the 1990–1994 Chenies data, most of the calibration factors will be close to $1/3$, and therefore also the smooth surface will be close to this value.

One would like the calibrated values R_r^* to be close to the corresponding gauge values. For $R_r^* = R_g$ to hold, we must have

$$R_g = R_r^* = R_r \, \frac{R_g + \varepsilon_g}{R_r + \varepsilon_r} \Rightarrow \frac{\varepsilon_g}{\varepsilon_r} = \frac{R_g}{R_r}. \tag{F.13}$$

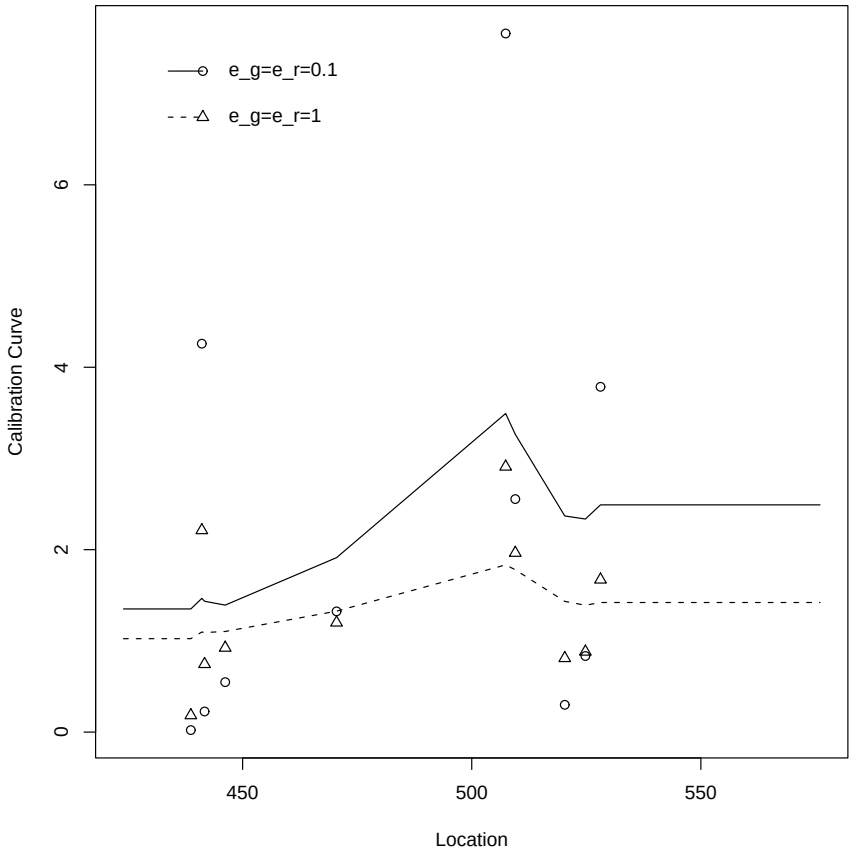

Figure F.27 Example calibration curves with different increment values

Therefore, it may be useful to define the ratio of the increments so that (F.13) is satisfied (approximately) on the average. One needs to specify in what sense the average is to be intended (e.g. in space, time or both, and whether we take the ratio of the averages rather or the average of the ratios – in the latter case there are again problems with the radar values equal to 0). In this report, it is taken to be a ratio of yearly space-time averages. This is to allow for some year to year variability, while still limiting the number of different values of ε_g and ε_r.

F.3 The Effect of K

As mentioned in the previous section, the parameter K determines how close the surface gets to the calibration factors. The simulated one-dimensional example of Figure F.38 helps to visualize the effect of K. The solid curve is the same as in Figure F.27 and is obtained using $K = -55$. The remaining two curves, obtained using $K = -15$ and $K = 0$ respectively, show that as K approaches 0, the resulting surface matches the individual calibration factors more closely. However, any calibration factor has an effect on a larger area (in space) when K is large (in absolute value). Overall, the effect of a large (in absolute value) K is a smoother surface. Note that the calibration factors, marked by circles, are the same in the 3 cases, since the values of ε_g and ε_r were kept fixed at 0.1.

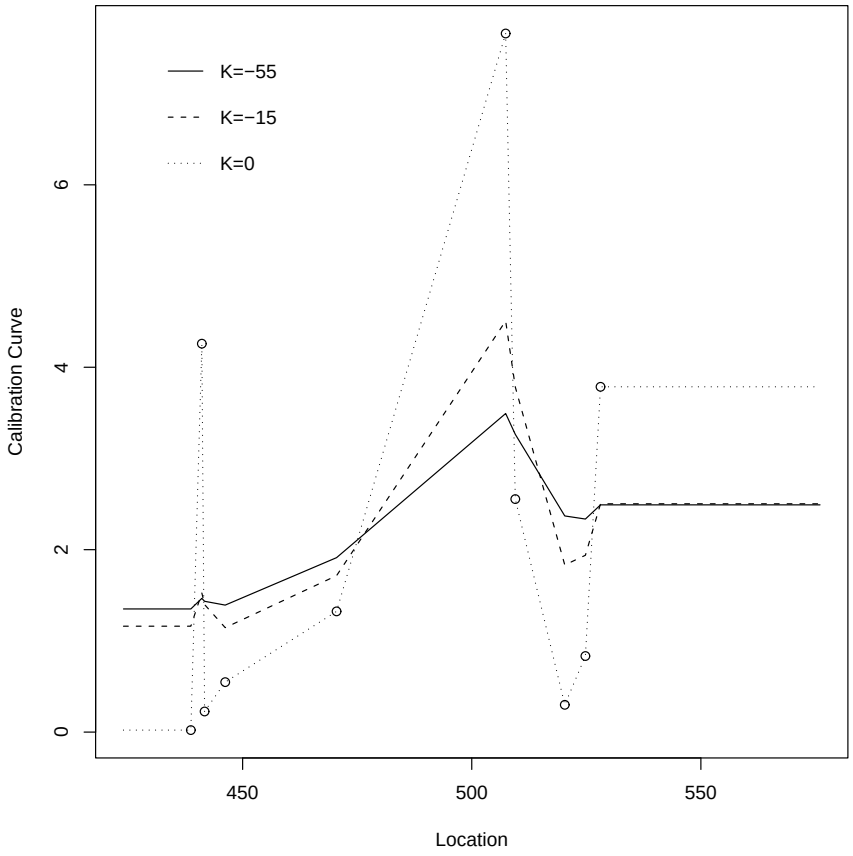

Figure F.38 Example calibration curves with different values of K

F.3.1 Recommended changes to the Parameter Values

The results of Section F.2.3 suggest that for application purposes the calibration performance needs to be improved, at least for the earlier part of the Chenies record. The values of ε_g, ε_r and K can be adjusted for this purpose, keeping in mind the effect of the different parameters on the calibration surface (Sections F.2.4 and F.3). Since the worst radar versus gauge comparisons were obtained in the period before 1994, the first modification of the procedure consisted in setting the corresponding increments to the same values as in the following years, i.e. both ε_g and ε_r were set to 1, giving significantly improved calibration results, as shown in Figures F.39 and F.310. However, the difference between the uncalibrated and calibrated radar data is not very large, at least for some of the gauges and for the spatial average.

The calibration results for several other sets of values for ε_g, ε_r and K were also investigated. The values of K considered for calibration were 0, -30 (the MMJB values for the period 1990-1994), and -55 (the MMJB value for the period following August 1994). Since K is substituted in G^* for the distance between a site and itself, $K = 0$ is a natural candidate. As mentioned in Section F.2.1, a non-zero value for K corresponds to a case when the correlation between the calibration factors does not tend to one as the distance goes to zero. It may be informative to fit a parametric curve to the correlation between calibration factors as a function of distance and extrapolate the value of K at which the correlation is equal to 1. Repeating this procedure for several time points may give some insight into the range of values for K to be considered in the calibration assessment. However, for the purpose of this exercise it is sufficient to focus on the three values mentioned above.

As for the increments, they have to be large enough not to result in very large values of the calibration

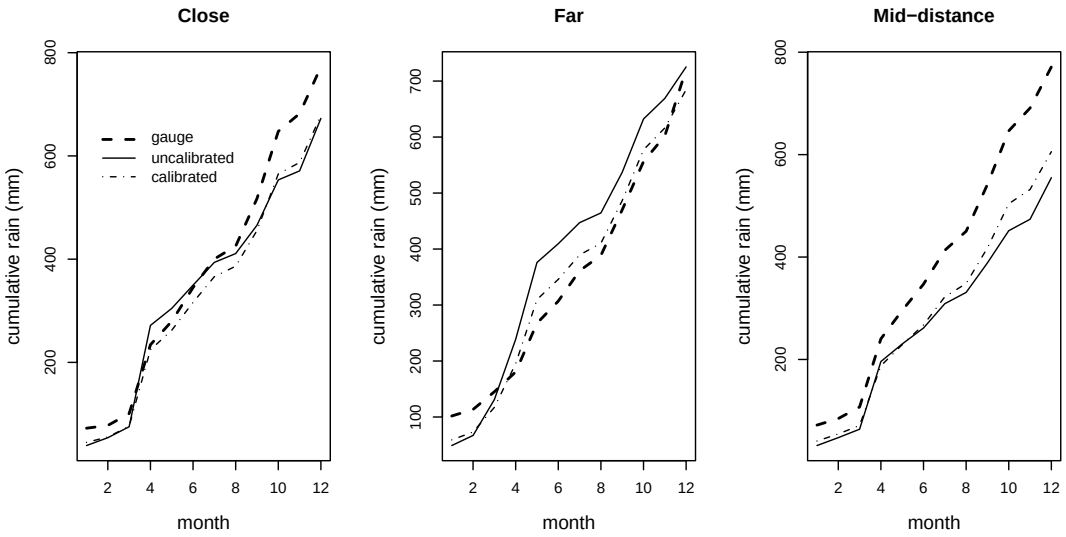

Figure F.39 1993: cumulative sums of gauge and radar monthly rainfall

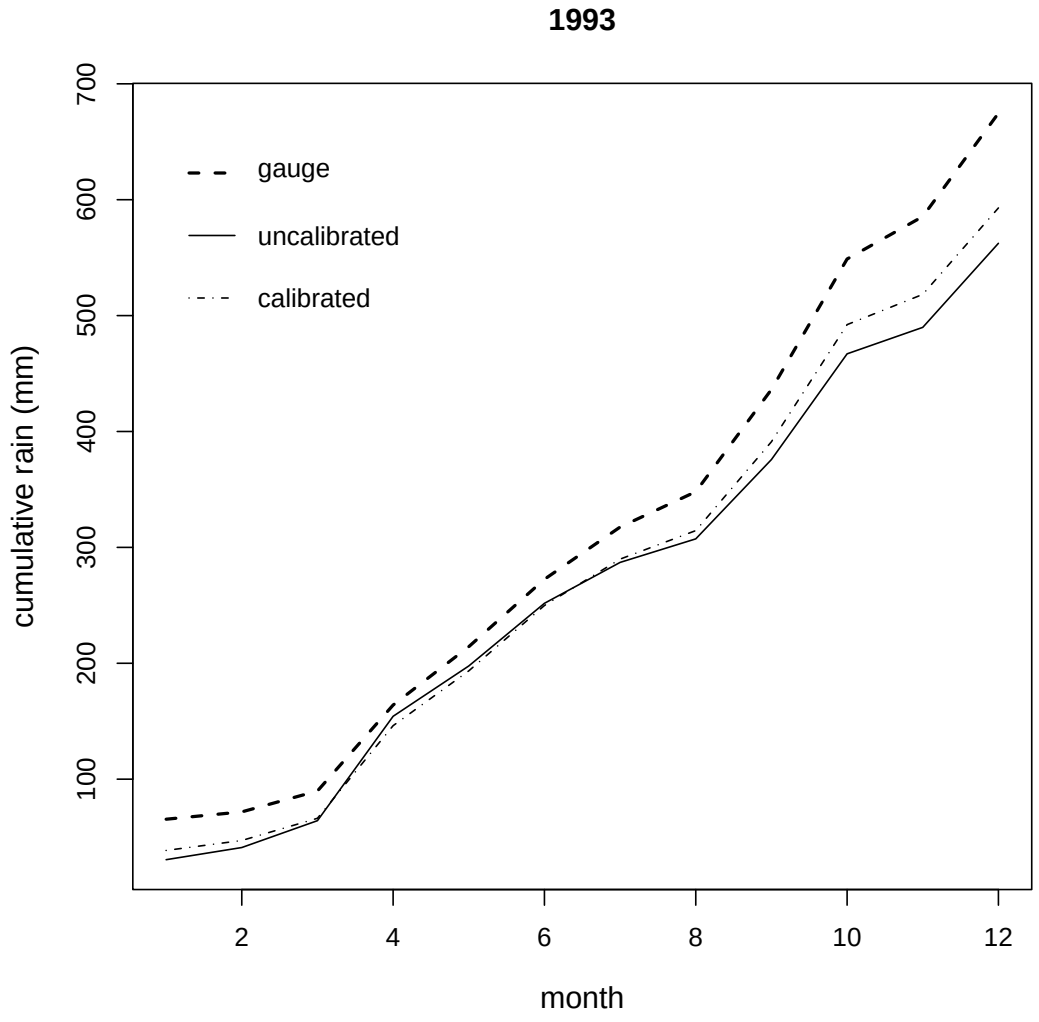

Figure F.310 Cumulative sum of gauge and radar spatially averaged rainfall

factors corresponding to zero or very small radar data. Also, the increments have to be small enough not to dominate most of the data, in which case the calibration factors would just be equal to the ratio of the increments. Several magnitudes of the increments were considered, from 0.0001 to 1. Also, two settings for the ratio $\varepsilon_g/\varepsilon_r$ were considered: 1, and $\overline{R}_g/\overline{R}_r$. Here $\overline{R}_g$ is taken to represent the average, over the whole network in space and over each year in time, of gauge rainfall. Similarly, $\overline{R}_r$ is the average of radar rainfall over all pixels that contain gauges and over each year.

Besides changing the values of K, ε_g and ε_r, another modification of the original procedure was investigated. Since histograms of the calibration factors at several time points showed a considerable skewness, it seemed reasonable to check whether the calibration results might be improved by smoothing the logarithm of the calibration factors, rather than the calibration factors themselves. This possibility was not exhaustively explored, in combination with different values of the calibration parameters, since the results were not very promising.

Not all the sets of values considered for K, ε_g and ε_r give reasonable results. For example, increments smaller than about 0.1 produce very poor agreement between calibrated radar and gauges due to extremely large calibration factors corresponding to zero radar values (see Section F.2.4). Also, the effect of increment size depends on the choice of K. Increments of the order of 0.1 result in poor performance when K is set to -30 or -55, while with $K = 0$ inadequate calibrated values result from increments around 1. Some calibration results from the better performing sets of values are included in Figure F.311. The plots show cumulative monthly totals of gauge and radar spatially averaged rainfall for the same years discussed in Section F.2.3.

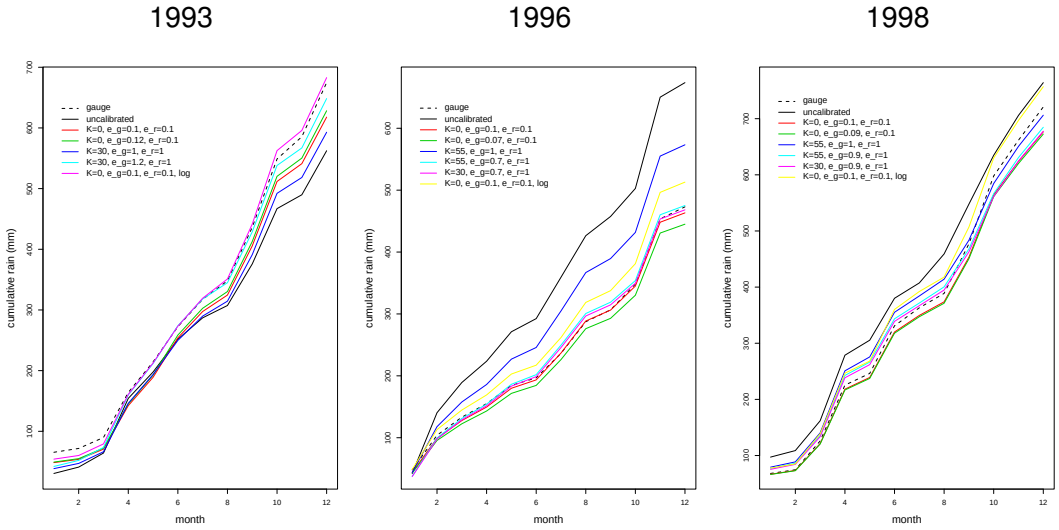

Figure F.311 Cumulative monthly rainfall for several sets of parameter values

The parameter set that provides the best calibration results, in terms of annual totals, varies with the years. For example, Figure F.311 suggests that in 1996 the radar records considerably more rainfall than the gauges, and the values $\{K = 0, \varepsilon_g = \varepsilon_r = 0.1\}$ provide better results than the original MMJB values. In 1998, when the uncalibrated radar data are relatively close to the gauge values, the original parameter set results in better agreement of annual totals between calibrated radar and gauges. A modified version of the MMJB values, with $\frac{\varepsilon_g}{\varepsilon_r} = \frac{\overline{R}_g}{\overline{R}_r}$, produces a very good agreement of annual totals in 1993 and 1996. In general, modifying the ratio of the increments to be equal to the ratio of average gauge to average radar, has a sizeable effect only when their order of magnitude is around 1 (and K is large in absolute value). When the order of magnitude of the increments is 0.1 (and $K = 0$), modifying the ratio does not change the curve of cumulative monthly totals by a noticeable amount. As for the effect of smoothing the logarithm of the calibration factors, the results vary depending on the year, but they seem to be reasonable in terms of yearly cumulative amounts. In summary, none of the parameter sets seems to be uniformly best over the whole period, but most produce at least acceptable annual totals.

While one would like the calibrated radar rainfall to be close to the gauge records in terms of monthly

and annual totals, it is even more important, for the purpose of this project, to ensure that the larger rainfall values are reproduced accurately. Quantile-quantile plots of hourly spatially averaged gauge amounts versus hourly spatially averaged radar amounts help to assess the agreement between the whole distributions of (spatially averaged) gauge and radar rainfall, including the extremes. The values of the calibration parameters used to obtain the Q-Q plots are the same as those of Figure F.311. The uncalibrated radar data always overestimate the larger hourly values, while the reproduction of the rest of the hourly rainfall distribution depends on the year. Although the performance of different sets of values for K, ε_r and ε_g also varies from year to year, setting K to zero and the increments to 0.1 seems consistently to provide the best agreement with the gauges. This is true also for years when other sets of parameter values result in closer annual totals with respect to the gauge average. In 1993, the modified version of the MMJB values (with $\frac{\varepsilon_g}{\varepsilon_r} = \frac{\overline{R}_g}{\overline{R}_r}$) provides a slightly better agreement with the gauges in terms of hourly maximum, but not in terms of the rest of the distribution. As for the ratio of the increments, the Q-Q plots corresponding to $K = 0$ and $\varepsilon_r = 0.1$, with either $\varepsilon_g/\varepsilon_r = 1$ or $\varepsilon_g/\varepsilon_r = \overline{R}_g/\overline{R}_r$, are practically indistinguishable. Another common feature is the poor performance of the calibration method that smoothes the logarithm of the z's, rather than the z's themselves. While this procedure gave reasonable results in terms of annual totals, the distribution of hourly calibrated radar values departs considerably from the gauge equivalent. Therefore, this option was not explored any further.

The results for 1993, 1996, and 1998, suggest that one may choose the set $\{K = 0, \varepsilon_g = 0.1, \varepsilon_r = 0.1\}$ and obtain a good agreement between the gauge and radar distributions of hourly values (averaged in space), while still maintaining a reasonable performance in terms of annual totals. Before adopting this as a general rule, it was necessary to check all available years using two sets of calibration values, i.e. $\{K = 0, \varepsilon_g = 0.1, \varepsilon_r = 0.1\}$ and $\{K = 0, \varepsilon_g = 0.1\overline{R}_g/\overline{R}_r, \varepsilon_r = 0.1\}$. The latter case was included in order to check whether, for years when the radar and gauge totals are substantially different, it would improve calibration results. In fact, this was not the case for any of the years considered. In general, the set $\{K = 0, \varepsilon_g = 0.1, \varepsilon_r = 0.1\}$ is reasonably good at reproducing the distribution of hourly totals observed at the gauges, with some variation in performance from year to year. For almost all years, the Q-Q plots involving calibrated values represent a substantial improvement over the uncalibrated case. Only in 1993 does the calibration result in a higher maximum hourly total with respect to both the gauges and the uncalibrated case. Even in this case, however, the comparison with the gauges in terms of the rest of the distribution improves with calibration. Figure F.312 shows two example plots, corresponding to 1991 and 1993. Both figures are obtained using radar data calibrated with the values $\{K = 0, \varepsilon_g = 0.1, \varepsilon_r = 0.1\}$ and show extreme examples of the calibration performance. In 1991 the uncalibrated radar data considerably overestimate the hourly totals observed at the gauges (corresponding Q-Q plot not shown here), while the agreement between the whole distributions of gauge and radar hourly totals is extremely good in the calibrated case. The 1993 plot represents the worst case of disagreement between calibrated radar and gauge distributions, at least in terms of the hourly maximum. The Q-Q plot obtained from uncalibrated radar data for the same year (not shown here), displays better agreement in terms of range, but not in terms of the shape of the distribution.

The performance of the chosen calibration procedure in terms of annual totals is shown for every year in Figure F.313, in order to check that the departure of the calibrated radar from the gauges remains within acceptable boundaries. Again, the effect of setting $\varepsilon_g/\varepsilon_r = \overline{R}_g/\overline{R}_r$ is very small. These findings, combined with the Q-Q plots, suggest that the further complication of updating the increment ratio every year is not worthwhile, and both ε_g and ε_r may be set to 0.1. Figure F.313 also shows that the calibrated radar always underestimates the gauge totals and, for some years, the uncalibrated totals are closer to the total gauge rainfall than the calibrated quantities. However, the relative difference in annual totals between calibrated radar and gauges, ranging from -2% (1996) to -12% (1991), seems acceptable for the purposes of this project.

F.3.2 The Effect of Calibration on Fitting Properties

The stochastic models for spatial-temporal rainfall are fitted to summary statistics calculated from the calibrated data from the Chenies radar. Simulated realisations from these models will then provide the artificial rainfall used as an input to the hydrological rainfall-runoff models. These properties of these simulations will reflect the characteristics of the data used for model fitting, and it is therefore important to check on the effects of the calibration procedures on the spatial and temporal properties of the radar data.

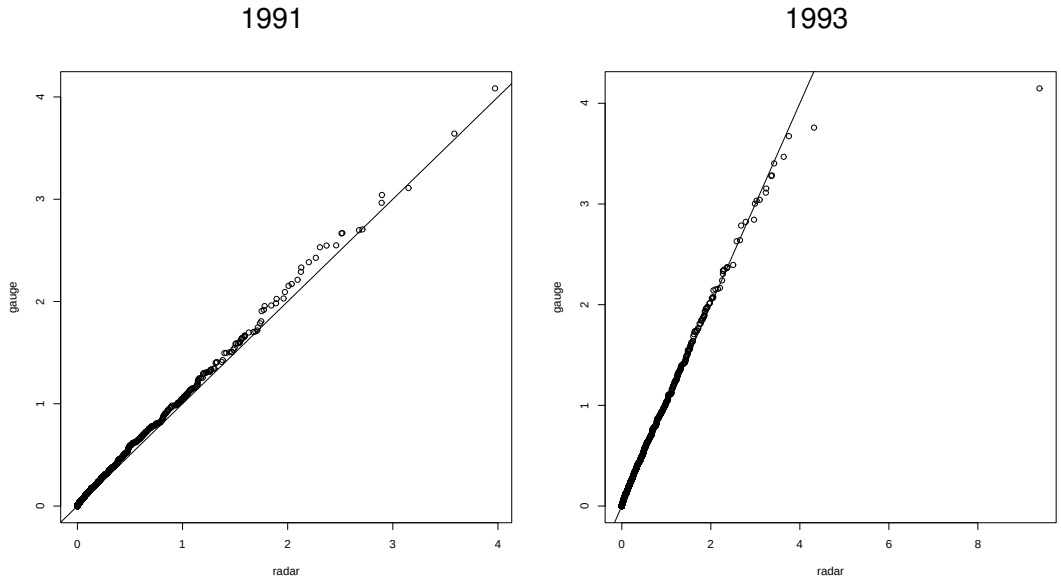

Figure F.312 Example Q-Q plots of gauge versus radar rainfall (spatially averaged hourly totals)

Visual inspection of sequences of calibrated radar images suggest the presence of temporal discontinuities that are not noticeable in the uncalibrated record. Comparison of figures F.314 and F.315 provides an example of this behaviour. Figure F.314 shows the uncalibrated radar sequence recorded between 15:15 and 15:40 on 26th February 1996, while Figure F.315 is the calibrated version at the same times. The images from 15:15 to 15:25 and from 15:30 to 15:40 are calibrated using the gauge values at two different time points (15:30 and 15:45 respectively). Correspondingly, the calibrated sequence shows the abrupt appearance of an area of increased intensity at 15:30, while the uncalibrated images change smoothly in time.

To address this issue, a modified calibration procedure was investigated, where R_g^j in (F.1) is the weighted average of the gauge j values at two time points (the 15-minute records immediately preceding and immediately following the time of the radar image). The results are not promising and are not reported here.

While investigating this problem, it became clear that calibration changes the temporal autocorrelation of the radar sequence, as well as other properties that are used for model fitting. As an example, Figures F.316 and F.317 show the temporal autocorrelation and two of its components for the rain events of 1st January 1998, 03:55–05:25, and 3rd January 1998, 16:45–19:40. The gauge-based values are represented by bold dashed lines that link the three correlations at 0, 15 and 30 minute lags. Since the temporal resolution of the radar is finer (5 minutes), seven correlations are computed at lags up to 30 minutes for the uncalibrated (solid lines) and calibrated (third line type) radar data. The autocorrelations are always higher for the gauges than the radar, while the autocorrelations based on calibrated images can be on either side of the uncalibrated equivalents, depending on the event. Figures F.316a and F.317a illustrate the two cases. On 1st January, calibration brings the temporal autocorrelation of the radar closer to the gauge-based quantities, while on 3rd January the opposite happens. Figures F.316b and F.317b show plots of $E(Y_t Y_{t-k} | Y_t > 0, Y_{t-k} > 0)$, for lags k of up to 30 minutes, where Y_t is the gauge (bold dashed lines), uncalibrated radar (solid lines), or calibrated radar (third line type) rainfall at time t. This component of the temporal autocorrelation behaves differently between the two January events. On 1st January, calibration lowers the radar-based line to match fairly closely the gauge-based one, which is much below the uncalibrated radar. On 3rd January, instead, calibration does not affect this quantity noticeably and both lines corresponding to uncalibrated and calibrated radar are below the gauge-based line. Another contributing factor to the temporal autocorrelation is $P(Y_t > 0, Y_{t-k} > 0)$ which is shown in Figures F.316c and F.317c. Since calibration does not affect the wet-dry properties of the radar images, there is only one line for the radar. In both January events, the radar-based line is above the corresponding gauge line. This may be due to the different resolution of the two measurements. .

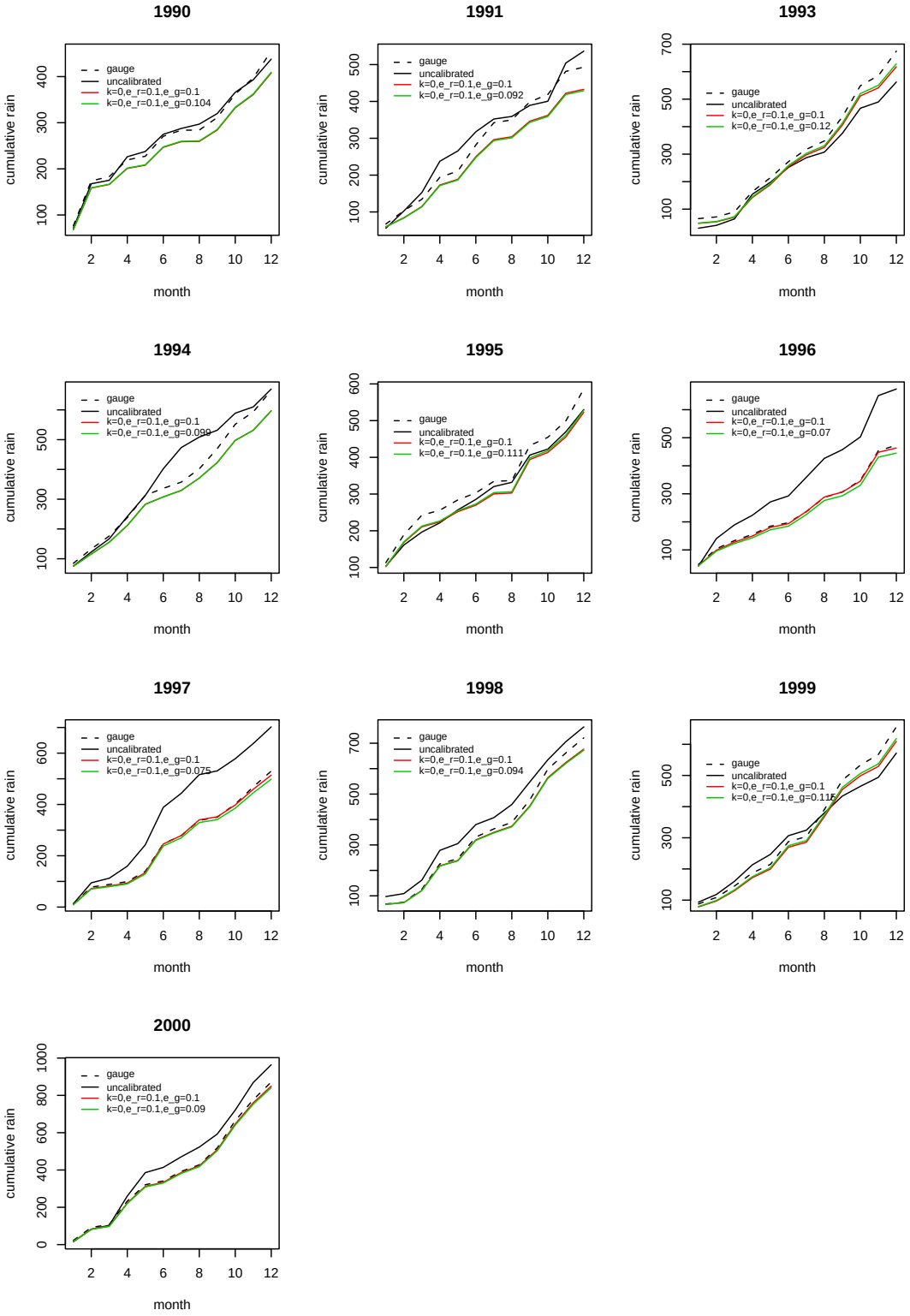

Figure F.313 Cumulative monthly rainfall for spatially averaged gauge and radar (all years)

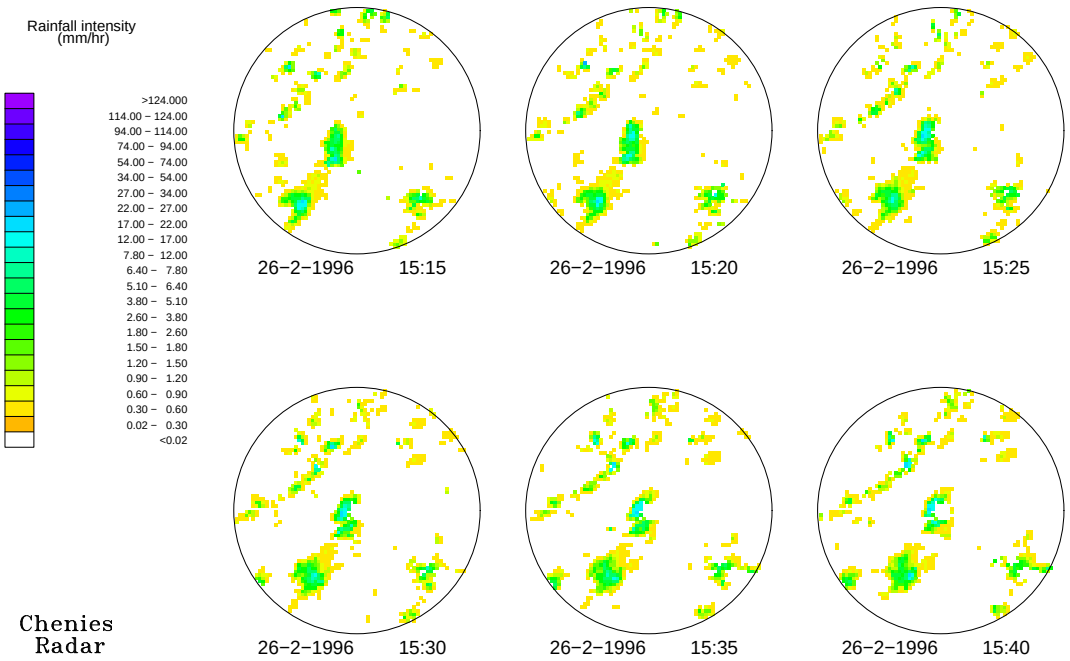

Figure F.314 Uncalibrated radar sequence for 26th February 1996, 15:15–15:40

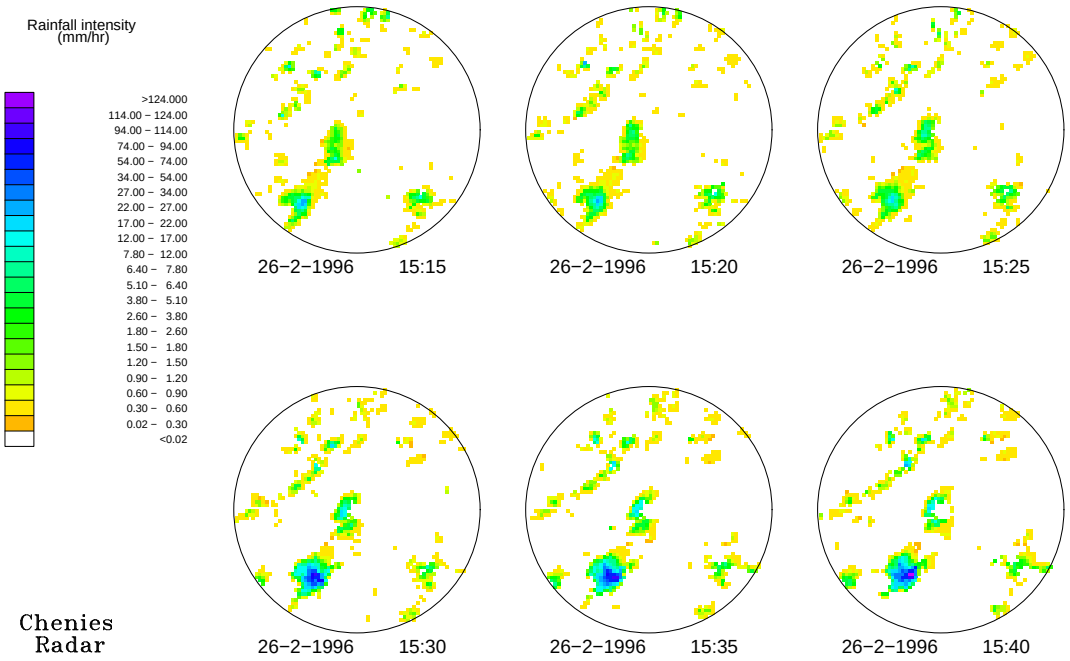

Figure F.315 Calibrated radar sequence for 26th February 1996, 15:15–15:40

Other properties of radar rainfall change with calibration. Several spatial correlations and variances at different levels of spatial aggregation are among the properties used to fit spatial-temporal models to radar data. Figure F.318 gives an example of how calibration affects these quantities. There is a considerable difference between the variance of calibrated and uncalibrated radar, at all levels of spatial aggregation. As for the spatial autocorrelation, the shape of the contours in Figure F.318b changes with calibration. However, plots of the spatial autocorrelation as a function of distance (not shown here) show

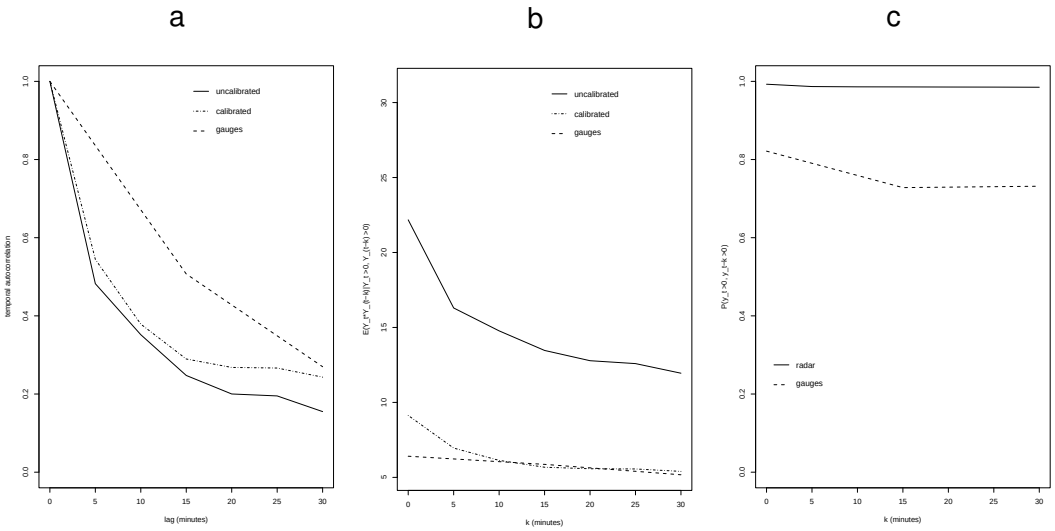

Figure F.316 Temporal properties for the event of 1st January 1998, 16:45–19:40

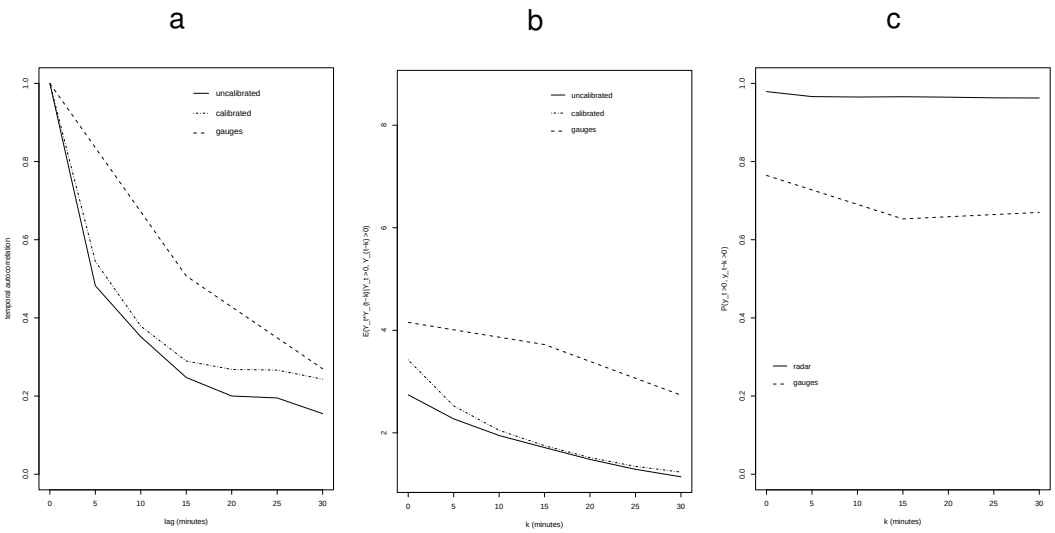

Figure F.317 Temporal properties for the event of 3rd January 1998, 03:55–05:25

close similarity in the behaviour of calibrated and uncalibrated data, while the gauge-based correlations are quite different from both the radar-based ones (note that for the gauge network the correlations are based on a very small number of data points).

It is not clear whether calibration should aim at reproducing all the characteristics of rainfall at the gauge network, since each gauge measures precipitation at a point, while a pixel in the radar image represents the average rainfall over an area of 4km^2. But it is important to notice that calibration affects several of the data properties that are used for model fitting, and that properties of the uncalibrated radar may be a closer match to their gauge equivalents. Changes in these properties will be reflected in the simulations from the fitted spatial-temporal models. The results of the previous sections suggest that the simulations will provide adequate monthly and yearly totals, as well as a satisfactory upper tail behaviour, and thus the current calibration scheme should be adequate for the purposes of this project. However, in the future it would be interesting to study the effects of calibration more extensively and possibly to implement different methods.

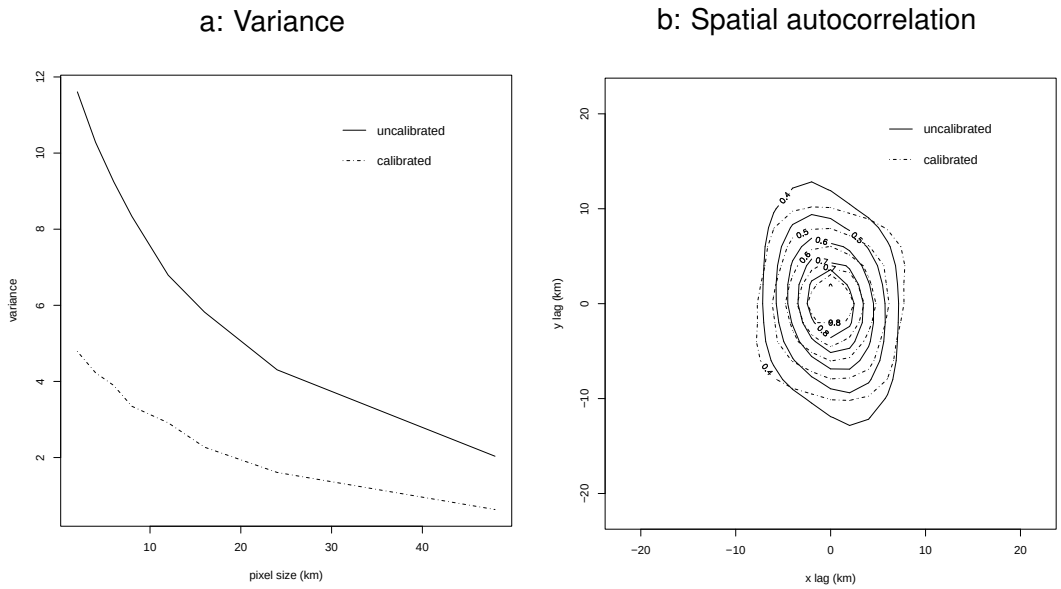

Figure F.318 Variances and spatial autocorrelations for the event of 1st January 1998, 16:45–19:40

F.3.3 Conclusions on calibration

The calibration technique recommended by Moore et al. (1994) was applied to the Chenies radar data. The results presented in the previous sections suggest that the original parameter values should be changed to obtain better agreement between the calibrated radar and the supporting gauge network. The best parameter set, among those considered in this note, is $\{K = 0, \varepsilon_g = 0.1, \varepsilon_r = 0.1\}$. In particular, over all available years, this parameter set best reproduces the larger hourly rainfall values (and the whole shape of the distribution) observed at the gauges. Since the larger rainfall amounts are particularly important for the purposes of this project, the same parameter set will be applied to all years, although for some periods, other values of the calibration parameters provide better agreement in annual totals with respect to the gauges.

It has not been within the scope of this project to develop the best possible calibration method for radar data. Rather, the goal has been to find a simple rule that can be readily applied to any radar image in the Chenies archive, and still provides good agreement between radar and gauges, especially in terms of the upper tail of the amount distribution. The overall procedure recommended in Moore et al. (1994), with the parameter values modified as suggested above, satisfies these requirements and has therefore been adopted as the current calibration rule for the Chenies radar. However, it is important to stress that calibration affects several of the properties used to fit spatial-temporal models to radar data, and Section F.3.2 includes a first illustration of this issue. The current calibration method appears suitable for the purposes of this project but further study is needed of the effects of the calibration of the radar data on the rainfall simulated from the stochastic model before a definitive conclusion on an optimum calibration scheme can be reached.

F.4 Fitting events in the Chenies radar data

F.4.1 Introduction

In this appendix, we have discussed the calibration of the Chenies radar data, so that the calibrated data are now available for model fitting. The next step is to select suitable sections of the radar record to which the stochastic model for event interiors can be fitted. The ‘event’ is the largest structure in the Poisson cluster spatial-temporal model for rainfall, and its arrival and departure from the fitting window

are marked by the upcrossing and downcrossing of a prespecified coverage threshold. This definition identifies a certain number of events in the records from the Chenies radar for the period 1990-2002. The number of such events by year is characterized by some unexpected features that may be 'real' or artefacts due to changes in the radar recording procedures. In particular, there appears to be a jump in the mid-nineties, as well as an overall increasing trend, in the number of events occurring. Section F.4.2 describes this behaviour in more detail.

A list of known changes in radar recording practices is reported in Section F.4.3. The characteristics of radar data that are affected by these changes are analysed to determine whether they explain all or part of the observed patterns in the number of events. The ability to detect light rain and the level of radar clutter present in the images are among the main factors that changed in the course of the 1990s. The declining number of missing images is another important improvement that at least partly accounts for the increase in the number of events.

Whether or not the trends in the observed radar data are genuine, it is relevant also to examine the records also in terms of event durations. The issue is to determine whether the radar identifies a small number of longer events at the beginning of the period, and a larger number of shorter events towards the end.

Finally, it should be noted that the various aspects of radar data that are analysed in the following sections are not completely unrelated to one another. Their effects on the number of events are therefore not easily distinguishable.

F.4.2 Identification of Events in Chenies Radar Data

The method of identifying events follows that used in the Wheater et al. (2000b), where discussion of the approach can be found. As described in Chapter 9.3, the duration of an event is defined as the time during which the coverage, C say, of the fitting window (that is, the proportion of the fitting window that is wet) is above a certain threshold, which we denote here by θ. The fitting window is a square region of 52×52 pixels (or 104×104 km) inside the Chenies 2km radar circle. The choice of θ is to an extent subjective, requiring a balance between the non-detection of events if it is too high, and mis-interpretation of radar noise as precipitation if it is too low. In addition, too low a threshold also allows the possibility that a sequence of events following each other rapidly across the window may be mis-identified as a single event, for which the stationarity assumption of the model for event interiors is likely to be questionable. Thus θ needs to be:

1. Low enough for the spells with $C \leq \theta$ to be reasonably labelled as *dry*.

2. High enough to separate radar clutter (i.e. background noise) from genuine rain.

3. High enough to break long events consisting of a sequence of rain "bands" into shorter ones that have homogeneous characteristics throughout their lives.

For Chenies radar, a threshold $\theta = 0.15$ was found a reasonable compromise with respect to these criteria.

Figure F.419 shows the number of events by year resulting from the use of a threshold level 15% coverage applied to the rainfall records from Chenies radar. A jump occurs around 1994 and an overall increasing trend characterizes the study period. These features may in part be explained by the much larger number of missing images in the earlier years (also shown in Figure F.419). In fact, in 1994 the number of missing images jumps down by about 10,000. However, there is no downward trend in the following years that would mirror the upward trend in the number of events. Note that the recorded number of events is a lower bound on the actual number, given that there are breaks in the series of radar images. In particular, if at least one of the coverages at the two ends of a series of missing images is above the threshold θ, one event is counted. Similarly, if both coverages are below θ, no event is counted, although it is unknown whether and how many times the threshold was crossed in the missing period.

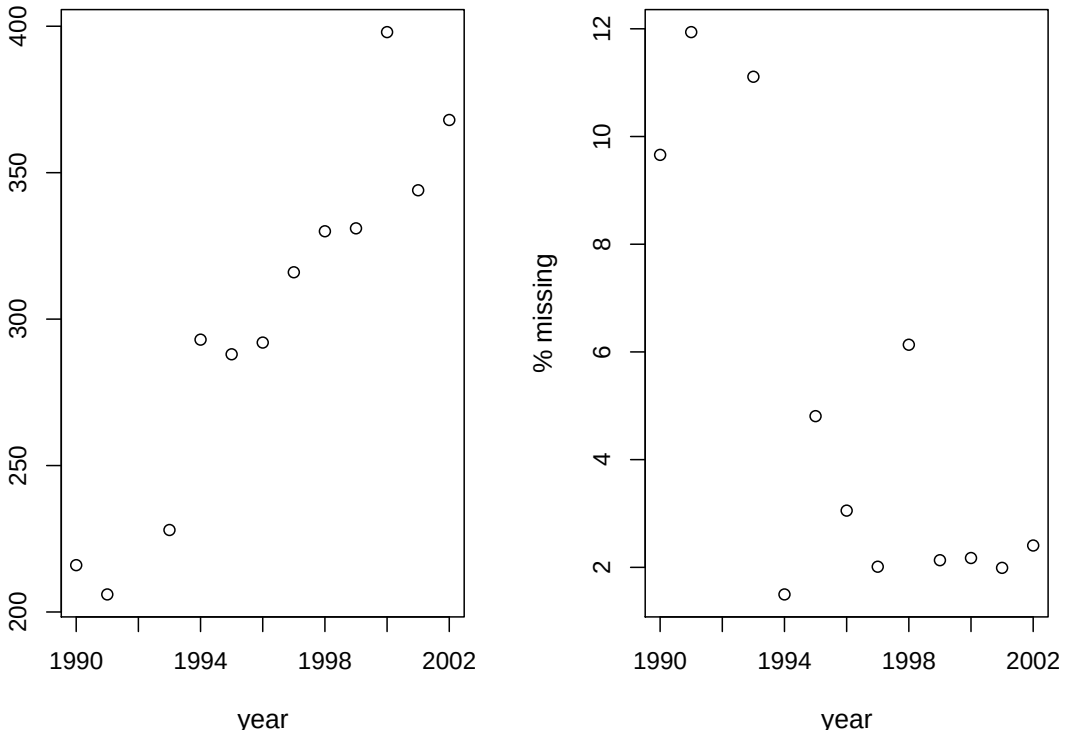

Figure F.419 Annual numbers of events and percentages of missing images

F.4.3 Changes in the Chenies Radar

The improvements in radar technology that led to a decrease in the number of missing radar images over the 1990s are not sufficient to explain all the patterns in the numbers of events noted in the previous Section. Further investigation highlighted some changes in the radar that may help explain the observed trends in the numbers of events in the 1990–2002 period (pers. comm. Malcolm Kitchen, Martin Crees)

1. In the mid 1990's the radar transmitters were modified to correct a design fault that limited the ability of the radar to detect light rain (installation of a modified transmitter in about 1996.)
2. As a consequence of the work in 1. above, the level of residual ground clutter may also have increased slightly.
3. In the mid to late 1990's the effective range of the radar was increased.
4. Again in 1999 and certainly before 2000 the radar was upgraded to see significantly further and improve quality.
5. In 2000, the signal processor at Chenies was upgraded, with associated changes in sensitivity and ground clutter frequency.
6. In 2002, changes were made in the method of filtering ground clutter.

Several features of the radar data need to be investigated to understand whether the points above explain the behaviour of the number of events, and whether there is a residual 'real' increase over the period. Note that points 3. and 4. above should not be relevant to the results shown in Figure F.419, since all the calculations on the number of events have been carried out on a square inscribed in the radar circle, i.e. a region which should be inside the effective range of the radar even in the first years of the record.

F.4.4 Radar Clutter

Points 2, 5 and 6 in Section F.4.3 suggest that some of the increase in the number of events over the 1990–2002 period may be due to an increase in surface clutter. Some of the clutter is removed by our practice of setting isolated wet pixels to zero. Also, the events of Figure F.419 are identified using a coverage threshold of 15%, so that images with a very low proportion of wet pixels are considered dry.

Visual inspection of some of the events in the Chenies radar record had suggested that 15% is high enough to avoid interpreting radar clutter as rain. However, it is worth investigating the issue further by computing, for each year, the number of events with maximum coverage above and below several thresholds. The event definition is kept unchanged, so that the threshold $\theta = 15\%$ separates wet and dry images, and a few values of a higher threshold θ' are used to determine which of the wet periods achieve a "high" coverage.

Definition An event starting at time t_s and ending at time t_e, achieves *high coverage* if

$$\max_{t \in [ts,te]}(C_t) \geq \theta'.$$

Figure F.420 shows the number of high and low coverage events by year, for four values of the threshold θ' that discriminates between high and low coverage. The highest threshold shown is $\theta' = 30\%$, since it does not seem realistic that ground clutter would cause more than 30% of the pixels to appear wet.

Since the number of high coverage events shows an upward trend (right-hand plots in Figure F.420) even with $\theta' = 30\%$, it is unlikely that an increase in residual clutter accounts for all the increase in the numbers of events between 1990 and 2002. However, for low thresholds like $\theta' = 18\%$, the left hand plots show a jump in the number of low coverage events around the mid-1990's. This feature may signify that an increase in clutter is partly responsible for the jump in Figure F.419.

To verify if the increase in low coverage events is due to radar clutter, it is necessary visually to inspect radar images with a proportion of wet pixels lower than, say, 18% in the later part of the 90's. A first check did not suggest that such events correspond to clutter rather than rainfall, although this issue may need further investigation. Given these preliminary results, the wet/dry threshold θ was not raised for the later years, that is, the event definition was left unchanged for the whole study period.

F.4.5 Light Rain

Problems with the detection of light rain led to changes in the radar transmitters in the mid 1990's. It is then reasonable to expect that at least part of the increase in the number of events in the 1990–2002 period is due to a larger number of light rain events in the later years. Here, light rain is defined in terms of the mean intensity of the event, although a check on the intensities at each single pixel is included later in this section.

Definition A wet spell is a light rain event if its mean intensity is smaller than a prespecified threshold μ, i.e.:

$$\frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it} < \mu,$$

where N is the number of pixels in the radar window, T is the number of time points spanned by the event, and Y_{it} is the rainfall intensity at pixel i and time t.

Figure F.421 shows the number of events by year that have mean intensity below or above several thresholds. The plots on the left confirm that a smaller number of light rain events were detected in the earlier part of the record, with no events corresponding to a mean intensity smaller than $\mu = 0.2$ mm/hr up to 1994. However, the number of events with larger mean intensities is also increasing over the period, at least for low thresholds (e.g. $\mu = 0.2$ and $\mu = 0.3$ mm/hr). When the threshold for the mean intensity is set to 0.5 mm/hr, the trend in the higher intensity events is less clear, and at $\mu = 1$ mm/hr the trend disappears.

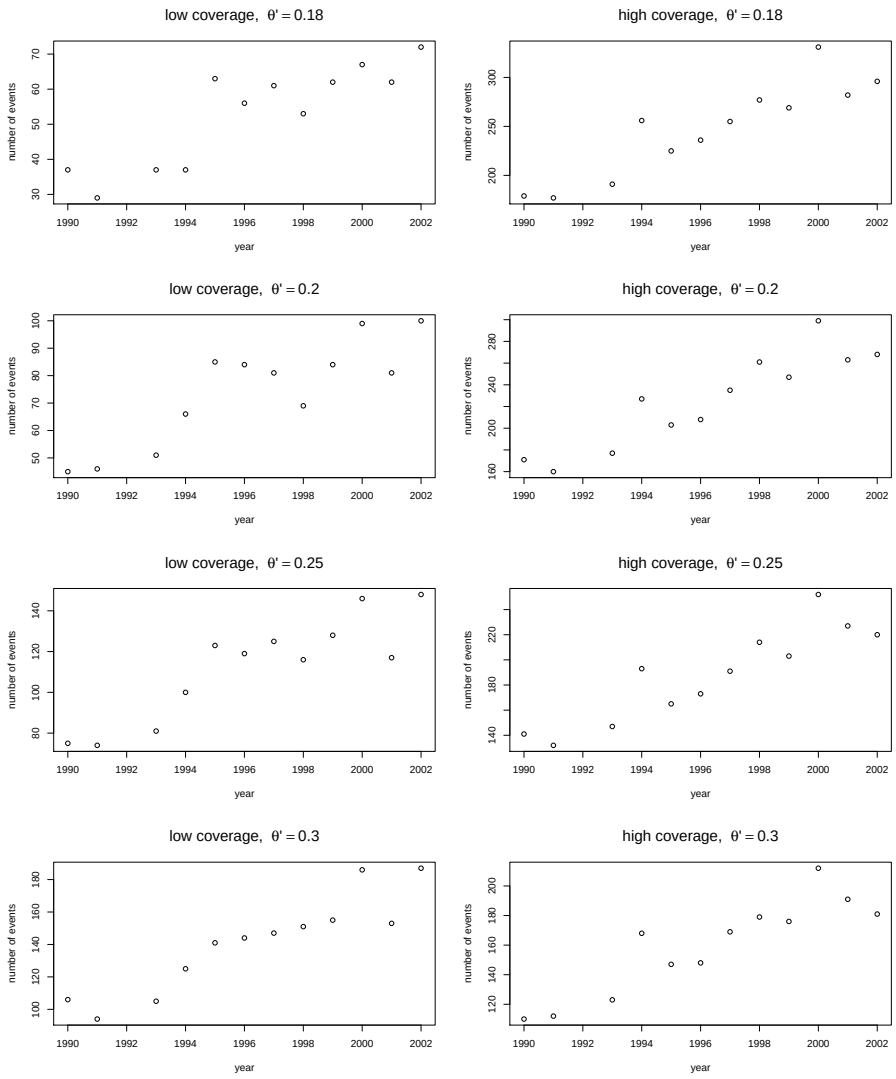

Figure F.420 Annual numbers of events with maximum coverage below or above several thresholds

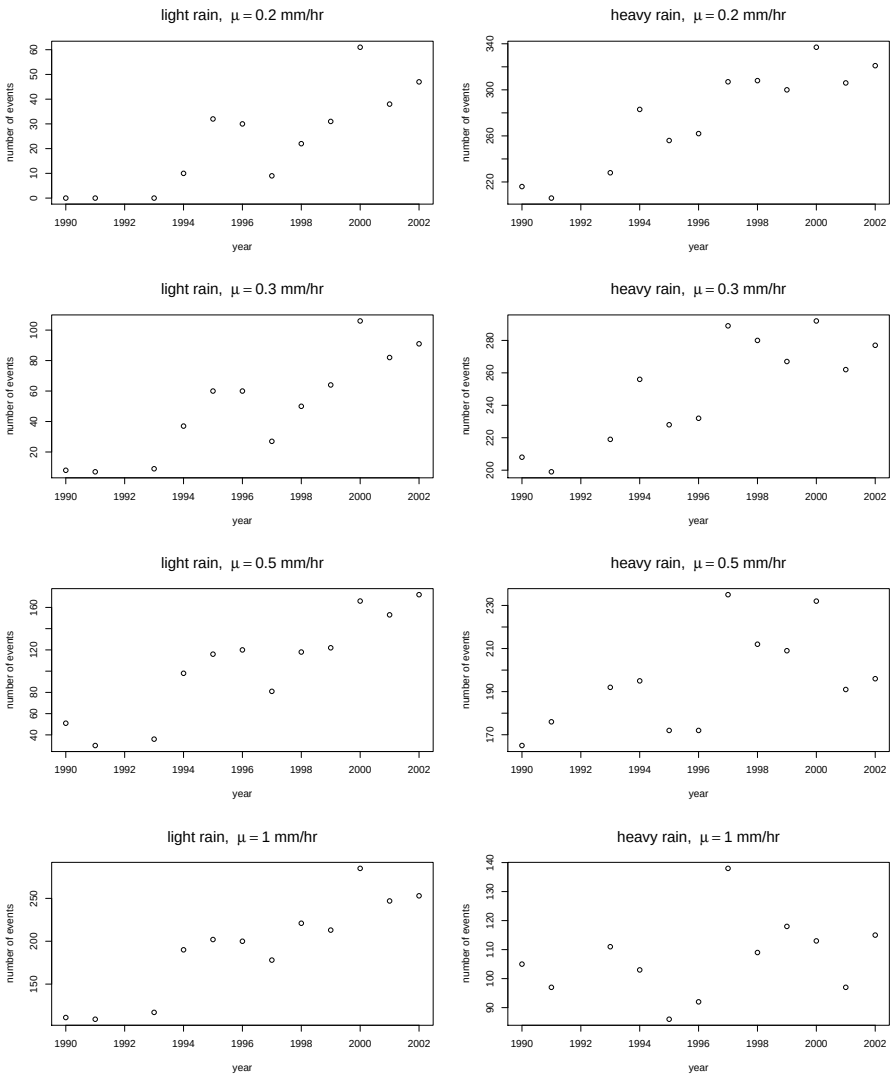

Figure F.421 Annual numbers of events with mean intensity below or above several thresholds

It should be noted that there is considerable overlap between light rain and low coverage events. Figure F.422 shows the maximum coverage corresponding to each event in the dataset, split according to the associated mean intensity. The lighter rain events, i.e. with mean intensity lower than 0.5 mm/hr (right hand boxplot), achieve lower coverages, usually below 50%, with the median at about 20%.

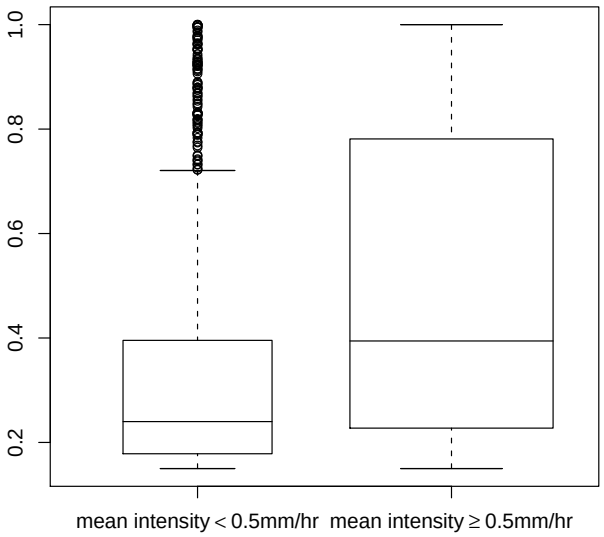

Figure F.422 Event coverages split by mean intensity

Further checks included setting radar pixels below several thresholds to zero before computing the rainfall coverage and identifying events. In this case, the light rain threshold affects the single pixels rather than the mean intensity of the event. Figure F.423 shows the number of events by year when the wet/dry threshold for each pixel is set to 0.5 mm/hr. The plot suggests a distinction between two groups of years. The number of events is lower before 1996 and higher since then. It is not clear, however, whether there is a trend in the latter part of the record. In general, although the increasing behaviour is not as marked as in Figure F.419, it is still present even after accounting for better detection of light rain. Higher thresholds of up to 1 mm/hr were also used. The corresponding plots of number of events by year (not shown here) are slightly different with respect to Figure F.423, as they display a fairly linear trend without a clear distinction between two groups of years. In general, even the highest threshold of 1mm/hr does not eliminate the increasing trend completely.

F.4.6 Event Duration

Since the number of events increases over the 1990-2002 period, it is important to verify whether the early part of the record is characterized by fewer, longer events, while the later part contains an increasing number of shorter events. The idea is that in the later part of the 1990s, the radar may be breaking long events into several shorter ones, rather than recording more rain.

As for the low/high coverage and light/heavy rain events, the definition of ‘short’ and ‘long’ events is based on a prespecified threshold τ.

Definition An event is labeled as ‘short’ if $D < \tau$, where D is the event duration, and τ is a threshold in minutes.

Figure F.424 shows plots of the number of events below (left hand) or above (right hand) four duration thresholds. Both types of events increase in number between 1990 and 2002, but following different patterns. The number of short events, for low values of τ (such as 30 minutes), jumps around 1995 and

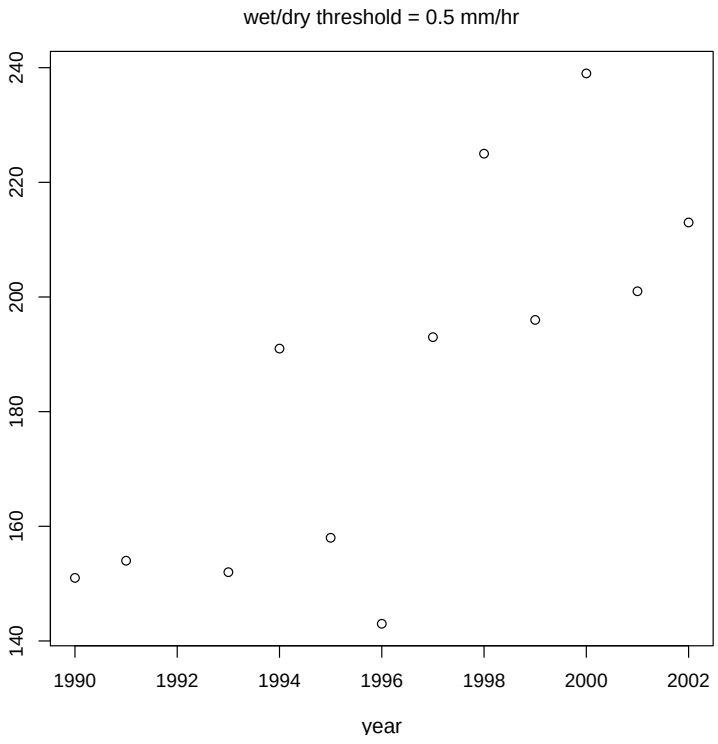

Figure F.423 Annual numbers of events when pixels under 0.5 mm/hr are set to 0

then seems to oscillate around a fairly constant level. The number of long events, instead, shows an upward trend, with no clear jumps.

The similarity between Figures F.420 and F.424 is due to the considerable overlap between short events and low coverage events. The boxplots in Figure F.425 show the coverages of all events in the Chenies radar dataset, split according to whether or not their duration is longer than an hour. The short events achieve fairly low coverages (mostly below 20%) while the longer events correspond to maximum coverages that span the whole (0.15, 1) interval.

F.4.7 Comparison of radar and rain gauge events

If the increases in the numbers of rain events recorded at the Chenies radar station are a 'real' phenomenon rather than an an artefact of changing recording practices, there should be a similar pattern in the rainfall measured by gauges in the same area. The comparison is not straightforward, however, since gauges measure rainfall at a point location rather than an areal average, and an accumulation over a period rather than an instantaneous intensity.

Data from a network of about 120 gauges, at a temporal resolution of 15 minutes, are available for the region covered by Chenies radar. Given the differences between the two types of measurements, the concept of 'event' needs to be redefined in the context of the gauge network. In principle, it is possible to standardise the two types of records by changing the event definition that has been used for the radar. For example, one may aggregate the images to the temporal scale of the gauges, say, change the spatial resolution to larger pixels that contain several gauges, and average the gauge data within each larger pixel. For the purposes of this study, however, a simpler approach was followed, whereby the definition of radar events is unchanged, and the definition of an event at the gauge network is similarly based on a coverage threshold.

Definition A rain event arrives and leaves a gauge network when the rainfall coverage C_g (the proportion of gauges with positive rainfall) rises above and falls below a fixed threshold θ_g.

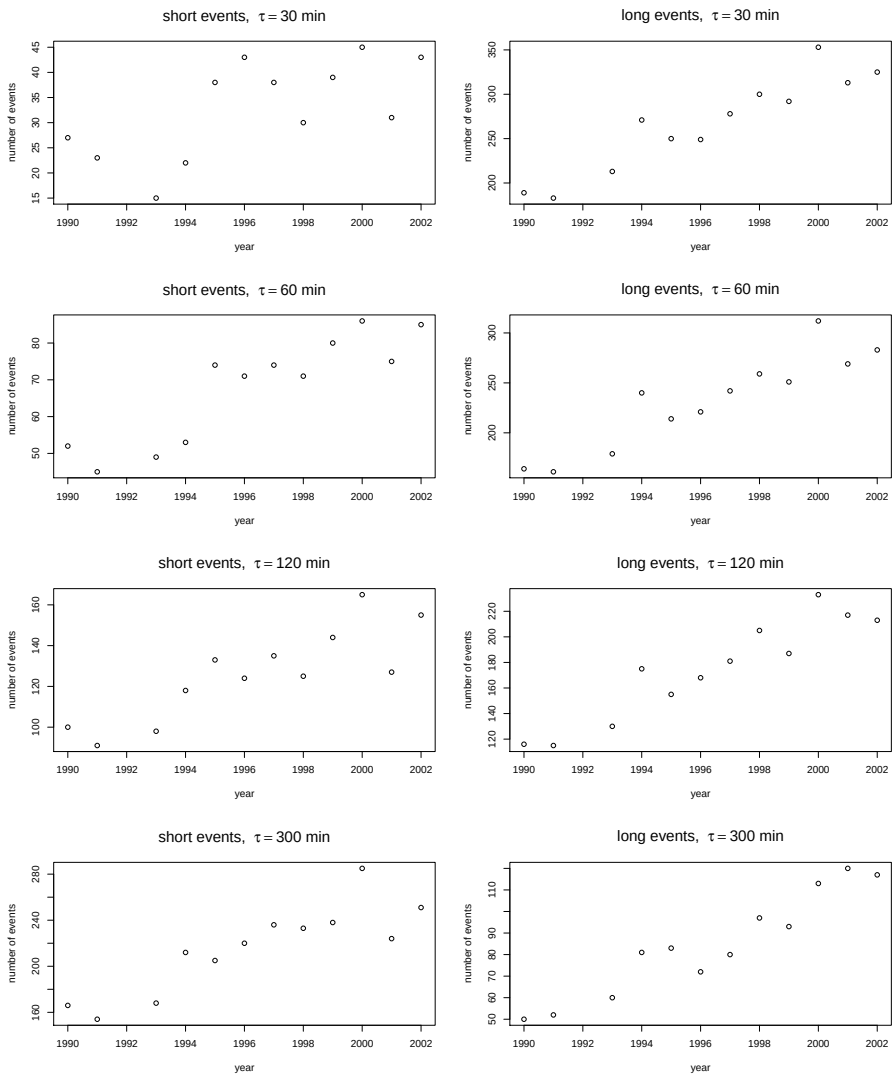

Figure F.424 Annual numbers of events with duration below or above several thresholds

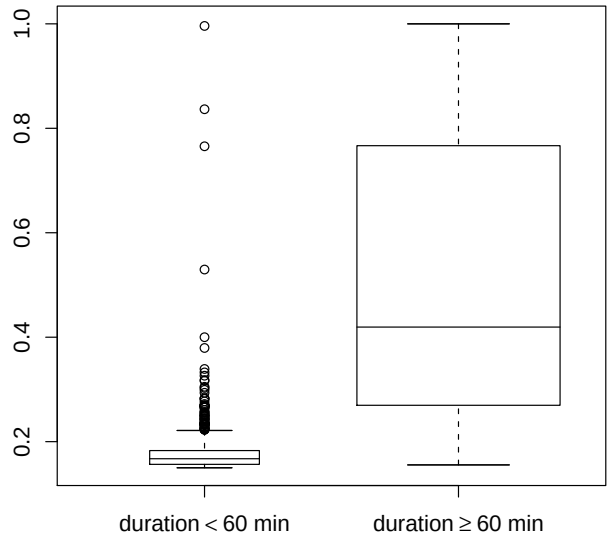

Figure F.425 Coverage split according to duration

In order to mirror the radar event definition as closely as possible, the threshold was set at $\theta_g = \theta = 0.15$.

Figure F.426 shows the corresponding number of events recorded by the gauges for each year between 1990 and 2002. An upward trend is apparent, although there is no indication of a jump after the first few years. The presence of a trend in the gauge data confirms that at least part of the increase in the number of events is genuine.

F.4.8 Conclusions on the number of fitting events

The rainfall data from Chenies radar station for the period 1990–2002 suggest a jump upwards in the number of wet events around 1994, and an increasing trend afterwards. Several changes in the radar recording practice have been found to account only partially for the patterns in the number of events.

The improvements in radar technology and consequent reduction in number of missing images, are responsible for some of the observed increase in the number of events. The improved detection of light rain following changes recording practice in the mid-90's accounts for some of the remainder, especially the difference between the first and second half of the decade. This consideration raised the issue of whether a threshold on the mean intensity should be imposed in the definition of an event, thereby excluding the light rain events that are recorded only in the later period. While a sufficiently high threshold would help homogenize the two parts of the record, for the purposes of this project light rain is also important. Therefore, the event definition was left unchanged.

A related problem is the possible increase in background noise in the second half of the 1990s. A symptom of this is that the number of events with low coverage (say smaller than 18%) increases sharply around 1995. However, a first visual inspection of the radar records does not suggest an increase in clutter, at least after applying the original 15% coverage threshold. Therefore, it does not seem necessary to raise the threshold in the later part of the record to exclude low coverage events that may correspond to clutter. However, this issue may need to be revisited after a more careful study of the low coverage events in the late 1990's.

The comparison with rainfall recorded at the gauges suggests that at least part of the increasing trend in number of wet events throughout the 1990s is genuine. This observation, and the fact that the changes in radar technology resulted in improvements in rainfall measurements that should be retained

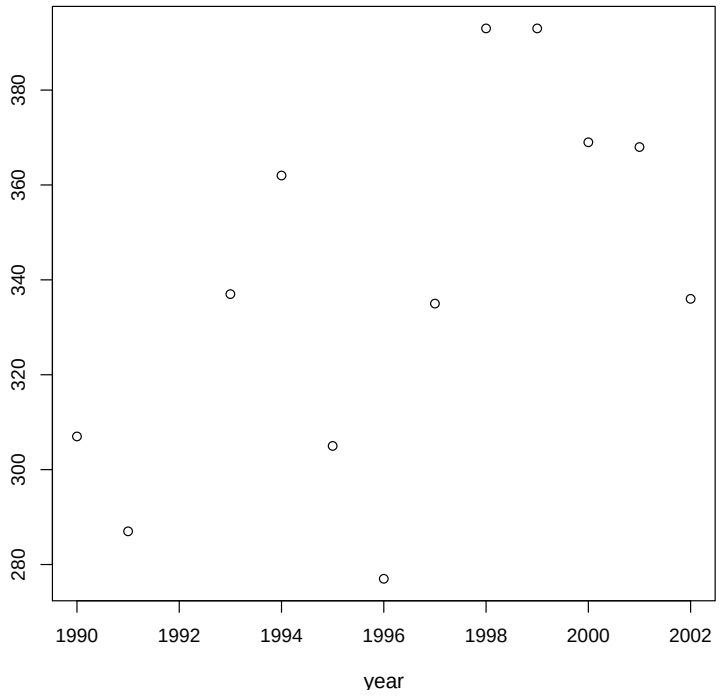

Figure F.426 Number of events at the gauge network

in the records used for modelling, led to the decision not to alter the adopted event definition.

When analysing the continuous simulations from the spatial-temporal model, it is important to take into account the findings about the trend in event numbers and the changing characteristics of the wet spells, such as coverage and mean intensity. The simulations should reproduce the overall characteristics of the radar data over the 12 years of records, exhibiting, for example, the correct proportion of short and long events, but trends will not be present in the simulated images. Keeping the threshold θ unchanged over the study period and simulating from the resulting Weibull parameters and Poisson model parameter library, has different implications in terms of event numbers and event type mix:

- If part of the trend were due to an increase in ground clutter, too many events would be simulated, with too high a proportion of low coverage and possibly low intensity events.
- If part of the increase in event numbers is due to better detection of light rain, the simulations from the 12 year parameter library would not include enough events, especially not enough low intensity ones.
- If there is a genuine trend, the effect on the simulations depends on its long term characteristics, which cannot be determined with only 12 years of data.

All these issues should be revisited in the future, especially once more data become available, when the possibility of discarding the part of the record that corresponds to worse recording practices could be considered. However, it is likely that the radar technology will keep changing, making even the latter part of the record 'suboptimal'. Several more years of radar data will also be helpful in establishing the characteristics of the genuine trend in event numbers.

F.5 Conclusions on data issues

In this appendix, we have discussed two important issues concerning the data from the Chenies radar: the calibration of the data to reflect ground truth, and the selection of events to which the stochastic

spatial-temporal model for the event interiors can be fitted.

With regard to calibration, we have explored the application of the UK standard procedure (Moore et al., 1994) and found that, for the Chenies data, it is possible to improve the agreement between the calibrated radar values and data from rain gauges under the radar region. As well as improving the agreement between the distributions as a whole, emphasis has been put on reproducing the larger hourly values which are particularly important for rainfall-runoff applications.

An important by-product of the calibration study was the discovery that different calibration schemes, and the use of different parameter sets within a particular method, can have a significant effect on the properties of the calibrated data. Since summary statistics from the calibrated radar data are used to fit the stochastic models described in Chapter 9, changes in the calibration scheme may substantially affect the properties of data simulated from the fitted model. In particular, each image is calibrated individually, with no temporal smoothing in the computation of the calibration factors. As a result, we have found that calibration often reduces the temporal autocorrelation in a sequence of radar images (temporal correlations play an important role in fitting model parameters). This sensitivity of statistical properties to the calibration procedure is a source of some concern and, at present, is a real impediment to the use of models based on radar data in flood risk assessment. A substantial further study is needed to gain a thorough understanding of the issues involved. Nonetheless, the scheme used in this paper is based on existing best practice in the UK, with some relatively minor modifications for our purposes.

A threshold rule has been applied to the calibrated Chenies radar series for the period 1990–2002 in order to select rainfall events to which the event interiors model can be fitted. This thresholding has resulted in the rather surprising fact that there appear to be an increasing number of rain events over the period; more specifically, there is a jump upwards in the number of wet events around 1994, and an increasing trend afterwards. In part, these observations can be explained by changes in radar recording practice: improved technology results in fewer missing images and better detection of light rain. However, inspection of rain gauge data suggests that at least part of the increasing trend in the number of wet events throughout the 1990s is genuine.

There are various implications for the continuous simulation of rainfall. The increased number of events fitted to the later part of the record as compared with the early part means that the parameter sets in the library of event parameters used for the continuous simulation will be weighted towards events in this period, reflecting any changing characteristics of the wet spells as regards coverage and mean intensity. The simulated realisations should reproduce the overall characteristics of the radar data over the whole period of the record, although of course the random sampling of parameter sets means that trends will not be present in the simulated images. If there are genuine changes in the numbers of events and their characteristics (and to determine long-term patterns far more than 12 years' data are needed), the need for nonstationary continuous simulation, as discussed in Section 9.5, becomes apparent.