Point Process based models for rainfall

Thesis submitted to the University of London for the degree
of Doctor of Philosophy in the Faculty of Science

by

Anastasia Kakou

Department of Statistical Science

University College London

April 1997

ProQuest Number: 10106696

ProQuest 10106696

Αφιερωμένο στους γονείς μου, Δημήτρη και Ευανθία.

Acknowledgements

Abstract

In this thesis, we study the evolution of rainfall at a single site and over a network of sites by generalising existing point process based models.

Stochastic models based on clustered point processes, such as the Neyman-Scott and the Bartlett-Lewis processes, have been used recently in the description of the behaviour of rainfall at a fixed point in space. In such models, storms are idealised as cluster origins that arrive in a Poisson process and are followed by a number of rain cell origins, the cluster members. A rectangular pulse is associated with each rain cell origin, having independent random duration and intensity. In this thesis, a class of models with rain cell duration and intensity being dependent random variables has been developed and the main properties have been derived.

For the description of the evolution of a rainfall event at many distinct spatial locations, we consider a master clustered point process which is decomposed into sub-processes according to a marking mechanism, depending on the location(s) that are affected by a storm and its rain cells. Each cluster of the sub-processes is randomly translated in time, in order to allow different sites to experience the same event at different times. Some of the model's parameters remain the same at all generated sub-processes, while others vary in a stochastic or deterministic way. We follow two approaches in modelling the probability that a storm or a rain cell affects a particular subset of sites. One is to describe the spatial structure of a rainfall event by assigning to each of its elements a band of random width, location and orientation, that intersects the catchment area. Alternatively, the probability that two sites experience the same event can be expressed as a deterministic function of the distance between the sites.

The models are fitted to raingauge data from the South-West of England.

3

Contents

Chapter 1

Stochastic Rainfall Models

1.1 Introduction

Rainfall modelling had traditionally been a research area of disciplines in the physical sciences like hydrology, meteorology, climatology and atmospheric physics. Over the last few decades however, after the fundamental work of Le Cam (1961), probabilitists and statisticians have been motivated to introduce a new way of approaching the evolution of rainfall events, by developing a stochastic approach.

Following Cox and Isham (1994), we distinguish three broad types of mathematical models for rainfall,

- the empirical statistical models in which the relationships between the explanatory variables are represented via empirical equations (Stein, 1986);

- the models of dynamic meteorology in which the physical processes are described using the theories of fluid dynamics and thermodynamics and large systems of nonlinear differential equations are solved numerically (Mason, 1986);

- the intermediate stochastic models in which a modest number of parameters, that are related to physical characteristics, are used to represent the evolution of the rainfall process.

All three types have an important role to play and the choice between them depends on the purpose of the analysis. The way we shall approach this area is to identify basic physical structures of the process and then to describe them with a stochastic model.

In this study, we start with an analysis of rainfall data collected over short recording intervals, and our aim is to describe the main properties of the observed process using stochastic models that have a small number of physically interpretable parameters. Such models can then be used to produce synthetic data for use as inputs for other hydrological processes, like rainfall run-off models in the design of storm-drainage systems.

Precipitation areas are usually classified according to their spatial extent. A hierarchical structure has been observed in most rainfall systems, where a precipitation area of a given scale has one or several smaller scale areas of more intense precipitation embedded within it. Waymire and Gupta (1981a) give a very good description of the structure of rainfall in space and time. According to this, areas greater than about 10^4 km^2 are referred to as synoptic areas, and generally have a lifetime of one to several days. Synoptic areas contain large subsynoptic precipitation areas, called large mesoscale areas (LMSA), that have a spatial extent ranging from 10^3 to 10^4 km^2, last for several hours and usually appear as elongated bands. The precipitation intensity inside an LMSA is always higher than the region surrounding it. One level down the hierarchy of the rainfall organisation are small mesoscale precipitation areas (SMSA), that build and dissipate within an LMSA, range from 10^2 to 10^3 km^2 in horizontal extent and have an average lifetime of a few hours. Both LMSA and SMSA contain identifiable regions of cumulus convective precipitation, known as convective cells, whose areas depend upon the storm and range from 10 to 30 km^2 and whose durations are of the order of several minutes to about half an hour. Convective cells within LMSA appear in clusters, which is a result of the movement of air masses of different temperatures. The rainfall intensity within a cell is always higher than that of the region surrounding it. Although the above classification is based on observations on cyclonic storms, a similar type of multi-level representation can also be used for convective rainfall, with a wide range of storm types exhibiting similar characteristics. So, a hierarchical structure has been observed in the spatial extent, the temporal development and the intensity of many rainfall systems.

Precipitation is a continuous time and space phenomenon. However, rainfall data are usually available in discrete form. For instance, gauge stations record the total rainfall intensity during a sequence of time periods, at a fixed spatial location, while radar beams scan the atmosphere above a large area, at regular times, and give spatial averages

of rainfall intensity over smaller sub-regions. Depending on the application, stochastic rainfall models may examine marginal behaviour and properties of the process, such as its temporal evolution at fixed points in space, or the spatial structure of a single rainfall event. Ideally, we expect that models that describe only the temporal or only the spatial behaviour of the process can be derived as special cases of fully spatial-temporal models. This is not always an easy task. Some of the problems were pointed out by Valdes, Rodriguez-Iturbe, and Gupta (1985), who used a multi-dimensional model that produces moving storms with realistic mesoscale features, to simulate rainfall in space and time. The rainfall intensities of the simulated process are measured at fixed gauge stations, and then three different one-dimensional temporal models are fitted. None of the three models reproduces in a satisfactory manner the characteristics of the multi-dimensional model. A question that arises is whether the spatial-temporal model that is used to simulate rainfall is indeed capable in reproducing all physical features of a precipitation event, and thus whether it should be used to assess the performance of the temporal models.

Although the spatial and temporal aspects of rainfall are linked in a dynamic way, models that describe one or the other or both of these aspects have been developed independently and in parallel. In what follows, we present some of the research that has been done during the last couple of decades, in modelling the various features of the evolution of precipitation fields.

1.2   Temporal evolution at a single location

One key issue in modelling the temporal evolution of rainfall at a fixed spatial location, is the representation of the process at various time scales using a single set of parameters. For example, one may wish to describe the physical attributes of the rainfall intensity at a fixed point for hourly, daily and weekly periods. It is desirable that the integration, over different time intervals, of the instantaneous process coming from a mathematical model, will represent realistically the characteristics of the integrated rainfall process.

Two very important features of the rainfall process, the first to be modelled, are the numbers of events that happen during various time intervals and the rainfall amount associated with each event. Waymire and Gupta (1981a) in a review paper present various

approaches to describing the above characteristics. Rodriguez-Iturbe, Cox, and Isham (1987a) start their study with a Poisson process based model, in which a rectangular pulse with random intensity and duration is associated with each point. The results of fitting such simple stochastic models to data suggest that there is a lot of dependence between rainfall intensities over disjoint time intervals, and a high probability of rainy periods occurring close to each other. One way to incorporate this dependence is to assume a high-order Markov chain model, either in discrete or in continuous time, that governs the alternation between wet and dry time intervals. Another way is to consider models based on cluster point processes that are overdispersed compared to a simple Poisson process.

Markov chain based models

Stern and Coe (1984) used a two state, discrete time, second-order Markov chain model for the analysis of daily rainfall data, where the states correspond to periods of rain below and above a certain threshold. The authors argue that the assumption of stationarity is inappropriate, even for periods as short as one month. For this reason, each day of the year is fitted separately, using data from a 53-year record, and a time series for each transition probability is produced. The amount of rain above a predetermined threshold, during a wet day, is assumed to have a gamma distribution with time dependent mean. Two cases are examined, one where the mean rainfall intensity per rainy day is independent of the history of the process, and the other where the mean depend stochastically on the state (i.e. 'dry' or 'wet') of the two previous days. Surprisingly, the former, rather than the latter, seems to be in accordance with the data, but this is not a typical result. In order to incorporate the temporal variability of the model's parameters, a curve, that has the form of either a generalised linear model or a Fourier series, is fitted to each time series. A disadvantage of this analysis is that the model can not easily be extented to cope with more frequent recording intervals. Also, its performance has not been tested on second order properties of the rainfall intensity.

Hutchinson (1990) examined a stationary, continuous time, three-state Markov process for the occurrence of rainfall, aiming to link some atmospheric conditions to the amount of rain at a fixed location. The three states of the Markov chain correspond to dry spell, transitional period and wet spell or shower. The second of the above states is

associated with clouds that have the potential, not always realised, to become disturbed and deposit rain on the ground. Wet and dry spells are always succeeded by transitional periods, so the transition matrix is specified by a single parameter. The model has three more parameters, the exponential sojourn rates for each state. A rainfall event is defined as a sequence of wet spells and transitional periods, thus allowing showers to occur in clusters but not to overlap. The distribution of the length of a rainfall event and of a dry period, which is a sequence of dry spells and transitional periods, have been derived. A shower's intensity is exponentially distributed, constant throughout its lifetime and independent of its duration. The author questions the assumption of independence between intensities of showers by examining a model with correlated intensities and a correlation function that decays exponentially with the time lag between the showers. The first and second order properties of the total rainfall intensity for models with independent and with correlated shower intensities are derived. The conclusion drawn from fitting the models is that the latter reproduces the observed properties, of hourly up to daily recording intervals, very accurately.

Cluster point process based models

Following a different approach, Kavvas and Delleur (1981) considered a cluster–based point process model, in continuous time, where at the primary level is the rainfall generating mechanism, for instance the fronts, and at the secondary level are the elements of the process that produce the actual rainfall, such as the shower cells. Cluster centres arrive in a Poisson process and each one generates a random number of shower cells. The time lag between the occurrence of a cluster centre and its members is exponentially distributed. The model, as defined above, has a Neyman–Scott (Neyman and Scott, 1958) structure in time and was used to study daily rainfall sequences. The interval length for this study is taken to be one day, mainly because of computational limitations. The log–survivor function between two consecutive rainfall occurrences and the rainfall counts spectrum, which is the Fourier transform of the covariance density of the rainfall counts process, are used to fit the above cluster–based model and also a simple Poisson model of rainfall arrivals. Both statistics show that the Neyman-Scott model preserves the behaviour of the daily data very well.

11

Poisson Process

{X_i} are iid exponential
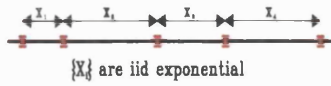
Clustering models

Storm origins ('parents') shown in blue
Cells in storms ('offspring') shown in red

1) Bartlett–Lewis

{X_i} & {Y_i} form finite renewal processes

2) Neyman–Scott

{X_i} and {Y_i} are iid

Rainfall intensity
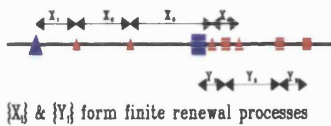
Rain cells
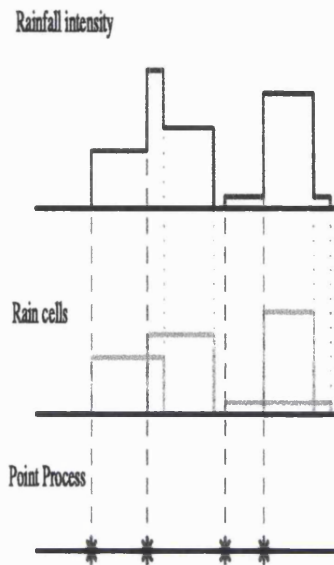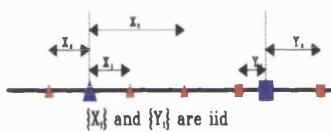
Point Process

Figure 1.1: Point process models in time; Poisson, Bartlett-Lewis and Neyman-Scott (left). Rectangular pulse rain cells as elements of rainfall models (right).

Rodriguez-Iturbe, Cox, and Isham (1987a) introduced two types of cluster point process models for the behaviour in time of the precipitation at a fixed point in space. A common feature of the two types is that cluster origins, called storms, arrive in a Poisson process and generate a number of rain cells, which are idealised as rectangular pulses having random duration and intensity. A graphical representation of the temporal structure of the Poisson process and two clustered point processes is given in Figure 1.1. In the Neyman-Scott based model, the number of rain cells per storm is a random variable, that follows a geometric or a Poisson distribution. The cell origins are independently displaced from the storm centre by distances which are exponentially distributed. In the Bartlett-Lewis based model, cell origins follow the storm centre in a renewal process,

12

centre in a renewal process, usually Poisson, which terminates after an exponential time. In both cases, the total rainfall intensity at a specific time is the sum of the intensities of all active rain cells. The assumption that rain cells have a finite lifetime, makes these models appropriate for the study of properties related to dry periods. Since rainfall data are available in aggregated form, the continuous–time process is discretised by taking the cumulative rainfall totals over disjoint time intervals of fixed length. The second–order properties for the continuous and the aggregated process, and the probability that an arbitrary time interval of fixed length is dry, have been derived. The Bartlett-Lewis based model is fitted to monthly rainfall data from Denver, Colorado, and is compared to a Poisson–based model in which storms have a rectangular profile. The results show that the former performs much better than the latter, and that it preserves the main properties of the process at all levels of aggregation, the only exception being the probability of zero rainfall during an arbitrary interval, which is overestimated.

A more detailed comparison between the three point process models with rectangular pulses was carried out by Rodriguez-Iturbe et al. (1987b). This analysis shows that the Poisson based model can be used for the description of the rainfall process only at a fixed level of aggregation, preferably larger than a typical duration of a storm. On the other hand, the Neyman-Scott and the Bartlett-Lewis based models are capable of reproducing, using a single set of parameters, the statistical properties of temporal rainfall at various scales of aggregation, ranging from one hour to one day. These models also perform well with regard to the extreme values of rainfall, but they give poor estimates of the probabiltiy of zero rainfall at an arbitrary time interval of fixed length.

In order to improve the performance of the cluster-based rectangular pulses models, Rodriguez-Iturbe, Cox, and Isham (1988) extended their earlier results by considering a Bartlett-Lewis model in which the cell duration parameter varies randomly between storms but is fixed for all cells within a cluster. Also, the parameters for the storm duration and the cell interarrival time are made scale–invariant, which corresponds to having a common structure for each storm but allowing distinct storms to occur at different timescales. So with only one extra parameter a continuum of storm types can be modelled successfully, and indeed the study shows that the probability that there is no rain in an arbitrary time interval is accurately reproduced at different time scales.

Entekhabi, Rodriguez-Iturbe, and Eagleson (1989) applied the above idea to a Neyman-Scott model and obtained similar results.

After the development and the analytical exploration of the modified Neyman-Scott and Bartlett-Lewis models, there was a need for a detailed and extensive comparison between all the cluster-based rectangular pulses models. Such a comparison was carried out by Velghe et al. (1994), who consider three classes of models, namely the Bartlett-Lewis, the Neyman-Scott with a geometric distribution and that with a Poisson distribution for the number of cells per storm. For each class, the original model, with constant parameter for the cell duration, and the modified one, with a random parameter, are examined. The rainfall data used are from Denver, Colorado. A reassuring result is that the modified models give better estimates, as compared to the original ones, of the zero depth probability, the correlation structure, several conditional distribution characteristics and of the extreme value properties. Minor differences in the performance are found between the original models, with the geometric Neyman-Scott one being slightly superior. Concerning the modified models, it is observed that the Bartlett-Lewis one is very sensitive to the sets of moments used in the parameter estimation, and that the geometric Neyman-Scott model, in general, performs better than the others.

A generalised Neyman-Scott rectangular pulses model has been developed recently by Cowpertwait (1994), that allows the rain cells to be of a number of different types, with a certain probability associated to the occurrence of each type. The distributions of cell intensity and duration depend on the cell type, so the generalised model provides a correlation between these two characteristics of the generated rain cells. However, given the type of the cell, all the variables involved are mutually independent. The case where cells can be of two types, 'heavy' or 'light', is examined in some detail and the model is fitted to rainfall data of different months separately. In order to ensure a smooth seasonal variation in the parameter estimates, it is assumed that each parameter varies across the year according to a harmonic relationship. As a result, the total number of parameters is reduced, but the estimates of the statistical properties of the rainfall intensity are less accurate. Compared to the original Neyman-Scott model, in terms of the extreme values distribution, the generalised model produces a better fit to the historical values.

It is worth mentioning that the idea of modelling the seasonal variability of parameters

was also explored by Onof and Wheater (1993), who fitted a Bartlett-Lewis based model with random rain cell duration to British rainfall data. It is suggested that it is possible to use a polynomial curve or a Fourier series for a smoother representation of the above variability, on a daily rather on a monthly basis, and to obtain satisfactory results.

In many hydrological applications, such as the rehabilitation of storm sewer systems, properties such as the probability of dry spell sequences of given length and the extreme values distribution are of particular interest. Cowpertwait et al. (1996) used a Neyman-Scott rectangular pulse model, with fixed rain cell duration parameter, and found that the two-parameter Weibull distribution for the rain cell intensity, instead of its special case the exponential distribution, gives good agreement between the simulated and observed daily annual maximum rainfalls. Also, it is observed that when the wet and dry spell transition probabilities, rather than the lag 1 autocorellations, are included in the generalised method of moments fitting procedure (a description of which can be found in Section 3.4), the estimates of the dry spell sequences are improved.

1.3 Spatial structure of a single event

Equally as interesting as the study of the temporal evolution of rainfall at a fixed point in space is the exploration of its spatial structure at a fixed point in time. Rodriguez-Iturbe, Cox, and Eagleson (1986) developed a rainfall model in continuous space, in which the temporal variability is ignored by examining individual storms separately and assigned to each point in space the total storm rainfall. Each storm consists of a number of rain cells, the rain-producing elements, that are distributed in the area of interest independently according to a two-dimensional Poisson process. The rainfall intensity at the cell centre is random and is distributed around the centre in a way specified by a spread function. Four circularly symmetrical spread functions are considered, and the second order properties of the total depth are derived using the moment generating function. When the spread function has a quadratic exponential form, and the intensity at the cell centre is exponentially distibuted, then the total storm depth at a given location has a gamma distribution and its properties are given in a closed form. Because the probability of an area being dry is non-zero only for spread functions with finite support, an arbitrary threshold is placed to the rainfall depth when the spread function decays

non-linearly with distance. A few generalisations of this model are examined. The first is to allow some parameters, such as the mean intensity at the cell centre, to vary randomly between storms. The second is to incorporate more flexible and realistic spread functions either by letting some of the parameters be random, varying from cell to cell, or by allowing for each rain cell a random fluctuation in intensity around the decaying trend from the cell centre. Finally, a clustering mechanism for the spatial distribution of rain cells is considered. In particular, a model with a Neyman-Scott two-dimensional process for the allocation of rain cells in space and with an isotropic quadratic exponential spread function is studied in more detail.

Poisson spatial models, such as the ones described above, were applied to air mass thunderstorm rainfall by Eagleson et al. (1987), who used observations from a dense network of raingauge stations. In order to produce a random rainfall field, the station observations are interpolated onto a rectangular grid. The bias of the correlation estimates that is produced from the interpolation is investigated and seems negligible. Also, the assumptions of homogeneity and isotropy of the random field are verified. Three different spread functions are tested, namely, a quadratic exponential, a simple exponential and a linear function. Storm days are identified and the three models are fitted to each storm separately. It is observed that the parameter estimates present a variability among storms and thus the mean rainfall intensity at the cell centre is assumed to be random. This generalisation improved significantly the fitting of the two models with exponential type spread function, for which the second-order properties can be derived analytically. An overall conclusion is that the Poisson process models developed by Rodriguez-Iturbe, Cox, and Eagleson (1986) are capable in reproducing important features of the spatial distribution of total storm precipitation, at least for storm types that are essentially stationary in space.

1.4 Model development in continuous space and time

A more complete study of the rainfall process is obtained via stochastic models that are built in continuous space and time. Waymire and Gupta (1981b) illustrated, with theorems and examples, the suitability of the point process based models, as compared to other stochastic models, to accommodate features of the physical structure exhibited

by space-time rainfall. They also stressed the fact that the tools of point process theory can facilitate algebraic operations and make further mathematical exploration possible. They formulated a quite general and abstract model that reflects the main features of the hierarchical rainfall structure. The large-scale elements of the process, the cluster potentials, arrive in the three-dimensional time-space according to a point process, and have a spatial extent of given radius. A similar generating mechanism exists for the smale-scale elements, the rain cells, that occur within the cluster potentials. The rainfall intensity at a point in space and time, due to a single cell, is a deterministic function of the age of the cell and the distance away from the cell centre. Cluster potentials and rain cells are allowed to move with different velocities. The total rainfall intensity at a specific time and ground location is the sum of the intensities contributed by all rain cells that arrived in the past over the region of interest.

Waymire, Gupta, and Rodriguez-Iturbe (1984) extended the above model to allow rainbands to occur, in time, in clusters over an arbitrary fixed geographical region, and made specific assumptions about the form of the functions involved. The extended model has a two–level clustering in space and one in time. It is assumed that rainbands arrive in time in a Poisson process and each rainband generates a number of cluster potentials that are spatially allocated according to a two-dimensional Poisson process. A number of rain cells, the elements that actually deposit rain on the ground, arrive within each cluster potential according to a point process in three–dimensional space–time. The point process of rain cell origins is assumed to have a Neyman-Scott structure in time and in space, with the two displacement components being statistically independent. Rain cells have a fixed intensity at the centre, at the time of birth, which decays exponentially with the age of the cell and the distance away from the centre. So, once a rain cell is born, it contributes to the total rainfall intensity at all spatial locations and at all times, which implies that the probability of a point being dry is always zero. In addition to the above specifications, rainbands and raincells move with fixed but different velocities. The authors, employing the tools of the point process theory, derived the mean and the covariance density in space and time of the rainfall intensity. A further important achievement of this study is that several qualitative features of the precipitation system follow as consequences of the above formulation and not as part of the initial assumptions. An example is the Taylor

or frozen field hypothesis, that states that the temporal autocovariance at a fixed point in space should be of the same form as the spatial autocovariance at a fixed point in time, with the 'time' argument of the former converted into the 'spatial' argument of the latter by multiplication by a typical velocity.

Phelan and Goodall (1990) introduced a generalised version of the previous model that has the same clustering mechanism, but allows some of the rain cell characteristics, such as the rain intensity at the cell origin, the aging rate, the planar velocity and the spatial extent, to be mutually independent random variables. The water content of a rain cell has now a Gaussian surface in space, and a double exponential shape in time, which captures both the intensive and the dissipative stages of a rain cell. The scale, orientation and anisotropy of the rain cells are estimated by applying a linear transformation that takes a rain cell of spherically symmetric covariance structure having unit scale to one with the estimated covariance structure. The model is fitted to individual tropical storms using hourly radar data and the analysis is focused on the geometry and kinetics of a sample intensity process. The rain cells are identified by direct inspection of the radar images. In order to estimate the parameters, the model is decoupled in time and each hour is analysed separately. At a fixed time, the intensity, spatial location and dispersion of each rain cell are estimated using the least squares method, yielding a time series of fitted characteristics for each one. Then, the model is recoupled and the remaining parameters, that is, the velocity and the aging rate, are estimated again by ordinary least squares. The conclusion drawn from the fitting is that the hypothesis of time invariant cell characteristics needs further investigation. In general, models that require manual identification of individual rain cells are not very promising in analysing long sequences of rainfall systems.

A different approach, based partly on the theory of Markov chains and partly on point processes, was followed by Smith and Karr (1985) in a study of daily rainfall sequences. Days are classified as 'wet' or 'dry' and the alternation between these two states is governed by a stationary transition matrix. For days in which rainfall occurs, the number of rain cell origins has a Poisson distribution with the rate being an exponential random variable, that is, the cells occur in a doubly stochastic Poisson process. Rain cells are idealised as cylinders that have fixed constant radius and exponential intensity,

and are located over the area of study in a spatial Poisson process. The model was fitted to rainfall data from a network of five raingauge stations. Two parameter estimation procedures are developed, one based on maximising the likelihood function, and the other on the method of moments. In both cases, the spatial and the temporal parameters are estimated separately. A serious disadvantage of this approach is that there is not a time continuum since spatial features of the rainfall are assumed independent between days. Thus, this model could not be used successfully in studying rainfall data of a small temporal resolution, such as hourly or five minutes, since it does not take into account the temporal correlation between successive periods.

Cox and Isham (1988) explored a simple point process based model, where storms consist of single rain cells that arrive in a Poisson process in space and time. Rain cells are idealised as cylinders of random radius that move with constant random velocity and die after a random time. The intensity of a cell is constant during its lifetime but varies stochastically between cells. The innovation of this formulation is that the rain cells have a finite spatial extent and duration which implies that the probability that there is no rainfall at a point in space and time is non-zero. The authors derived a general expression for the spatial-temporal covariance function under the assumption that all the variables involved in the model are mutually independent. Further to this result and under the additional assumptions that the rain cell's duration and radius are exponentially distributed and the velocity has a gamma distribution, several statistical properties of the process are determined, including the spatial covariance at a fixed point in time and the temporal covariance at a fixed point in space, as well as properties related to the probability of non-zero rainfall at various points in space and time. Cox and Isham (1988) also considered a cluster-based spatial-temporal model in which storm centres arrive in a Poisson process in space and time, and cells occur in a Bartlett-Lewis process in time, at the same spatial location as the storm origin. An analytic expression for the spatial-temporal covariance function is not available, but other properties such as the variance of the marginal rainfall intensity can be derived approximately under reasonable assumptions.

A far more realistic and mathematically elegant model has been developed recently by Northrop (1996). Storm centres arrive in a homogeneous Poisson process in three-

dimensional space-time. Following each storm centre, rain cell origins arrive in a temporal Bartlett-Lewis-type cluster, that is, they arrive in a Poisson process, that starts with a cell located in time at the storm centre, and terminates after an exponential time. All rain cell origins within the storm are displaced from the storm centre according to a spatial displacement distribution. Two such distributions have been studied — one where the cell origins are uniformly distributed over a random ellipse centred on the storm centre, the other where the displacements follow a bivariate Gaussian distribution. Both these variants of the model have the flexibility to produce a wide range of storm structures, ranging from rain bands to more widespread rainfall. Each cell is elliptical in shape, with a random major axis, and deposits rain at a constant intensity on all points in space covered by its defining ellipse during its lifetime. The elliptical peripheries of the rain cells are scaled versions of the storm ellipse, or the elliptical contours of the bivariate Gaussian distribution. The intensity, the duration and the major axis of a cell are mutually independent random variables. All cells within a storm and the storm centre itself, move with the same random velocity. Cell clusters belonging to distinct storms are independent. Northrop (1996) derived explicit expressions for the second order properties — mean, variance and covariance density — of the process, and an approximate expression for the probability that a randomly chosen pixel, of arbitrary size, is dry at a given time. The models are fitted to radar data from Wardon Hill, South-West England, and have proved to be capable of reproducing the main spatial-temporal features of the rainfall field.

1.5 Multi-site models

It is often desirable to study the temporal evolution of rainfall at various distinct spatial locations simultaneously. Although fully spatial-temporal models describe the behaviour of the rainfall process in a continuous time and space domain, it is usually difficult to derive from such models analytical expressions about marginal properties at specific points in space. Also, when data are available as time series of rainfall intensities at various sites, then some features of the process, such as the velocity or the spatial extent of the event, can not be estimated directly from the data, which makes the spatial-temporal models inappropriate for such analysis. For these reasons, it is important to develop

multivariate models, that describe the relationships between rainfall intensity at different sites and also preserve the structure of the single site models for their marginal processes at each site. The literature on multivariate or *multi-site* models is limited. This is partly because research effort has concentrated on the development of satisfactory single site models, and partly because progress in multi-site modelling requires an extensive analysis of data collected from a carefully designed network of raingauge stations. The network should ideally consist of a large number of stations, located close enough together to capture the within rain cell dependences, and should occupy a large area, so that several spatial features of the rainfall structure can be observed.

Among the first who explored this area of rainfall modelling are Cox and Isham (1994), who used some rectangular pulses point process models for a single site and modified them to allow the study of more than one, say n, sites. More specifically, it is assumed that there is a master point process of storms which evolve independently of each other, and that each storm can affect a subset of the n sites with a certain probability. From the between sites characteristics, for simplicity, only second order properties between pair of sites are considered. So, if initially the analysis is restricted to two sites, three subprocesses are generated from the master Poisson process, and each storm origin belongs to one of them depending on the site(s) it affects. A key issue is how the subprocesses are related, in other words, what assumptions should be made about the dependence between the evolution of the rainfall processes at distinct sites.

One extreme case is to have identical storms at all affected sites. Cox and Isham (1994) examined a slightly modified model, by allowing the rain cell intensity to be scaled by a different constant at each site. The theoretical cross-correlation function of the cumulative rainfall intensity at the two sites is derived, and its form does not depend on the level of aggregation, a result which is inconsistent with the empirical data. Another extreme case is to suppose that a storm evolves idependently at all affected sites. Cowpertwait (1994) suggested the use of a model with a number of different cell types (mentioned in Section 1.2), where each type corresponds to rain cells affecting a different subset of a network of sites. Although the model is not fitted to rainfall data, it is unlikely that it will give satisfactory results, since the assumptions of independence between the variables of different types usually produce smaller cross-correlations than

observed. A more realistic, intermediate model that has the Bartlett-Lewis structure in time was constructed by Cox and Isham (1994), who assumed that the Poisson process of cell origins is the same at all affected sites but the storm truncation mechanism acts independently at each site. Also, the intensities of cells with a common origin but at different sites are independent, but the durations of such cells are dependent and are scaled versions of the same exponential variable. The common parameter for the cell duration can either be constant or vary randomly for different storms. The theoretical cross-correlation function is derived, and allows a varying degree of dependence between the rainfall intensity at the two sites for different parameter sets. Such an approach, where some of the rainfall characteristics are kept the same for all subprocesses and others vary in a stochastic or deterministic way, is very promising. However, a detailed data analysis is a prerequisite to provide guidelines for appropriate assumptions.

1.6   The project

The work presented in this thesis has being carried out as part of the Hydrological Radar Experiment (HYREX) project, funded by the Natural Environment Research Council. A carefully designed dense network of 49 raingauge stations has been installed in the Brue river catchment, in South-West England, covering an area of roughly 140 km^2. The gauges are 0.2 mm tipping buckets and the way they operate is that rain is accumulated in each bucket until it is filled out. Then, the water is released, the time is recorded and the bucket is ready for the next measurement. So, the raingauge data available are sequences of times when a bucket tipped, from which rainfall intensity aggregated over disjoint time intervals can be calculated. A radar station, located close to this network, is monitoring the region every 5 minutes. Radar data are collected for 5×5 km^2 pixels up to a distance of 210 km from the station, and for 2×2 km^2 pixels up to a distance of 75 km. Both types of data are available from September 1993.

Our part of the project involves analysis of the raingauge data and modelling of the evolution of rainfall over the network of gauges. In Chapter 2, we discuss the results of the data analysis that provide guidelines for the theoretical developments. A class of point process models for the temporal evolution of rainfall at a single site, with rain cell characteristics being dependent random variables, is developed in Chapter 3. The

modelling of the precipitation process simultaneously at a small set of sites, and the exploration of different ways in which the inter-site dependence can be incorporated is presented in Chapter 4. The results of fitting the multi-site models to raingauge data are shown in Chapter 5, while the conclusions of this research are given in Chapter 6.

2	KILKENNY FIELDS
3	ALFORD FIELDS
4	DITCHET HILL FARM
5	KITE LANE
6	MEADOW BUILDINGS
7	SNAGG FARM FIELDS
8	BRIDGE FARM ORCHARD
9	OPP SPRINGFIELD COTTS
10	MOOR LANE PADDOCK
11	JACOBS LANE FIELD
12	SHIELDSVILLE FARM
13	LOWER FARM SUTTON
14	INDUST PARK CC
15	SPRING STONY STRATTON
16	FIELDS IN EVERCREECH
17	RODMOOR FARM FIELDS
18	WADDON FARM ORCHARD
19	EAST HILL LANE FIELDS
20	KNAPP HOLLOW WYKE LANE
21	CHESTERBLADE BANKS FM
22	BADGER SET MANOR FM
23	MILTON WOOD CORN FIELD
24	GREENSCOME FIELD
25	CRABTREE LANE FIELD
26	PITCOMBE FIELD
27	HADSPEN HOUSE FIELDS
28	FRANCOMBE LANE FIELDS
29	BATCOMBE HILL FIELDS
30	QUARRY FIELD BATCOMBE
31	SEAT LANE FIELD
32	WHADDONN FARM FIELDS
33	SEPTIC TANK COMPOUND
34	NEW HOUSE FARM FIELD
35	GODMINSTER WOODS
36	DROPPING LANE FARM
37	GODMINSTER FARM
38	REDLYNCH FARM STUMP
39	DODDS CORNER FIELD
40	TOWERS FIELD
41	POND FIELD BY TOWERS
42	KNOWLE FIELD BY STREAM
44	GOODEDGE FARM FIELDS
45	COGLY WOOD FIELDS
46	HORSLEY FARM FIELDS
47	MOOR WOOD FIELDS
48	WALTERS FARM FIELD
49	COOKS FARM FIELD
50	CRAWLY FARM FIELD
51	GLADWELL FARM FIELD

Figure 1.2: The map of the Brue catchment area with the locations and names of the raingauge stations.

Chapter 2

Data Analysis

Our research interest is the stochastic modelling of the temporal evolution of rainfall at a single point in space and also at a number of spatially separated locations. Before doing any theoretical work, it is important to carry out a substantial amount of data analysis, which will provide guidelines for the theoretical model development. The data we have available are from a network of raingauge stations (Figure 1.2), which covers an area of about 140 km^2, expanded around the basin of the river Brue and has an altitude range from 35 to 193 meters. Since the area is topographically inhomogeneous, it is crucial to know whether the statistical properties of the rainfall intensity are influenced by the spatial location and the elevation of the sites. This is one of the three broad areas of our data analysis. The second area involves the investigation of the temporal and seasonal variation of the rainfall intensity at a fixed point in space. Finally, we are interested in the between-site properties and how they are related to the separation distance and the orientation of the sites. This will provide information about some spatial characteristics of the rainfall, such as the storm velocities in different seasons.

As raingauge data become available, a routine analysis is carried out that includes the production of tables of monthly summary statistics (Appendix A) and of plots of cumulative rainfall intensity during a month (Figure 2.1). Also, a quality control process is applied, which involves the inspection of the monthly cumulative hyetographs and the elimination of bad data. For instance, with reference to the cumulative plots in Figure 2.1, it is easy to identify the problematic gauges no. 28 in June and no. 35 in December. The process of identifying bad data is particularly difficult in summer months because
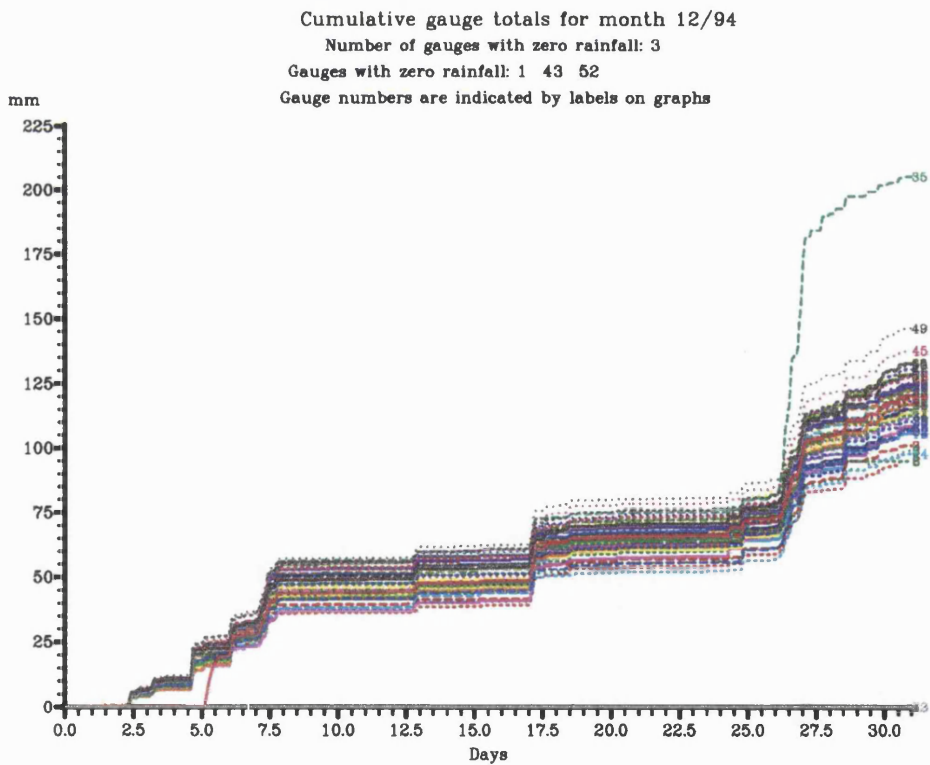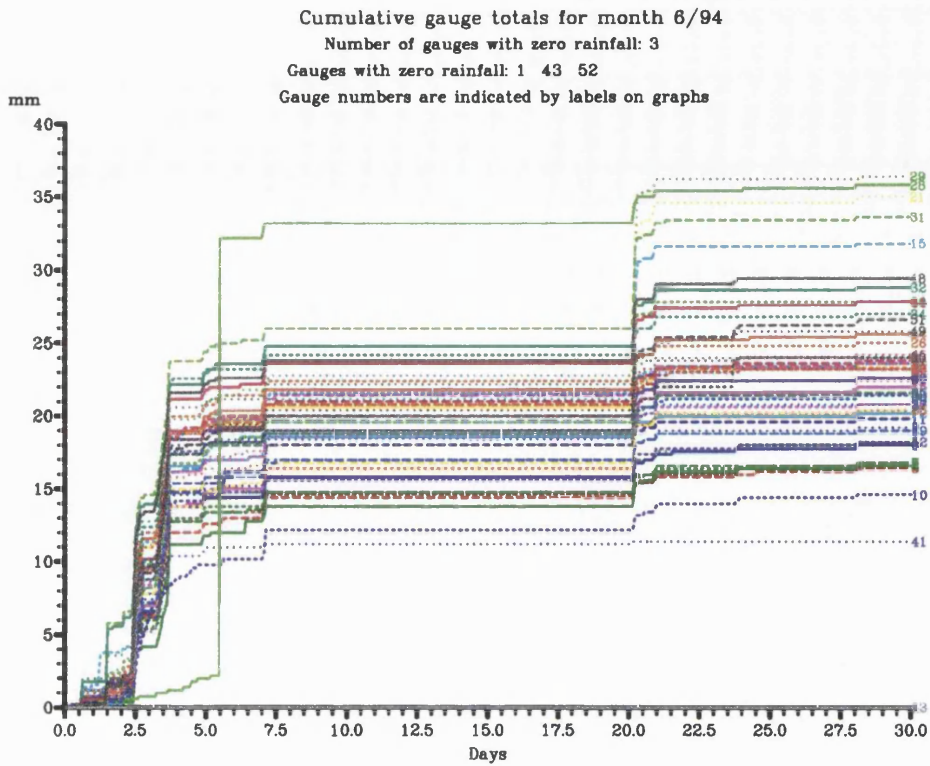
Figure 2.1: Hyetographs of cumulative rainfall intensity for June 1994 (top) and December 1994 (bottom).

25

there is more variability in the rainfall intensity between gauges. In addition to our gauge quality assessment, reports have been supplied by the Institute of Hydrology for each station, with a list of the periods during which a technical problem was reported. Suspect gauges have been excluded from any subsequent data analysis.

2.1 Temporal correlations

We examine the temporal characteristics of rainfall at a single site, and demonstrate the results for June 1994 and December 1994, a typical summer and winter month respectively, and for one raingauge of the whole network. In this context the rainfall intensity is to be regarded as an instantaneous quantity, measured in millimetres per hour. The data available are measurements of the total rainfall intensity during disjoint intervals of fixed length h. Let us denote by t_n the n-th time interval of a sequence and by Y_n the aggregated rainfall intensity during t_n. With reference to the hourly time series in Figure 2.2, we notice the clustered structure of the process, as well as the differences in the arrival rate and the intensity of a rainfall event between the two months. When the data are aggregated over larger time intervals, the time series become smoother, there are more rainy intervals and the accumulated rainfall intensity is higher during rainy periods.

In addition to the time series, one can plot the aggregated rainfall intensity, Y_n, during interval t_n against that of t_{n+k}. In such graphs the origin has a high frequency, that corresponds to transition between dry intervals, and there are fewer points further away from it. The time-ordering is lost, but it is easy to observe the relationship between rainfall intensities at time intervals separated by lag k, which is related to the lag k correlation function. The plots of the rainfall intensities during successive intervals, shown in Figure 2.2 for the same months as previously, are fairly symmetric relative to the diagonal, suggesting that the rates at which storms build and dissipate are similar. Because near the origin there are multiple points, in this graph, different symbols represent points with different multiplicity, while the atom at the origin has a frequency of 671 in June and 552 in December.

A more detailed analysis of the data involves the calculation, for individual months, of some correlation measures, such as the lag k correlation function, $\gamma(k) = \mathrm{corr}\{Y_n, Y_{n+k}\}$, and the probability that two time intervals of fixed length h, separated by lag k, are both
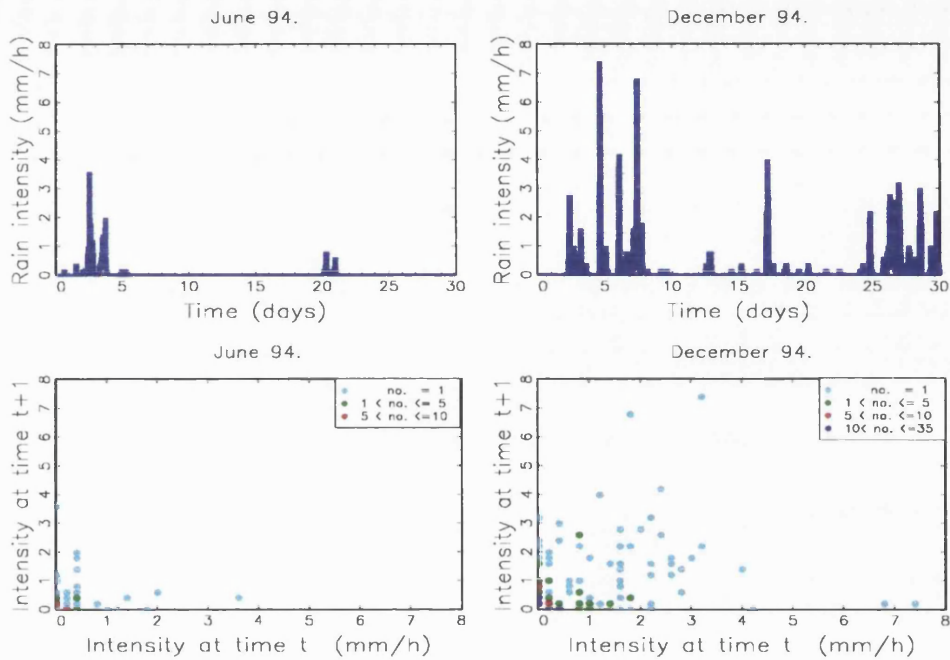
Figure 2.2: Hourly time series of rainfall intensities (top), plots of hourly rainfall intensities during successive intervals (bottom).

dry, $p_{00}(k) = P(Y_n = Y_{n+k} = 0)$, and their graphical representation for different k. With reference to Figure 2.3, we notice that the autocorrelation $\gamma(k)$ decays fairly rapidly for small lags and stabilizes at values close to zero, but the probability $p_{00}(k)$ remains almost constant for all time lags, and exhibits only a slight drop from the value resulting from lag zero. The latter suggests that dry intervals appear almost randomly in time and thus the conditional probability $P(Y_{n+k} = 0|Y_n = 0)$ is essentially independent of k, for positive values of k, and asymptotically we have $p_{00}(k) = p_{00}(0)^2$, where $p_{00}(0) = P(Y_n = 0)$ is the probability that an arbitrary time interval of a given length is dry. A similar pattern (not presented here), is exhibited by the probability $p_{++}(k) = P(Y_n > 0, Y_{n+k} > 0)$, which takes very similar values for all positive k and it is approximately equal to the square of $p_{++}(0) = P(Y_n > 0)$.

A correlation measure that, in general, provides more information about the structure of the rainy periods is the conditional lag k autocorrelation, which is calculated keeping the temporal spacing fixed but only using pairs where both values are positive, that is
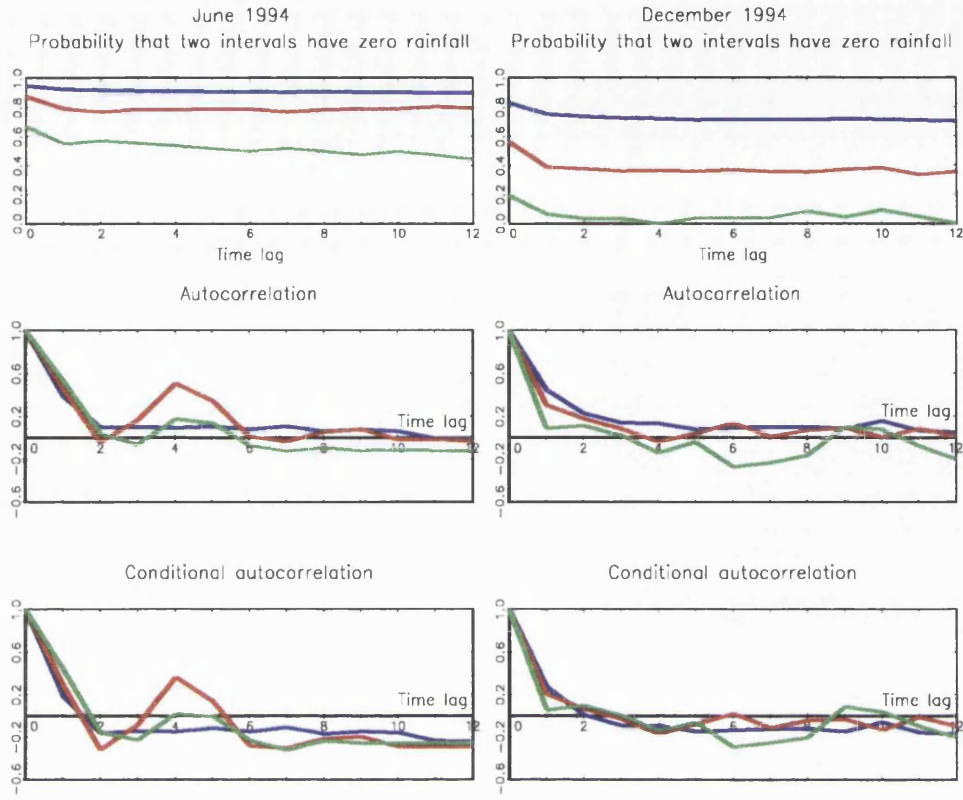
Figure 2.3: Correlation measures against time lag, of hourly (blue), 6-hourly (red) and daily (green) rainfall data, for June 1994 (left) and December 1994 (right).

$\tilde{\gamma}(k) = \text{corr}\{Y_n, Y_{n+k} | Y_n > 0, Y_{n+k} > 0\}$. However, since we observed that for all positive time lags, $p_{++}(k)$ is approximately independent of k and equal to $P(Y_n > 0)^2 = p_+^2$, we expect that $\tilde{\gamma}(k)$ is approximately proportional to $\gamma(k)$, and can be written as

$$
\begin{aligned}
\tilde{\gamma}(k) &= \frac{p_{++}(k)\ E(Y_n Y_{n+k}) - p_+^2\ E(Y_n)^2}{p_+ E(Y_n^2) - p_+^2 E(Y_n)^2} \\
&\approx \frac{p_+^2\ \text{cov}\{Y_n, Y_{n+k}\}}{\text{var}(Y_n | Y_n > 0)} \\
&= \frac{p_+^2\ \text{var}(Y_n)}{\text{var}(Y_n | Y_n > 0)}\ \gamma(k).
\end{aligned}
$$

The plots of $\tilde{\gamma}(k)$ against the time lag k, for three levels of aggregation h, are illustrated in Figure 2.3. The patterns of the two correlation functions are indeed very similar and the closer to one is the probability p_+, the closer are the values of $\gamma(k)$ and $\tilde{\gamma}(k)$.

In June, the correlation of hourly data decays sharply, indicating a fairly short lifetime of the rainfall events. For the 6-hourly data of this month, an interesting feature is the
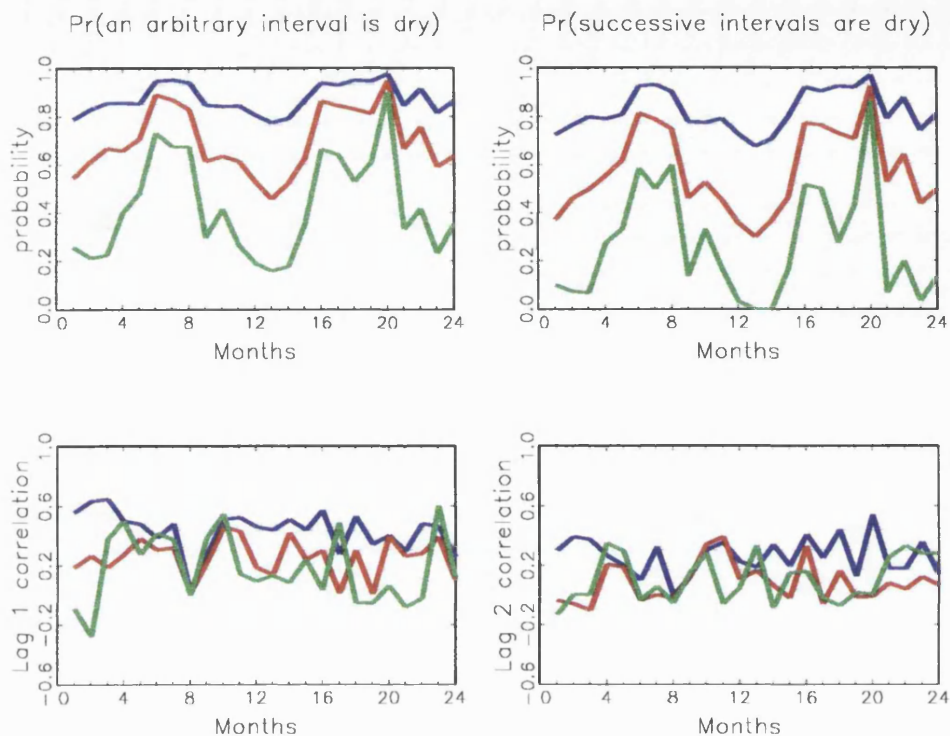
Figure 2.4: Temporal autocorrelation measures at a single site for 24 months, of hourly(blue), 6-hourly (red) and daily (green) rainfall data: probabilities $p_{00}(0)$ (top left) and $p_{00}(1)$ (top right) and correlations $\gamma(1)$ (bottom left) and $\gamma(2)$ (bottom right)

very steep drop of $\gamma(k)$ and $\tilde{\gamma}(k)$ to just below zero for small time lags, and then the following peak at lag 4, which corresponds to 24 hours. Such behaviour is quite common in summer months, when rain is falling during certain hours of the day, usually in the afternoon. The lag 1 correlation of daily data is also very high, which points to the conclusion that dry periods tend to last a long time and are interrupted by clusters of rainy days. On the other hand, the correlation functions produced from the December data set are quite smooth, and in contrast with the summer months, once they reach values close to zero they remain fairly stable. Also, the rainfall intensities between successive days are almost uncorrelated.

Figure 2.4 illustrates the seasonal variability of the probabilities $p_{00}(0)$ and $p_{00}(1)$ and of the temporal autocorrelations $\gamma(1)$ and $\gamma(2)$, for a period of two years, starting

from January 1994, labeled no. 1. The former exhibit an obvious seasonal pattern, with high values in summer and low values in winter, while the latter do not show any dependence with season. It might be interesting to note that in August both lag 1 and lag 2 correlations, for all three levels of aggregation (1, 6 and 24 hours), are very close to zero, mainly because rainfall events during this month are sporadic and have very short lifetimes.

2.2 Between-sites correlations

In the previous section we looked at some temporal correlation measures of the rainfall intensity at a fixed point in space. The main purpose, however, of planning and setting up the network of raingauge stations is to study the joint properties of rainfall at several locations and to gain insight into the spatial structure of an event. The quantities we calculate are between-site properties only, and more specifically, our attention is focused on the following three measures:

- the lag k cross-correlation, $\gamma_{ij}(k)$, between the rainfall intensity at site i, $Y_n^{(i)}$, observed during an arbitrary time interval of fixed length h, and the one at site j, $Y_{n+k}^{(j)}$, observed kh time units later during a period of the same length;

- the probability that two sites, i and j, are both dry, $p_{00}^{(ij)}(k) = P(Y_n^{(i)} = Y_{n+k}^{(j)} = 0)$, or both rainy, $p_{++}^{(ij)}(k) = P(Y_n^{(i)} > 0, Y_{n+k}^{(j)} > 0)$, during arbitrary periods of the same fixed length h separated by time lag k;

- the lag k cross-correlation, calculated from the pairs that have both observations positive, $\tilde{\gamma}_{ij}(k) = \text{corr}\{Y_n^{(i)}, Y_{n+k}^{(j)} | Y_n^{(i)} > 0, Y_{n+k}^{(j)} > 0\}$, usually referred to as conditional cross-correlation.

We investigate the effect of the distance and orientation of the pairs, on the cross-correlation measures, at various time scales and time lags. As before, we present the results for a dry summer month, June 1994, and a winter month with high rainfall, December 1994. Data are used from all gauges that are functioning properly during the examined periods.

2.2.1 Zero time lag

We start using hourly data, and group the pairs of sites according to the angle, ϕ, that is formed between the line that joins them and the east-west axis. When two sites are examined simultaneously ($k = 0$) the ordering is irrelevant and the value of ϕ, in radians, belongs to one of the sets $\mathcal{S}_1 = [-\pi/2, -\pi/4)$, $\mathcal{S}_2 = [-\pi/4, 0)$, $\mathcal{S}_3 = [0, \pi/4)$ and $\mathcal{S}_4 = [\pi/4, \pi/2)$.

The graphs of $\gamma_{ij}(0)$ plotted against the actual distance between the sites, are quite erratic, since many pairs with close distances produce fairly different values of the property, and so the underlying shape is not easily seen. In order to smooth the curves, the distance is calculated to the nearest kilometer, and the plotted value is an average, from the values of all pairs that have the same distance (in km).
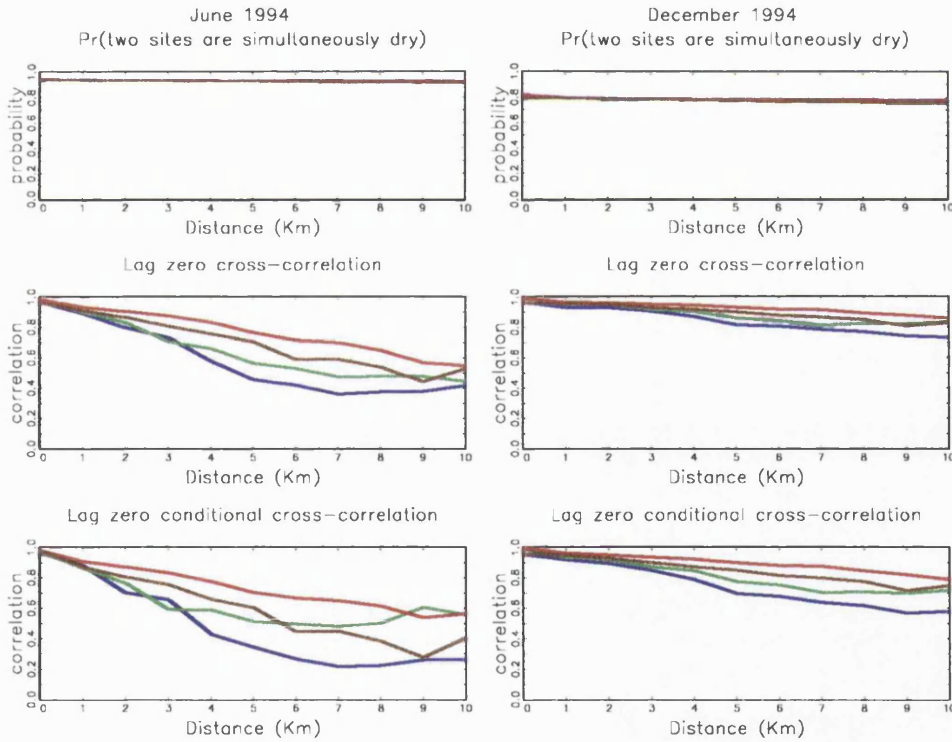
Figure 2.5: Cross-correlation measures at zero time lag, for pair of sites with angular separation in $\mathcal{S}_1 = [-\pi/2, -\pi/4)$ (blue), $\mathcal{S}_2 = [-\pi/4, 0)$ (green), $\mathcal{S}_3 = [0, \pi/4)$ (red) and $\mathcal{S}_4 = [\pi/4, \pi/2)$ (brown), for hourly data, for June 1994 (left) and December 1994 (right).

Figure 2.5 shows the plots of some cross-correlation measures against distance, where the four curves correspond to the different angular separation between sites. The probabilities $p_{00}^{(ij)}(0)$ and $p_{++}^{(ij)}(0)$ (the latter not presented here), do not change appreciably with the distance or angle, reflecting the fact that the Brue area is fairly small compared to the spatial extent of a rainfall event and once a storm hits the catchment, it visits all locations more or less simultaneously. On the other hand, the cross-correlation function, $\gamma_{ij}(0)$, decreases with distance, implying the existence of small-scale elements within the rainfall event that do not cover the whole area, and thus produce high correlations for nearby sites. These elements have a smaller size during summer months than during winter months, so that $\gamma_{ij}(0)$ decays more sharply in the former than in the latter case. Also, the cross-correlation, $\tilde{\gamma}_{ij}(0)$, that is derived conditionally upon the pair of non-zero observations, has the same pattern as the unconditional one, but lower values. One of the most interesting features of the graphs is that the cross-correlation of the pairs that have an angle, ϕ, in the set $\mathcal{S}_3 = [0, \pi/4)$, decays less slowly with distance, compared to the other sets, while the pair of sites with ϕ in $\mathcal{S}_1 = [-\pi/2, -\pi/4)$, which is orthogonal to $\mathcal{S}_3$, has the most rapidly decreasing pattern. This suggests the existence of a prevailing orientation, but before drawing any further conclusion we should look at the non-zero lag cross-correlations, which provide more information about the velocity of a storm's movement.

2.2.2 Nonzero time lags

When two sites are examined in different periods ($k \neq 0$), the ordering is imporatant and the angle ϕ takes values in the interval $[-\pi, \pi)$, which is partitioned into eight intervals each of length $\pi/4$. The cross-correlation function for each one of the eight groups is calculated, with distances rounded to the nearest kilometer. Instead of having a graph with eight curves, which would be rather messy to read, we keep the four sets $\mathcal{S}_1, \ldots, \mathcal{S}_4$ defined earlier, and we assign to each pair of sites a directional distance, which is the actual Euclidean distance if ϕ belongs in $[-\pi/2, \pi/2]$, or its negative value if $\phi + \pi$ belongs in $[-\pi/2, \pi/2]$. So for instance, two points with negative directional distance that belong in group $\mathcal{S}_3$, have an angular separation between $-\pi$ and $-3\pi/4$, in radians.

In Figure 2.6 we present the conditional cross-correlation function $\tilde{\gamma}_{ij}(k)$ of hourly
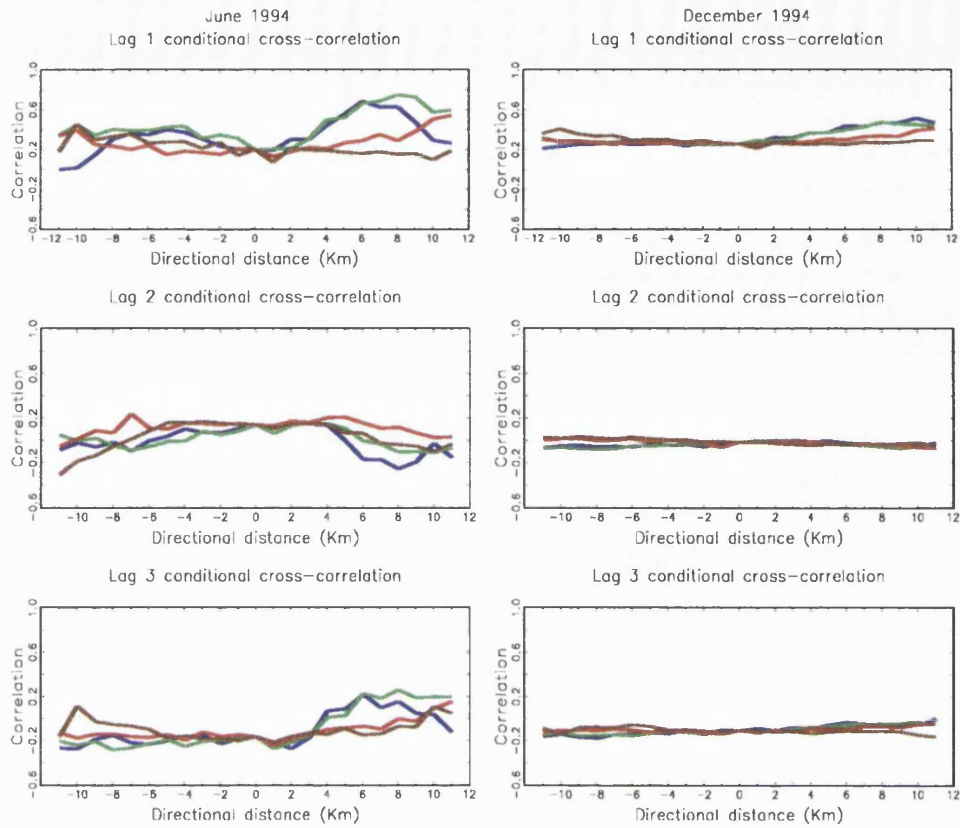
Figure 2.6: Conditional cross-correlations, for pair of sites with angular separation in $\mathcal{S}_1 = [-\pi/2, -\pi/4]$ (blue), $\mathcal{S}_2 = [-\pi/4, 0)$ (green), $\mathcal{S}_3 = [0, \pi/4)$ (red) and $\mathcal{S}_4 = [\pi/4, \pi/2]$ (brown), for hourly data at time lags of 1 hour (top), 2 hours (middle), 3 hours (bottom).

data for time lags 1, 2 and 3. The unconditional properties have a similar pattern, but the differences between the curves corresponding to the four directions are smaller. In June 1994, $\tilde{\gamma}_{ij}(1)$ increases for small absolute distances, it reaches a maximum value and then it starts dropping. From the location and the height of the peak, it is possible to obtain information about storm movement and the degree of correlation between the rainfall intensity at various sites. The curve that corresponds to the lag 1 conditional cross-correlation of group $\mathcal{S}_3$ takes its maximum value for the pairs that are furthest apart in the Brue area, and it would be very interesting to know if it would continue to increase for larger distances. The graphs show that storms with direction from north-west to south-east move with the highest speed, since the peak of $\tilde{\gamma}_{ij}(1)$ corresponds to pairs

that are a distance of roughly 8km apart.

It is believed that the initially unexpected increase in the lag 3 cross-correlation for pairs of sites with positive directional distance, is due to the clustering of storms and the arrival of a new storm, that has a general direction from west to east.

Having examined the cross-correlation functions of several other summer months, the conclusion we draw is that because of the relatively few storm arrivals during this season, the shapes of the functions are determined by a small number of events, and thus detailed inference about the movement of rainfall can not be made. Indeed, summer months with similar total rainfall intensity as June 1994, can produce different patterns for the joint-properties between pair of sites.

The cross-correlation functions for December 1994, on the other hand, remain fairly constant with distance, and there are no obvious differences in the pattern between the various groups of pairs, which is evidence of the very slow movement and wide-spread nature of winter rainfall. Similar things apply to the cross-correlation function of most of the winter months we have examined.

Overall, the conclusion from this analysis is that the velocity (speed and direction) of rainfall influences the joint-properties between sites, and thus should be included in the model construction.

2.3 Topographic effects

The topographic effects of the Brue area on the behaviour of the rainfall process have been examined using basic statistical tools and presented graphically. Our analysis is focused on the examination of two effects, the ground altitude and the location within the catchment area.

2.3.1 Ground altitude

The evolution of the process at the various sites for monthly periods, is described via some summary statistics (Appendix A), such as the mean, $\hat{\mu}$, and standard deviation, $\hat{s}$, of the rainfall depth, the probability, p_+, that an arbitrary interval of given length is wet, and the conditional mean, $\hat{\mu}_+$, and standard deviation, $\hat{s}_+$, calculated using the nonzero

observations only. The time resolution of the data used is 1 hour, since aggregation over larger time periods would smooth the overall pattern and reduce any difference between the sites. Because the variable p_+ is constrained to lie between the values 0 and 1, the logistic transformation, $p_+^* = \ln(\frac{p_+}{1-p_+})$, is used as a response variable in the regression model. Similarly, the other four statistics that take non-negative values, are transformed using the natural logarithmic function. Each transformed statistic, y^* say, is initially plotted against the altitude, x, of the ground in metres, for a visual inspection of the underlying pattern. For most months, the properties that show a linear relationship with the ground elevation are the mean, $\hat{\mu}^* = \ln(\hat{\mu})$, and the probability, p_+^*, while there seems to be no correlation between the other three statistics and the elevation. So, a regression line, $y^* = \alpha + \beta\, x$, is fitted to the transformed mean and probability, and the results for June and December are presented here.

With reference to Figure 2.7 and Tables 2.1 – 2.2, we notice that the regression lines fitted to the data have statistically significant slopes and explain between about 30% and 50% of the total variation between the gauges, of each one of these statistics. So, in describing the behaviour of the rainfall process at several points in space, one should keep in mind that part of the spatial variation of the process is due to the variation in altitude. The decision, however, of whether to include the elevation of a site in the model building, should be made only with reference to a specific model. In many cases, the simplicity of a model is more desirable than the incorporation of a factor that is not absolutely necessary.

An alternative way of illustrating the relationships between the ground elevation and the mean rainfall intensity, $\hat{\mu}$, and the probability of a wet interval, p_+, is via scatter plots, where the range of values of each property is split into a number of levels, say five, and each level is represented by different colour. The topography of the Brue area is illustrated via this type of map. With reference to Figure 2.8, it is clear from the scatter plots of the four statistics, that the ones that look close to the pattern of the Brue topography are those of the mean, $\hat{\mu}$, in June 1994 and of the probability, p_+, in December 1994. Indeed, in the regression analysis we found that these two properties have heigher R^2 values than the others.

$$\hat{\mu}^* = -3.7822 + 0.0034\,x \qquad R^2 = 49.21\%$$

Predictor	Coef.	St.error	t-ratio	F-ratio
Constant	-3.7822	0.0554	-68.2211	41.67
Elevation	0.0034	0.0005	6.4552	

$$p_+^* = -3.0016 + 0.0014\,x \qquad R^2 = 29.45\%$$

Predictor	Coef.	St.error	t-ratio	F-ratio
Constant	-3.0016	0.0353	-85.1129	17.95
Elevation	0.0014	0.0003	4.2362	

Table 2.1: Regression statistics for the log of the mean and the logistic of the probability that an arbitrary interval is rainy, for June 1994.

$$\hat{\mu}^* = -1.9778 + 0.0014\,x \qquad R^2 = 36.79\%$$

Predictor	Coef.	St.error	t-ratio	F-ratio
Constant	-1.9778	0.0285	-69.3779	6.20
Elevation	0.0014	0.0003	5.1182	

$$p_+^* = -1.6862 + 0.0015\,x \qquad R^2 = 45.6\%$$

Predictor	Coef.	St.error	t-ratio	F-ratio
Constant	-1.6862	0.0255	-66.0585	36.91
Elevation	0.0015	0.0002	6.0753	

Table 2.2: Regression statistics for the log of the mean and the logistic of the probability that an arbitrary interval is rainy, for December 1994.
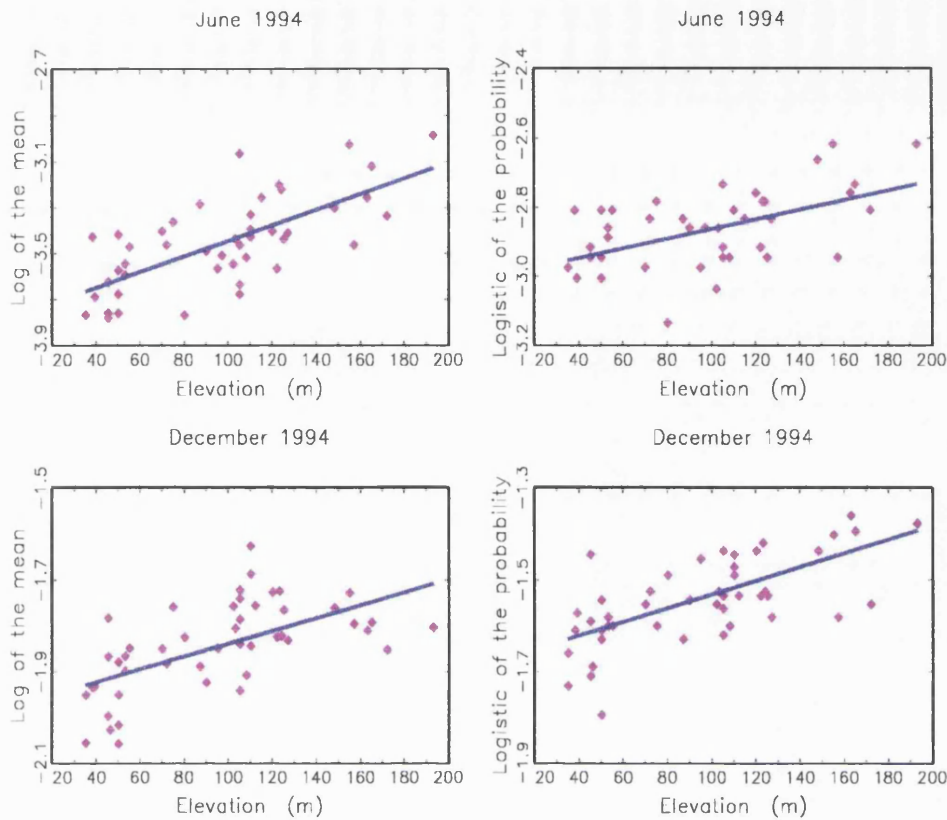
Figure 2.7: Plots and regression lines of the log mean rainfall intensity, $\hat{\mu}^*$, (left) and the transformed probability, p_+^*, of a wet interval (right) against the elevation of the sites for June 1994 (top) and December 1994 (bottom).

2.3.2 Spatial characteristics

Next, we investigate the spatial association of the rainfall process within the Brue catchment. Six transects are formed, each of 2 km width, that cover the main part of the Brue area, three having a North-South and three a East-West orientation, (see Figure 1.2) and they are examined separately. A few joint properties are calculated for all pairs of sites within each transect, such as the zero lag cross-correlation, $\gamma_{ij}(0)$, between the rainfall intensities at two sites, say i and j, and the probability, $p_{00}^{(ij)}(0)$, that these sites are simultaneously dry. In order to get smoothed curves, the inter-site distance is rounded to the nearest kilometer, and the derived value is an average, from the values of all pairs that have the same distance (in km).
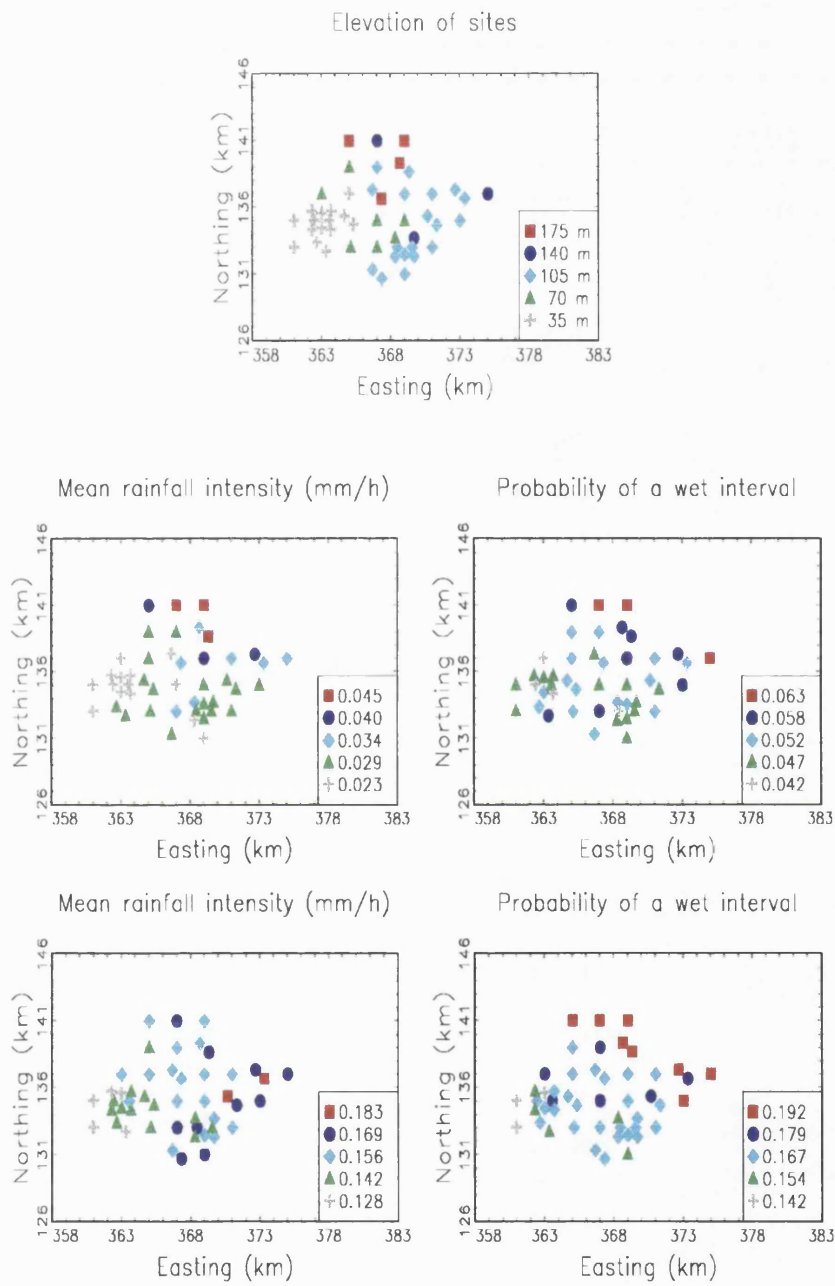
Figure 2.8: Topography of the Brue area (top), elevation effects for June 1994 (middle) and December 1994 (bottom) and hourly data. The number assossiated to each symbol is the minimum value of the corresponding level.
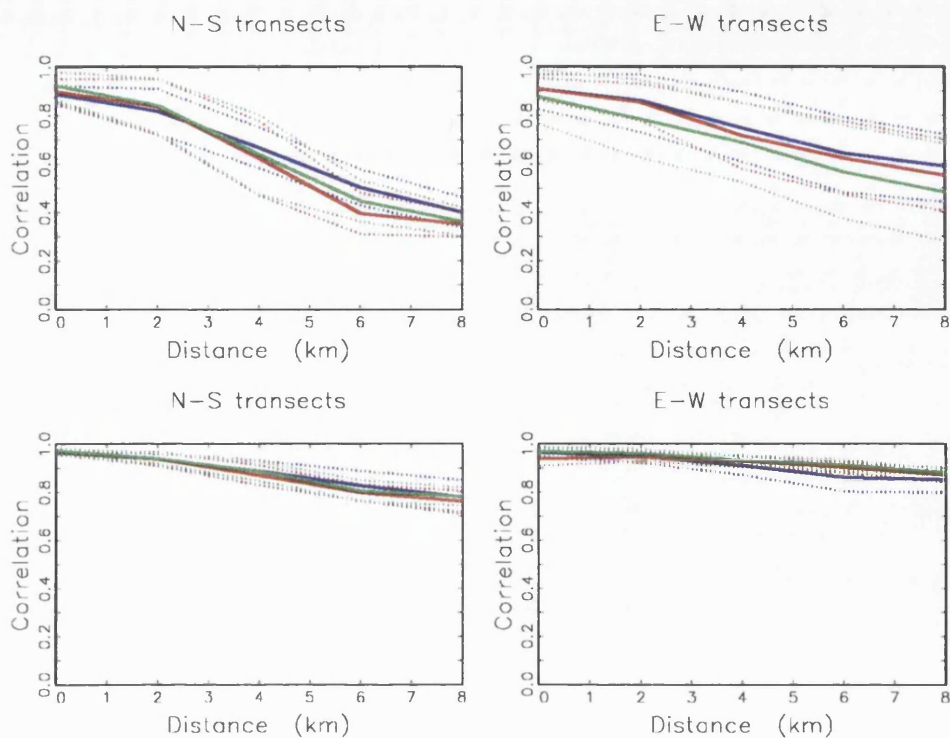
Figure 2.9: Lag zero average cross-correlation functions (solid lines) of the various transects and a confidence zone of ± one standard deviation (dotted lines) for June 1994 (top) and December 1994 (bottom).

The plots of $\gamma_{ij}(0)$ for June 1994 and December 1994, are illustrated in Figure 2.9. It might be worth clarifying that the value of $\gamma_{ij}(0)$ at zero distance is the mean cross-correlation for pairs that are less than 0.5 kilometer apart, which is why the value is a little less than one. For a given month, the cross-correlation functions that correspond to transects with the same orientation are very similar and show a decreasing spatial association with inter-site distance. A difference is observed between the quantities that correspond to North-South and East-West orientation. The former take lower values than the latter, which is evidence of the presence of a prevailing direction in the storm's movement. Also, the localised summer storms result in rapidly decaying and very variable functions, while in uniform winter systems the effect of the distance upon the spatial correlation is small. Comparable plots for the probability $p_{00}^{(ij)}(0)$ show no dependence

either on inter-site distance or on the orientation, reflecting the fact that the size of the Brue area is small compared to the size of a rainfall event and all sites experience an event almost simultaneously.

2.4   Conclusions

From the temporal properties of rainfall at a single site, only the ones that are related to the proportion of wet and dry time intervals exhibit a systematic seasonal pattern, while the cross-correlation measures, calculated at various levels of aggregation and time lags, give a lot of information about the structure of the rainfall process at small time scales, for the different seasons.

The analysis of the spatial association of rainfall has shown that the zero time lag cross-correlation functions decay with inter-site distance, with a slope that depends on the orientation of the pair of sites. The cross-correlation at non-zero time lags is also influenced by the distance and the orientation of the sites. This reflects the presence of a prevailing direction which should be taken into consideration in the multi-site model development. Also, the between-site properties depend on the orientation of the pairs of sites rather than their absolute location.

The investigation of the orographic effects of the Brue area on the behaviour of the rainfall process has revealed that the ground elevation explains about $30\% - 50\%$ of the variation in the mean rainfall intensity and the probability of a wet interval. Thus, the inclusion of this variable has the potential to improve the modelling of the process, even at this very local scale.

Chapter 3

Single-site models

3.1 Introduction

The temporal evolution of rainfall events at a fixed point in space has been extensively studied during the last few decades and several models based on the theory of stochastic point process have been developed. A few interesting results from research in this area have already been discussed in Section 1.2. One common feature of these models is that the random variables involved are mutually independent. However, the duration and intensity of rainfall are often inversely related. For example, rainfall is often classified into two broad categories, convective, which is very intense short-lived rain (e.g. summer thunderstorms), and non-convective, which is lighter and longer lasting (e.g. stratiform rain). Also, Waymire and Gupta (1981a) mention that cyclonic storms contain two distinct types of precipitation areas, namely the large mesoscale areas, which have large spatial extent, move slowly and have low intensity, and the small mesoscale areas, which cover a smaller region, move faster and have higher intensity compared to the former. So, if rainfall events are observed at a fixed point in space, it is thought that the components of the process with high intensity will tend to have short lifetimes and vice versa. One way to model the presence of distinct types of rainfall components, suggested by Cowpertwait (1994), is to have within the same storm many types of small scale elements, each with independent duration and intensity. This idea was applied using two types of rain cells, 'heavy' and 'light', with the former having generally shorter lifetimes and higher intensities than the latter. A disadvantage of this approach is that the number of

types in which rain cells are classified, and the ranges of values for the parameters of the duration and intensity of each type, have to be specified in advance by the researcher.

In this chapter we study the relationship between duration and intensity of rainfall. The stochastic models that we develop are based on the theory of clustered point processes and in particular on the Neyman-Scott and Bartlett-Lewis processes. In such models, the large-scale elements, called storms, are generators of the small-scale elements, called rain cells. At this point, it should be mentioned that the concept of rain cell is not precise. Although we regard it as a mathematical device in the clustering mechanism, rather than as a physical entity, it is desirable that the parameters that describe its characteristics have physical interpretations and take realistic values. In general, the ranges of the parameter estimates depend on the data resolution and the type of the model that is used (e.g. spatial-temporal, single-site, multi-site, etc.) and thus it is rather dangerous and arbitrary to have a fixed idea about the rain cell's features.

We shall follow the general approach and notation of Rodriguez-Iturbe, Cox, and Isham (1987a) (RICI), who discuss two models based on a stationary Poisson process of storm origins of rate λ. In each case, with each storm origin a random number of rain cells with independent rectangular profiles is associated. The duration, L, of a cell is exponential with parameter η and survivor function $\mathcal{F}_L$, and its intensity, X, is also random, independent of L. In the Neyman-Scott model the positions of the rain cells relative to the storm origin are determined by a set of independent and identically distributed random variables, and it is assumed that there is no cell in the storm origin. Natural candidates for the distribution of the distance between the cluster centre and its members are the Normal and the exponential distributions, the latter is used more often in rainfall modelling and in this particular model. The number of rain cells per storm, C, has either a Poisson or a geometric distribution. In the Bartlett-Lewis model, the cluster members follow the cluster centre in a finite renewal process. In the special version that was examined by RICI, the renewal process is Poisson with rate β, and terminates after a time that is exponentially distributed with parameter γ. In this model, it is assumed that there is a rain cell located at the storm origin. Under these assumptions, the number of rain cells per storm has a geometric distribution with mean $E(C) = 1 + \beta/\gamma$.

The total rainfall intensity, $Y(t)$, at a time t is the sum of the intensities of all rain

cells active at that time,

$$Y(t) = \int_0^\infty X_{t-u}(u)\, dN(t-u), \tag{3.1}$$

where $N(t)$ counts arrivals in the clustered point process of rain cells, and $X_{t-u}(u)$ is the intensity, at time t, of a rain cell having origin at $t - u$, so that

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

The mean of the total depth process is

$$E\{Y(t)\} = \lambda\, E(C)\, E(X)\, \eta^{-1},$$

and the autocovariance at lag τ ($\tau \geq 0$) is given by

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

The covariance of $dN(t)$ of a clustered point process, can be expressed (Cox and Isham, 1980, §2.5) in terms of the conditional intensity function, $h(\tau)$, as

$$c(\tau) = \lim_{d\tau, dt \to 0} \operatorname{cov}\{dN(t), dN(t+\tau)\} = \lambda\, E(C)\, \{\delta(\tau) + h(\tau) - \lambda\, E(C)\}, \tag{3.3}$$

where δ denotes the Dirac delta function. The function $h(\tau)$, which is the limiting rate at which points occur in $(\tau, \tau + d\tau)$ conditionally upon there being a point at the origin, is different for the two clustered point process. Specifically, (Cox and Isham, 1980, §3.4) for the Neyman-Scott process specified above, $h(\tau)$ is given by

$$\begin{aligned} h(\tau) &= \lambda\, E(C) + \frac{E\{C(C-1)\}}{E(C)} \int_0^\infty \beta^2\, e^{-\beta x}\, e^{-\beta(x+\tau)}\, dx \\ &= \lambda\, E(C) + \frac{E\{C(C-1)\}}{2\, E(C)}\, \beta\, e^{-\beta\tau}, \end{aligned}$$

and when this is substituted in (3.3), it results in

$$c(\tau) = \lambda\, E(C)\, \delta(\tau) + \frac{\lambda\, \beta}{2} E\{C(C-1)\}\, e^{-\beta\tau}. \tag{3.4}$$

In a Bartlett-Lewis process on the other hand, the conditional intensity function is

$$h(\tau) = \lambda\, E(C) + \beta\, e^{-\gamma\tau},$$

43

and from (3.3) we get

$$c(\tau) = \lambda\, E(C)\, \delta(\tau) + \lambda\, E(C)\, \beta\, e^{-\gamma\tau}. \tag{3.5}$$

It is then straightforward to derive the second-order properties of the total rainfall depth for the Neyman-Scott and the Bartlett-Lewis based models, after substitution of (3.4) and (3.5), respectively, into (3.2).

3.2 Theoretical model

3.2.1 Description

We shall now develop some clustered point process models for the temporal evolution of rainfall at a fixed spatial location, which also capture the inverse relationship between the intensity and duration of the phenomenon. From the descriptions of the Bartlett-Lewis and the Neyman-Scott models in the previous section, it is clear that the difference between the two concerns the precise way in which the cluster members are distributed in time, which is expressed through the covariance function $c(\tau)$. As far as the derivation of the second-order properties of the total rainfall intensity is concerned, the same procedure can be followed for both clustered processes. So far, comparisons between the two types of models have not provided significant evidence in favour of one or the other, and the choice between them is, most of the time, rather arbitrary. In the following description and in the model fitting, the Bartlett-Lewis based model is used, because the probability of zero rainfall during an arbitrary time interval of fixed length can be given algebraically in a closed form, in contrast with the Neyman-Scott based model where simulated data are needed to provide estimates of this property.

We consider a Bartlett-Lewis model, in which storm origins arrive in a Poisson process of rate λ and each is followed by a Poisson process of rain cell origins, of rate β, that terminates after a time that is exponentially distributed with rate γ. Each rain cell has a rectangular profile, with intensity, X, and duration, L, where now, these are assumed to be dependent random variables, with joint probability density function f_{XL}, independent of the Poisson process of storm origins. The marginal density functions of X and L are denoted by f_X and f_L, respectively. Distinct cells have mutually independent durations and intensities.

44

The total rainfall intensity at time t, $Y(t)$, is given by the integral representation

$$Y(t) = \int_0^\infty X_{t-u}(u) \, dN(t-u) \,, \tag{3.6}$$

where the intensity, at time t, of a rain cell of age u, $X_{t-u}(u)$, is now defined as

$$X_{t-u}(u) = \begin{cases} x & \text{with probability density } \mathcal{F}_L(x,u) \text{ for } x > 0, \\ 0 & \text{with probability } 1 - \int_0^\infty dx \, \mathcal{F}_L(x,u) \,, \end{cases} \tag{3.7}$$

where $\mathcal{F}_L(x,u) = \int_u^\infty dl \, f_{XL}(x,l)$ is the survivor function of the rain cell duration.

3.2.2 Second-order properties

In the present analysis, we are mainly interested in the second order properties of $Y(t)$. Higher order moments can be derived in a similar way. The expected value of the total rainfall intensity is:

$$
\begin{aligned}
E\{Y(t)\} &= \int_0^\infty E\{X_{t-u}(u) dN(t-u)\} \\
&= \lambda\, E(C) \int_0^\infty du \, E\{X_{t-u}(u)\} \\
&= \lambda\, E(C) \int_0^\infty du \int_0^\infty dx \, x \int_u^\infty dl \, f_{XL}(x,l) \\
&= \lambda\, E(C) \int_0^\infty dl \, l \, f_L(l) \, E_X[X \mid l] \tag{3.8} \\
&= \lambda\, E(C) \, E(XL), \tag{3.9}
\end{aligned}
$$

which is the product of the storm arrival rate, the mean number of rain cells per storm and the expected 'size' of a rain cell.

The lag τ autocovariance ($\tau \geq 0$) of the intensity function $Y(t)$ is derived by substituting (3.5) and (3.7) into (3.2), to give:

$$
\begin{aligned}
c_Y(\tau) &= \int_0^\infty \int_0^\infty E\{X_{t-u}(u) X_{t+\tau-v}(v)\} \mathrm{cov}\{dN(t-u), dN(t+\tau-v)\} \\
&= \lambda\, E(C) \int_0^\infty du \, E\{X_{t-u}(u) \, X_{t-u}(u+\tau)\} \\
&\quad + \lambda\, \beta\, E(C) \int_0^\infty du \int_0^\infty dv \, E\{X_{t-u}(u) X_{t+\tau-v}(v)\} \, e^{-\gamma|\tau+u-v|} \\
&= \lambda\, E(C) \int_0^\infty du \int_{u+\tau}^\infty dl \, E[X^2 \mid l] \, f_L(l) + \lambda\, \beta\, E(C) \\
&\quad \times \left\{ e^{-\gamma\tau} \int_0^\infty du \int_u^\infty dl \, E[X \mid l] \, f_L(l) \, e^{-\gamma u} \int_0^{u+\tau} dv \int_v^\infty dl' \, E[X' \mid l'] \, f_L(l') \, e^{\gamma v} \right. \\
&\quad \left. + e^{\gamma\tau} \int_0^\infty du \int_u^\infty dl \, E[X \mid l] \, f_L(l) \, e^{\gamma u} \int_{u+\tau}^\infty dv \int_v^\infty dl' \, E[X' \mid l'] \, f_L(l') \, e^{-\gamma v} \right\},
\end{aligned}
$$

where X and X' are independent and identically distributed random variables. The first term of $c_Y(\tau)$, gives the 'within cell' contribution to the covariance, while the second term gives the contribution from different rain cells of the same storm. Becausfse storms arrive and develop independently of each other, rainfall intensities due to rain cells that belong to different storms are uncorrelated.

After routine integration the 4-fold integrals are reduced to double integrals resulting in the following formula for the lag τ autocovariance of the total rainfall intensity,

$$
\begin{aligned}
c_Y(\tau) = \ & \lambda\, E(C) \int_\tau^\infty dl\, (l - \tau) E[X^2 \mid l] f_L(l) + \frac{\lambda\beta}{\gamma^2}\, E(C) \\
& \times \Big\{ 2\gamma\, D(0,\tau,l+\tau,0) + 2\gamma\, B(0,l+\tau,\infty,0) - 2\gamma\tau\, A(0,\tau,l+\tau,0) \\
& - e^{-\gamma\tau}\, A(0,0,\infty,0) + e^{\gamma\tau}\, A(0,\tau,\infty,\beta) - e^{-\gamma\tau}\, A(-\gamma,l+\tau,\infty,\gamma) \\
& + e^{-\gamma\tau}\, A(\gamma,0,\infty,0) - e^{-\gamma\tau}\, A(\gamma,0,l+\tau,-\gamma) + e^{-\gamma\tau}\, A(0,0,\tau,-\gamma) \Big\}, \quad (3.10)
\end{aligned}
$$

where

$$
A(\theta,a,b,\zeta) = \int_0^\infty dl \int_a^b dl'\, E[X \mid l]\, f_L(l)\, E[X' \mid l']\, f_L(l')\, e^{-\theta l}\, e^{-\zeta l'},
$$

and

$$
B(\theta,a,b,\zeta) = \int_0^\infty dl \int_a^b dl'\, E[X \mid l]\, f_L(l)\, E[X' \mid l']\, f_L(l')\, l\, e^{-\theta l}\, e^{-\zeta l'},
$$

and

$$
D(\theta,a,b,\zeta) = \int_0^\infty dl \int_a^b dl'\, E[X \mid l]\, f_L(l)\, E[X' \mid l']\, f_L(l')\, l'\, e^{-\theta l}\, e^{-\zeta l'}.
$$

The variance of the process $Y(t)$ is deduced from the covariance $c_Y(\tau)$, by setting τ equal to zero,

$$
\begin{aligned}
\mathrm{var}\{Y(t)\} = \ & \lambda\, E(C)\, E(X^2 L) + \frac{\lambda\beta}{\gamma} E(C)\Big\{ 2\gamma\, D(0,0,l,0) + 2\gamma\, B(0,l,\infty,0) \\
& + 2\, A(\gamma,0,\infty,0) - A(0,0,\infty,0) - A(\gamma,0,l,-\gamma) - A(-\gamma,l,\infty,\gamma) \Big\}. \quad (3.11)
\end{aligned}
$$

It is not difficult to verify that when X and L are independent, Equation (3.10) simplifies to the corresponding expression for the independent case (Rodriguez-Iturbe, Cox, and Isham, 1987a, Eq.4.12)

$$
c_Y(\tau) = \frac{\lambda}{\eta}\, E(C)\, E(X^2) + \frac{\lambda\,\beta\, E(C)\, E(X)^2}{\eta\,(\gamma^2 - \eta^2)}\{\gamma\, e^{-\eta\tau} - \eta\, e^{-\gamma\tau}\}.
$$

3.2.3 Further properties

Since rainfall data are usually observed in aggregated form, we consider the cumulative rainfall intensity in disjoint time intervals of fixed length h,

$$Y_{t_i}^{(h)} = \int_{(t_{i-1})h}^{t_i h} Y(u) du \ .$$

The expected value of $Y_{t_i}^{(h)}$ is easily derived from the continuous time process,

$$E\{Y_{t_i}^{(h)}\} = h \ E\{Y(t)\} = h \ \lambda \ E(C) \ E(XL) \ . \tag{3.12}$$

The second order properties of $Y_{t_i}^{(h)}$ are obtained using the expressions

$$\text{var}\{Y_{t_i}^{(h)}\} = 2 \int_0^h (h - \tau) \ c_Y(\tau) \ d\tau \ , \tag{3.13}$$

and for $k \geq 1$

$$\text{cov}\{Y_{t_i}^{(h)}, Y_{t_{i+k}}^{(h)}\} = \int_{-h}^h (h - |\tau|) \ c_Y(kh + \tau) \ d\tau \ , \tag{3.14}$$

where $c_Y(\tau)$ is the covariance of the continuous process, given in Equation (3.10) for the model with dependent rain cell duration and intensity.

Other properties of interest, such as the probability of zero rainfall intensity during an arbitrary time interval, the transition probabilities between dry and wet intervals, and the mean duration of a dry and a wet spell, are not functions of the rain cell intensity, and in particular are the same whether or not X and L are dependent. Expressions for such properties can be found in Rodriguez-Iturbe et al. (1987a) and Onof et al. (1994) for the Bartlett-Lewis based models and in Cowpertwait (1991) for the Neyman-Scott based models.

3.3 Exponential dependence

The motivation for the current research is to explore the behaviour of rainfall models where the intensity and duration of rain cells are negatively correlated. From the wide range of models that satisfy this requirement, we consider some that are reasonably plausible and mathematically tractable in order to be able to derive analytic expressions for the statistical properties of the process.

In the present study, the exponential distribution for the cell duration, L, with parameter η say, is retained since it has been used in the past to give satisfactory results.

Concerning the conditional distribution of the intensity X, it is assumed that it has an exponential form with mean depending on L, that is $E(X \mid L) = g(L)$. A first thought is to take an inverse function for the conditional mean, $g(L) = 1/L$. This choice however is mathematically inconvenient because when $g(L)$ is substituted in (3.10) it gives rise to incomplete Gamma functions that have to be integrated. A class of functions that is both realistic and mathematically tractable is $g(L) = f\ L^d\ e^{-cL}$, where f, c and d are non-negative scalars. Note that the second order properties involve X only through $E(XL)$ and $E(X^2L)$, so the explicit distributional form is unimportant once these moments are fixed. In what follows, we describe the cases $d = 0$ and $d = 1$. When d is fixed, the model is specified by six parameters, one extra compared to the original Bartlett-Lewis model.

3.3.1 Model with $E(X \mid L) = f\ e^{-cL}$

The mean total rainfall intensity is obtained by substituting $E(X \mid L) = fe^{-cL}$ and $f_L(l) = \eta e^{-\eta l}$ in equation (3.8), which gives

$$E\{Y(t)\} = \lambda\ E(C)\ \frac{f\ \eta}{(c+\eta)^2}. \tag{3.15}$$

We assume an exponential distribution for the cell intensity given the duration, thus $E(X^2|L) = 2f^2e^{-2cL}$, which is substituted in (3.10) and results in the following expression for the lag τ covariance of $Y(t)$:

$$c_Y(\tau) = \lambda\ E(C)\ f^2\ \frac{2\ \eta\ e^{-(2c+\eta)\tau}}{(2c+\eta)^2} + \lambda\ E(C)\ f^2\ \beta\ \eta^2\ \frac{\{\gamma\ e^{-(c+\eta)\tau} - (c+\eta)\ e^{-\gamma\tau}\}}{(c+\eta)^3\ \{\gamma^2 - (c+\eta)^2\}}. \tag{3.16}$$

The mean and second order properties for the aggregated rainfall process, $Y_{t_i}^{(h)}$, are derived by combining equations (3.13), (3.14) and (3.16) and the details are given in Appendix B.

The variables X and L are always negatively correlated (Figure 3.1), with correlation function

$$
\begin{aligned}
\mathrm{corr}(X, L) &= \frac{\mathrm{cov}(X, L)}{\sqrt{\mathrm{var}(X)\ \mathrm{var}(L)}} \\[2mm]
&= \frac{-\frac{c}{(c+\eta)^2}}{\sqrt{\frac{\eta\{c^2+(c+\eta)^2\}}{(2c+\eta)\ (c+\eta)^2}\ \frac{1}{\eta^2}}} \\[2mm]
&= \frac{-\zeta}{(\zeta+1)}\ \sqrt{\frac{2\zeta+1}{2\zeta^2+2\zeta+1}}, \tag{3.17}
\end{aligned}
$$

where $\zeta = c/\eta$. The correlation tends to zero as ζ tends to infinity, that is as either the mean cell duration, η^{-1}, tends to infinity, or the conditional cell intensity, $E(X|L)$, tends to zero.

The unconditional mean of X is $E(X) = f\eta/(c + \eta)$ and its marginal probability density function is

$$f_X(x) = \int_0^\infty f_L(l) \, f_{X|L}(x|l) \, dl = \int_0^\infty \frac{\eta}{f} \, \exp\{(c - \eta) \, l - f^{-1} \, e^{cl} \, x\} \, dl.$$

Some plots of f_X, using estimated parameter values after fitting the model to rainfall data, are presented in Figure 3.2 (at the end of this chapter). Compared to the exponential density function of the original model, f_X has a larger mode at the origin and decays faster, and thus generates cells with small intensity.
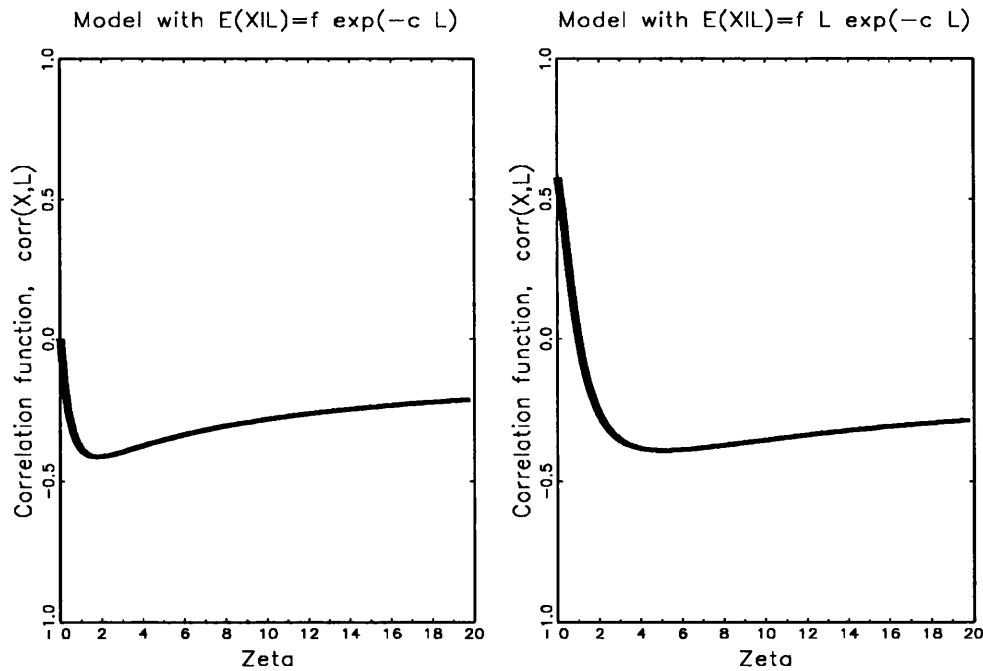
Figure 3.1: Correlation between rain cell intensity, X, and duration, L, as a function of $\zeta = c/\eta$.

49

3.3.2 Model with $E(X \mid L) = f\, L\, e^{-cL}$

The exponential distribution for the cell duration is retained in this model, so we have $E(XL) = 2\eta/(c+\eta)^3$ and thus the mean of $Y(t)$ is

$$E\{Y(t)\} = \lambda\, E(C)\, \frac{2\, f\, \eta}{(c+\eta)^3}. \tag{3.18}$$

Under these assumptions, the covariance function of the total rainfall intensity becomes

$$c_Y(\tau) = \frac{2\,\lambda\, E(C)\,\eta\, f^2}{(2c+\eta)^4}\{6 + 4(2c+\eta)\tau + (2c+\eta)^2\tau^2\}e^{-(2c+\eta)\tau} + \frac{\lambda\, E(C)\, f^2\,\beta\,\eta^2}{(c+\eta)^4\,\{(c+\eta)^2 - \gamma^2\}^2}$$
$$\times \left\{ \{4(c+\eta) - \gamma\}^2\, e^{-\gamma\tau} + \frac{5\gamma^2 - 11(c+\eta)^2 + 3\tau(c+\eta)\{\gamma^2 - (c+\eta)^2\}}{2(c+\eta)}\,\gamma\, e^{-(c+\eta)\tau} \right\}.$$

The mean and second order properties for the aggregated rainfall process, $Y_{t_i}^{(h)}$, are given in Appendix B.

The correlation function between rain cell intensity and duration is

$$\begin{aligned}
\operatorname{corr}(X, L) &= \frac{-\dfrac{c-\eta}{(c+\eta)^3}}{\sqrt{\left\{\dfrac{4\eta}{(2c+\eta)^3} - \dfrac{\eta^2}{(c+\eta)^4}\right\}\dfrac{1}{\eta^2}}} \\
&= \frac{(1-\zeta)}{(1+\zeta)}\sqrt{\frac{(2\,\zeta+1)^3}{4\,(\zeta+1)^4 - (2\,\zeta+1)^3}}. \tag{3.19}
\end{aligned}$$

When $\zeta = c/\eta > 1$, X and L are negatively correlated, while when $\zeta = 1$, they are uncorrelated but not independent. As with previous model, when ζ tends to infinity, the correlation between X and L tends to zero.

The unconditional variable X has mean $E(X) = f\eta/(c+\eta)^2$ and probability density

$$f_X(x) = \int_0^\infty f_L(l)\, f_{X\mid L}(x\mid l)\, dl = \int_0^\infty \frac{\eta}{f\, l}\, \exp\left\{(c-\eta)\, l - \frac{x\, e^{cl}}{f\, l}\right\}\, dl.$$

The density function f_X, plotted using some parameter estimates (Figure 3.2), has a mode at the origin higher than the exponential one in April and June and lower in December. This suggests that it is more likely to have cells with high intensities during rainy months and low intensities during summer.

3.4 Model fitting

We use two methods for fitting the models presented in the previous sections to raingauge data: a generalised method of moments and a spectral approach. Let $\theta = \{\theta_1, \ldots, \theta_m\}$

be the parameter vector of a model.

3.4.1 Generalised method of moments

Suppose that m distinct population moments of a random variable Y_t can be calculated as a function of θ, such as

$$E(Y_t^i) = \mu_i(\theta) \quad \text{for} \quad i = 1, 2, \ldots, m.$$

The classical method of moments estimate of θ is the value $\widehat{\theta}$ for which these population moments are equated to the observed sample moments; that is $\widehat{\theta}$ is the value for which

$$\mu_i(\widehat{\theta}) = \widehat{\mu}_i = \frac{1}{N} \sum_{t=1}^{N} y_t^i \quad \text{for} \quad i = 1, 2, \ldots, m,$$

where $\mathbf{y} = (y_1, \ldots, y_N)$ is a series of data.

In the generalised method of moments (GMM) (Hamilton, 1994) one might try to estimate θ so as to be as close as possible to $p \geq m$ moments, by minimising with respect to θ a criterion function

$$[\mathbf{g}(\theta; \mathbf{y})]' \, \mathbf{W} \, [\mathbf{g}(\theta; \mathbf{y})] \tag{3.20}$$

where $[\mathbf{g}(\theta; \mathbf{y})]' = [\{\mu_1(\theta) - \widehat{\mu}_1\} \ldots \{\mu_p(\theta) - \widehat{\mu}_p\}]$ and $\mathbf{W}$ is a $(p \times p)$ positive defined symmetric weighting matrix reflecting the importance given to matching each of the moments. The weighting matrix $\mathbf{W}$ may be a function of the data $\mathbf{y}$, and its optimal value is given by $\mathbf{S}^{-1}$, the inverse of the asymptotic variance-covariance matrix.

In order to estimate the parameter vector θ of the rainfall models discussed in this chapter, we use a version of the GMM, whose description is as follows. Suppose that a set $\Psi = \{\psi_1, \ldots, \psi_k\}$ of primary features of the data is selected, for which the corresponding theoretical expressions, $\Phi(\theta) = \{\phi_1(\theta), \ldots, \phi_k(\theta)\}$, have been derived. Typical examples of such features are the mean, the variance, the correlation and the probability of no rainfall, at various levels of aggregation. We select a number $p \geq m$ of the observed and the corresponding theoretical values of these features, to form the vector $[\mathbf{g}(\theta; \mathbf{y})]' = [\{\phi_1(\theta) - \psi_1\} \ldots \{\phi_p(\theta) - \psi_p\}]$. To avoid bias from large magnitude components, we set $\mathbf{W}$ to be a diagonal matrix, with its (j, j) element equal to $1/\psi_j^2$. The variance-covariance matrix $\mathbf{S}$ would require enormous algebraic effort to be derived, thus it is not being used in this application. So, the objective function that is minimised

over the parameter space is the weighted sum of squares of the differences between the observed and the theoretical model properties,

$$\sum_{j=1}^{p} \left(\frac{\phi_j(\theta)}{\psi_j} - 1 \right)^2.$$

(3.21)

The goodness of fit is assessed by comparing observed and fitted values of the complete set of features in Ψ.

A key advantage of the GMM is that it requires specification only of certain moment conditions rather than the full density function. This can also be a drawback, in that GMM often does not make efficient use of all the information in the sample. In addition, the estimate $\widehat{\theta}$ depends on the features used in the fitting procedure, and indeed can vary greatly. It has been observed (e.g. Rodriguez-Iturbe, Cox, and Isham (1988), Onof (1992)) that when this version of GMM is used in fitting a five or six parameter point process model for rainfall, an appropriate set of six primary features that gives good estimates includes the mean, correlation lag 1, variance, and probability of dry interval at 1 hour level of aggregation, and the last two properties at 24 hours level of aggregation. These conclusions have been confirmed using the dependent intensity-duration models, but the details will not be given here. These six properties were used to obtain results via the GMM fitting that are reported in the rest of this chapter. In general, the primary features used in the fitting should be chosen so as to capture the small scale, within storm structure of the process as well as the large scale, between storms' dependencies. For instance, the correlation of the rainfall intensity provides information about the temporal distribution of rain cells and the patchiness of the storm, since it has contributions from rain cells that belong to the same storm. So, this property should be used at small levels of aggregation. Summary statistics calculated from daily data, such as the probability that an arbitrary daily interval is dry and the variance of the rainfall intensity, are important in estimating the storm arrival rate.

3.4.2 Spectral method

The spectral method of parameter estimation is based on Whittle's method for time series, and has recently been applied (Chandler, 1997) in fitting point process models to rainfall data. One of the main advantages of this approach is that it makes use of the

whole series of data and, specifically, of its Fourier transform. The assumptions necessary in order to apply the spectral method are that the stochastic process Y_t is stationary with mean $\mu_Y(\boldsymbol{\theta})$, and that in any realisation of Y_t, observations that are far apart will be approximately independent. Under these assumptions, the second-order spectral density of the process exists and is defined by

$$h_Y(\omega; \boldsymbol{\theta}) = \int_{-\infty}^{\infty} c_Y(\tau; \boldsymbol{\theta}) \, e^{-i\omega\tau} \, d\tau \qquad \omega \in (-\pi/\Delta, \pi/\Delta) \tag{3.22}$$

where $c_Y(.; \boldsymbol{\theta})$ is the covariance function of the rainfall model with, now, its dependence on $\boldsymbol{\theta}$ made explicit and Δ is the level of aggregation. Given a series of data $y_1, \ldots, y_N$, the sample Fourier coefficients for every possible frequency ω can be calculated and combined to form the periodogram $I(\omega)$. If only a single realisation of the process is available, then the series of data can be split into a number, S, of segments long enough to be treated as (approximately) independent realisations (Chandler, 1997). The fragmentation of the time series can also be applied with multiple realisations of the process, and the advantage is that the final periodogram is smoother than the original and thus the computational time required for the convergence of the optimisation procedure is reduced substantially. When the series of data give a total of S segments, we use the mean of the S periodograms, $\bar{I}(\omega_{S_p})$, evaluated at frequencies

$$\omega_{S_p} = \frac{2\pi p S}{N\Delta} \qquad p = 0, 1, \ldots, [N/2S].$$

The approximate log-likelihood function, which is maximised to give an estimate $\widehat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$, is written as

$$\ln L_S(\boldsymbol{\theta}) = -\frac{S}{2}(\ln 2 + \ln h_Y(0; \boldsymbol{\theta})) - S \sum_{p=1}^{[N/2S]} \left(\frac{\bar{I}(\omega_{S_P})}{4\pi h_Y(\omega_{S_P}; \boldsymbol{\theta})} + \ln h_Y(\omega_{S_P}; \boldsymbol{\theta}) \right)$$
$$+ S\left(\left[\frac{N}{2S}\right] + \frac{1}{2} \right) \ln \left(\frac{[N/S]}{8\pi^2} \right) - \frac{[N/S]}{4\pi h_Y(0; \boldsymbol{\theta})} \sum_{s=1}^{S} (\bar{y}_s - \mu_Y(\boldsymbol{\theta}))^2$$

where $\bar{y}_s$ is the sample mean rainfall intensity from the S segments, and with the following correction term added whenever $[N/S]$ is even,

$$\frac{S}{2} \left(\ln h_Y(\pi/\Delta; \boldsymbol{\theta}) - \ln \left(\frac{[N/S]}{8\pi^2} \right) \right).$$

When the spectral method for parameter estimation is applied, one way of comparing different models is to use the Akaike Information Criterion (AIC) (Akaike, 1974) which

53

is given by

$$\text{AIC}(\widehat{\theta}) = -2\ln L_S(\widehat{\theta}) + 2m$$

where m is the number of parameters being estimated. The model with the lowest AIC value is judged as providing the best fit to the data. In addition, it is possible to assess the adequacy of fit by evaluating the set of features in $\Phi(\widehat{\theta})$ and comparing them with the corresponding observed values of Ψ, which in our application is more useful than the AIC.

The second-order spectral densities of the rainfall models that were introduced in the previous section are derived using (3.22) and the corresponding covariance functions. So, the spectrum of the model with the conditional mean of X given by fe^{-cL} is

$$h_Y(\omega) = \frac{2\,\lambda\,\eta\,f^2\,(1+\frac{\beta}{\gamma})}{\pi\,(2c+\eta)\,\{(2c+\eta)^2+\omega^2\}} + \frac{\lambda\,\beta\,\gamma\,\eta^2\,f^2(1+\frac{\beta}{\gamma})}{\pi(c+\eta)^2\,(\gamma^2+\omega^2)\,\{(c+\eta)^2+\omega^2\}}. \qquad (3.23)$$

For the model with $E(X\mid L) = fLe^{-cL}$, we have

$$h_Y(\omega) = \frac{4\,\lambda\,\eta\,f^2\,(1+\frac{\beta}{\gamma})\,\{6\,(2c+\eta)^4 + 3\,(2c+\eta)^2\,\omega^2 + \omega^4\}}{\pi\,(2c+\eta)^3\,\{(2c+\eta)^2+\omega^2\}^3}$$
$$+\frac{2\,\lambda\,\beta\,\gamma\,\eta^2\,f^2\,(1+\frac{\beta}{\gamma})\,\{4\,(c+\eta)^2+\omega^2\}}{\pi\,(\gamma^2+\omega^2)\,(c+\eta)^4\,\{(c+\eta)^2+\omega^2\}^2}. \qquad (3.24)$$

3.5 Discussion of the results

Tha data available have been collected at a network of raingauge stations, located in the South-West England, during the period January 1994-December 1996. Since the models described in this chapter are built for the analysis of the temporal evolution of rainfall events at a fixed point in space, in the fitting we use data from a single site and we present here the results for gauge no. 37.

One requirement is that the time series of rainfall intensities does not have any trends (including seasonal trends), since all models assume a stationary point process. On the other hand, if a very short run of data is used, say weekly, that satisfies temporal homogeneity, it is very unlikely that any general conclusions can be drawn about the models' behaviour, since the results will not be representative. In fitting rainfall models it is common to examine each month of the year separately, and to consider the time series of a particular month from all annual sequences, say all Januaries, as independent

realisations of the same process. Then, the observed properties are calculated as averages of the corresponding values obtained from each individual month. In this case, it is possible to get sets of typical parameter values for every month that can be used in other hydrological studies. Of course, one hopes to have as much data as possible but even with three years rainfall sequences, as in the present study, a natural grouping is via months. In this section, we present and discuss the fitting results for December, April and June data sets, which are representative winter, spring and summer months, respectively. The first part of the discussion concerns the comparison between the methods used for the estimation of the models' parameters, namely the generalised method of moments and the spectral method.

During the model fitting procedure, we have introduced two dimensionless parameters $\kappa = \beta/\eta$ and $\phi = \gamma/\eta$. Also, when the rain cell duration parameter η becomes random (Rodriguez-Iturbe, Cox, and Isham, 1988), it is assumed that it follows a gamma distribution with index parameter α and scale parameter ν.

The spectral method has been successfully used in the past to provide parameter estimates of point process models (Chandler, 1997). It also provides criteria for comparison between models with different number of parameters. However, when this approach is used in fitting rainfall models, it fails to give close estimates of the probability that an arbitrary time interval of fixed length is dry, which is a property of great importance for many hydrological applications. Also up to now, this method can not be applied to fit the Bartlett-Lewis model with random cell duration parameter, denoted by $\mathcal{M}_2$, but research is currently in progress (R.Chandler, personal communication) in order to derive the loglikelihood function of this model. With reference to Table 3.1, that gives the fitting results for the December data set, we notice that the second order-properties of the data are predicted fairly accurately by the original Bartlett-Lewis model, denoted by $\mathcal{M}_1$ (Rodriguez-Iturbe et al., 1987a), and the one with dependent rain cell duration and intensity of the type $E(X|l) = fe^{-cl}$, denoted by $\mathcal{M}_3$. In contrast, the estimated values of the probability of zero rainfall are often far from the observed ones. The AIC value for these models are very close, suggesting very similar performances concerning the first and second order properties. In comparison, the model with $E(X|l) = fle^{-cl}$, denoted by $\mathcal{M}_4$, has a relatively high AIC value, indicating a poor performance, which is also

evident from the estimated properties. Similar results are obtained when the spectral method is applied to other data sets, but are not presented here.

With the GMM on the other hand, because the probability of zero rainfall of hourly and daily data is used during the parameter estimation procedure, the estimates are 'forced' to give good fitted values for these properties. So, the question in this case is whether a model is able to reproduce the observed properties at levels other than the ones used in the fitting. The results for months December, April and June are presented in Tables 3.2, 3.3 and 3.4, respectively.

Starting with model $\mathcal{M}_4$, with dependent rain cell intensity and duration of the form $E(X|l) = fle^{-cl}$, we notice that it produces very poor estimates of the lag one correlation of hourly data, and more specifically, this property is overestimated by $10 - 33\%$, although it is included in the objective function (Equation (3.21)). In addition, the lag one correlation is overestimated at higher levels of aggregation and, the probability of zero rainfall at an arbitrary time interval of length 24 hours is underestimated up to about 10%. These results give a rather disappointing overall picture of model $\mathcal{M}_4$, thus we exclude it from the remaining of this discussion.

Having in mind that the rectangular pulse rainfall model with random cell duration, denoted by $\mathcal{M}_2$ (Rodriguez-Iturbe et al., 1988), improved the performance of the original model, $\mathcal{M}_1$, at the expense of one extra parameter, it is interesting to see how model $\mathcal{M}_3$ performs as compared to $\mathcal{M}_1$ and $\mathcal{M}_2$. With reference to Table 3.4 that gives the results for month June, we notice that although all three models produce similar values for the first and second order properties, $\mathcal{M}_2$ and $\mathcal{M}_3$ give better estimates of the probability of zero rainfall as compared to $\mathcal{M}_1$, with the former slightly underestimating and the latter slightly overestimating this property. Similar conclusions are drawn from fitting these rainfall models to data from April, presented in Table 3.3.

The results from the December data set, shown in Table 3.2, suggest that the probability of zero rainfall, at levels of aggregation not used during the parameter estimation, is predicted more accurately by model $\mathcal{M}_2$ than $\mathcal{M}_3$, with the latter slightly underestimating this property. As before, the shape of the correlation functions of models $\mathcal{M}_2$ and $\mathcal{M}_3$ are very similar and closer to the observed, as compared to $\mathcal{M}_1$.

So, from the assessment of fit we conclude that models $\mathcal{M}_2$ and $\mathcal{M}_3$, defined by six

parameters, are more successful, as compared to the original five parameter model $\mathcal{M}_1$, in reproducing observed rainfall properties at a range of time scales, especially concerning the probability of zero rainfall during an arbitrary time interval.

Now, for a discussion about the parameter estimates, we concentrate on the three rectangular pulse rainfall models $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$, whose performances are satisfactory. The values of parameters λ, $E(C)$ and η show that model $\mathcal{M}_2$ generates many storms, in some cases nearly twice as many compared to $\mathcal{M}_3$, each storm consists of many rain cells that have high intensity and extremely short average lifetime, ranging from about 25 minutes in June, to 8.4 minutes in December. The mean storm duration for $\mathcal{M}_2$ is less than 3 hours, and the parameters suggest that storm lifetime is longer during summer than during winter months, which is the opposite of what we expected.

On the other hand, model $\mathcal{M}_3$ generates a small number of storms, with rain cells lasting several hours. Storms have an average lifetime ranging from about 7 to 12 hours, and generate more rain cells in December than in June. It is quite interesting to note that the correlation, ρ, between the rain cell duration and intensity for this model is always negative with value close to -0.38.

Finally, the original Bartlett-Lewis model, $\mathcal{M}_1$, generates storms with mean duration about 10 hours for December, and 4 hours for June and rain cells with mean duration about 40 minutes. The number of cells per storm in the summer months is double the number we get in winter, but the rain cell intensity is lower in the former than in the latter case.

On the basis of this analysis, we conclude that model $\mathcal{M}_3$, in which the mean rain cell intensity is given by $E(X|l) = fe^{-cl}$, provides an alternative way of improving the prediction of the probability of zero rainfall of the original Bartlett-Lewis model $\mathcal{M}_1$, and also captures the inverse relationship between rain cell intensity and duration. The results of the fitting are encouraging but before drawing more general conclusions, the model should be tested on rainfall data collected at other geographical regions that have weather types different from that of South-West England, and also to longer sequences of rainfall data. A disadvantage of model $\mathcal{M}_3$ is that the expressions for the second-order propeties are rather complicated, so it is unlikely that it can be used in the study of rainfall process at several spatial locations.

Fitting results using the **spectral method** for parameter estimation.

Table 3.1: Observed and estimated properties for December rainfall data.

level of aggregation		Mean	Probability of no rain	Variance	Correlation		
					lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	0.168	0.808	0.363	0.487	0.249	0.172
	$\mathcal{M}_1$	0.168	0.704	0.301	0.404	0.203	0.181
	$\mathcal{M}_3$	0.168	0.834	0.365	0.471	0.250	0.186
	$\mathcal{M}_4$	0.136	0.873	0.298	0.713	0.120	0.037
6 h	$\mathcal{H}$	1.007	0.558	5.282	0.312	0.166	0.120
	$\mathcal{M}_1$	1.006	0.366	4.145	0.432	0.319	0.246
	$\mathcal{M}_3$	1.010	0.666	5.347	0.272	0.059	0.013
	$\mathcal{M}_4$	0.813	0.671	3.284	0.115	0.002	0.000
12 h	$\mathcal{H}$	2.014	0.397	13.856	0.276	0.128	0.100
	$\mathcal{M}_1$	2.013	0.254	11.873	0.460	0.270	0.161
	$\mathcal{M}_3$	2.020	0.527	13.601	0.158	0.008	0.000
	$\mathcal{M}_4$	1.626	0.491	7.094	0.055	0.000	0.000
24 h	$\mathcal{H}$	4.029	0.222	37.004	0.223	0.070	0.131
	$\mathcal{M}_1$	4.026	0.148	34.668	0.397	0.140	0.049
	$\mathcal{M}_3$	4.041	0.329	31.505	0.075	0.000	0.000
	$\mathcal{M}_4$	3.252	0.262	14.737	0.026	0.000	0.000

$\mathcal{H}$: Observed properties,
$\mathcal{M}_1$: Original Bartlett-Lewis model,
$\mathcal{M}_3$: Bartlett-Lewis model with $E(X|l) = fe^{-cl}$,
$\mathcal{M}_4$: Bartlett-Lewis model with $E(X|l) = fle^{-cl}$.

Model's parameters						Derived parameters				
	λ	κ	ϕ	η	$E(X)$		β	γ	$E(C)$	
$\mathcal{M}_1$	0.043	0.100	0.017	2.555	1.436		0.255	0.043	6.894	

	λ	κ	ϕ	η	c	f	β	γ	$E(C)$	$E(X)$	ρ
$\mathcal{M}_3$	0.039	0.945	0.279	0.885	0.802	3.158	0.836	0.247	4.384	1.657	-0.378

	λ	κ	ϕ	η	c	f	β	γ	$E(C)$	$E(X)$	ρ
$\mathcal{M}_4$	0.052	0.055	0.889	0.609	1.400	16.292	0.034	0.541	1.062	4.936	-0.302

Model	# of param.	AIC value
$\mathcal{M}_1$	5	-11675.8
$\mathcal{M}_3$	6	-11673.8
$\mathcal{M}_4$	6	-11509.6

Fitting results using the **generalised method of moments**.

Table 3.2: Observed and estimated properties for December rainfall data.

level of aggregation		Mean	Probability of no rain	Variance	Correlation lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	**0.168**	**0.808**	**0.363**	**0.487**	0.249	0.172
	$\mathcal{M}_1$	**0.167**	**0.791**	**0.377**	**0.508**	0.232	0.162
	$\mathcal{M}_2$	**0.168**	**0.808**	**0.363**	**0.487**	0.282	0.209
	$\mathcal{M}_3$	**0.168**	**0.808**	**0.363**	**0.487**	0.240	0.185
	$\mathcal{M}_4$	**0.164**	**0.813**	**0.394**	**0.603**	0.168	0.139
6 h	$\mathcal{H}$	1.007	0.558	5.282	0.312	0.166	0.120
	$\mathcal{M}_1$	1.004	0.546	5.534	0.310	0.149	0.081
	$\mathcal{M}_2$	1.007	0.600	5.562	0.326	0.135	0.078
	$\mathcal{M}_3$	1.007	0.548	5.381	0.345	0.161	0.081
	$\mathcal{M}_4$	0.986	0.575	4.982	0.368	0.226	0.155
12 h	$\mathcal{H}$	2.014	0.397	13.856	0.276	0.128	0.100
	$\mathcal{M}_1$	2.008	0.410	14.497	0.263	0.074	0.022
	$\mathcal{M}_2$	2.015	0.431	14.756	0.254	0.082	0.043
	$\mathcal{M}_3$	2.015	0.379	14.474	0.278	0.068	0.017
	$\mathcal{M}_4$	1.972	0.394	13.440	0.362	0.164	0.077
24 h	$\mathcal{H}$	4.029	**0.222**	**37.004**	0.223	0.070	0.131
	$\mathcal{M}_1$	4.016	**0.240**	**36.624**	0.171	0.015	0.001
	$\mathcal{M}_2$	4.030	**0.222**	**37.004**	0.183	0.046	0.022
	$\mathcal{M}_3$	4.030	**0.222**	**37.004**	0.169	0.011	0.001
	$\mathcal{M}_4$	3.943	**0.198**	**36.418**	0.283	0.062	0.014

The properties in **bold** have been used in the fitting procedure.

$\mathcal{H}$: Observed properties,
$\mathcal{M}_1$: Original Bartlett-Lewis model,
$\mathcal{M}_2$: Bartlett-Lewis model with random cell duration,
$\mathcal{M}_3$: Bartlett-Lewis model with $E(X|l) = fe^{-cl}$,
$\mathcal{M}_4$: Bartlett-Lewis model with $E(X|l) = fle^{-cl}$.

Model's parameters						Derived parameters				
	λ	κ	ϕ	η	$E(X)$	β	γ	$E(C)$		
$\mathcal{M}_1$	0.044	0.198	0.065	1.556	1.450	0.308	0.101	4.040		

	λ	κ	ϕ	α	ν	$E(X)$	$E(\beta)$	$E(\gamma)$	$E(C)$	$E(\eta)$
$\mathcal{M}_2$	0.055	0.439	0.062	2.821	0.394	1.746	3.142	0.444	8.068	7.162

	λ	κ	ϕ	η	c	f	β	γ	$E(C)$	$E(X)$	ρ
$\mathcal{M}_3$	0.030	3.238	0.592	0.194	0.834	4.695	0.627	0.115	6.465	0.885	-0.368

	λ	κ	ϕ	η	c	f	β	γ	$E(C)$	$E(X)$	ρ
$\mathcal{M}_4$	0.053	0.053	0.109	0.575	1.933	28.549	0.031	0.063	1.487	2.609	-0.370

Table 3.3: Observed and estimated properties for April rainfall data.

level of aggregation		Mean	Probability of no rain	Variance	Correlation		
					lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	**0.088**	**0.871**	**0.134**	**0.526**	0.297	0.251
	$\mathcal{M}_1$	**0.088**	**0.860**	**0.146**	**0.564**	0.272	0.176
	$\mathcal{M}_2$	**0.088**	**0.871**	**0.134**	**0.526**	0.320	0.242
	$\mathcal{M}_3$	**0.088**	**0.871**	**0.134**	**0.527**	0.283	0.220
	$\mathcal{M}_4$	**0.084**	**0.874**	**0.149**	**0.708**	0.212	0.159
6 h	$\mathcal{H}$	0.527	0.707	2.095	0.394	0.129	0.071
	$\mathcal{M}_1$	0.525	0.696	2.289	0.296	0.130	0.068
	$\mathcal{M}_2$	0.527	0.737	2.195	0.346	0.133	0.069
	$\mathcal{M}_3$	0.528	0.693	2.140	0.362	0.152	0.068
	$\mathcal{M}_4$	0.505	0.710	2.045	0.376	0.209	0.132
12 h	$\mathcal{H}$	1.054	0.583	6.272	0.219	0.034	0.046
	$\mathcal{M}_1$	1.050	0.588	5.934	0.241	0.061	0.017
	$\mathcal{M}_2$	1.055	0.610	5.909	0.253	0.067	0.030
	$\mathcal{M}_3$	1.056	0.570	5.830	0.270	0.052	0.010
	$\mathcal{M}_4$	1.010	0.570	5.535	0.342	0.130	0.052
24 h	$\mathcal{H}$	2.109	**0.417**	**14.809**	0.181	0.091	0.073
	$\mathcal{M}_1$	2.101	**0.426**	**14.727**	0.153	0.011	0.001
	$\mathcal{M}_2$	2.110	**0.417**	**14.809**	0.166	0.029	0.011
	$\mathcal{M}_3$	2.111	**0.417**	**14.802**	0.151	0.006	0.000
	$\mathcal{M}_4$	2.020	**0.382**	**14.765**	0.245	0.038	0.006

The properties in **bold** have been used in the fitting procedure.

$\mathcal{H}$: Observed properties,
$\mathcal{M}_1$: Original Bartlett-Lewis model,
$\mathcal{M}_2$: Bartlett-Lewis model with random cell duration,
$\mathcal{M}_3$: Bartlett-Lewis model with $E(X|l) = fe^{-cl}$,
$\mathcal{M}_4$: Bartlett-Lewis model with $E(X|l) = fle^{-cl}$.

	Model's parameters						Derived parameters				
	λ	κ	ϕ	η	$E(X)$		β	γ	$E(C)$		
$\mathcal{M}_1$	0.027	0.267	0.087	1.237	0.994		0.330	0.107	4.077		
	λ	κ	ϕ	α	ν	$E(X)$	$E(\beta)$	$E(\gamma)$	$E(C)$	$E(\eta)$	
$\mathcal{M}_2$	0.032	0.661	0.070	3.476	0.632	1.034	3.637	0.385	10.509	5.500	
	λ	κ	ϕ	η	c	f	β	γ	$E(C)$	$E(X)$	ρ
$\mathcal{M}_3$	0.021	2.911	0.553	0.243	0.769	2.863	0.707	0.134	6.261	0.687	-0.393
	λ	κ	ϕ	η	c	f	β	γ	$E(C)$	$E(X)$	ρ
$\mathcal{M}_4$	0.031	0.086	0.143	0.535	1.589	14.957	0.046	0.076	1.602	1.775	-0.353

Table 3.4: Observed and estimated properties for June rainfall data.

level of aggregation		Mean	Probability of no rain	Variance	Correlation lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	**0.035**	**0.957**	**0.071**	**0.643**	0.476	0.339
	$\mathcal{M}_1$	**0.035**	**0.955**	**0.074**	**0.684**	0.446	0.326
	$\mathcal{M}_2$	**0.035**	**0.957**	**0.071**	**0.643**	0.402	0.295
	$\mathcal{M}_3$	**0.035**	**0.957**	**0.071**	**0.644**	0.373	0.265
	$\mathcal{M}_4$	**0.035**	**0.957**	**0.075**	**0.855**	0.417	0.296
6 h	$\mathcal{H}$	0.212	0.889	1.645	0.195	0.072	0.035
	$\mathcal{M}_1$	0.212	0.909	1.466	0.332	0.064	0.013
	$\mathcal{M}_2$	0.212	0.905	1.326	0.370	0.153	0.090
	$\mathcal{M}_3$	0.212	0.877	1.295	0.378	0.179	0.096
	$\mathcal{M}_4$	0.210	0.899	1.391	0.378	0.105	0.032
12 h	$\mathcal{H}$	0.423	0.836	3.547	0.204	0.084	0.054
	$\mathcal{M}_1$	0.423	0.859	3.906	0.177	0.007	0.000
	$\mathcal{M}_2$	0.424	0.852	3.632	0.280	0.093	0.049
	$\mathcal{M}_3$	0.424	0.826	3.568	0.302	0.082	0.024
	$\mathcal{M}_4$	0.419	0.845	3.788	0.227	0.020	0.002
24 h	$\mathcal{H}$	0.847	**0.756**	**9.297**	0.109	0.166	0.063
	$\mathcal{M}_1$	0.847	**0.766**	**9.195**	0.081	0.000	0.000
	$\mathcal{M}_2$	0.847	**0.756**	**9.297**	0.201	0.052	0.026
	$\mathcal{M}_3$	0.848	**0.759**	**9.289**	0.188	0.015	0.001
	$\mathcal{M}_4$	0.839	**0.747**	**9.255**	0.110	0.001	0.000

The properties in **bold** have been used in the fitting procedure.

$\mathcal{H}$: Observed properties,
$\mathcal{M}_1$: Original Bartlett-Lewis model,
$\mathcal{M}_2$: Bartlett-Lewis model with random cell duration,
$\mathcal{M}_3$: Bartlett-Lewis model with $E(X|l) = fe^{-cl}$,
$\mathcal{M}_4$: Bartlett-Lewis model with $E(X|l) = fle^{-cl}$.

|	Model's parameters						Derived parameters				
	λ	κ	ϕ	η	$E(X)$		β	γ	$E(C)$		
$\mathcal{M}_1$	0.009	1.254	0.163	1.659	0.708		2.081	0.270	8.711		
	λ	κ	ϕ	α	ν	$E(X)$	$E(\beta)$	$E(\gamma)$	$E(C)$	$E(\eta)$	
$\mathcal{M}_2$	0.010	0.623	0.152	2.799	1.174	1.068	1.485	0.362	5.091	2.383	
	λ	κ	ϕ	η	c	f	β	γ	$E(C)$	$E(X)$	ρ
$\mathcal{M}_3$	0.005	2.929	0.691	0.150	0.441	3.149	0.439	0.104	5.236	0.798	-0.398
	λ	κ	ϕ	η	c	f	β	γ	$E(C)$	$E(X)$	ρ
$\mathcal{M}_4$	0.010	0.401	0.348	0.572	1.181	7.437	0.229	0.199	2.153	1.384	-0.273
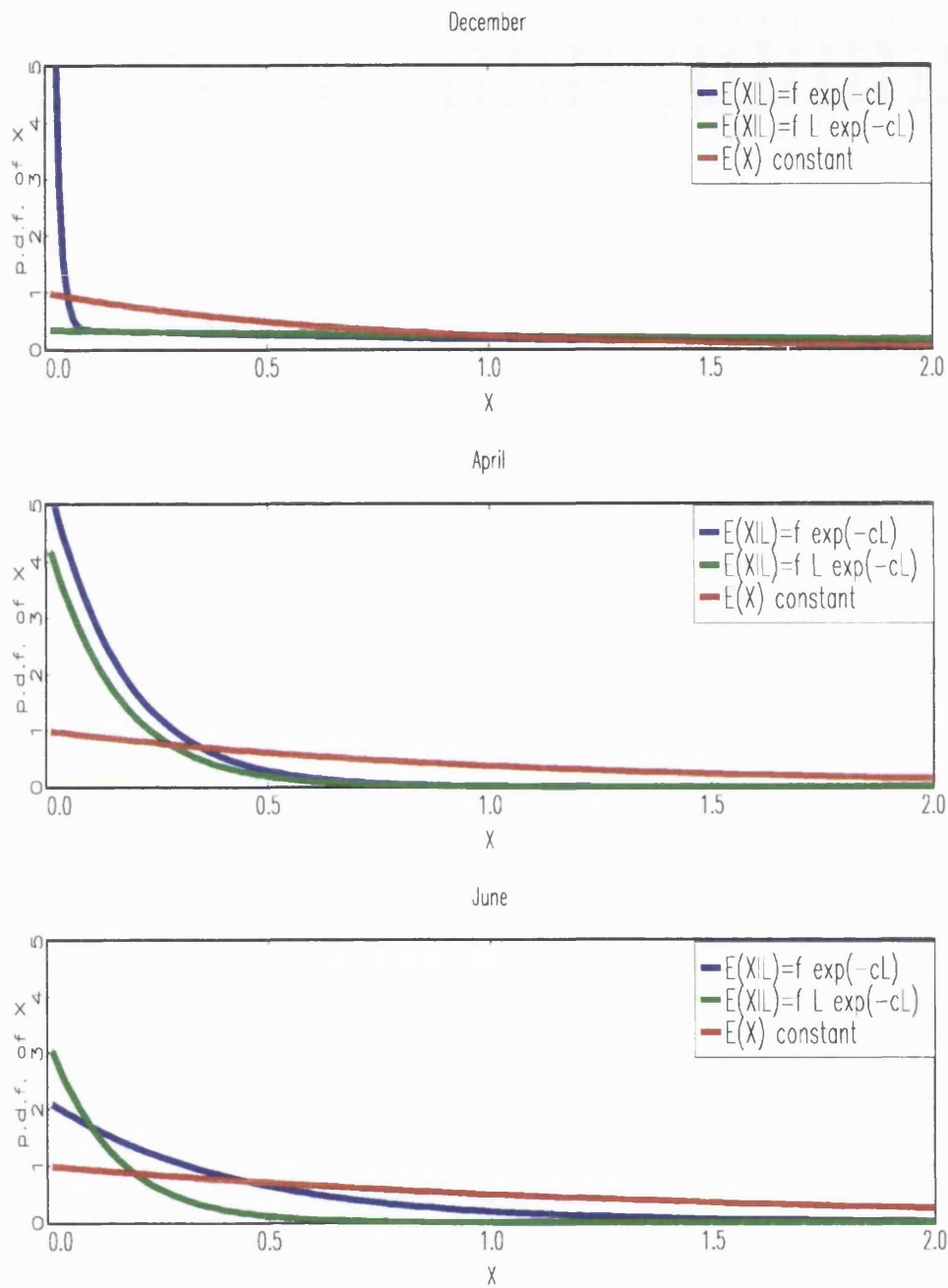
Figure 3.2: Marginal probability density function of rain cell intensity, X, for three rainfall models; two with dependent cell duration intensity of the form $E(X|l) = fe^{-cl}$ (blue) and $E(X|l) = fle^{-cl}$ (green) and one with exponential cell intensity and independent of the duration (red).

Chapter 4

Multi-site analysis of rainfall

4.1 Introduction

In regions where data are available at a small number of spatially separated points, it is useful to develop models that can be calibrated from a network of raingauges. In principle, it is possible to apply a full spatial-temporal model, such as the one developed by Northrop (1996), and to derive some properties of the model at a discrete collection of points in space. In practice however, this is complicated and an alternative approach is to develop models that express directly the between-site interactions, in a manner which reflects the underlying spatial-temporal structure of rainfall. These models are called *multi-site models*, and should have inter-site properties that depend on the distance between the examined sites and possibly on topography, and should preserve the structure of single-site models for their marginal processes at each site.

The models we consider are generalisations of those proposed by Cox and Isham (1994), whose approach is as follows: when rainfall is studied over a network of sites, a storm or a cell may be categorised according to the subset of sites which it affects. The basic idea is that storms arrive at the study area in a point process called the *master process*, and the type of the storm is determined by some random mechanism. In order to decide upon a suitable function for the probabilities of the different storm/cell types, we incorporate assumptions about the spatial structure of the rainfall event.

4.2 Fixed time displacement of clusters

We start by examining the evolution of rainfall at two sites, say i and j, from a network of n raingauge stations. In this analysis, we consider only storms that affect at least one of the two sites, and assume that these arrive in a stationary Poisson process in time, of rate λ. Associated with each storm origin is a random number, C, of rain cells that are independently and identically distributed around the storm origin, forming a temporal Neyman-Scott process. Each rain cell has a rectangular profile with random intensity, X, and exponential duration, L. The different intensities and durations are mutually independent and independent of all other variables.

Storms can be classified into three types or classes, say 'i', 'j' and 'ij', depending on the site(s) they affect. We define the probabilities:

$\tilde{p}_i = \mathrm{P}(\text{storm affects site } i \text{ only})$,

$\tilde{p}_j = \mathrm{P}(\text{storm affects site } j \text{ only})$,

$\tilde{p}_{ij} = \mathrm{P}(\text{storm affects both sites } i \text{ and } j) = 1 - \tilde{p}_i - \tilde{p}_j$,

and attach to each storm origin a multivariate random variable, M, called mark, independent of all other variables, to indicate its class, according to the probabilities $\tilde{p}_i$, $\tilde{p}_j$ and $\tilde{p}_{ij}$. The master point process of storm origins generates two subprocesses, one consists of the storms that affect site i and the other of these affecting site j. This is the result of the following two operations on point processes, the *decomposition* of the master process of storm origins into three subprocesses according to the value of the mark, M, and the *superposition* of the subprocesses of types 'i' and 'ij' and those of types 'j' and 'ij'. Since the original process of storm origins is Poisson, both subprocesses are also Poisson (in time), with rates $\bar{\lambda}_i = \lambda(\tilde{p}_i + \tilde{p}_{ij})$ and $\bar{\lambda}_j = \lambda(\tilde{p}_j + \tilde{p}_{ij})$, respectively. In addition, it is assumed that all rain cells within a storm belong to the same class as the storm itself. In order to allow the two sites to experience the same storm at different times, we introduce the displacement factors t_κ ($\kappa = 1, \ldots, n$) which are real scalars. So, for instance, storm and rain cell origins of the subprocess that corresponds to site κ, are translated in time by the amount t_κ, which is fixed for all storms and rain cells affecting that site. The following diagram illustrates the generation of the two subprocesses, each one corresponding to a different site. The large bullets in the diagram denote the storm origins and the small ones the rain cell origins.
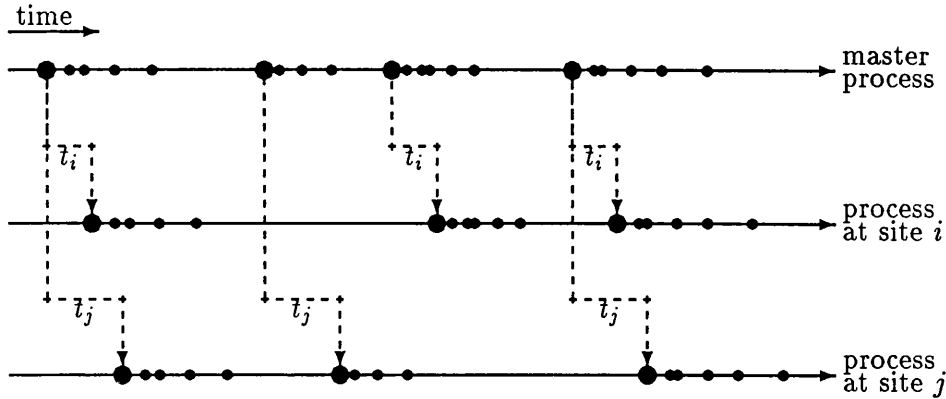
In order to introduce some variation between the rainfall intensities at different sites, we allow some of the model's parameters to vary between the generated subprocesses. In a catchment area as small as the Brue, where the largest distance between sites is about 12 km, it is reasonable to start with a model in which the storm-related parameters are the same for all sites, and to let only the rain cell characteristics vary between sites. Initially, in the master process we assume that the numbers of rain cells, C, per storm are independent and identically distributed random variables, regardless of storm type, and that the distances (in time) between the storm and rain cell origins are independent and exponential distributed with parameter β, common for all storm types. It is also assumed that there is no rain cell at the storm origin. Under this formulation, all affected sites have identical storm structures.

Concerning the rest of the model's parameters, let us suppose that the durations of rain cells experienced at site κ ($\kappa = 1,\ldots,n$), are scaled by a fixed positive factor l_κ. In other words, the duration of the same rain cell experienced at different sites is a scaled version of the same random variable. If the rain cell duration, L, in the master process is exponential with parameter η, then the duration at site κ is defined as $L_\kappa = l_\kappa L$, and is also exponential with parameter $\eta_\kappa = \eta/l_\kappa$. In order to avoid over-parameterisation and without loss of generality we set $\eta = 1$. Since the second order properties involve the rain cell intensity, X, only through the moments $E(X)$ and $E(X^2)$, it is not necessary to specify here its distributional form. In general, it is possible to apply a similar scaling mechanism to the intensities of a rain cell as experienced at different sites, but for simplicity we assume that these are the same at all affected sites. This

65

assumption can be easily removed if it is necessary to improve the model's performance.

The marginal properties of the above model at a single site are the same as those given in Section 3.1, so our interest in this chapter is focussed on the between-site properties, starting with the lag τ cross-covariance function, $c_{ij}(\tau)$, between the total rainfall intensities at site i at time t, $Y_i(t)$, and at site j at time $t + \tau$, $Y_j(t+\tau)$. Let us denote by $\widetilde{Y}_m(t)$ the rainfall intensity at time t, due to storms of type $m \in \{i, j, ij\}$. Since storms evolve independently, only rain cells that belong to the same storm contribute to the cross-covariance function, which for $\tau \geq 0$ is written as

$$
\begin{aligned}
c_{ij}(\tau) &= \text{cov}\{Y_i(t), Y_j(t+\tau)\} \\
&= \text{cov}\{\widetilde{Y}_i(t) + \widetilde{Y}_{ij}(t), \widetilde{Y}_j(t+\tau) + \widetilde{Y}_{ij}(t+\tau)\} \\
&= \text{cov}\{\widetilde{Y}_{ij}(t), \widetilde{Y}_{ij}(t+\tau)\} \\
&= \int_0^\infty \int_0^\infty E\{X_{t-u}^{(i)}(u) X_{t+\tau-v}^{(j)}(v)\} \, \text{cov}\{dN_{ij}^{(i)}(t-u), dN_{ij}^{(j)}(t+\tau-v)\} \quad (4.1)
\end{aligned}
$$

where for $\kappa = i, j$,

$$
X_{t-u}^{(\kappa)}(u) = \begin{cases} X & \text{if} \quad L_\kappa > u \\ 0 & \text{otherwise}, \end{cases}
$$

$$
dN_{ij}^{(\kappa)}(t) = \begin{cases} 1 & \text{if there is a cell origin of type '}ij\text{' at time } t \\ & \text{at the subprocess corresponding to site } \kappa, \\ 0 & \text{otherwise}. \end{cases}
$$

A cell origin at time t at site κ, corresponds to a point at time $t - t_\kappa$ in the master point process thus, the covariance of the counting measures $dN_{ij}^{(i)}(t)$ and $dN_{ij}^{(j)}(t+\tau)$ becomes

$$
\text{cov}\{dN_{ij}^{(i)}(t-u), dN_{ij}^{(j)}(t+\tau-v)\} = \lambda_{ij} E(C)\delta(u-v+\tau+t_{ij}) + \frac{\lambda_{ij}}{2} E(C^2-C)\beta e^{-\beta|u-v+\tau+t_{ij}|}
$$

$$
(4.2)
$$

where $\lambda_{ij} = \lambda \, \tilde{p}_{ij}$ and $t_{ij} = t_i - t_j$. After substituting (4.2) into (4.1) we get

$$
\begin{aligned}
c_{ij}(\tau) &= \lambda_{ij} E(C) \int_{\max\{0,-\tau-t_{ij}\}}^{\infty} du \, E(X^2) P(L_i > u, L_j > u + \tau + t_{ij}) \\
&+ \frac{\lambda_{ij}\beta}{2} E(C^2 - C) E(X)^2 \int_0^\infty du \int_0^\infty dv \, e^{-\beta|u-v+\tau+t_{ij}|} P(L_i > u) P(L_j > v)
\end{aligned}
$$

66

$$= \lambda_{ij} E(C) E(X^2) \int_{\max\{0,-\tau-t_{ij}\}}^{\infty} du\; e^{-\eta \max\{\frac{u}{l_i}, \frac{u+\tau+t_{ij}}{l_j}\}}$$

$$+ \frac{\lambda_{ij}\beta}{2} E(C^2 - C)E(X)^2 \left\{ \int_{\max\{0,-\tau-t_{ij}\}}^{\infty} du \int_{0}^{u+\tau+t_{ij}} dv\; e^{-\beta(u-v+\tau+t_{ij})} e^{-\frac{\eta}{l_i}u}\, e^{-\frac{\eta}{l_j}v} \right.$$

$$\left. + \int_{0}^{\infty} du \int_{\max\{0,u+\tau+t_{ij}\}}^{\infty} dv\; e^{\beta(u-v+\tau+t_{ij})}\, e^{-\frac{\eta}{l_i}u}\, e^{-\frac{\eta}{l_j}v} \right\}$$

$$= \lambda_{ij} E(C) E(X^2)\, \mathcal{I}_w + \frac{\lambda_{ij}\beta}{2} E(C^2 - C)E(X)^2\, \mathcal{I}_b. \tag{4.3}$$

The first term in the above expression gives the contribution from a single rain cell (within cell correlation), while the second term gives the contribution due to different rain cells of the same storm (between cells correlation).

The integral $\mathcal{I}_w$ in Equation (4.3) depends on the signs of $(l_i - l_j)$ and $(\tau + t_{ij})$, and in order to calculate it we distinguish four cases.

- If $l_i - l_j \geq 0$ and $\tau + t_{ij} \geq 0$ then,

$$\max\{\tfrac{u}{l_i}, \tfrac{u+\tau+t_{ij}}{l_j}\} = \tfrac{u+\tau+t_{ij}}{l_j}, \quad \text{for } u \geq 0$$

and

$$\mathcal{I}_w = \int_{0}^{\infty} du\; e^{-\eta(u+\tau+t_{ij})/l_j} = \frac{l_j}{\eta}\, e^{-\eta(\tau+t_{ij})/l_j} = \frac{e^{-\eta_j(\tau+t_{ij})}}{\eta_j}.$$

- If $l_i - l_j \geq 0$ and $\tau + t_{ij} < 0$ then,

$$\max\{\tfrac{u}{l_i}, \tfrac{u+\tau+t_{ij}}{l_j}\} = \tfrac{u+\tau+t_{ij}}{l_j}, \quad \text{for } u \geq \tfrac{l_i\,(\tau+t_{ij})}{(l_j-l_i)}$$

and

$$\mathcal{I}_w = \int_{-\tau-t_{ij}}^{\frac{l_i\,(\tau+t_{ij})}{(l_j-l_i)}} du\; e^{-\eta u/l_i} + \int_{\frac{l_i\,(\tau+t_{ij})}{(l_j-l_i)}}^{\infty} du\; e^{-\eta(u+\tau+t_{ij})/l_j}$$

$$= \frac{l_i}{\eta}\, e^{\eta(\tau+t_{ij})/l_i} - \frac{(l_i - l_j)}{\eta}\, e^{\eta(\tau+t_{ij})/(l_i-l_j)}$$

$$= \frac{e^{\eta_i(\tau+t_{ij})}}{\eta_i} - \frac{e^{\eta_{ij}(\tau+t_{ij})}}{\eta_{ij}},$$

where $\eta_{ij} = \frac{\eta}{l_i-l_j} = -\frac{\eta}{l_j-l_i} = -\eta_{ji}.$

- If $l_i - l_j \leq 0$ and $\tau + t_{ij} \geq 0$ then,

$$\max\{\tfrac{u}{l_i}, \tfrac{u+\tau+t_{ij}}{l_j}\} = \tfrac{u+\tau+t_{ij}}{l_j}, \quad \text{for } u \leq \tfrac{l_i\,(\tau+t_{ij})}{(l_j-l_i)}$$

and

$$
\begin{aligned}
\mathcal{I}_w &= \int_0^{\frac{l_i\,(\tau+t_{ij})}{(l_j-l_i)}} du\; e^{-\,\eta\,(u+\tau+t_{ij})/l_j} + \int_{\frac{l_i\,(\tau+t_{ij})}{(l_j-l_i)}}^{\infty} du\; e^{-\,\eta\,u/l_i} \\
&= \frac{l_j}{\eta}\, e^{-\eta(\tau+t_{ij})/l_j} - \frac{(l_j-l_i)}{\eta} e^{-\eta(\tau+t_{ij})/(l_j-l_i)} \\
&= \frac{e^{-\eta_j(\tau+t_{ij})}}{\eta_j} - \frac{e^{-\eta_{ji}(\tau+t_{ij})}}{\eta_{ji}}.
\end{aligned}
$$

- If $l_i - l_j \leq 0$ and $\tau + t_{ij} < 0$ then,

$$\max\{\tfrac{u}{l_i}, \tfrac{u+\tau+t_{ij}}{l_j}\} = \tfrac{u}{l_i} \quad \text{for } u \geq 0$$

and

$$\mathcal{I}_w = \int_{-\tau-t_{ij}}^{\infty} du\; e^{-\eta u/l_i} = \frac{l_i}{\eta}\, e^{\eta(\tau+t_{ij})/l_i} = \frac{e^{\eta_i(\tau+t_{ij})}}{\eta_i}.$$

To summarise the above results we write

$$
\int_{\max\{0,-\tau-t_{ij}\}}^{\infty} du\; e^{-\eta\max\{\frac{u}{l_i},\frac{u+\tau+t_{ij}}{l_j}\}} =
\begin{cases}
\dfrac{e^{\eta_i\,(\tau+t_{ij})}}{\eta_i} - \left(\dfrac{1}{\eta_{ij}}\right)^+ e^{\eta_{ij}\,(\tau+t_{ij})} & \text{if } \tau+t_{ij} < 0, \\[3mm]
\dfrac{e^{-\eta_j\,(\tau+t_{ij})}}{\eta_j} - \left(\dfrac{1}{\eta_{ji}}\right)^+ e^{-\eta_{ji}\,(\tau+t_{ij})} & \text{if } \tau+t_{ij} \geq 0,
\end{cases}
$$

where $x^+ = \max\{0,x\}$.

Similarly, $\mathcal{I}_b$ in Equation (4.3), depends on the sign of $(\tau+t_{ij})$, and so we distinguish two cases.

- If $\tau + t_{ij} \geq 0$ then $u + \tau + t_{ij} > 0$ for $u > 0$, and we get

$$
\begin{aligned}
\mathcal{I}_b &= \int_0^{\infty} du \int_0^{u+\tau+t_{ij}} dv\; e^{-\beta(u-v+\tau+t_{ij})}\, e^{-\eta_i u}\, e^{-\eta_j v} \\
&\quad + \int_0^{\infty} du \int_{u+\tau+t_{ij}}^{\infty} dv\; e^{\beta(u-v+\tau+t_{ij})}\, e^{-\eta_i u}\, e^{-\eta_j v} \\
&= \frac{e^{-\beta(\tau+t_{ij})}}{(\eta_i+\beta)\,(\eta_j-\beta)} + 2\beta\,\frac{e^{-\eta_j(\tau+t_{ij})}}{(\eta_i+\eta_j)\,(\beta^2-\eta_j^2)}.
\end{aligned}
$$

- If $\tau + t_{ij} < 0$, then,

$u + \tau + t_{ij} > 0$ for $u > -\tau - t_{ij}$, and we get

$$
\begin{aligned}
\mathcal{I}_b \;=\; & \int\limits_{0}^{-\tau-t_{ij}} du \int\limits_{0}^{\infty} dv \; e^{\beta(u-v+\tau+t_{ij})} \, e^{-\eta_i u} \, e^{-\eta_j v} \\[2mm]
& + \int\limits_{-\tau-t_{ij}}^{\infty} du \int\limits_{0}^{u+\tau+t_{ij}} dv \; e^{-\beta(u-v+\tau+t_{ij})} \, e^{-\eta_i u} \, e^{-\eta_j v} \\[2mm]
& + \int\limits_{-\tau-t_{ij}}^{\infty} du \int\limits_{u+\tau+t_{ij}}^{\infty} dv \; e^{\beta(u-v+\tau+t_{ij})} \, e^{-\eta_i u} \, e^{-\eta_j v} \\[2mm]
\;=\; & \frac{e^{\beta(\tau+t_{ij})}}{(\eta_j+\beta)\,(\eta_i-\beta)} + 2\beta \frac{e^{\eta_i(\tau+t_{ij})}}{(\eta_i+\eta_j)\,(\beta^2-\eta_i^2)} .
\end{aligned}
$$

Finally, the lag τ cross-covariance function $(\tau \geq 0)$ becomes

$$
c_{ij}(\tau) \;=\;
\begin{cases}
\lambda_{ij} E(C)E(X^2) \left\{ \dfrac{e^{\eta_i(\tau+t_{ij})}}{\eta_i} - (\tfrac{1}{\eta_{ij}})^+ e^{\eta_{ij}(\tau+t_{ij})} \right\} + \dfrac{\lambda_{ij}\beta}{2} E(C^2-C)E(X)^2 \\[3mm]
\quad \times \left\{ \dfrac{e^{\beta(\tau+t_{ij})}}{(\eta_j+\beta)\,(\eta_i-\beta)} + 2\beta \dfrac{e^{\eta_i(\tau+t_{ij})}}{(\eta_i+\eta_j)\,(\beta^2-\eta_i^2)} \right\} = c_{ij,1}(\tau) \qquad \text{if } \tau < -t_{ij} \\[6mm]
\lambda_{ij} E(C)E(X^2) \left\{ \dfrac{e^{-\eta_j(\tau+t_{ij})}}{\eta_j} - (\tfrac{1}{\eta_{ji}})^+ e^{-\eta_{ji}(\tau+t_{ij})} \right\} + \dfrac{\lambda_{ij}\beta}{2} E(C^2-C)E(X)^2 \\[3mm]
\quad \times \left\{ \dfrac{e^{-\beta(\tau+t_{ij})}}{(\eta_i+\beta)\,(\eta_j-\beta)} + 2\beta \dfrac{e^{-\eta_j(\tau+t_{ij})}}{(\eta_i+\eta_j)\,(\beta^2-\eta_j^2)} \right\} = c_{ij,2}(\tau) \qquad \text{if } \tau \geq -t_{ij} .
\end{cases}
$$

4.3 Aggregated process

Since rainfall data are available in aggregated form, we consider the cumulative rainfall totals in disjoint time intervals of fixed length h. Let $Y_{i,n}^{(h)} = \int_{(n-1)h}^{nh} Y_i(t)dt$ be the total rainfall intensity experienced at site i during the n-th interval of length h of the aggregated process. The lag k cross-covariance function $(k > 0)$, $C_{ij}^{(h)}(t_{ij}, k)$, of the rainfall intensity of the aggregated process, depends on the value of t_{ij} relative to $-h(k+1)$, $-hk$ and $-h(k-1)$. So, we take the value of t_{ij} to lie in each of the four intervals and calculate the cross-correlation in each case separately.

$$
C_{ij}^{(h)}(t_{ij}, k) = \text{cov} \left\{ \int\limits_{0}^{h} Y_i(t)dt, \int\limits_{kh}^{(k+1)h} Y_j(s)ds \right\} = \int\limits_{-h}^{h} c_{ij}(kh + \tau)(h - |\tau|)d\tau
$$

$$
= \begin{cases}
\int\limits_{-h}^{h} c_{ij,2}(kh+\tau)\,(h-|\tau|)d\tau = C_{ij,1}^{(h)}(t_{ij},k) & \text{if } -h(k-1) \le t_{ij}, \\[2ex]
\int\limits_{-h}^{-t_{ij}-kh} c_{ij,1}(kh+\tau)\,(h+\tau)d\tau + \int\limits_{-t_{ij}-kh}^{h} c_{ij,2}(kh+\tau)\,(h-|\tau|)d\tau = C_{ij,2}^{(h)}(t_{ij},k) \\[1ex]
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } -kh \le t_{ij} < -h(k-1), \\[2ex]
\int\limits_{-h}^{-t_{ij}-kh} c_{ij,1}(kh+\tau)\,(h-|\tau|)d\tau + \int\limits_{-t_{ij}-kh}^{h} c_{ij,2}(kh+\tau)\,(h-\tau)d\tau = C_{ij,3}^{(h)}(t_{ij},k) \\[1ex]
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } -h(k+1) < t_{ij} \le -kh, \\[2ex]
\int\limits_{-h}^{h} c_{ij,1}(kh+\tau)\,(h-|\tau|)d\tau = C_{ij,4}^{(h)}(t_{ji},k) & \text{if } t_{ij} \le -h(k+1),
\end{cases}
$$

$$
= \lambda_{ij}E(C)E(X^2)\,V_1(t_{ij},k) + \frac{\lambda_{ij}\beta}{2}E(C^2-C)E(X)^2\,V_2(t_{ij},k). \tag{4.4}
$$

The functions $V_1(t_{ij},k)$ and $V_2(t_{ij},k)$ are the results of the integration of the exponential functions that are involved in $c_{ij}(\tau)$, over the interval $[-h,h]$, for the four cases specified above and are derived as follows.

- If $\quad -h(k-1) \le t_{ij}, \quad$ then

$$
V_1(t_{ij},k) = \frac{W_1(\eta_j,t_{ij},k)}{\eta_j} - \left(\frac{1}{\eta_{ji}}\right)^+ W_1(\eta_{ji},t_{ij},k),
$$

$$
V_2(t_{ij},k) = \frac{W_1(\beta,t_{ij},k)}{(\eta_i+\beta)\,(\eta_j-\beta)} - \frac{2\beta W_1(\eta_j,t_{ij},k)}{(\eta_i+\eta_j)\,(\eta_j^2-\beta^2)},
$$

where

$$
W_1(\zeta,t_{ij},k) = \int\limits_{-h}^{h} e^{-\zeta|t_{ij}+kh+\tau|}(h-|\tau|)d\tau = \frac{1}{\zeta^2}e^{-\zeta|t_{ij}+kh|}(e^{-\zeta h}-2+e^{\zeta h}). \tag{4.5}
$$

- When $\quad -kh \le t_{ij} < -h(k-1), \quad$ then

$$
V_1(t_{ij},k) = \frac{W_2(\eta_i,t_{ij},k)}{\eta_i} + \frac{W_3(\eta_j,t_{ij},k)}{\eta_j} - \left(\frac{1}{\eta_{ij}}\right)^+ W_2(\eta_{ij},t_{ij},k) - \left(\frac{1}{\eta_{ji}}\right)^+ W_3(\eta_{ji},t_{ij},k),
$$

$$
V_2(t_{ij},k) = \frac{W_2(\beta,t_{ij},k)}{(\eta_j+\beta)(\eta_i-\beta)} + \frac{W_3(\beta,t_{ij},k)}{(\eta_i+\beta)(\eta_j-\beta)} - \frac{2\beta}{(\eta_i+\eta_j)}
$$
$$
\times \left\{ \frac{W_2(\eta_i,t_{ij},k)}{(\eta_i^2-\beta^2)} + \frac{W_3(\eta_j,t_{ij},k)}{(\eta_j^2-\beta^2)} \right\}.
$$

where

$$
W_2(\zeta,t_{ij},k) = \begin{cases}
\int\limits_{-h}^{-t_{ij}-kh} e^{\zeta(t_{ij}+kh+\tau)}\,(h+\tau)\,d\tau & \text{if } t_{ij}+kh \ge 0 \\[2ex]
\int\limits_{-t_{ij}-kh}^{h} e^{-\zeta(t_{ij}+kh+\tau)}\,(h-\tau)\,d\tau & \text{if } t_{ij}+kh \le 0
\end{cases}
$$

$$= \frac{1}{\zeta^2} \left\{ e^{-\zeta(h-|t_{ij}+kh|)} - 1 + \zeta(h- \mid t_{ij} + kh \mid) \right\}, \qquad (4.6)$$

$$W_3(\zeta, t_{ij}, k) = \begin{cases} \int\limits_{-h}^{-t_{ij}-kh} e^{\zeta(t_{ij}+kh+\tau)} \left(h - |\tau| \right) d\tau & \text{if } t_{ij} + kh \le 0 \\ \int\limits_{-t_{ij}-kh}^{h} e^{-\zeta(t_{ij}+kh+\tau)} \left(h - |\tau| \right) d\tau & \text{if } t_{ij} + kh \ge 0 \end{cases}$$

$$= \frac{1}{\zeta^2} \left\{ (e^{-\zeta h} - 2) e^{-\zeta|t_{ij}+kh|} + 1 + \zeta(h- \mid t_{ij} + kh \mid) \right\}. \qquad (4.7)$$

- When $\quad -h(k+1) < t_{ij} \le -kh \quad$ then

$$V_1(t_{ij}, k) = \frac{W_3(\eta_i, t_{ij}, k)}{\eta_i} + \frac{W_2(\eta_j, t_{ij}, k)}{\eta_j} - (\frac{1}{\eta_{ij}})^+ W_3(\eta_{ij}, t_{ij}, k) - (\frac{1}{\eta_{ji}})^+ W_2(\eta_{ji}, t_{ij}, k),$$

$$V_2(t_{ij}, k) = \frac{W_3(\beta, t_{ij}, k)}{(\eta_j + \beta)(\eta_i - \beta)} + \frac{W_2(\beta, t_{ij}, k)}{(\eta_i + \beta)(\eta_j - \beta)} - \frac{2\beta}{(\eta_i + \eta_j)}$$
$$\times \left\{ \frac{W_3(\eta_i, t_{ij}, k)}{(\eta_i^2 - \beta^2)} + \frac{W_2(\eta_j, t_{ij}, k)}{(\eta_j^2 - \beta^2)} \right\}.$$

- Finally, if $\quad t_{ij} \le -h(k+1), \quad$ then

$$V_1(t_{ij}, k) = \frac{W_1(\eta_i, t_{ij}, k)}{\eta_i} - (\frac{1}{\eta_{ij}})^+ W_1(\eta_{ij}, t_{ij}, k),$$

$$V_2(t_{ij}, k) = \frac{W_1(\beta, t_{ij}, k)}{(\eta_j + \beta)(\eta_i - \beta)} - \frac{2\beta W_1(\eta_i, t_{ij}, k)}{(\eta_i + \eta_j)(\eta_i^2 - \beta^2)}.$$

The lag zero cross-covariance function between the rainfall intensities at two sites of the aggregated process is obtained using the expression

$$C_{ij}^{(h)}(t_{ij}, 0) = \text{cov} \left\{ \int_0^h Y_i(t)dt, \int_0^h Y_j(s)ds \right\} = \int_0^h c_{ij}(\tau)(h - \tau)d\tau + \int_0^h c_{ji}(\tau)(h - \tau)d\tau,$$

which, after calculating the integrals for each of the four cases specified earlier, results in Equation (4.4) with $k = 0$.

It may be worth mentioning that if the cell duration is the same at all affected sites, $(l_\kappa = l, \forall \kappa)$, then $\eta_{ij}^{-1} = \eta_{ji}^{-1} = 0$ and thus the cross-covariance function $C_{ij}^{(h)}(t_{ij}, k)$, at time lags $k \ge 0$, simplifies to

$$C_{ij}^{(h)}(t_{ij}, k) = \frac{\lambda_{ij}}{\eta} E(C)E(X^2)W(\eta, t_{ij}, k) + \frac{\lambda_{ij}\beta}{2\eta(\eta^2 - \beta^2)} E(C^2 - C)E(X)^2$$
$$\times \{\eta W(\beta, t_{ij}, k) - \beta W(\eta, t_{ij}, k)\},$$

where

$$W(\zeta, t_{ij}, k) = \begin{cases} W_1(\zeta, t_{ij}, k) & \text{if } |\, t_{ij} + kh \,| \geq h, \\ W_2(\zeta, t_{ij}, k) + W_3(\zeta, t_{ij}, k) & \text{if } |\, t_{ij} + kh \,| < h. \end{cases}$$

The cross-correlation function between the rainfall intensities at sites i and j, is derived by dividing the covariance $C_{ij}^{(h)}(t_{ij}, k)$ by the square roots of the variances of the aggregated process of total rainfall intensity at the two sites. Storms at site i arrive at a rate $\tilde{\lambda}_i = \lambda(\tilde{p}_i + \tilde{p}_{ij})$, and the duration of a rain cell experienced at that site is scaled by l_i, giving an average value of η_i^{-1}. So, the variance of the total rainfall intensity of the aggregated process at site i, for this Neyman-Scott based rainfall model, is (Rodriguez-Iturbe et al., 1987a)

$$\begin{aligned} \text{var}\{Y_{i,n}^{(h)}\} &= \frac{\tilde{\lambda}_i}{\eta_i^3} 2E(C)E(X^2)\,(\eta_i h - 1 + e^{-\eta_i h}) + \frac{\tilde{\lambda}_i\,\beta\,E(X)^2}{2\eta_i\,(\beta^2 - \eta_i^2)} \\ &\quad \times E(C^2 - C)\,\{\beta\,(\eta_i h - 1 + e^{-\eta_i h}) - \eta_i\,(\beta h - 1 + e^{-\beta h})\}. \end{aligned} \quad (4.8)$$

Similarly, the variance of the aggregated rainfall intensity at site j, $\text{var}\{Y_{j,n}^{(h)}\}$, is obtained from (4.8) by replacing the parameters $\tilde{\lambda}_i$ and η_i by $\tilde{\lambda}_j = \lambda(\tilde{p}_j + \tilde{p}_{ij})$ and η_j, respectively.

4.4   Random displacement

In the previous section, a rainfall model is considered in which storm origins, the cluster centres, are classified according to which sites they affect, and are translated in time by fixed amounts t_κ ($\kappa = 1 \ldots n$) at the subprocess corresponding to site κ. So far, the time displacement factors are assumed to be independent of any locational characteristics of the examined sites and of the storm's structure in space. In this section, some of these rather unrealistic assumptions are removed and a parameterisation is made that allows the displacements to depend on some measure of distance between the examined sites, and to vary randomly between storms.

Initially, a two-dimensional Euclidean space is defined and the region of the network of raingauge stations is inscribed in a disc, $\mathcal{D}$, of (finite) radius R_D, and whose centre is taken to be the origin of the coordinate system. In the following analysis, each site is represented by its corresponding coordinates. As already mentioned, storm origins are assumed to occur in time according to a stationary Poisson process of rate λ. Only storms that affect parts of the disc, $\mathcal{D}$, are considered in the analysis and thus contribute to the

storm arrival rate. The value of λ is, in general, related to the size of the examined area, which is constant in this study.
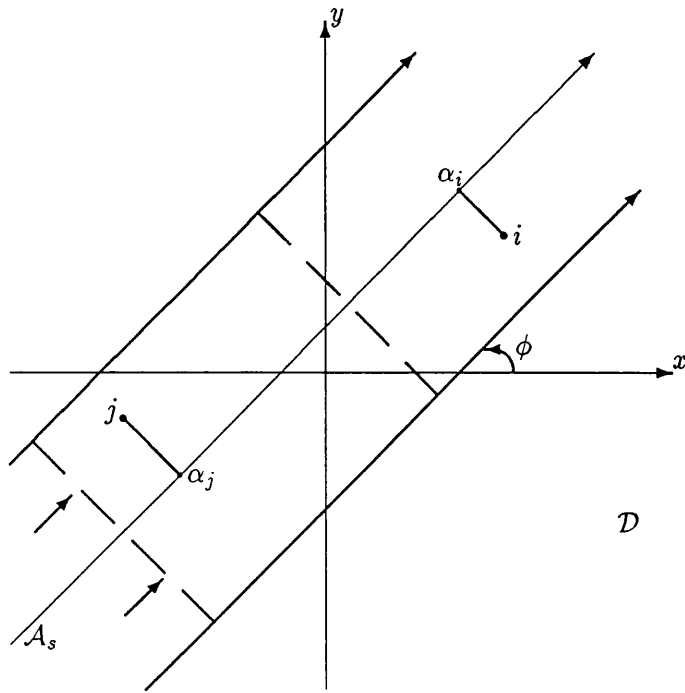
Storms and their associated rain cells are assumed to move across the disc, $\mathcal{D}$, with constant velocity, and affect a subset of sites of the network. The trace of a particular storm's movement over the catchment area is envisaged as a band of given width, location and orientation. The movement and the spatial extent of a rainfall band is illustrated schematically in the above diagram. The assumption that a storm and all its rain cells affect the same sites is kept, implying that all cluster members have essentially equivalent spatial extents, and thus their traces at the examined area coincide with that of the storm. Additionally, we assume that as the cluster moves along the band, it hits the disc, $\mathcal{D}$, in fronts that are perpendicular to the direction of its movement (dashed lines in the diagram). Locations within the disc $\mathcal{D}$ that are covered by the band, experience rainfall that has a duration and intensity specified by the rain cells' characteristics. Since the size of the catchment area is relative small, we assume that storms and rain cells do not start or terminate within the disc $\mathcal{D}$.

Suppose that the main axis, $\mathcal{A}_s$, of a storm's band has a fixed direction ϕ, which is the angle formed with the x-axis. Let the point with coordinates (x_i, y_i), which corresponds

to site i of the network, be projected on to the band's main axis, $\mathcal{A}_s$, at the point α_i. The difference of the displacement factors, $t_{ij} = t_i - t_j$, now has a physical interpretation, since it represents the lag between the times that rainfall elements reach the two sites i and j. Obviously, this time lag depends on the speed, $\mathcal{V}$, of the storm movement and on the distance, d_{ij}, between the points α_i and α_j, which in turns depend on ϕ. A realistic form of dependence is to have

$$t_{ij} = \frac{d_{ij}}{\mathcal{V}},$$

where, the signed distance d_{ij} may take any real value and is given by

$$d_{ij} = (x_i - x_j)\cos\phi + (y_i - y_j)\sin\phi. \tag{4.9}$$

It might be worth mentioning that conditioning on a storm affecting two sites i and j, the distance d_{ij} and consequently the time lag t_{ij} are independent of the width and the location of the band associated with that storm.

In a more realistic situation, one expects to have different storms moving with different speeds. We define the variable $\mathcal{W} = 1/\mathcal{V}$ and assume it follows a gamma distribution, with index θ and scale parameter ν. The density function of $\mathcal{W}$ is $f_{\mathcal{W}}(t) = \nu(\nu t)^{\theta-1}e^{-\nu t}/\Gamma(\theta)$, so the mean and variance of the speed, $\mathcal{V}$, are respectively

$$E(\mathcal{V}) = \frac{\nu}{(\theta - 1)} \quad \text{and} \quad \text{var}(\mathcal{V}) = \frac{\nu^2}{(\theta - 1)^2\,(\theta - 2)}.$$

Also, let all cells within a storm move with the same speed as the storm, which ensures that the displacement factor is the same for all members of a cluster. Since $\mathcal{W}$ varies between storms but remains constant throughout a storm's lifetime, the cross-covariance function of the aggregated process of the new model can be obtained by replacing in Equation (4.4) t_{ij} by $\mathcal{W}d_{ij}$ and taking the expected value of $C_{ij}^{(h)}(\mathcal{W}d_{ij}, k)$ with respect to $\mathcal{W}$,

$$\begin{aligned}
E_{\mathcal{W}}[C_{ij}^{(h)}(\mathcal{W}d_{ij}, k)] &= \lambda_{ij}\, E(C)\, E(X^2)\, E_{\mathcal{W}}[V_1(\mathcal{W}d_{ij}, k)] \\
&+ \frac{\lambda_{ij}\,\beta}{2}\, E(C^2 - C)\, E(X)^2\, E_{\mathcal{W}}[V_2(\mathcal{W}d_{ij}, k)]. \tag{4.10}
\end{aligned}$$

The details of this calculation are given in Appendix C.

So far we have allowed for the effect of the spatial separation between the sites on the time lag t_{ij}. But the expression of $E_{\mathcal{W}}[C_{ij}^{(h)}(\mathcal{W}d_{ij}, k)]$ also involves the probability, $\tilde{p}_{ij}$,

74

that a storm affects both sites i and j, which again depends on the distance between the sites and probably also on the storm's features. We follow two approaches in modelling these dependences, a deterministic and a stochastic one.

4.5 Empirical modelling of some between-site interactions

One way to express the dependence with distance of the probability that two sites are affected by the same storm or cell is to use a deterministic function that describes the underlying physical behaviour. Although such properties cannot be observed directly from the data, because rainfall elements may overlap in time and in space so that individual effects cannot be distinguished, it is reasonable to assume that the further away are two locations, the less likely it is that they get rain from the same element. Also, as the distance tends to zero, this joint probability should collapse to the marginal probability that a single location is affected by a storm or cell.

More specifically, suppose that storms affect part of the catchment area and all rain cells within a storm affect the same sites. An intuitively realistic form of dependence is to have

$$\tilde{p}_{ij} = \min\{p_i, p_j\}\, e^{-\kappa_s\, \tilde{d}_{ij}} + p_i\, p_j\, (1 - e^{-\kappa_s\, \tilde{d}_{ij}}), \tag{4.11}$$

where p_i = P(storm affects site i), $\quad p_j$ = P(storm affects site j), $\quad$ and $\tilde{d}_{ij}$ is the Euclidean distance between sites i and j. So, when the distance is small, the dominant part in Equation (4.11) is the first one, where it is expected that p_i and p_j are very similar, and that $\tilde{p}_{ij}$ is essentially equal to the probability that at least one of the two sites is affected by the storm. For large distances, it is assumed that a storm affects site i independently of site j, and thus $\tilde{p}_{ij}$ is given by the product of p_i and p_j.

It is possible to express the probability, $\tilde{q}_{ij}$, that a rain cell affects both sites i and j with a similar function,

$$\tilde{q}_{ij} = \min\{q_i, q_j\}\, e^{-\kappa_c\, \tilde{d}_{ij}} + q_i\, q_j\, (1 - e^{-\kappa_c\, \tilde{d}_{ij}}), \tag{4.12}$$

where q_i = P(cell affects site i), and q_j = P(cell affects site j). Then, we expect that in general $\kappa_s < \kappa_c$, in other words the probability $\tilde{q}_{ij}$ decays more rapidly with distance as compared to $\tilde{p}_{ij}$.

75

The advantages of this formulation are that it is both intuitively and mathematically simple, and it can be expanded easily to model the probability that a rainfall element affects more than two sites. Among the disadvantages is that it does not include any information about the spatial structure of the rainfall event. As a result, the probabilities p_i are model parameters that are estimated directly from the data. Thus, for every new site that is included in the analysis, one extra probability has to be estimated and so the dimensionality of the parameter space increases linearly with the number of the examined sites.

4.6 Random rainfall bands

In Section 4.4, it is assumed that each storm moves with random speed and the trace of its passage over the catchment area is a band of fixed width and location. In this section we allow some of the features of the rainfall bands to vary randomly between storms. As a result, the probability that a storm affects a subset of sites of the network is no longer a constant parameter but a random quantity. Furthermore, a substantial reduction on the total number of parameters of the model is achieved when the rainfall process is examined at more than two locations.

The problem of locating *at random* a line in a plane, or in a bounded region, is not a simple one. Kendall and Moran (1963) give a few paradoxes that arise when a probability measure is ascribed to geometrical objects, such as points, lines and rotations. The confusion is usually caused by the failure to distinguish the proper reference set. In general, the distribution of a geometric object is defined by determining first a system of coordinates which define the object uniquely, and then a probability distribution on the range of these coordinates. Kendall and Moran (1963) mention that if we write the equation of the line as

$$x \sin \phi - y \cos \phi = \alpha, \qquad (4.13)$$

then the differential element $d\alpha \, d\phi$ is the only measure that remains invariant under the group of all translations and rotations. In our application, the examined region is a disc of (finite) radius, R_D, and a line is specified by the angle, ϕ, it makes with a fixed axis, say the x-axis, and the distance, α, from the origin. Its equation is given by (4.13),

and the case where α and ϕ are independent random variables is now considered.

Suppose that the rainfall band associated with a storm has a width $2R_s$, a direction ϕ, and the signed distance between its main axis, $\mathcal{A}_s$, and the origin is α_s. The spatial extent and the location of the band is fully determined by ϕ, α_s and R_s, and in order to describe the stochastic nature of the rainfall process, these variables should be random. The values of α_s and R_s affect only the probability that a subset of sites is affected by a storm, while the direction of the storm's movement, ϕ, also affects the lags between the times that rainfall hits the various sites. The following diagram illustrates the spatial features of the rainfall band.
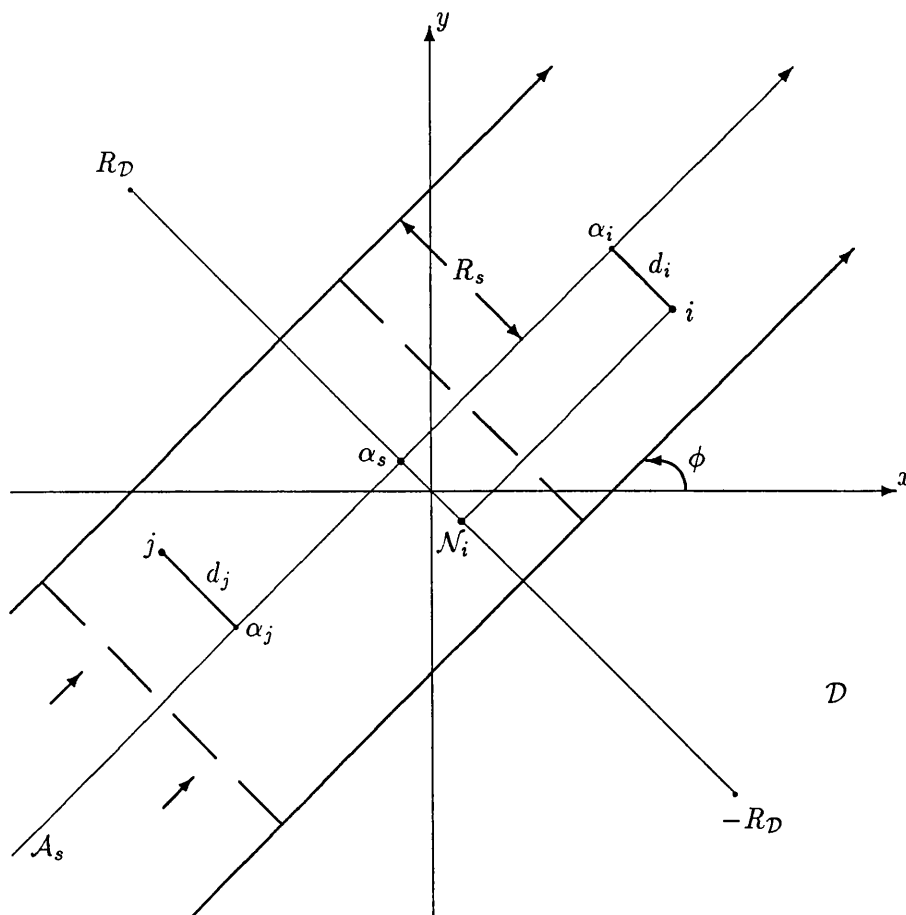
So, at this stage, only α_s and R_s are allowed to vary between storms, while the randomisation of ϕ is left to a later stage. Two natural candidates for the distribution of the width of a rainfall band are the gamma distribution and its special case, the exponential distribution. The latter is mathematically more tractable and in modelling

the spatial-temporal rainfall fields, when storms are envisaged as cylinders (Cox and Isham, 1988) it has been used for the distribution of storm radius. So, it is assumed here that R_s is exponentially distributed with parameter ρ_s, and with survivor function $\mathcal{F}_{R_s}$. A particular site experiences a storm if its distance from the storm's main axis is smaller than the value of R_s. The signed distance, d_i, between the axis $\mathcal{A}_s : y = x \, \tan \phi + \frac{\alpha_s}{\cos \phi}$, and site i with coordinates (x_i, y_i) is

$$
\begin{aligned}
d_i &= \alpha_s - y_i \cos \phi + x_i \sin \phi \\
&= \alpha_s - \mathcal{N}_i \, .
\end{aligned}
$$

For given values of α_s and ϕ, the probability, $p_i(\phi, \alpha_s)$, that a storm affects site i is

$$
p_i(\phi, \alpha_s) = \mathcal{F}_{R_s}(|d_i|) = \exp\{-\rho_s |\alpha_s - \mathcal{N}_i|\}, \tag{4.14}
$$

and so for the point process that corresponds to site i, the storm arrival rate is $\lambda \, p_i(\phi, \alpha_s)$.

Similarly, the probability, $p_{ij}(\phi, \alpha_s)$, that a single storm affects two particular sites i and j is expressed in terms of the survivor function $\mathcal{F}_{R_s}$. More specifically, the value of R_s should be greater than both absolute distances $|d_i|$ and $|d_j|$, in order that both sites experience the storm, that is

$$
p_{ij}(\phi, \alpha_s) = \mathcal{F}_{R_s}(\max\{|d_i|, |d_j|\}) = \exp\{-\rho_s \, \max\{|d_i|, |d_j|\}\}. \tag{4.15}
$$

In general, different storms may affect different parts of the area $\mathcal{D}$, so it is more realistic to let the position of a storm's rainfall band, specified by the signed distance α_s, be random. In the absence of any topographical effects to indicate a preferred route for the storms, (Section 2.3.2) we assume that α_s is uniformly distributed in $[-R_D, R_D]$.

Specifically, the probability $p_i(\phi, \alpha_s)$ in equation (4.14) is now replaced by its mean value, $p_i(\phi)$, given by

$$
\begin{aligned}
p_i(\phi) &= \frac{1}{2R_D} \int_{-R_D}^{R_D} \exp\{-\rho_s |\alpha_s - \mathcal{N}_i|\} \, d\alpha_s \\
&= \frac{1}{2R_D \rho_s} \Big\{ 2 - \exp\{-\rho_s (R_D + \mathcal{N}_i)\} - \exp\{-\rho_s (R_D - \mathcal{N}_i)\} \Big\}. \tag{4.16}
\end{aligned}
$$

Similarly, the expected value with respect to α_s, of the probability $p_{ij}(\phi, \alpha_s)$ that a rain cell affects two sites i and j is

$$p_{ij}(\phi) = \frac{1}{2R_D} \int\limits_{-R_D}^{R_D} \exp\{-\rho_s \, \max\{|d_i|, |d_j|\}\} \, d\alpha_s$$

$$= \frac{\exp\{-\frac{\rho_s}{2}|\mathcal{N}_i - \mathcal{N}_j|\}}{2 \, R_D \, \rho_s} \Big\{2 - \exp\{-\rho_s(R_D + \mathcal{N}_{ij})\} - \exp\{-\rho_s(R_D - \mathcal{N}_{ij})\}\Big\} \quad (4.17)$$

where $\mathcal{N}_{ij} = (\mathcal{N}_i + \mathcal{N}_j)/2$.

It is straightforward to obtain the cross-covariance function of the model with random rainfall bands from the expression of $E_{\mathcal{W}}[C_{ij}^{(h)}(\mathcal{W}d_{ij}, k)]$ given in Equation (4.10), by replacing parameter λ_{ij} by $\lambda p_{ij}(\phi)$, resulting in

$$\Sigma_{ij,s}^{(h)}(\phi, k) = \lambda \, p_{ij}(\phi) \, E(C) \, E(X^2) \, E_{\mathcal{W}}[V_1(\mathcal{W}d_{ij}, k)]$$

$$+ \frac{\lambda \, p_{ij}(\phi) \, \beta}{2} \, E(C^2 - C) \, E(X)^2 \, E_{\mathcal{W}}[V_2(\mathcal{W}d_{ij}, k)]. \quad (4.18)$$

In addition, the expression of the variance of the total rainfall intensity at a site, say i, is easily derived from Equation (4.8) by replacing $\tilde{\lambda}_i = \lambda(\tilde{p}_i + \tilde{p}_{ij})$ by $\lambda p_i(\phi)$.

4.7 Cell classification only

So far, it is has been assumed that different storms affect different parts of the catchment area and that all rain cells within a storm have the same spatial characteristics (location and spatial extent) as the storm, thus affecting the same sites. However, results from fitting the previous model to observed rainfall sequences in the Brue area, showed values close to one for both $p_i(\phi)$ and $p_{ij}(\phi)$, suggesting that once a storm hits this area, it affects almost all sites. Also, for that model the estimated cross-correlation function between the rainfall intensities at two sites takes higher values as compared to the observed ones, which implies that rain cells have smaller spatial extent than the storm, and each one affects a subset of the sites affected by the storm. Thus, a marking mechanism similar to the one discussed in Section 4.2, is applied to the rain cell origins rather than to the storm origins. So now, cell origins are classified into the three types 'i', 'j' and 'ij', depending on the site(s) they affect. We define the probabilities:

$\tilde{q}_i = $ P(cell affects site i **only**),

$\tilde{q}_j = $ P(cell affects site j **only**),

$\tilde{q}_{ij} = $ P(cell affects **both** sites i and j).

Initially, the master point process of cell origins is decomposed into three subprocesses,

according to the probabilities $\tilde{q}_i$, $\tilde{q}_j$ and $\tilde{q}_{ij}$. Then, the subprocesses of types 'i' and 'ij' and those of types 'j' and 'ij' are superposed to generate two processes, one for each affected site. The Poisson process of storm origins, of rate λ, is the same at all generated subprocesses. At this stage, we specify the distribution of the total number of rain cells per storm, C, and assume it is Poisson with mean μ_c. Thus, the number of rain cells per storm that affect site i is also Poisson with mean $\mu_c(\tilde{q}_i + \tilde{q}_{ij})$.

A schematic representation of the generation of the two subprocesses of rain cell origins from the master process is given in the following diagram, where the large bullet denotes the storm origin and the small ones the cell origins.
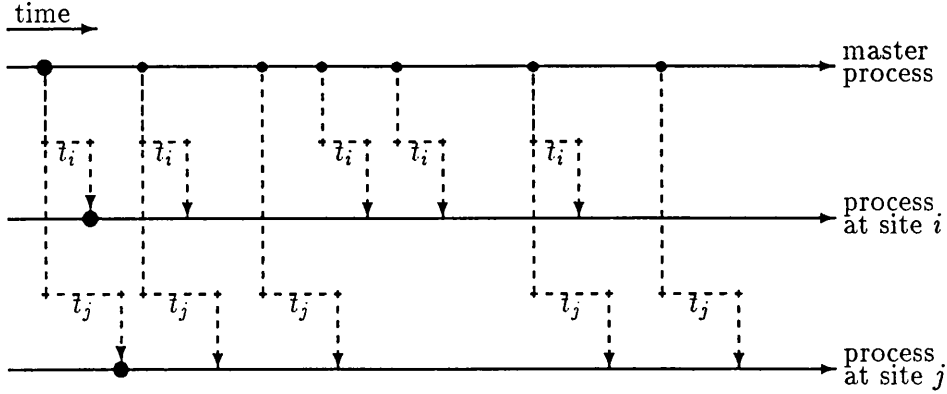
In addition to the above specifications of the model, suppose that each storm affects the whole catchment area, it moves with a speed, $\mathcal{V}$, that is constant during its lifetime but varies randomly between storms, and with a direction, ϕ, that is kept fixed at this stage. Each storm has a main axis, $\mathcal{A}_s$, whose signed distance from the origin is α_s. It is assumed that rain cells move with the same speed and direction as the storm, and associated with each one is a rainfall band whose width and position relative to $\mathcal{A}_s$ are random variables, and whose main axis is denoted by $\mathcal{A}_c$. These bands are related to the paths of the rain cells as they move across the disc $\mathcal{D}$. Suppose that the distance between the parallel axes $\mathcal{A}_c$ and $\mathcal{A}_s$ follows a two-sided exponential distribution with parameter β_c, that is, the density function of the signed distance between the cell axis and the origin is $f(w) = \frac{\beta_c}{2}e^{-\beta_c|\alpha_s - w|}$. Each cell band has a width $2R_c$, where R_c is exponential with mean ρ_c^{-1}.

For given values of ϕ and α_s, the probability, $q_i(\phi, \alpha_s)$, that a rain cell affects site i is

$$q_i(\phi, \alpha_s) = \frac{\beta_c}{2} \int_{-\infty}^{\infty} \exp\left\{-\rho_c|\alpha_s - \mathcal{N}_i - w|\right\} \exp\left\{-\beta_c|\alpha_s - w|\right\} dw \qquad (4.19)$$

$$= \frac{\beta_c}{\beta_c^2 - \rho_c^2}\left\{\beta_c \exp\{-\rho_c|\alpha_s - \mathcal{N}_i|\} - \rho_c \exp\{-\beta_c|\alpha_s - \mathcal{N}_i|\}\right\}, \qquad (4.20)$$

where $\mathcal{N}_i = y_i \cos\phi - x_i \sin\phi$. The first exponential term in (4.19) gives the probability that the absolute distance between site i, with coordinates (x_i, y_i), and $\mathcal{A}_c$ is less than R_c. Similarly, conditioning on the values of ϕ and α_s, the probability, $q_{ij}(\phi, \alpha_s)$, that a cell affects both sites i and j is

$$q_{ij}(\phi, \alpha_s) = \frac{\beta_c \exp\{-\frac{\rho_c}{2}|\mathcal{N}_i - \mathcal{N}_j|\}}{\beta_c^2 - \rho_c^2} \left\{\beta_c \exp\{-\rho_c|\alpha_s - \mathcal{N}_{ij}|\} - \rho_c \exp\{-\beta_c|\alpha_s - \mathcal{N}_{ij}|\}\right\}.$$
$$\qquad (4.21)$$

If now we let α_s be uniformly distributed in the interval $[-R_D, R_D]$, then the expected value of the probability $q_i(\phi, \alpha_s)$ is given by

$$q_i(\phi) = \frac{1}{2R_D} \int_{-R_D}^{R_D} q_i(\phi, \alpha_s) \, d\alpha_s$$

$$= \frac{\beta_c^2 \, \mathcal{K}(\mathcal{N}_i, \rho_c, R_D) - \rho_c^2 \, \mathcal{K}(\mathcal{N}_i, \beta_c, R_D)}{2 \, R_D \, \rho_c \, (\beta_c^2 - \rho_c^2)}, \qquad (4.22)$$

where

$$\mathcal{K}(\mathcal{N}, \zeta, R_D) = 2 - \exp\{-\zeta(R_D + \mathcal{N})\} - \exp\{-\zeta(R_D - \mathcal{N})\}$$

$$= 2\left(1 - e^{-\zeta R_D} \cosh \zeta \mathcal{N}\right). \qquad (4.23)$$

The probability $q_{ij}(\phi, \alpha_s)$ is replaced by its expected value with respect to α_s,

$$q_{ij}(\phi) = \frac{1}{2R_D} \int_{-R_D}^{R_D} q_{ij}(\phi, \alpha_s) \, d\alpha_s$$

$$= \frac{\exp\{-\frac{\rho_c}{2}|\mathcal{N}_i - \mathcal{N}_j|\}}{2R_D \, \rho_c \, (\beta_c^2 - \rho_c^2)}\{\beta_c^2 \, \mathcal{K}(\mathcal{N}_{ij}, \rho_c, R_D) - \rho_c^2 \, \mathcal{K}(\mathcal{N}_{ij}, \beta_c, R_D)\}. \qquad (4.24)$$

In order to derive the lag k cross-covariance function ($k \geq 0$) between the rainfall intensities at two sites, we refer to the corresponding function, $E_{\mathcal{W}}[C_{ij}^{(h)}(\mathcal{W}d_{ij}, k)]$, given in Equation (4.10). In the model with rain cell classification only, the storm arrival rate is λ, while the numbers of rain cells that affect site i, site j and both sites, are

Poisson variables with means $\mu_c q_i(\phi)$, $\mu_c q_j(\phi)$ and $\mu_c q_{ij}(\phi)$, respectively. So, the lag k cross-covariance function ($k \geq 0$) becomes

$$\begin{aligned}
\Sigma_{ij,c}^{(h)}(\phi,k) &= \lambda\, q_{ij}(\phi)\, \mu_c\, E(X^2)\, E_{\mathcal{W}}[V_1(\mathcal{W}d_{ij},k)] \\
&+ \frac{\lambda\beta}{2}\, q_i(\phi)\, q_j(\phi)\, \mu_c^2\, E(X)^2\, E_{\mathcal{W}}[V_2(\mathcal{W}d_{ij},k)],
\end{aligned} \tag{4.25}$$

where the first term gives the 'within cell' contribution to the covariance, and the second term gives the contribution from different rain cells of the same storm.

4.8 Random direction

In the previous sections it has been assumed that the direction, ϕ, of the storm movement is fixed for all storms. This is not realistic, even for relatively short periods, such as one month. The study of the observed cross-correlation functions for various angular separations between pair of sites (Section 2.2), has revealed that although there exists a prevailing direction, the contributions from other directions are also very significant. In this section, we investigate ways in which ϕ can vary randomly between storms.

The main problem of randomising ϕ is the complexity of the expressions in which it is involved, which limits any further algebraic manipulation. For instance, for a fixed pair of sites, the direction of the storm movement influences the probability that a rain cell affects the pair of sites, and also the lag between the time it hits each one of them. Although ϕ can take any real value, we restrict its domain to the interval $(0, 2\pi]$, since the values ϕ and $\phi + 2k\pi$, for any integer k and any real ϕ, correspond to the same point on the circumference of a circle.

So, a possibility is to model ϕ using a continuous circular distribution. The between-site properties of the aggregated process under the new assumption are derived by taking the expected value, with respect to ϕ, of the corresponding expressions with fixed ϕ's. This requires integration of some functions that do not have an analytic closed form and have to be evaluated numerically. One way to overcome this computational obstacle is to consider ϕ as a discrete random variable, in which case one needs to evaluate a relatively small number of sums, as compared to the numerous iterations required by the numerical methods of integration. Theoretically, given any distribution on the line, it is possible to wrap it around the circumference of a circle of unit radius, and to construct a circular

distribution (Mardia, 1972). However, the literature on discrete circular distributions is very limited, mainly because their properties are not as tractable as their analogues on the real line. Thus, in this analysis we start with a continuous distribution and then discretise it, by integrating its density function over a number of intervals.

The most common of the continuous circular distributions is the von Mises, which plays a key role in statistical inference on the circle, and has probability density function

$$g(\phi; \phi_0, \kappa) = \frac{e^{\kappa \cos(\phi - \phi_0)}}{2\pi I_0(\kappa)}, \quad 0 < \phi \leq 2\pi, \quad \kappa > 0, \quad 0 < \phi_0 \leq 2\pi,$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero,

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{1}{r!^2} \left(\frac{k}{2}\right)^{2r}.$$

The von Mises distribution is unimodal, symmetric around the mean direction ϕ_0 and the larger the value of the concentration parameter, κ, the greater the clustering around the mode, (see Figure 4.1). Although the von Mises distribution seems, intuitively, a suitable model for the direction of the storm movement, the discretisation of its density function does not lead to mathematically tractable expressions and thus this choice is inappropriate because of computational constraints.

Another continuous circular distribution is the cardioid, with probability density function

$$g(\phi) = \frac{1}{2\pi} \{1 + 2\kappa \cos(\phi - \phi_0)\}, \quad 0 < \phi \leq 2\pi, \quad |\kappa| \leq \frac{1}{2}. \tag{4.26}$$

With reference to Figure 4.1, we see that the cardioid distribution is also symmetrical and unimodal, with mean direction ϕ_0, and as compared to the von Mises distribution, it can not produce density functions with very high concentration around the mean. However, since the data analysis did not reveal such a large clustering, this weakness is of little practical significance. On the other hand, if ϕ follows the cardioid distribution, then the probability $P(\phi_1 < \phi \leq \phi_2)$, $(0 < \phi_1 < \phi_2 \leq 2\pi)$, is given by the simple analytic form $\frac{1}{2\pi} \{(\phi_2 - \phi_1) + 2\kappa \{\sin(\phi_2 - \phi_0) - \sin(\phi_1 - \phi_0)\}\}$. Thus, if $(0, 2\pi]$ is segmented into s disjoint intervals, each of length $2\pi/s$, then the probability, g_r, that ϕ belongs to the r-th interval, $(r = 1, ..., s)$, is easily calculated. The function defined by

$$g_r = \frac{1}{s} + \frac{\kappa}{\pi} \left\{\sin(\phi_r - \phi_0 + \frac{\pi}{s}) - \sin(\phi_r - \phi_0 - \frac{\pi}{s})\right\}, \quad \phi_r = \frac{\pi}{s}(2r - 1), \quad r = 1, ..., s, \tag{4.27}$$

83
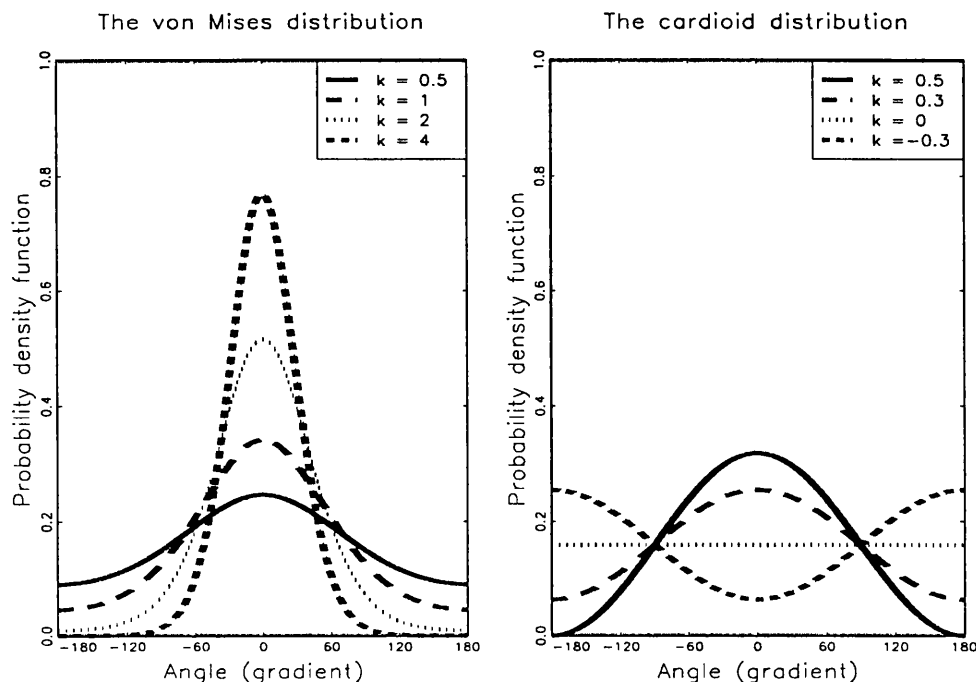
Figure 4.1: Probability density function of the von Mises distribution (left) and the cardioid distribution (right), both for mean direction $\phi_0 = 0°$.

satisfies the usual properties of a probability density function of a discrete distribution, and is used as the probability distribution of the direction, ϕ, of the storm movement.

So, when ϕ is random, combining Equations (4.22), (4.24) and (4.27), we get the probabilities that a rain cell affects a given site, say i, or a pair of sites, say i and j, respectively

$$q_i = \sum_{r=1}^{m} g_r \, q_i(\phi_r) \quad \text{and} \quad q_{ij} = \sum_{r=1}^{m} g_r \, q_{ij}(\phi_r).$$

Also, the cross-covariance function is derived using Equations (4.25) and (4.27) as follows,

$$
\begin{aligned}
E_\phi[\Sigma_{ij,c}^{(h)}(\phi, k)] &= \lambda \, \mu_c \, E(X^2) \, E_\phi[q_{ij}(\phi) \, E_W[V_1(\mathcal{W}d_{ij}(\phi), k)]] \\
&+ \frac{\lambda\beta}{2} \, \mu_c^2 \, E(X)^2 \, E_\phi[q_i(\phi) \, q_j(\phi) \, E_W[V_2(\mathcal{W}d_{ij}(\phi), k)]].
\end{aligned}
\quad (4.28)
$$

The marginal properties of the rainfall intensity at a single site are also functions of the direction, ϕ, and we take the expected value with respect to ϕ of the corresponding expressions.

4.9 Probability of zero rainfall at two sites.

In modelling the evolution of rainfall at more than one distinct point in space, it is important to study the probability that two (or more) sites have zero rainfall during either the same arbitrary time interval or different periods, and to explore how these properties depend on the distance between the examined sites. If $Y_{i,n}^{(h)} = \int_{(n-1)h}^{nh} Y_i(t)dt$ is the total rainfall intensity experienced at site i during the n-th interval of length h of the aggregated process, then ideally, we would like to derive all conditional probabilities of the form $P(Y_{i,n}^{(h)} = y_i | Y_{j,n+k}^{(h)} = y_j)$, but from our experience in modelling rainfall at a single site, we know that those for $y_i \neq 0$ and $y_j \neq 0$ are going to be difficult to obtain analytically. So, we focus on the probability of zero rainfall at two sites i and j simultaneously, $P(Y_{i,n}^{(h)} = Y_{j,n}^{(h)} = 0)$, for an arbitrary period of fixed length h, say $[0, h]$.

4.9.1 Storms with fixed time displacements at various sites.

We examine the evolution of rainfall over a network of sites, and consider a model in which storms arrive in a Poisson process (in time) of rate λ and generate a number, C, of rain cells, that occur in a Neyman-Scott process. It is assumed that storm and rain cells move across the catchment area with a fixed speed and direction, which implies that the lag $t_{ij} = t_i - t_j$ between the times that a rainfall element reaches two sites i and j is the same for all storms and all rain cells. The traces that are formed as rain cells move across the area, are visualised as parallel bands that have width and location that varies randomly between cells. Since the catchment area is relatively small, we assume that all storms affect the whole area, thus the rainfall band associated with each storm always has a width that is greater than the examined area. The specifications of the spatial structure of rainfall are similar to the ones described in Section 4.7, so when the direction ϕ is fixed, the probabilities $q_i(\phi)$ and $q_{ij}(\phi)$, that a rain cell affects a particular site, say i, or a pair of sites, say i and j, are given in Equations (4.22) and (4.24), respectively. In addition, we let the number of rain cells, C, per storm follow a Poisson distribution with mean μ_c, and as before the duration of rain cells is exponential with parameter η, and is scaled by a constant factor l_i at site i giving a mean lifetime of $\eta_i = \eta/l_i$. We remind that it is assumed that there is no rain cell attached to the storm origin.

Examining the rainfall intensities at two sites i and j at times t and $t+\tau$ respectively,

is equivalent to looking at the master point process and considering the rainfall intensity at time $t - t_i$ due to elements that affect site i, and the intensity at time $t + \tau - t_j$ due to elements that affect site j. So, the probability that the subprocesses corresponding to sites i and j are simultaneously dry during the interval $[0, h]$, is the same as the probability that in the master process, there are no active rain cells of type 'i' or 'ij' during the interval $[-t_i, -t_i + h]$ and of type 'j' or 'ij' during $[-t_j, -t_j + h]$. Consequently, $P(Y_{i,n}^{(h)} = Y_{j,n}^{(h)} = 0)$ can be written as the product of three terms, the probability, $\varpi_h^{(ij)}(t_{ij})$, that in the master process, there is no rainfall during $[-t_i, -t_i + h]$ and $[-t_j, -t_j + h]$ due to rain cells that affect both sites, and the probabilities that in the master process, an arbitrary interval of length h is dry, due to rain cells that affect site i only, $\varpi_h^{(i-)}$, and due to the ones affecting site j only, $\varpi_h^{(-j)}$.

Initially, we derive the probability $\varpi_h^{(ij)}(t_{ij})$ and thus consider only rain cells that affect both sites i and j. Since C is Poisson, the number of type 'ij' rain cells per storm is also Poisson with mean $\mu_{ij}(\phi) = \mu_c \, q_{ij}(\phi)$. In what follows the explicit dependence of $\mu_{ij}(\phi)$ on ϕ is dropped for the sake of notational simplicity. As in the derivation of the cross-covariance function when the displacement factors are fixed, we distinguish four cases depending on the value of $t_{ij} = t_i - t_j$ relative to $-h$, 0 and h.

Case 1: $t_{ij} > h$.
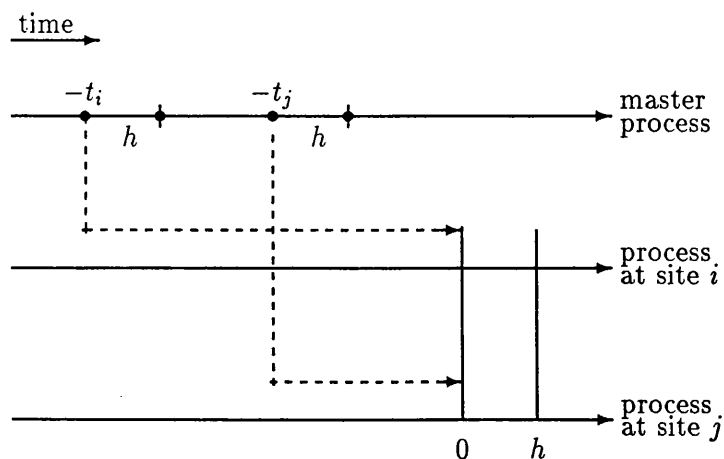
This diagram illustrates how the interval $[0, h]$ of the subprocesses corresponding to sites i and j is translated in time to the intervals $\mathcal{E}_i = [-t_i, -t_i + h]$ and $\mathcal{E}_j = [-t_j, -t_j + h]$, respectively, in the master process. In this diagram both displacement factors are positive,

but the same results apply to any values of t_i and t_j, such that $t_{ij} > h$. In the following analysis we focus on the master point process of rain cell origins and we explore the contribution to the probability $P(Y_{i,n}^{(h)} = Y_{j,n}^{(h)} = 0)$ of the type 'ij' rain cells only.

First let us consider a single storm with origin at time y. If $y < -t_i$ then $\mathcal{E}_i$ is dry if there are no rain cell arrivals during this interval and if all cells with origin, say u, before $-t_i$ have duration less than $-t_i - u$ at site i and less than $-t_j - u$ at site j, with probability

$$P\Big(L_i < -t_i - u, L_j < -t_j - u\Big) = P\Big(L < \min\Big\{\frac{-t_i - u}{l_i}, \frac{-t_j - u}{l_j}\Big\}\Big).$$

We distinguish the following two cases.

If $\quad l_i \geq l_j \quad$ then $\quad \min\{\frac{-t_i-u}{l_i}, \frac{-t_j-u}{l_j}\} \quad = \quad \frac{-t_i-u}{l_i} \quad$ for $\quad y < u < -t_i$. $\qquad$ (4.29)

If $\quad l_i < l_j \quad$ then $\quad \min\{\frac{-t_i-u}{l_i}, \frac{-t_j-u}{l_j}\} \quad = \quad \begin{cases} \frac{-t_j-u}{l_j} & \text{if} \quad y < u < \frac{l_j t_i - l_i t_j}{l_i - l_j} = \mathcal{T} \\ \frac{-t_i-u}{l_i} & \text{if} \quad \mathcal{T} < u < -t_i. \end{cases}$ $\quad$ (4.30)

Also, $\mathcal{E}_j$ is dry if no rain cells arrive during this interval, and if all cells with origin, u, after $-t_i + h$ and before $-t_j$ have duration less than $-t_j - u$ at site j.

When $l_i \geq l_j$, the probability, $\Pi_1^{(ij)}(y, t_{ij})$, of zero rainfall during intervals $\mathcal{E}_i$ and $\mathcal{E}_j$ due to a single storm with origin at $y < -t_i$ is

$$\Pi_1^{(ij)}(y, t_{ij})$$

$$= \sum_{n=0}^{\infty} \frac{\mu_{ij}^n}{n!} e^{-\mu_{ij}} \sum_{k=0}^{n} \binom{n}{k} \left\{ \int_y^{-t_i} \beta e^{-\beta(u-y)} \Big(1 - e^{\eta_i(t_i+u)}\Big) du \right\}^k$$

$$\times \sum_{m=0}^{n-k} \binom{n-k}{m} \left\{ \int_{-t_i+h}^{-t_j} \beta e^{-\beta(u-y)} \Big(1 - e^{\eta_j(t_j+u)}\Big) du \right\}^m \left\{ \int_{-t_j+h}^{\infty} \beta e^{-\beta(u-y)} du \right\}^{n-k-m}$$

$$= \sum_{n=0}^{\infty} \frac{\mu_{ij}^n}{n!} e^{-\mu_{ij}} \sum_{k=0}^{n} \binom{n}{k} \left\{ 1 - \frac{1}{\beta - \eta_i} \Big\{ \beta\, e^{\eta_i(t_i+y)} - \eta_i\, e^{\beta(t_i+y)} \Big\} \right\}^k$$

$$\times \sum_{m=0}^{n-k} \binom{n-k}{m} \left\{ 1 - \frac{1}{\beta - \eta_j} \Big\{ \beta e^{-\eta_j(t_i-t_j-h)} - \eta_j\, e^{-\beta(t_i-t_j-h)} \Big\} \right\}^m$$

$$\times e^{-\beta(h-t_i-y)\, m}\, e^{-\beta(h-t_j-y)\,(n-k-m)}$$

$$= e^{-\mu_{ij}} \sum_{k=0}^{\infty} \frac{\mu_{ij}^k}{k!} \left\{ 1 - \frac{1}{\beta - \eta_i} \Big\{ \beta\, e^{\eta_i(t_i+y)} - \eta_i\, e^{\beta(t_i+y)} \Big\} \right\}^k$$

$$\times \sum_{m=0}^{\infty} \frac{\mu_{ij}^m}{m!} \left\{ 1 - \frac{1}{\beta - \eta_j} \left\{ \beta \, e^{-\eta_j(t_i - t_j - h)} - \eta_j \, e^{-\beta(t_i - t_j - h)} \right\} \right\}^m$$

$$\times \, e^{-\beta(h - t_i - y)} m \sum_{n=k+m}^{\infty} \frac{\mu_{ij}^{n-k-m}}{(n-k-m)!} \, e^{-\beta(h - t_j - y)(n-k-m)}$$

$$= \exp\left\{ -\mu_{ij} \left\{ \frac{1}{\beta - \eta_i} \{ \beta \, e^{\eta_i(t_i + y)} - \eta_i \, e^{\beta(t_i + y)} \} - e^{\beta(t_i + y)} \right. \right.$$

$$\left. \left. \times \left\{ e^{-\beta h} + e^{-\beta(t_{ij} + h)} - \frac{e^{-\beta h}}{\beta - \eta_j} \{ \beta \, e^{-\eta_j(t_{ij} - h)} - \eta_j \, e^{-\beta(t_{ij} - h)} \} \right\} \right\} \right\}.$$

When $l_i < l_j$, the probability $\Pi_1^{(ij)}(y, t_{ij})$ is given by

$$\Pi_1^{(ij)}(y, t_{ij})$$

$$= \sum_{n=0}^{\infty} \frac{\mu_{ij}^n}{n!} \sum_{k=0}^{n} \binom{n}{k} \left\{ \int_y^T \beta \, e^{-\beta(u-y)} \left(1 - e^{\eta_j(t_j + u)} \right) du + \int_T^{-t_i} \beta \, e^{-\beta(u-y)} \left(1 - e^{\eta_i(t_i + u)} \right) du \right\}^k$$

$$\times \, e^{-\mu_{ij}} \sum_{m=0}^{n-k} \binom{n-k}{m} \left\{ \int_{-t_i + h}^{-t_j} \beta e^{-\beta(u-y)} \left(1 - e^{\eta_j(t_j + u)} \right) du \right\}^m \left\{ \int_{-t_j + h}^{\infty} \beta e^{-\beta(u-y)} du \right\}^{n-k-m}$$

$$= \exp\left\{ -\mu_{ij} \left\{ \frac{\beta}{\beta - \eta_j} \, e^{\eta_j(t_i + y)} \, e^{-\eta_j t_{ij}} - e^{\beta(t_i + y)} \left\{ \frac{\eta_i}{\beta - \eta_i} + e^{-\beta h} + e^{-\beta(t_{ij} + h)} \right. \right. \right.$$

$$\left. \left. \left. + \frac{\eta_j}{\beta - \eta_j} e^{-\beta t_{ij}} - \frac{\beta}{\beta - \eta_j} e^{-\beta h} e^{-\eta_j(t_{ij} - h)} - \frac{\beta(\eta_i - \eta_j)}{(\beta - \eta_j)(\beta - \eta_i)} e^{-(\beta l_i - \eta)\frac{t_{ij}}{(l_i - l_j)}} \right\} \right\} \right\}.$$

Since storms arrive in a Poisson process of rate λ, the mean number of storm origins during an interval of length x is also Poisson with mean λx and, conditioning on the number of points in that interval, their times are uniformly distributed. So, if we consider all storms that arrived before time $-t_i$, then the probability that intervals $\mathcal{E}_i$ and $\mathcal{E}_j$ are dry, due to type 'ij' rain cells only, is

$$\tilde{\Pi}_1^{(ij)}(t_{ij})$$

$$= \lim_{x \to -\infty} \sum_{k=0}^{\infty} \frac{\lambda^k(-t_i - x)^k}{k!} \, e^{-\lambda(-t_i - x)} \left\{ \int_x^{-t_i} \frac{\Pi_1^{(ij)}(y, t_{ij})}{(-t_i - x)} \, dy \right\}^k$$

$$= \exp\left\{ -\lambda \int_{-\infty}^0 \left\{ 1 - \exp\left\{ -\mu_{ij} \, \mathcal{A}_1(\eta_i, \eta_j, y) + \mu_{ij} \, e^{\beta y} \left\{ \frac{\eta_i}{\beta - \eta_i} + e^{-\beta h} + e^{-\beta(t_{ij} + h)} \right. \right. \right. \right.$$

$$\left. \left. \left. \left. + \frac{\eta_j}{\beta - \eta_j} \, e^{-\beta t_{ij}} - \frac{\beta}{\beta - \eta_j} e^{-\beta h} e^{-\eta_j(t_{ij} - h)} - \frac{\beta(\eta_i - \eta_j)^+}{(\beta - \eta_i)(\beta - \eta_j)} e^{-(\beta l_i - \eta)\frac{t_{ij}}{(l_i - l_j)}} \right\} \right\} \right\} dy \right\},$$

where

$$\mathcal{A}_1(\eta_i, \eta_j, y) = \begin{cases} \frac{\beta}{\beta - \eta_i} \, e^{\eta_i y} & \text{if } \eta_i \le \eta_j \\ \frac{\beta}{\beta - \eta_j} \, e^{-\eta_j(t_{ij} - y)} & \text{if } \eta_i > \eta_j \end{cases} \tag{4.31}$$

and $z^+ = \max\{z, 0\}$.

Suppose that a single storm has origin, y, in $\mathcal{E}_i$. Then, in order intervals $\mathcal{E}_i$ and $\mathcal{E}_j$ to be dry there must be no cell arrival before time $-t_i + h$ or during interval $\mathcal{E}_j$, and all cells with origin, $u < -t_j$ must have duration less than $-t_j - u$ at site j. In this case, the probability, $\Pi_2^{(ij)}(y, t_{ij})$, of zero rainfall during $\mathcal{E}_i$ and $\mathcal{E}_j$ due to a single storm, is

$$\Pi_2^{(ij)}(y, t_{ij})$$
$$= \sum_{n=0}^{\infty} \frac{\mu_{ij}^n}{n!} e^{-\mu_{ij}} \sum_{k=0}^{n} \binom{n}{k} \left\{ \int_{-t_i+h}^{-t_j} \beta e^{-\beta(u-y)} \left(1 - e^{\eta_j(t_j+u)}\right) du \right\}^k \left\{ \int_{-t_j+h}^{\infty} \beta e^{-\beta(u-y)} du \right\}^{n-k}$$
$$= \exp\left\{ -\mu_{ij}\left\{ 1 + e^{-\beta(h-t_i-y)} \left(\frac{\beta \, e^{-\eta_j(t_{ij}-h)} - \eta \, e^{-\beta(t_{ij}-h)}}{\beta - \eta_j} - 1 - e^{-\beta t_{ij}} \right) \right\} \right\}.$$

The number of storms with origin in $\mathcal{E}_i$ is Poisson with mean λh and conditioning on the number of storm arrivals, their times are uniformly distributed in the corresponding interval. So, the probability of zero rainfall in the master process during $\mathcal{E}_i$ and $\mathcal{E}_j$ due to all storms with origin in $\mathcal{E}_i$, is

$$\widetilde{\Pi}_2^{(ij)}(t_{ij})$$
$$= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda h} \left\{ \int_{-t_i}^{-t_i+h} \Pi_2^{(ij)}(y, t_{ij}) dy \right\}^k$$
$$= \exp\left\{ -\lambda \int_0^h 1 - \exp\left\{ -\mu_{ij}\left(1 + e^{-\beta(h-y)}\left(\frac{\beta e^{-\eta_j(t_{ij}-h)} - \eta_j e^{-\beta(t_{ij}-h)}}{\beta - \eta_j} - 1 - e^{-\beta t_{ij}}\right)\right) \right\} dy \right\}.$$

Next, we consider a single storm that arrived during $[-t_i + h, -t_j]$. Obviously, this storm does not affect interval $\mathcal{E}_i$, and for $\mathcal{E}_j$ to be dry, there should be no cell arrivals during $\mathcal{E}_j$ and all rain cells that arrived before time $-t_j$ should terminate at site j before $-t_j$. The probability, $\Pi_3^{(ij)}(y, t_j)$, of this event is

$$\Pi_3^{(ij)}(y, t_j)$$
$$= \sum_{n=0}^{\infty} \frac{\mu_{ij}^n}{n!} e^{-\mu_{ij}} \sum_{k=0}^{n} \binom{n}{k} \left\{ \int_y^{-t_j} \beta e^{-\beta(u-y)} \left(1 - e^{\eta_j(t_j+u)}\right) du \right\}^k \left\{ \int_{-t_j+h}^{\infty} \beta e^{-\beta(u-y)} du \right\}^{n-k}$$
$$= \exp\left\{ -\mu_{ij}\left\{ \frac{\beta}{\beta - \eta_j} e^{\eta_j(t_j+y)} - \left(\frac{\eta_j}{\beta - \eta_j} + e^{-\beta h}\right) e^{\beta(t_j+y)} \right\} \right\},$$

and as before, taking into account all storms that arrived in $[-t_i+h, -t_j]$, the probability

89

that $\mathcal{E}_i$ and $\mathcal{E}_j$ are dry is

$$\widetilde{\Pi}_3^{(ij)}(t_{ij})$$

$$= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda(t_{ij}-h)}}{k!} \left\{ \int_{-t_i+h}^{-t_j} \Pi_3^{(ij)}(y,t_j)\, dy \right\}^k$$

$$= \exp\left\{ -\lambda \int_0^{t_{ij}-h} 1 - \exp\left\{ -\frac{\mu_{ij}\,\beta}{\beta-\eta_j}\, e^{-\eta_j y} + \mu_{ij}\left(\frac{\eta_j}{\beta-\eta_j} + e^{-\beta h}\right) e^{-\beta y} \right\} dy \right\}.$$

Finally, when a single storm has origin in $\mathcal{E}_j$, then all rain cells of this storm should arrive after time $-t_j + h$, with probability

$$\Pi_4^{(ij)}(y) = \sum_{n=0}^{\infty} \frac{\mu_{ij}^n\, e^{-\mu_{ij}}}{n!} \left\{ \int_{-t_j+h}^{\infty} \beta e^{-\beta(u-y)}\, du \right\}^n$$

$$= \exp\{-\mu_{ij}(1 - e^{-\beta(h-t_j-y)})\},$$

and the corresponding probability due to all storms that arrived in $\mathcal{E}_j$ is

$$\widetilde{\Pi}_4^{(ij)} = \sum_{k=0}^{\infty} \frac{\lambda^k\, e^{-\lambda h}}{k!} \left\{ \int_{-t_j}^{-t_j+h} \Pi_4^{(ij)}(y)\, dy \right\}^k$$

$$= \exp\left\{ -\lambda \int_0^h 1 - \exp\{-\mu_{ij}\,(1 - e^{-\beta(h-y)})\}\, dy \right\}.$$

So, when $t_{ij} > h$, and considering only type 'ij' rain cells, the probability that in the master process, the intervals $\mathcal{E}_i = [-t_i, -t_i + h]$ and $\mathcal{E}_j = [-t_j, -t_j + h]$ have zero rainfall, is

$$\varpi_h^{(ij)}(t_{ij}) = \widetilde{\Pi}_1^{(ij)}(t_{ij})\, \widetilde{\Pi}_2^{(ij)}(t_{ij})\, \widetilde{\Pi}_3^{(ij)}(t_{ij})\, \widetilde{\Pi}_4^{(ij)}. \tag{4.32}$$

Case 2: $0 \le t_{ij} \le h$.

When the absolute difference between the time displacement factors, is less than the level of aggregation h, the intervals $\mathcal{E}_i$ and $\mathcal{E}_j$ that are formed by translation of $[0, h]$ from the two subprocesses corresponding to sites i and j back to the master process, overlap and their union is one interval of legth $h + |t_{ij}|$, as it is illustrated in the following diagram. As before, we focus on the master point process and examine only the contribution of the type 'ij' rain cells to the probability that interval $\mathcal{E}_{ij} = [-t_i, -t_j + h]$ is dry.
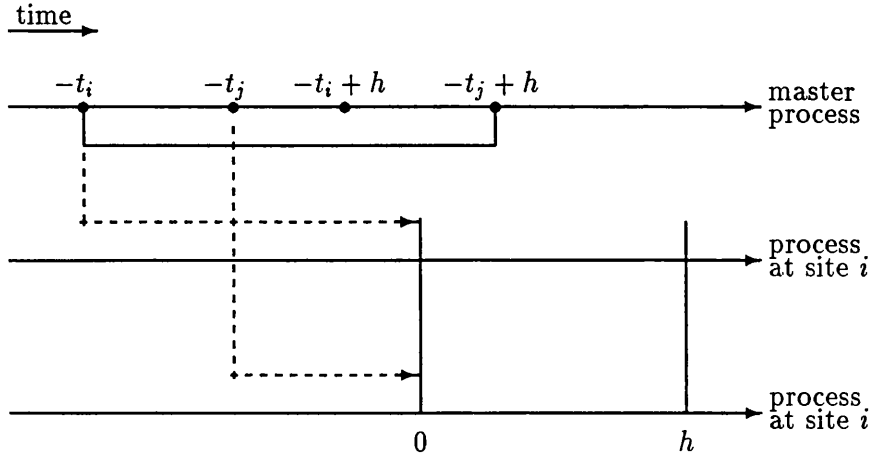
We start by considering a single storm with origin, y, before time $-t_i$. For interval $\mathcal{E}_{ij}$ in the master process to be dry, all rain cells with origin $u < -t_i$ must have lifetimes less than $t_i - u$ when experienced at site i and less than $t_j - u$ at site j. The probability $P(L < \min\{\frac{-t_i-u}{l_i}, \frac{-t_j-u}{l_j}\})$ depends on the sign of $(l_i - l_j)$, and its expression is given in Equations (4.29)–(4.30). If $l_i \geq l_j$, then the probability that in the master process the interval $\mathcal{E}_{ij}$ has zero rainfall due to a single storm with origin $y < -t_i$ is

$$\Pi_5^{(ij)}(y, t_{ij})$$

$$= \sum_{n=0}^{\infty} \frac{\mu_{ij}^n}{n!} e^{-\mu_{ij}} \sum_{k=0}^{n} \binom{n}{k} \left\{ \int_{-t_j+h}^{\infty} \beta e^{-\beta(u-y)} du \right\}^{n-k}$$

$$\times \left\{ \int_{y}^{-t_i} \beta e^{-\beta(u-y)} \left(1 - e^{\eta_i(t_i+u)}\right) du \right\}^{k}$$

$$= \exp\left\{ -q_{ij}\, \mu_c \left\{ \frac{\beta}{\beta - \eta_i} e^{\eta_i(t_i+y)} - \left(\frac{\eta_i}{\beta - \eta_i} + e^{-\beta(t_{ij}+h)} \right) e^{\beta(t_i+y)} \right\} \right\},$$

while, when $l_i < l_j$ we have instead

$$\Pi_5^{(ij)}(y, t_{ij})$$

$$= \sum_{n=0}^{\infty} \frac{\mu_{ij}^n}{n!} e^{-\mu_{ij}} \sum_{k=0}^{n} \binom{n}{k} \left\{ \int_{-t_j+h}^{\infty} \beta e^{-\beta(u-y)} du \right\}^{k}$$

$$\times \left\{ \int_{y}^{T} \beta\, e^{-\beta(u-y)} \left(1 - e^{\eta_j(t_j+u)}\right) du + \int_{T}^{-t_i} \beta\, e^{-\beta(u-y)} \left(1 - e^{\eta_i(t_i+u)}\right) du \right\}^{n-k}$$

$$= \exp \left\{ -\mu_{ij} \left\{ \frac{\beta \, e^{-\eta_j t_{ij}}}{\beta - \eta_j} \, e^{\eta_j(t_i + y)} \right. \right.$$

$$\left. \left. - e^{\beta(t_i + y)} \left\{ \frac{\eta_i}{\beta - \eta_i} + e^{-\beta(t_{ij} + h)} - \frac{\beta(\eta_i - \eta_j)}{(\beta - \eta_i)(\beta - \eta_j)} \, e^{-(\beta l_i - \eta) \frac{t_{ij}}{(l_i - l_j)}} \right\} \right\} \right\} .$$

If we consider all storms that arrived before $-t_i$, then the corresponding probability becomes

$$\tilde{\Pi}_5^{(ij)}(t_{ij})$$

$$= \lim_{x \to -\infty} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{\lambda(t_i + x)} \left\{ \int_x^{-t_i} \Pi_5^{(ij)}(y, t_{ji}) \, dy \right\}^k$$

$$= \exp \left\{ -\lambda \int_{-\infty}^{0} 1 - \exp \left\{ - \mu_{ij} \, \mathcal{A}_1(\eta_i, \eta_j, y) - \mu_{ij} \, e^{\beta y} \right. \right.$$

$$\left. \left. \times \left\{ \frac{\eta_i}{\beta - \eta_i} + e^{-\beta(t_{ij} + h)} - \frac{\beta(\eta_i - \eta_j)^+}{(\beta - \eta_i)(\beta - \eta_j)} \, e^{-(\beta l_i - \eta) \frac{t_{ij}}{(l_i - l_j)}} \right\} \right\} dy \right\} ,$$

where $\mathcal{A}_1(\eta_i, \eta_j, y)$ is given in (4.31).

Similarly, the probability of zero rainfall in the master process during $\mathcal{E}_{ij}$ due to all storms with origins in this interval is

$$\tilde{\Pi}_6^{(ij)}(t_{ij}) = \exp \left\{ -\lambda \int_0^{t_{ij} + h} 1 - \exp\{ -\mu_{ij} \, (1 - e^{-\beta(t_{ij} + h - y)}) \} \, dy \right\} .$$

Finally, when $0 < t_{ij} < h$ the probability that the interval $\mathcal{E}_{ij}$ of the master process has zero rainfall, due to all storms and considering only rain cells that affect both sites, is

$$\varpi_h^{(ij)}(t_{ij}) = \tilde{\Pi}_5^{(ij)}(t_{ij}) \, \tilde{\Pi}_6^{(ij)}(t_{ij}) \tag{4.33}$$

Case 3: $-h \le t_{ij} \le 0$.

In this case, the difference between the time displacements is less than the length h, with time $-t_j$ proceeding time $-t_i$ and because of symmetry with Case 2, we have

$$\varpi_h^{(ij)}(t_{ij}) = \tilde{\Pi}_5^{(ji)}(t_{ji}) \, \tilde{\Pi}_6^{(ji)}(t_{ji}). \tag{4.34}$$

Case 4: $t_{ij} < -h$.

Similarly, this case is symmetric to Case 1, with respect to i and j, so the probability that in the master process there is no rainfall during the intervals $[-t_j, -t_j + h]$ and

$[-t_i, -t_i + h]$, due to rain cells of type 'ij' only is

$$\varpi_h^{(ij)}(t_{ij}) = \widetilde{\Pi}_1^{(ji)}(t_{ji})\, \widetilde{\Pi}_2^{(ji)}(t_{ji})\, \widetilde{\Pi}_3^{(ji)}(t_{ji})\, \widetilde{\Pi}_4^{(ji)}. \tag{4.35}$$

So far, for the calculation of $P(Y_i^{(h)} = Y_j^{(h)} = 0)$ we have considered only rain cells that affect both sites i and j, and thus the contributions to this probability due to cells that affect only one of the two sites remain to be examined. We now focus on rain cells that affect site i only and derive the probability that in the master process, an arbitrary interval of length h, say $\mathcal{E}_i = [-t_i, -t_i + h]$, has zero rainfall intensity. The number of type 'i' rain cells per storm is a Poisson variable with mean $\mu_i = \mu_i(\phi) = \mu_c\, \{q_i(\phi) - q_{ij}(\phi)\}$.

A single storm with origin $y < -t_i$ has no active rain cells of type 'i' during $\mathcal{E}_i$ with probability

$$\Pi_1^{(i-)}(y, t_i)$$

$$= \sum_{n=0}^{\infty} \frac{\mu_i^n}{n!}\, e^{-\mu_i} \sum_{k=0}^{n} \binom{n}{k} \left\{ \int_y^{-t_i} \beta e^{-\beta(u-y)} (1 - e^{\eta_i(t_i+u)})\, du \right\}^k \left\{ \int_{-t_i+h}^{\infty} \beta e^{-\beta(u-y)}\, du \right\}^{n-k}$$

$$= \exp\left\{ -\mu_i \left\{ \frac{\beta}{\beta - \eta_i} e^{\eta_i(t_i+y)} - \left(\frac{\eta_i}{\beta - \eta_i} + e^{-\beta h} \right) e^{\beta(t_i+y)} \right\} \right\}.$$

If we consider all storms that arrived before $-t_i$, then the probability that $\mathcal{E}_i$ is dry is

$$\widetilde{\Pi}_1^{(i-)} = \lim_{x \to -\infty} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda(t_i+x)} \left\{ \int_x^{-t_i} \Pi_1^{(i-)}(y, t_i)\, dy \right\}^k$$

$$= \exp\left\{ -\lambda \int_{-\infty}^{0} \left\{ 1 - \exp\left\{ -\frac{\mu_i\, \beta}{\beta - \eta_i} e^{\eta_i y} + \mu_i \left(\frac{\eta_i}{\beta - \eta_i} + e^{-\beta h} \right) e^{\beta y} \right\} \right\} dy \right\},$$

which does not depend on the displacement factor t_i, but only on the length, h, of the interval.

A storm with origin, y, in $\mathcal{E}_i$ leaves this interval dry if all of its rain cells arrive after time $-t_i + h$. Storms arrive in a Poisson process of rate λ and conditional on the number of arrivals during $\mathcal{E}_i$, their times are uniformly distributed in $\mathcal{E}_i$. So, the probability that in the master process, the interval $\mathcal{E}_i$ is dry due to all storms that arrived during this interval, is

$$\widetilde{\Pi}_2^{(i-)} = \sum_{k=0}^{\infty} \frac{\lambda^k\, e^{-\lambda h}}{k!} \left\{ \int_{-t_i}^{-t_i+h} \sum_{n=0}^{\infty} \frac{e^{-\mu_i}\, \mu_i^n}{n!} \left\{ \int_{-t_i+h}^{\infty} \beta e^{-\beta(u-y)}\, du \right\}^n dy \right\}^k$$

93

$$= e^{-\lambda h} \exp \left\{ \lambda \int_0^h \exp\{-\mu_i \left(1 - e^{-\beta(h-y)}\right)\} \, dy \right\}.$$

So, the probability that in the master process an arbitrary interval of length h is dry due to rain cells that affect site i only is

$$\varpi_h^{(i-)} = \tilde{\Pi}_1^{(i-)} \, \tilde{\Pi}_2^{(i-)}. \tag{4.36}$$

Similarly, we derive the corresponding probability, $\varpi_h^{(-j)}$, due to rain cells that affect site j only.

Finally, we are in the position to calculate the probability that sites i and j are dry simultaneously during an arbitrary period of length h, when storms and rain cells in the master process are displaced by a constant factor at the various sites. This probability is a function of $t_{ij} = t_i - t_j$ and is given from

$$P(Y_i^{(h)} = Y_j^{(h)} = 0) = \varpi_h^{(ij)}(t_{ij}) \, \varpi_h^{(i-)} \, \varpi_h^{(-j)}. \tag{4.37}$$

4.9.2 Random storm features

We now allow the displacement factors to depend on the location of the examined sites and to vary randomly between storms, and calculate the probability that two sites are simultaneously dry during an arbitrary interval. Following similar steps as in the calculation of the cross-covariance function (Section 4.4), we express t_{ij} as the product of the distance, $d_{ij} = (x_i - x_j)\cos\phi + (y_i - y_j)\sin\phi$, between sites i and j, and the random variable $\mathcal{W} = 1/\mathcal{V}$, where $\mathcal{V}$ is the speed of the storm movement. As before, $\mathcal{W}$ has a gamma distribution with density function $f_{\mathcal{W}}(t) = \nu(\nu t)^{\theta-1}e^{-\nu t}/\Gamma(\theta)$, so the probability of the two sites being simultaneously dry during an arbitrary period of length h becomes

$$E_{\mathcal{W}}[P(Y_i^{(h)} = Y_j^{(h)} = 0)] \tag{4.38}$$

$$= \varpi_h^{(i-)} \, \varpi_h^{(-j)} \left\{ \int_0^{h/|d_{ij}|} f_{\mathcal{W}}(w) \, \varpi_h^{(ij)}(w|d_{ij}| < h) dw + \int_{h/|d_{ij}|}^{\infty} f_{\mathcal{W}}(w) \, \varpi_h^{(ij)}(h < w|d_{ij}|) dw \right\},$$

which is calculated numerically.

In theory, it is possible to extend the above result and derive with the same way, the probability that two sites are dry for periods of the same length h, lagged by k time units. However, because Cases 1 and 4, as well as Cases 2 and 3 will no longer be symmetric,

the complication of the formulae increases dramatically and they become impractical to use.

The previous expression involves the probabilities q_i, q_j and q_{ij} and the signed distance d_{ij}, which are all functions of the direction ϕ of the storm movement. Suppose that ϕ is a discrete random variable, that varies between storms and its density function is given in Equation (4.27). Then the expression of the probability that two sites are simultaneously dry, is derived by taking the expected value, with respect to ϕ, of $E_W[P(Y_i^{(h)} = Y_j^{(h)} = 0)]$, which is not given in an analytic closed form, but it is possible to evaluate it evaluated using numerical methods of integration.

4.10 Alternative multi-site models

4.10.1 Storm and rain cell classification

In modelling the evolution of rainfall over an area of the size of Brue catchment, it is reasonable to assume that the storms considered in the analysis affect the whole area and that rain cells may be classified according to the locations they affect. In other words, storms arrive at the catchment area in a Poisson process of rate λ and the master point process of rain cell origins is decomposed into a number of subprocesses, each one corresponding to a different subset of sites. However, when the network of raingauge stations covers a much larger area, some of these assumption are unlikely to be correct and will need to be changed. Using a similar framework to the above, it is possible to build models that could be applied when storms have a mean spatial extent that is small compared to the examined area, which is inscribed in a disc $\mathcal{D}$ of radius R_D. Specifically, let us assume that storms move with random speed, $\mathcal{V}$, that remains constant throughout their lifetime, and with direction, ϕ, that is kept fixed at this stage. The trace of a storm movement is visualised as a band of width $2R_s$, that intersects the catchment area, and has a main axis that is a signed distance α_s away from the origin. Each storm generates a random number of rain cells that arrive in time in a Neyman-Scott process and move with the same speed and direction as the storm. The trace from a cell's passage over the disc $\mathcal{D}$, which is the area that experiences rain due to that cell, is envisaged as a band with main axis $\mathcal{A}_c$ and random width $2R_c$. We assume that R_c has an exponential

distribution with parameter ρ_c and that the signed distance, α_c, between the axis $\mathcal{A}_c$ and the origin of the coordinate system, is uniformally distributed in $[\alpha_s - R_s, \alpha_s + R_s]$. The assumption that storms and rain cells do not start or terminate at a location of the catchment area is retained, for the sake of simplicity, although it is very likely that when the area is large it should be removed.

This model is based on the idea that both storms and rain cells may be classified according to the sites they affect, and the probability that a rainfall element affects a subset of sites reflects the spatial structure of the event and possibly the topography of the catchment area.

For this model, conditioning on the values of ϕ, α_s and R_s, the probability that a rain cell affects a particular site, say i with coordinates (x_i, y_i), is

$$
q_i(\phi, \alpha_s, R_s) = \frac{1}{2R_s} \int\limits_{\alpha_s - R_s}^{\alpha_s + R_s} \exp\{-\rho_c |\alpha_c - y_i \cos\phi + x_i \sin\phi|\} \, d\alpha_c
$$

$$
= \begin{cases} \frac{1}{2R_s \rho_c} \left\{ 2 - e^{-\rho_c(R_s + \alpha_s - \mathcal{N}_i)} - e^{-\rho_c(R_s - \alpha_s + \mathcal{N}_i)} \right\} & \text{if } |\alpha_s - \mathcal{N}_i| < R_s \\ \\ \frac{1}{2R_s \rho_c} \left\{ e^{-\rho_c(|\alpha_s - \mathcal{N}_i| - R_s)} - e^{-\rho_c(|\alpha_s - \mathcal{N}_i| + R_s)} \right\} & \text{if } |\alpha_s - \mathcal{N}_i| \geq R_s \end{cases}
\tag{4.39}
$$

where $\mathcal{N}_i = y_i \cos\phi - x_i \sin\phi$.

If there is no evidence to support the existance of topographic effects, we may assume that the signed distance, α_s, of the storm axis from the origin is uniformally distributed in $[-R_D, R_D]$. Also, if R_s is a positively defined random variable with probability density function f_{R_s}, then the probability that a rain cell affects site i becomes

$$
\begin{aligned}
q_i(\phi) &= E_{\alpha_s}[E_{R_s}[\, q_i(\phi, \alpha_s, R_s)\,]\,] \\
&= \frac{1}{2R_D} \int\limits_{-R_D}^{R_D} d\alpha_s \int\limits_{0}^{\infty} dR_s \, f_{R_s}(R_s) \, q_i(\phi, \alpha_s, R_s) \\
&= \mathcal{B}(\mathcal{N}_i) + \mathcal{B}(-\mathcal{N}_i),
\end{aligned}
\tag{4.40}
$$

where

$$
\begin{aligned}
\mathcal{B}(\mathcal{N}) &= \frac{1}{2\rho_c R_D} \Bigg\{ \int\limits_{0}^{|R_D + \mathcal{N}|} f_{R_s}(R_s) \left\{ 1 - \frac{1}{2\rho_c R_s}(e^{\rho_c R_s} - e^{-\rho_c R_s}) \, e^{-\rho_c(R_D + \mathcal{N})} \right\} dR_s \\
&\quad + \int\limits_{|R_D + \mathcal{N}|}^{\infty} \frac{f_{R_s}(R_s)}{R_s} \left\{ R_D + \mathcal{N} - \frac{e^{-\rho_c R_s}}{2\rho_c}\{e^{\rho_c(R_D + \mathcal{N})} - e^{-\rho_c(R_D + \mathcal{N})}\} \right\} dR_s \Bigg\}.
\end{aligned}
$$

Under the same assumptions about the distribution of the storm and rain cell characteristics, and for a fixed direction ϕ, the probability that a rain cell affects two sites, say i and j, is

$$
\begin{aligned}
q_{ij}(\phi) &= \frac{1}{2\,R_D} \int\limits_{-R_D}^{R_D} d\alpha_s \int\limits_{0}^{\infty} dR_s \, \frac{f_{R_s}(R_s)}{2\,R_s} \int\limits_{\alpha_s-R_s}^{\alpha_s+R_s} d\alpha_c \, \exp\{-\rho_c \max\{|\alpha_c - \mathcal{N}_i|, |\alpha_c - \mathcal{N}_j|\}\} \\
&= \exp\left\{ -\frac{\rho_c}{2}|\mathcal{N}_i - \mathcal{N}_j| \right\} \{\mathcal{B}(\mathcal{N}_{ij}) + \mathcal{B}(-\mathcal{N}_{ij})\}, \quad\quad (4.41)
\end{aligned}
$$

where $\mathcal{N}_{ij} = (\mathcal{N}_i + \mathcal{N}_j)/2$.

Natural candidates for the distribution of R_s are the gamma and, its special case, the exponential distribution. The latter, specified by a single parameter, has the mode at the origin and thus it is expected to produce storms with small mean width, in contrast with the former, which is a two parameter distribution, and thus its density function can take a wider range of shapes. The gamma distribution would be a more appropriate model for storms with large mean spatial extent.

Let us now suppose that the number of rain cells, C, per storm is Poisson with mean μ_c. Then, the probability, $p_i(\phi)$, that a storm affects a particular site i, is written as

$$
\begin{aligned}
p_i(\phi) &= 1 - \text{P(storm does NOT affect site } i) \\
&= 1 - \text{P(NO rain cell affects site } i) \\
&= 1 - \sum_{k=0}^{\infty} \frac{\mu_c^k e^{-\mu_c}}{k!} \, (1 - q_i(\phi))^k \\
&= 1 - \exp\{-\mu_c \, q_i(\phi)\}. \quad\quad (4.42)
\end{aligned}
$$

Similarly, for the probability, $p_{ij}(\phi)$, that a storm affects two sites i and j, we have

$$
\begin{aligned}
p_{ij}(\phi) &= 1 - \text{P(storm does NOT affect sites } i \text{ AND } j) \\
&= 1 - \text{P(NO rain cell affects site } i \text{ OR site } j) \\
&= 1 - \sum_{k=0}^{\infty} \frac{\mu_c^k e^{-\mu_c}}{k!} \, (1 - q_i(\phi) - q_j(\phi) + q_{ij}(\phi))^k \\
&= 1 - \exp\{-\mu_c(q_i(\phi) + q_j(\phi) - q_{ij}(\phi))\}. \quad\quad (4.43)
\end{aligned}
$$

The lag k cross-covariance function of the aggregated process between the rainfall intensity at two sites is

$$
\begin{aligned}
\Sigma_{ij}^{(h)}(\phi, k) &= \lambda\, p_{ij}(\phi)\, \mu_c\, q_{ij}(\phi)\, E(X^2)\, E_{\mathcal{W}}[V_1(\mathcal{W}d_{ij}, k)] \\
&\quad + \frac{\lambda\, p_{ij}(\phi)\, \beta}{2}\, \mu_c^2\, q_i(\phi)\, q_j(\phi)\, E(X)^2\, E_{\mathcal{W}}[V_2(\mathcal{W}d_{ij}, k)], \quad\quad (4.44)
\end{aligned}
$$

where the expressions of $E_{\mathcal{W}}[V_m(\mathcal{W}d_{ij}, k)]$, $m = 1, 2$ are given in Appendix C, if the assumptions about the storm and rain cell movement are the same as in the previous sections. Furthermore, it is possible to allow the direction, ϕ, to vary randomly between storms according to the discretised cardioid distribution g_r given in Equation (4.27).

4.10.2 Bartlett-Lewis based multi-site models

As already mention, in modelling the evolution of rainfall at a single point in space, the Neyman-Scott and Bartlett-Lewis based models with rectangular pulse rain cells (Rodriguez-Iturbe, Cox, and Isham, 1987a), are fairly tractable and intuitively realistic and when fitted to rainfall data, both give estimated properties that are very close to the observed ones. So, although the temporal structure of clusters is different in the two processes, their statistical properties, when they are used as rainfall models and under the assumptions specified by these authors, are very similar.

Having developed some Neyman-Scott based multi-site models, it would be interesting to see if it is possible to apply the same ideas to a Bartlett-Lewis process. Cox and Isham (1994) examined the properties of some Bartlett-Lewis based multi-site rainfall models, and we try here to expand their results by incorporating some spatial features of the rainfall event, following the same steps as with the Neyman-Scott process.

As before, we assume that there is a master process of storms that is Poisson (in time) of rate λ, and storms evolve independently of each other. For a particular storm we have

P(storm affects site i) $= p_i$,   where $\lambda_i = \lambda p_i$,

P(storm affects sites i and j) $= p_{ij}$,   where $\lambda_{ij} = \lambda p_{ij}$.

Let us suppose that rain cells follow a storm, whose origin is at time u, in a Poisson process of rate β and counting measure N_u, that is the same at all affected sites. Conventionally, it is assumed that a rain cell is attached to each storm origin. The storm truncation mechanism acts independently at each site and at a site, say i, the mean storm lifetime is exponentially distributed with parameter γ_i. We also assume that the duration and intensity of a particular cell at a particular site, are independent variables, that intensities of cells with common origin but at different sites are identical, but that durations of such cells are dependent. Specifically, we assume that the duration L_i at site i is given by $L_i = l_i L$, where the l_i are scalar constants and L is exponential with parameter η. In

addition to the above specifications, in order to allow different sites to experience the same storm and rain cell at different times, the processes of storm and of rain cell origins are translated in time by a fixed amount t_i at site i, different for each site. The lag τ cross-covariance function $(\tau > 0)$ between rainfall intensity at sites i and j is

$$
c_{ij}(\tau) = \lambda_{ij} \int_{-\infty}^{\min\{t-t_i, t-t_j+\tau\}} du\, E\left[\int_u^{t-t_i} dN_u(v)\, X_{t-v-t_i}^{(i)}(v)\, e^{-\gamma_i(v-u)} \right.
$$
$$
\left. \times \int_u^{t-t_j+\tau} dN_u(w)\, X_{t+\tau-w-t_j}^{(j)}(w)\, e^{-\gamma_j(w-u)} \right], \tag{4.45}
$$

where $X_{t-v-t_i}^{(i)}(v)$ is the intensity at site i at time t, of a cell whose origin at that site is at time $v + t_i$. For $t_i = t_j = 0$ the expression is given in Cox and Isham (1994, Equation(1.3.4)). Then, by expressing $E[dN_u(v)]$ as $[\delta(u-v) + \beta]dv$ and following the same steps as in Section 4.2, it is straightforward to show that $c_{ij}(\tau)$ is written as

$$
c_{ij}(\tau) = \begin{cases}
\lambda_{ij}E(X^2)\left(1 + \frac{\beta}{\gamma_i+\gamma_j}\right)\left\{ \frac{e^{\eta_i(\tau+t_{ij})}}{\eta_i} - (\frac{1}{\eta_{ij}})^+ e^{\eta_{ij}(\tau+t_{ij})} \right\} + \lambda_{ij}\,\beta\,E(X)^2 \\
\quad \times \frac{(\gamma_i+\gamma_j+\beta)}{(\eta_i-\gamma_i)}\left\{ \frac{e^{\gamma_i(\tau+t_{ij})}}{(\gamma_i+\gamma_j)(\gamma_i+\eta_j)} - \frac{e^{\eta_i(\tau+t_{ij})}}{(\eta_i+\eta_j)(\eta_i+\gamma_j)} \right\} = c_{ij,1}(\tau) \qquad \text{if } \tau < -t_{ij}, \\
\\
\lambda_{ij}E(X^2)\left(1 + \frac{\beta}{\gamma_i+\gamma_j}\right)\left\{ \frac{e^{-\eta_j(\tau+t_{ij})}}{\eta_j} - (\frac{1}{\eta_{ji}})^+ e^{-\eta_{ji}(\tau+t_{ij})} \right\} + \lambda_{ij}\,\beta\,E(X)^2 \\
\quad \times \frac{(\gamma_i+\gamma_j+\beta)}{(\eta_j-\gamma_j)}\left\{ \frac{e^{-\gamma_j(\tau+t_{ij})}}{(\gamma_i+\gamma_j)(\eta_i+\gamma_j)} - \frac{e^{-\eta_j(\tau+t_{ij})}}{(\eta_i+\eta_j)(\eta_j+\gamma_i)} \right\} = c_{ij,2}(\tau) \qquad \text{if } \tau \geq -t_{ij}.
\end{cases}
$$

Note that the factor $1 + \beta/(\gamma_i + \gamma_j)$ is the mean number of cells per storm that are common to the two sites, and that for $i = j$ the expression is given in Rodriguez-Iturbe et al. (1987a, Equation (4.12)).

As expected, the cross-covariance function of this Bartlett-Lewis based model has the same structure as the corresponding function of the Neyman-Scott based model considered in Section 4.2. Thus, the cross-covariance of the aggregated process, $C_{ij}^{(h)}(t_{ij}, k)$, of this model can be derived using the functions $W_m(\zeta, t_{ij}, k)$, $m = 1, 2, 3$, given in Equations (4.5)-(4.7), in each of the four cases specified in Section 4.3. Let us suppose that storms move with fixed direction, ϕ, and random speed, $\mathcal{V}$. As before, it is possible to write the difference between two displacement factors, $t_{ij} = t_i - t_j$, as a function of a measure of distance between the examined sites and the speed, $\mathcal{V}$.

The cross-covariance function of this model is calculated in the same way as in Appendix C. The site-dependent parameters are those of the rain cell duration, η_i, at a given site say i, $(i = 1, \ldots, n)$, and of the storm duration, γ_i. Now, suppose that γ_i is expressed as a function of the site's location and some features of the storm, such as its position and spatial extent, and let us assume that these features vary randomly between storms. Unlike the Neyman-Scott based models in which the mean number of cells per storm appears in the expressions of the second-order properties as a 'scalar' that can be modelled separately, in the Bartlett-Lewis based models, γ_i determines not only the mean number of rain cells per storm, experienced at a subset of sites including i, but is also involved in some denominators and exponents in the cross-covariance function (Equation (4.46)). So, even under very simple assumptions about the storm characteristics, not only would it not be possible to get an analytic closed form for the cross-covariance function for this Bartlett-Lewis model, but its expression is too complex for straightforward numerical evaluation.

A similar problem arises in the calculation of the probability that two sites are simultaneously dry during an arbitrary period of length h. The calculation of this property, even at the early stage when the displacement factors are assumed constant, involves integrals that cannot be evaluated numerically with reasonable accuracy. The idea of approximating some parts by series of polynomials has not simplified the expressions substantially.

We conclude that it is not possible to make much useful progress in developing a Bartlett-Lewis based multi-site model, analogous to the Neyman-Scott one considered in this chapter, and with the specifications described above.

4.11   Summary

In this chapter we have developed some multi-site models that aim to describe the stochastic nature of the rainfall process, as it is observed at a network of raingauge stations. The basic concept is that there is a master point process of storm or rain cell origins which is decomposed into a number of subprocesses, each one corresponding to a different site. A key issue in the modelling is that the dependences between the generated subprocesses should reflect the between-site correlations and the spatial structure of the rainfall event.

In making the sensible assumptions, we have in mind the results from the data analysis (Chapter 2) and the research that has already been done in this and in relevent areas (Sections 1.4 and 1.5).

The models we have developed have a Neyman-Scott structure in time, which is kept the same at all generated subprocesses. In order to allow different sites to experience the same rainfall event at different times, it is assumed that storms and their associated rain cells move with the same velocity (speed and direction), which remains constant throughout their lifetime but varies randomly between storms. Since the data available are rainfall sequences at a small number of distinct points in space, it is not possible to make inferences about the shape of the storm or the rain cells. Instead, the traces that are formed as the rainfall elements move across the examined area, are visualised as bands, parallel to the direction of the movement, with random width and position. So, the probability that a storm and/or a rain cell affects two sites depends on these spatial features of the associated band and on some measure of distance between the examined sites. An alternative approach, where the above probability is modelled in a deterministic way, is also studied. The cross-covariance function between rainfall intensities at two sites and the probability that these sites are simultaneously dry for an arbitrary period of given length, have been derived.

The models we built describe the main properties of the evolution of the rainfall process. However, because they are mathematically complex it is not possible to make any further assumption about the way some of their parameters, such as the cell duration, depend on topography. Ideally, we would like to assess the performance of a model, in which the cell duration scaling factors vary either randomly or deterministically with the elevation of the ground, but this is difficult because of computational constraints.

The multi-site model we developed has one site-dependent parameter when a stochastic approach is followed in modelling the dependencies between the rainfall intensities at various sites. When several sites are examined simultaneously, one way to reduce the dimensionality of the parameter space is to group the sites according to their location and to have the same rain cell duration parameter for nearby sites.

Chapter 5

Fitting of multi-site models

5.1 Methods applied

In this chapter, we start by outlining the method used for the parameter estimation and the optimisation procedure applied and then we present the results of fitting some multi-site models developed in Chapter 4, to rainfall data from the Brue catchment area.

The estimation of the model's parameters is based on the generalised method of moments, discussed in Section 3.4. In brief, let the m−dimensional parameter vector of the model be $\theta = (\theta_1, \ldots, \theta_m)$, and select a set of k features of the data, ψ_i ($i = 1, \ldots, k$), for which the corresponding model expressions, $\phi_i(\theta)$, have been obtained. We form the sum of squared differences between the $\{\psi_i\}$ and the $\{\phi_i(\theta)\}$, weighting each term appropriately, and we find the vector $\widehat{\theta}$ which minimises the objective function

$$\min_{\theta} \sum_{i=1}^{k} \left(\frac{\phi_i(\theta)}{\psi_i} - 1 \right)^2. \tag{5.1}$$

We now have the possibility of including more than m features in the minimisation ($k \geq m$). Assessment of fit of the model to the data is achieved by comparing the estimated values of features not used in the fitting procedure with those of the data.

A key issue when the model fitting is based on the GMM is the selection of the properties to be included in (5.1), which is essentially subjective. Some general guidelines are that the chosen features should a) contain sufficient information that all the model parameters can be reliably estimated, b) not be highly mutually correlated and c) have relatively small sampling errors. However, even when these guidelines are followed,

because of the high dimensionality of the parameter space and the non-linearity of the model expressions, the parameter estimates depend very much upon the initial values given to the numerical optimisation algorithm, which complicates the choice of an appropriate set of features still further. There is also a problem of parameter identification, since it is possible to get similar estimated properties for very different sets of parameter values. This is usually evidence of the existence of many local minima. We have done a substantial amount of analysis in order to find proper sets of features to be used in the parameter estimation procedure of some multi-site models developed in Chapter 4.

The choice of the optimisation method is crucial, especially when both the objective function and its derivatives are not available in an analytic closed form. The success of a numerical method depends mostly upon the number of calculations it makes at each iteration, the computer memory it requires to store the intermediate results, and the rate of convergence to the optimal. The method we use is called BFGS and is a quasi-Newton method. This means that both first and second order derivative information is used, but the Hessian matrix $H(\mathbf{x})$ is approximated via an iterative scheme, reducing considerably the computational requirements. The basic idea and the steps of this method (Georgiou and Vassiliou, 1993) are described below. Let us denote by $f(\mathbf{x})$, $\nabla f(\mathbf{x})$ and h_k the objective function, the vector of its first derivatives and the approximation of the inverse of the Hessian matrix at the k-th iteration, respectively, and by $\mathbf{x}^*$ the minimum of the objective function $f(\mathbf{x})$. The algorithm starts at the initial value $\mathbf{x}_0$, provided by the researcher, and sets $h_0 = I$, the identity matrix. Let the approximation of $\mathbf{x}^*$ after the k-th iteration be $\mathbf{x}_k$. The estimate of $H^{-1}(\mathbf{x}_k)$ is obtained iteratively by

$$h_k = h_{k-1} + \mathcal{A}_{k-1} - \mathcal{B}_{k-1},$$

where $\mathcal{A}_{k-1}$ and $\mathcal{B}_{k-1}$ are particular functions of $\mathbf{x}_{k-1}$, $f(\mathbf{x}_{k-1})$ and h_{k-1}. The direction of search for the next approximation of the minimum is $d_k = -h_k \nabla f(\mathbf{x}_k)$, and the size of the step, s_k, along d_k, is obtained via minimisation of $f(\mathbf{x}_k + s d_k)$ with respect to s, using an iterative search method, the golden section. Then, x_{k+1} is given by

$$x_{k+1} = x_k - s_k\, h_k\, \nabla f(\mathbf{x}_k). \tag{5.2}$$

Convergence criteria, applied to each coordinate separately and to the objective function, determine the termination of the algorithm. The BFGS method has a hyper-linear

convergence rate, that is, there exists a sequence of non-negative real numbers $\{a_k\}_{k=1}^{\infty}$ such that for all $k \geq k_0$, where $k_0 > 0$, we have $\| \mathbf{x}_{k+1} - \mathbf{x}^* \| \leq a_k \| \mathbf{x}_k - \mathbf{x}^* \|$.

The model fitting analysis has been performed using the computer software **Gauss 3.0** and its procedure *OPTIMUM*.

5.2 Model fitting to three sites

As the multi-site model development was progressing, the results of fitting the various intermediate models to rainfall data provided useful information about the dependencies and the assumptions that should be made. As an example, at an early stage of the modelling we studied a Neyman-Scott based model in which storm and rain cells affect two sites and we assumed that the evolutions of the process at the two sites are identical, with the only exception being a different scaling of the rain cell durations at the two sites. When only the temporal properties of the model at a single site are studied, this idea is very similar (both conceptually and mathematically) to having a single-site model, that is fitted to more than one site simultaneously using only marginal properties. However, although in the single-site Neyman-Scott based model, both the Poisson and the geometric distribution for the number of rain cells per storm give satisfactory results, the conclusion we came to, after fitting the above simple multi-site model, is that the Poisson distribution is much superior to the geometric one. Similarly, at a later stage, the fitting results suggested that a multi-site model in which the durations of the same rain cell at different sites are dependent and scaled versions of the same random variable, is more successful in reproducing observed properties of the data, as compared to models with either independent or identical rain cell durations.

Initially, the models were fitted to raingauge data from only two sites, but as they became more complicated, information about the behaviour of the rainfall process at more than two locations was needed for reliable estimation of the parameters. In this section, we present the results of fitting two multi-site models to three sites simultaneously, using monthly sequences of rainfall intensities of hourly resolution.

The models we consider are the one with random rain cell characteristics (Section 4.8), denoted by $\mathcal{M}_1$, and the one in which the probability that a cell affects a given site is a model parameter (Section 4.5), denoted by $\mathcal{M}_2$. In both models, storm and rain cells

move with the same random speed. In general, the examined sites should be chosen to provide as much information as possible about the spatial structure of the rainfall events. A good scheme is one where the sites form a triangle with sides of similar lengths, which covers a large part of the catchment area. Obviously, the choice of the sites influences the estimates of some parameters, especially when there is a lot of spatial variation between the observed properties.

Another crucial factor in the parameter estimation procedure is the choice of the model properties to be included in the objective function (5.1). A selection of some marginal properties of the three sites and some joint properties of the three pairs is required for a satisfactory representation of the spatial structure of the rainfall process. Also, properties at various levels of aggregation should be used in order to capture the temporal behaviour of the process. In general, observed properties of sequences of rainfall intensities aggregated over small time intervals, say 1 hour, provide more information about the within storm structure, while the between storms dependencies are captured by rainfall data aggregated over longer periods.

Based on our experience of the single-site model fitting, the marginal properties of each site that are included in this analysis are the mean, variance, lag one correlation and probability that an interval is dry at the 1 hour level of aggregation, and the variance at the 24 hours level of aggregation. The choice of an appropriate set of joint properties is more difficult, due to the absence of any previous experience. Initially, from each pair of sites, we used two zero lag cross-correlations, one at an aggregation level of 1 hour and one of higher level, for instance 6, or 12, or 24 hours, but the fitting results were poor. Then the latter statistic was replaced by the lag one cross-correlation of hourly data, which seemed to provide additional information about the movement of rainfall event, and improved significantly the estimated properties. Thus, throughout the analysis presented here, we use the lag zero and lag one cross-correlation for each of the three pairs of sites. The same features are used in the parameter estimation procedure of both models $\mathcal{M}_1$ and $\mathcal{M}_2$.

In this section, we present the fitting results for June 1994 and December 1994, representative summer and winter months, respectively.

5.2.1 Parameter estimates

Initially, we specify the distribution of some model variables. For both models, it is assumed that for a particular rain cell at a particular site, the intensity and duration are independent variables and that intensities of cells with common origin but at different sites are identical. Throughout the rest of this chapter the cell intensity, X, is assumed to have an exponential distribution, so that $E(X^2) = 2E(X)^2$. It is straightforward to remove this assumption, with the introduction of a second moment parameter for X. Also, as already mentioned in Chapter 4, the cross-correlation between the total rainfall intensities at two sites can be reduced by letting the rain cell intensities at different sites be either independent or scaled versions of the same random variable. Some of these cases may be considered if the fitting results suggest so.

We recall that in modelling the direction, ϕ, of the storm movement, we have assumed that it is a discrete random variable with a two-parameter probability density function (Equation 4.27). The data analysis presented in Chapter 2, shows that there exists a prevailing orientation for the rainfall movement (see Table 2.5), and there is more variability of this in June than in December. It is possible to use these results in fitting model $\mathcal{M}_1$, by keeping the mean direction of the storm movement, ϕ_0, fixed and equal to $\pi/8$, the midpoint of interval $[0, \pi/4]$. Then, from the fitting procedure we will get an estimate of the concentration parameter, κ, around ϕ_0.

The parameter estimates of models $\mathcal{M}_1$ and $\mathcal{M}_2$ are given in Tables 5.1 and 5.3, respectively. Of all the parameters, the most robust to choice of initial values, is the storm arrival rate, λ, which in our experience always takes the same value for both multi-sites models. In addition, these estimates of λ are the same as those obtained from fitting the Neyman-Scott single site model (Rodriguez-Iturbe et al., 1987a) to each of the three sites separately. The estimates of β and $E(X)$ for models $\mathcal{M}_1$ and $\mathcal{M}_2$ are in general fairly similar, but not the same, and also they vary slightly from the corresponding values obtained from the single site fitting.

These results have motivated us to split the fitting procedure into two parts. In the first part, the relates to $\mathcal{M}_1$ single-site model is fitted to all three sites simultaneously, and its parameters are estimated using only marginal properties. If n sites are fitted, then the number of parameters to be estimated at this stage is $4 + n$, since the model has

Model parameters	Estimate for	
	June 94	December 94
Storm arrival rate, λ	$0.006\ \mathrm{hr}^{-1}$	$0.020\ \mathrm{hr}^{-1}$
Mean rain cell intensity, $E(X)$	$0.81\ \mathrm{mm/h}$	$0.97\ \mathrm{mm/hr}$
Mean temporal displacement of cells from storm origin, β^{-1}	$5.78\ \mathrm{hr}$	$4.37\ \mathrm{hr}$
Mean no. of rain cells per storm, $E(C)$	21.1	14.9
Mean speed of storm/cell movement, $E(\mathcal{V})$	$15.3\ \mathrm{km/h}$	$8.4\ \mathrm{km/hr}$
Std. dev. of storm/cell movement	3.4	2.8
Mean width of a rain cell, $2\rho^{-1}$	$14\ \mathrm{km}$	$45.5\ \mathrm{km}$
Mean distance of cell from storm axis, β_c^{-1}	$1.5\ \mathrm{km}$	$1.4\ \mathrm{km}$
Mean direction of storm movement, ϕ_0	$\pi/8$	$\pi/8$
Concentration parameter of direction, κ	0.43	0.37
Mean cell duration site i, η_i^{-1}	$37.3\ \mathrm{min}$	$34.6\ \mathrm{min}$
Mean cell duration site j, η_j^{-1}	$27.0\ \mathrm{min}$	$49.9\ \mathrm{min}$
Mean cell duration site k, η_k^{-1}	$43.6\ \mathrm{min}$	$41.2\ \mathrm{min}$

Table 5.1: Parameter estimates of model $\mathcal{M}_1$. The data used are from sites: $i = 13$, $j = 42$, $k = 30$ in June, and $i = 5$, $j = 42$, $k = 29$ in December.

one site-dependent parameter, the scalar of the cell duration. At the next step, we use the values of λ, β and $E(X)$ obtained at the first stage and estimate all the remaining $6 + n$ parameters, using both marginal and joint properties. Note that when the rain cell depth has an exponential distribution, $E(X)$ cancels out in the expression of the cross-correlation. The model features that are included in the objective function are the same for the two methods, so that a comparison of the results can be made directly. An advantage of the two-step procedure is that in the second stage the dimension of the parameter space is reduced by three, which speeds up computation especially when n is large. A disadvantage is that this reduction usually limits the searching space for the minimum point, but since our objective function is highly nonlinear it is practically

Derived model parameters	June 94	December 94
P(cell affects site i)	0.467	0.750
P(cell affects site j)	0.453	0.739
P(cell affects site k)	0.468	0.750
P(cell affects sites i and j)	0.300	0.634
P(cell affects sites i and k)	0.454	0.750
P(cell affects sites j and k)	0.309	0.634

Table 5.2: Additional parameters of model $\mathcal{M}_1$, estimated using the values of the previous table. The sites examined are: $i = 13$, $j = 42$, $k = 30$ in June, and $i = 5$, $j = 42$, $k = 29$ in December.

impossible to find the global optimum. When three sites are examined simultaneously, the two-step procedure produces estimates of the properties that are very similar to the ones obtained with the one-step fitting procedure, and for this reason, they are not presented here.

With reference to Tables 5.1 and 5.3, the means and standard deviations of the speed of the storm/cell movement for the two models are fairly close, which indicates that the parameterisation of the speed is fairly satisfactory. For model $\mathcal{M}_1$, the mean width of the rainfall band, associated with each rain cell, is greater than the largest distance between the three examined sites. However, because the mean distance of the cell from the storm axis is about 1.5 km, the mean value of the probability that a rain cell affects a site is less than one (Table 5.2). It is very interesting to note that this probability takes similar values at all three sites, as a result of the spatially homogeneous structure of the process. In contrast, in model $\mathcal{M}_2$ these probabilities are site-dependent parameters, they are not related to any spatial characteristic of the process and, in general, their values are very different depending on the marginal properties used in the fitting procedure.

Concerning the mean number of cells per storm, the two multi-site models produce on average more rain cells per storm, compared to the Neyman-Scott single site model, but each one affects a site with a certain probability, and thus the expected number of

Model parameters		Estimate for	
		June 94	**December 94**
Storm arrival rate, λ		$0.006\ \mathrm{hr}^{-1}$	$0.020\ \mathrm{hr}^{-1}$
Mean rain cell intensity, $E(X)$		$0.84\ \mathrm{mm/h}$	$1.32\ \mathrm{mm/hr}$
Mean temporal displacement of cells from storm origin, β^{-1}		$7.46\ \mathrm{hr}$	$4.93\ \mathrm{hr}$
Mean no. of rain cells per storm, $E(C)$		22.0	26.3
Mean speed of storm/cell movement, $E(\mathcal{V})$		$14.3\ \mathrm{km/h}$	$9.1\ \mathrm{km/hr}$
Std. dev. of storm/cell movement		4.2	3.1
Probability that a cell affects	site i, q_i	0.264	0.377
	site j, q_j	0.268	0.359
	site k, q_k	0.671	0.998
t_c		0.119	0.057
Mean cell duration at	site i, η_i^{-1}	59.1 min	28.2 min
	site j, η_j^{-1}	38.2 min	42.2 min
	site k, η_k^{-1}	28.5 min	14.1 min

Table 5.3: Parameter estimates of model $\mathcal{M}_2$. The data used are from sites: $i = 13$, $j = 42$, $k = 30$ in June, and $i = 5$, $j = 42$, $k = 29$ in December.

these rain cells that each site experiences is reduced.

Having fitted model $\mathcal{M}_1$ to data from different months, we noticed that in general, during the dry summer months, storms arrive less frequently than in winter, and they consist of more rain cells, which have a shorter mean duration, intensity and spatial extent. Storm and rain cells move on average faster in summer than in winter.

5.2.2 Assesment of fit

A successful multi-site model should be capable of describing the main between-site dependencies of the rainfall process and also of preserving its temporal behaviour at a fixed location. The adequacy of the fit is assessed via comparison between observed and es-

timated properties, both marginal and joint, that have not been used in the parameter estimation procedure.

Marginal properties

With reference to tables 5.4 and 5.6, both multi-site models reproduce fairly accurately most of the marginal properties of all three sites. This is achieved by having the same storm structure at all sites and letting only the number of rain cells per storm that affect a given site and the mean cell duration vary with the location. Models $\mathcal{M}_1$ and $\mathcal{M}_2$ share the same advantages and disadvantages as the corresponding single-site model, which has a Neyman-Scott structure in time (Rodriguez-Iturbe et al., 1987a). That is, their estimates of the first and second order statistics are very close to the observed ones, but the probability of zero rainfall during an arbitrary time interval tends to be systematically overestimated. Because none of the between-site properties of the multi-site models is given by an analytic closed form, the randomisation of the rain cell duration parameter, η, in a way similar to that applied by Rodriguez-Iturbe et al. (1988) that would probably improve the fitting, is computationally implausible. Another drawback of these types of models is that they can not capture some seasonal characteristics of the rainfall sequence, such as the daily effect that is usually present in summer weather systems and causes substantial correlations at relatively high levels of aggregation. By contrast, during winter months the correlation of the rainfall intensity during two time intervals is a decreasing function of the time lag between the intervals, and this decaying pattern is well reproduced by the single-site model and its extensions, models $\mathcal{M}_1$ and $\mathcal{M}_2$.

Joint properties

The between-site properties that are estimated (via numerical methods) for the assessment of fit, are the cross-correlation function at various time lags and the probability that two sites are simultaneously dry during an arbitrary time interval of given length. With reference to tables 5.5 and 5.7, we notice that model $\mathcal{M}_1$ is better than $\mathcal{M}_2$ in reproducing the observed properties of the data. Indeed, for December 1994, and in general for most rainy months, the zero lag cross-correlation of $\mathcal{M}_1$, for all three pair

of sites and at all levels of aggregation, is estimated very accurately. The non-zero lag statistics are systematically slightly underestimated, while the probability that both sites are dry is well estimated at small levels of aggregation but is overestimated for time intervals of 12 and 24 hours. This however is related to, and might be caused by, the poor performance of the model concerning the marginal probability of a dry interval. Similar remarks apply to the fitting results of $\mathcal{M}_1$, for June 1994. An additional comment is that the observed lag one cross-correlation function, like the marginal one, increases for large levels of aggregation, and this behaviour is not captured by the theoretical function. Also, the probability of no rainfall at two sites, is reproduced quite satisfactorily not only for small levels of aggregation but also for large ones. Concerning the fitting results of model $\mathcal{M}_2$, its main disadvantage is that it does not describe very successfully the dependencies between the rainfall intensity at two sites, as they are summarised by the cross-correlation functions.

level of aggregation		Mean	Probability of no rain	Variance	Correlation		
					lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	**0.032**	**0.943**	**0.041**	**0.576**	0.228	0.228
	$\mathcal{M}_1$	**0.031**	**0.941**	**0.043**	**0.545**	0.280	0.199
	$\mathcal{M}_2$	**0.030**	**0.948**	**0.046**	**0.620**	0.326	0.206
6 h	$\mathcal{H}$	0.195	0.875	0.690	0.424	-0.030	0.101
	$\mathcal{M}_1$	0.185	0.882	0.681	0.305	0.098	0.035
	$\mathcal{M}_2$	0.181	0.889	0.780	0.287	0.101	0.045
12 h	$\mathcal{H}$	0.390	0.783	1.446	0.470	0.388	0.314
	$\mathcal{M}_1$	0.370	0.842	1.777	0.205	0.024	0.003
	$\mathcal{M}_2$	0.362	0.847	2.008	0.208	0.037	0.007
24 h	$\mathcal{H}$	0.780	0.667	**5.129**	0.432	0.008	-0.007
	$\mathcal{M}_1$	0.741	0.779	**4.282**	0.106	0.002	0.000
	$\mathcal{M}_2$	0.723	0.783	**4.850**	0.119	0.004	0.000

level of aggregation		Mean	Probability of no rain	Variance	Correlation		
					lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	**0.025**	**0.950**	**0.019**	**0.431**	0.388	0.232
	$\mathcal{M}_1$	**0.021**	**0.947**	**0.025**	**0.448**	0.212	0.161
	$\mathcal{M}_2$	**0.020**	**0.955**	**0.025**	**0.490**	0.205	0.133
6 h	$\mathcal{H}$	0.152	0.850	0.350	0.471	-0.022	0.185
	$\mathcal{M}_1$	0.129	0.884	0.341	0.286	0.094	0.033
	$\mathcal{M}_2$	0.119	0.892	0.349	0.253	0.096	0.043
12 h	$\mathcal{H}$	0.303	0.767	0.721	0.632	0.393	0.212
	$\mathcal{M}_1$	0.257	0.844	0.876	0.197	0.024	0.003
	$\mathcal{M}_2$	0.237	0.848	0.875	0.195	0.036	0.007
24 h	$\mathcal{H}$	0.607	0.667	**2.759**	0.413	0.018	0.005
	$\mathcal{M}_1$	0.515	0.781	**2.096**	0.103	0.002	0.000
	$\mathcal{M}_2$	0.474	0.784	**2.090**	0.114	0.004	0.000

level of aggregation		Mean	Probability of no rain	Variance	Correlation		
					lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	**0.039**	**0.940**	**0.059**	**0.450**	0.195	0.124
	$\mathcal{M}_1$	**0.036**	**0.939**	**0.054**	**0.587**	0.316	0.221
	$\mathcal{M}_2$	**0.037**	**0.925**	**0.048**	**0.496**	0.269	0.215
6 h	$\mathcal{H}$	0.232	0.850	0.785	0.252	0.001	0.072
	$\mathcal{M}_1$	0.215	0.881	0.903	0.314	0.100	0.035
	$\mathcal{M}_2$	0.221	0.853	0.749	0.365	0.155	0.069
12 h	$\mathcal{H}$	0.463	0.783	1.835	0.187	0.208	0.198
	$\mathcal{M}_1$	0.429	0.841	2.373	0.209	0.025	0.003
	$\mathcal{M}_2$	0.442	0.811	2.046	0.272	0.053	0.011
24 h	$\mathcal{H}$	0.927	0.700	**5.161**	0.281	0.029	-0.042
	$\mathcal{M}_1$	0.859	0.778	**5.736**	0.108	0.002	0.000
	$\mathcal{M}_2$	0.884	0.750	**5.207**	0.153	0.006	0.000

Table 5.4: June 1994. Marginal properties of sites 13 (top), 42 (middle), 30 (bottom). $\mathcal{H}$ are the observed values. Properties in **bold** have been used in the model fitting.

level of aggregation		probability of both sites dry	Cross-correlation			
			lag 0	lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	0.928	**0.618**	**0.494**	0.409	0.372
	$\mathcal{M}_1$	0.907	**0.723**	**0.302**	0.212	0.142
	$\mathcal{M}_2$	0.920	**0.582**	**0.486**	0.243	0.155
6 h	$\mathcal{H}$	0.833	0.840	0.522	-0.017	0.172
	$\mathcal{M}_1$	0.766	0.825	0.260	0.180	0.116
	$\mathcal{M}_2$	0.790	0.726	0.262	0.102	0.045
12 h	$\mathcal{H}$	0.750	0.827	0.660	0.617	0.479
	$\mathcal{M}_1$	0.671	0.859	0.182	0.102	0.053
	$\mathcal{M}_2$	0.696	0.771	0.201	0.037	0.007
24 h	$\mathcal{H}$	0.667	0.886	0.690	-0.009	-0.008
	$\mathcal{M}_1$	0.537	0.880	0.096	0.000	0.000
	$\mathcal{M}_2$	0.561	0.804	0.118	0.004	0.000

level of aggregation		probability of both sites dry	Cross-correlation			
			lag 0	lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	0.924	**0.494**	**0.426**	0.239	0.158
	$\mathcal{M}_1$	0.904	**0.548**	**0.402**	0.224	0.132
	$\mathcal{M}_2$	0.893	**0.478**	**0.499**	0.266	0.179
6 h	$\mathcal{H}$	0.800	0.766	0.311	-0.009	0.101
	$\mathcal{M}_1$	0.767	0.807	0.284	0.176	0.103
	$\mathcal{M}_2$	0.748	0.696	0.324	0.131	0.059
12 h	$\mathcal{H}$	0.750	0.806	0.361	0.277	0.177
	$\mathcal{M}_1$	0.673	0.855	0.194	0.087	0.002
	$\mathcal{M}_2$	0.658	0.756	0.247	0.047	0.009
24 h	$\mathcal{H}$	0.633	0.889	0.318	0.022	0.007
	$\mathcal{M}_1$	0.539	0.881	0.102	0.000	0.000
	$\mathcal{M}_2$	0.534	0.796	0.142	0.005	0.000

level of aggregation		probability of both sites dry	Cross-correlation			
			lag 0	lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	0.924	**0.666**	**0.387**	0.150	0.129
	$\mathcal{M}_1$	0.929	**0.704**	**0.394**	0.263	0.194
	$\mathcal{M}_2$	0.887	**0.520**	**0.458**	0.281	0.190
6 h	$\mathcal{H}$	0.825	0.716	0.234	-0.008	0.124
	$\mathcal{M}_1$	0.859	0.952	0.270	0.111	0.020
	$\mathcal{M}_2$	0.744	0.701	0.309	0.129	0.057
12 h	$\mathcal{H}$	0.733	0.722	0.292	0.384	0.374
	$\mathcal{M}_1$	0.805	0.974	0.184	0.065	0.004
	$\mathcal{M}_2$	0.655	0.757	0.235	0.045	0.009
24 h	$\mathcal{H}$	0.633	0.772	0.524	0.002	0.002
	$\mathcal{M}_1$	0.724	0.984	0.096	0.006	0.000
	$\mathcal{M}_2$	0.533	0.795	0.135	0.005	0.000

Table 5.5: June 1994. Cross-properties between sites 13 and 42 (top), 42 and 30 (middle), 13 and 30 (bottom). $\mathcal{H}$ are the observed values. Properties in **bold** have been used in the model fitting.

level of aggregation		Mean	Probability of no rain	Variance	Correlation		
					lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	**0.133**	**0.836**	**0.216**	**0.487**	0.250	0.177
	$\mathcal{M}_1$	**0.127**	**0.832**	**0.229**	**0.558**	0.301	0.214
	$\mathcal{M}_2$	**0.123**	**0.841**	**0.247**	**0.477**	0.236	0.174
6 h	$\mathcal{H}$	0.798	0.581	3.362	0.223	0.145	0.016
	$\mathcal{M}_1$2	0.763	0.692	3.719	0.290	0.068	0.017
	$\mathcal{M}_2$	0.738	0.689	3.585	0.280	0.077	0.023
12 h	$\mathcal{H}$	1.597	0.435	8.185	0.206	0.071	0.029
	$\mathcal{M}_1$	1.526	0.603	9.598	0.172	0.011	0.001
	$\mathcal{M}_2$	1.477	0.597	9.177	0.178	0.015	0.001
24 h	$\mathcal{H}$	3.194	0.258	**22.517**	0.057	0.070	-0.026
	$\mathcal{M}_1$	3.051	0.472	**22.498**	0.083	0.000	0.000
	$\mathcal{M}_2$	2.954	0.466	**21.627**	0.089	0.001	0.000

level of aggregation		Mean	Probability of no rain	Variance	Correlation		
					lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	**0.176**	**0.835**	**0.427**	**0.452**	0.258	0.172
	$\mathcal{M}_1$	**0.181**	**0.819**	**0.399**	**0.656**	0.389	0.270
	$\mathcal{M}_2$	**0.175**	**0.831**	**0.438**	**0.585**	0.313	0.215
6 h	$\mathcal{H}$	1.053	0.573	6.256	0.331	0.203	0.048
	$\mathcal{M}_1$	1.085	0.688	7.348	0.311	0.071	0.018
	$\mathcal{M}_2$	1.053	0.688	7.264	0.294	0.078	0.023
12 h	$\mathcal{H}$	2.106	0.403	16.869	0.271	0.064	0.085
	$\mathcal{M}_1$	2.170	0.600	19.264	0.180	0.011	0.001
	$\mathcal{M}_2$	2.105	0.597	18.802	0.183	0.015	0.001
24 h	$\mathcal{H}$	4.213	0.226	**48.993**	0.081	0.093	0.034
	$\mathcal{M}_1$	4.340	0.470	**45.454**	0.086	0.000	0.000
	$\mathcal{M}_2$	4.211	0.466	**44.481**	0.091	0.001	0.000

level of aggregation		Mean	Probability of no rain	Variance	Correlation		
					lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	**0.165**	**0.798**	**0.284**	**0.477**	0.250	0.151
	$\mathcal{M}_1$	**0.151**	**0.826**	**0.302**	**0.606**	0.341	0.238
	$\mathcal{M}_2$	**0.163**	**0.777**	**0.279**	**0.478**	0.325	0.264
6 h	$\mathcal{H}$	0.989	0.484	4.397	0.207	0.159	0.044
	$\mathcal{M}_1$	0.908	0.690	5.209	0.299	0.070	0.018
	$\mathcal{M}_2$	0.978	0.626	4.519	0.365	0.106	0.031
12 h	$\mathcal{H}$	1.977	0.306	10.635	0.210	0.143	0.076
	$\mathcal{M}_1$	1.817	0.601	13.538	0.176	0.011	0.001
	$\mathcal{M}_2$	1.956	0.544	12.336	0.223	0.019	0.002
24 h	$\mathcal{H}$	3.955	0.165	**30.655**	0.118	0.044	-0.006
	$\mathcal{M}_1$	3.633	0.471	**31.828**	0.084	0.000	0.000
	$\mathcal{M}_2$	3.912	0.425	**30.177**	0.108	0.001	0.000

Table 5.6: December 1994. Marginal properties of sites 5 (top), 42 (middle), 29 (bottom). $\mathcal{H}$ are the observed values. Properties in **bold** have been used in the model fitting.

level of aggregation		probability of both sites dry	Cross-correlation			
			lag 0	lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	0.793	**0.778**	**0.637**	0.248	0.183
	$\mathcal{M}_1$	0.754	**0.772**	**0.561**	0.341	0.221
	$\mathcal{M}_2$	0.736	**0.522**	**0.733**	0.414	0.271
6 h	$\mathcal{H}$	0.500	0.957	0.259	0.152	0.034
	$\mathcal{M}_1$	0.517	0.923	0.298	0.135	0.025
	$\mathcal{M}_2$	0.469	0.835	0.382	0.094	0.028
12 h	$\mathcal{H}$	0.339	0.965	0.219	0.070	0.051
	$\mathcal{M}_1$	0.368	0.945	0.174	0.044	0.000
	$\mathcal{M}_2$	0.319	0.883	0.232	0.018	0.002
24 h	$\mathcal{H}$	0.161	0.977	0.059	0.109	-0.013
	$\mathcal{M}_1$	0.199	0.955	0.083	0.000	0.000
	$\mathcal{M}_2$	0.161	0.907	0.114	0.001	0.000

level of aggregation		probability of both sites dry	Cross-correlation			
			lag 0	lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	0.753	**0.743**	**0.447**	0.232	0.158
	$\mathcal{M}_1$	0.728	**0.721**	**0.358**	0.128	0.102
	$\mathcal{M}_2$	0.656	**0.605**	**0.547**	0.376	0.221
6 h	$\mathcal{H}$	0.411	0.902	0.261	0.153	0.031
	$\mathcal{M}_1$	0.508	0.920	0.229	0.123	0.029
	$\mathcal{M}_2$	0.443	0.818	0.348	0.214	0.029
12 h	$\mathcal{H}$	0.210	0.912	0.224	0.041	0.079
	$\mathcal{M}_1$	0.362	0.943	0.137	0.023	0.000
	$\mathcal{M}_2$	0.333	0.859	0.216	0.018	0.002
24 h	$\mathcal{H}$	0.132	0.947	0.036	-0.005	0.069
	$\mathcal{M}_1$	0.196	0.954	0.066	0.000	0.000
	$\mathcal{M}_2$	0.208	0.881	0.106	0.001	0.000

level of aggregation		probability of both sites dry	Cross-correlation			
			lag 0	lag 1	lag 2	lag 3
1 h	$\mathcal{H}$	0.769	**0.885**	**0.513**	0.239	0.177
	$\mathcal{M}_1$	0.787	**0.811**	**0.336**	0.246	0.112
	$\mathcal{M}_2$	0.666	**0.601**	**0.558**	0.491	0.241
6 h	$\mathcal{H}$	0.419	0.952	0.193	0.126	0.029
	$\mathcal{M}_1$	0.676	0.921	0.223	0.113	0.015
	$\mathcal{M}_2$	0.446	0.825	0.355	0.100	0.030
12 h	$\mathcal{H}$	0.242	0.958	0.179	0.069	0.033
	$\mathcal{M}_1$	0.590	0.967	0.134	0.003	0.000
	$\mathcal{M}_2$	0.335	0.866	0.221	0.019	0.002
24 h	$\mathcal{H}$	0.132	0.971	0.025	-0.001	0.013
	$\mathcal{M}_1$	0.471	0.985	0.065	0.000	0.000
	$\mathcal{M}_2$	0.209	0.888	0.108	0.001	0.000

Table 5.7: December 1994. Cross-properties between sites 5 and 42 (top), 42 and 29 (middle), 5 and 29 (bottom). $\mathcal{H}$ are the observed values. Properties in **bold** have been used in the model fitting.

Chapter 6

Conclusions

In this thesis, we have studied the evolution of rainfall at a single site and over a network of sites using clustered point process based models.

We have developed some single-site models with dependent rain cell duration, L, and intensity, X. The analysis that is carried out shows that the 6-parameter Bartlett-Lewis model in which the mean cell intensity is given by $E(X|L) = fe^{-cL}$ produces accurate estimates of the first and second order properties of the aggregated rainfall intensity, at several levels of aggregation, and also improves the prediction of the probability that an arbitrary time interval of fixed length is dry, as compared to the original model with indepedent cell duration and intensity.

The multi-site models we have considered are based on the idea that when rainfall is studied over a network of sites, a storm or a rain cell may be categorised according to the subset of sites which it affects. In particular, we assume that the point process of storm origins is the same at all sites, while the process of rain cell origins, that has a Neyman-Scott structure in time, is decomposed to generate a number of subprocesses, one for each examined site. In modelling the lag between the time that different sites experience the same rainfall elements, it is assumed that all cells within a storm and the storm itself move with the same random velocity, which remains constant throughout their lifetime. Associated with each rain cell is a band, parallel to the direction of the movement, whose width and position are random, and which determines the probability that a particular cell affects a given site. This model preserves the structure of the Neyman-Scott model for its marginal processes at each site, and we have derived its cross-correlation function

and the probability that two sites are simultaneously dry. The overall performance of this model, when it is fitted simultaneously to data from three sites, is satisfactory and encouraging.

Appendix A

Tables of monthly summary statistics

As part of a routine data analysis, tables of monthly summary statistics are calculated and stored in an archive, accessible by all members of the project. These tables include, for each gauge, properties like the monthly total and hourly mean rainfall intensity, and at levels of aggegation of 1 hour, 6 hours and 24 hours the probability that a time interval is rainy, the standard deviation and the conditional mean and standard deviation. The last two properties are calculated using non-zero observations only. We present part of the tables with summary statistics for June 1994 and December 1994.

SUMMARY STATISTICS, JUNE 94.

SITE	Rainfall (mm) total	mean 1h	Probability wet 1 h	6 h	24 h	Unconditional std. dev. 1 h	6 h	24 h	Conditional Mean 1 h	6 h	24 h	Conditional std. dev. 1 h	6 h	24 h
2	16.600	0.023	0.049	0.117	0.300	0.178	0.589	1.500	0.474	1.186	1.844	0.659	1.315	2.262
3	16.400	0.023	0.051	0.150	0.367	0.123	0.512	1.424	0.443	0.911	1.491	0.326	1.022	2.030
4	16.600	0.023	0.042	0.125	0.333	0.183	0.595	1.458	0.553	1.107	1.660	0.715	1.326	2.130
5	18.200	0.025	0.050	0.133	0.333	0.185	0.645	1.715	0.506	1.138	1.820	0.662	1.414	2.573
6	16.800	0.023	0.050	0.125	0.300	0.164	0.595	1.590	0.467	1.120	1.867	0.573	1.316	2.448
7	16.800	0.023	0.050	0.117	0.300	0.163	0.579	1.521	0.467	1.200	1.867	0.572	1.267	2.296
8	18.000	0.025	0.047	0.125	0.333	0.167	0.605	1.656	0.529	1.200	1.800	0.572	1.292	2.463
9	19.200	0.027	0.050	0.108	0.267	0.190	0.683	1.744	0.533	1.477	2.400	0.675	1.536	2.680
10	14.600	0.020	0.072	0.167	0.333	0.089	0.365	1.170	0.281	0.730	1.460	0.189	0.594	1.640
11	19.800	0.027	0.053	0.117	0.300	0.156	0.647	1.753	0.521	1.414	2.200	0.453	1.349	2.618
12	20.200	0.028	0.047	0.117	0.300	0.177	0.714	1.894	0.594	1.443	2.244	0.569	1.590	2.904
13	23.400	0.032	0.057	0.125	0.333	0.202	0.830	2.265	0.571	1.560	2.340	0.641	1.841	3.426
14	23.600	0.033	0.068	0.192	0.300	0.172	0.696	1.869	0.482	1.026	2.622	0.468	1.295	2.614
15	31.800	0.044	0.061	0.158	0.300	0.266	1.060	2.838	0.723	1.674	3.533	0.817	2.177	4.255
16	22.000	0.031	0.054	0.133	0.267	0.213	0.705	1.820	0.564	1.375	2.750	0.733	1.444	2.622
17	22.400	0.031	0.057	0.158	0.400	0.225	0.712	1.911	0.546	1.179	1.867	0.782	1.424	2.654
18	20.800	0.029	0.054	0.133	0.333	0.179	0.695	1.855	0.533	1.300	2.080	0.570	1.468	2.726
19	23.600	0.033	0.057	0.150	0.300	0.211	0.873	2.238	0.576	1.311	2.622	0.683	1.903	3.447
20	22.600	0.031	0.056	0.117	0.267	0.183	0.779	2.172	0.565	1.614	2.825	0.551	1.701	3.442

Figure A.1: Part of a table with summary statistics of June 1994.

SUMMARY STATISTICS, DECEMBER 94.

SITE	Rainfall (mm) total	Rainfall (mm) mean 1h	Probability wet 1 h	Probability wet 6 h	Probability wet 24 h	Unconditional st. dev. 1 h	Unconditional st. dev. 6 h	Unconditional st. dev. 24 h	Conditional Mean 1 h	Conditional Mean 6 h	Conditional Mean 24 h	Conditional st. dev. 1 h	Conditional st. dev. 6 h	Conditional st. dev. 24 h
2	95.200	0.128	0.151	0.395	0.774	0.465	1.827	4.620	0.850	1.943	3.967	0.906	2.483	4.901
3	101.000	0.136	0.153	0.387	0.645	0.497	1.941	5.096	0.886	2.104	5.050	0.974	2.649	5.586
4	120.000	0.161	0.184	0.468	0.839	0.547	2.164	5.524	0.876	2.069	4.615	1.000	2.781	5.740
5	99.000	0.133	0.164	0.419	0.742	0.464	1.834	4.745	0.811	1.904	4.304	0.875	2.432	5.057
6	95.000	0.128	0.142	0.371	0.710	0.486	1.956	5.200	0.896	2.065	4.318	0.983	2.762	5.717
7	115.000	0.155	0.169	0.419	0.742	0.563	2.125	5.468	0.913	2.212	5.000	1.086	2.817	5.818
8	107.800	0.145	0.172	0.435	0.806	0.486	1.941	5.009	0.842	1.996	4.312	0.887	2.530	5.245
9	125.200	0.168	0.191	0.468	0.806	0.617	2.381	6.002	0.882	2.159	5.008	1.170	3.104	6.310
10	105.800	0.142	0.160	0.395	0.774	0.567	2.137	5.101	0.889	2.159	4.408	1.159	2.955	5.406
11	115.200	0.155	0.168	0.427	0.774	0.584	2.221	5.677	0.922	2.174	4.800	1.149	2.973	6.036
12	105.800	0.142	0.167	0.427	0.806	0.485	1.934	5.030	0.853	1.996	4.232	0.897	2.544	5.283
13	107.600	0.145	0.167	0.427	0.839	0.501	1.980	5.214	0.868	2.030	4.138	0.938	2.610	5.445
14	98.000	0.132	0.156	0.387	0.742	0.466	1.859	5.168	0.845	2.042	4.261	0.890	2.525	5.596
15	124.000	0.167	0.199	0.460	0.774	0.526	2.088	5.851	0.838	2.175	5.167	0.911	2.631	6.181
16	108.600	0.146	0.176	0.435	0.742	0.488	1.941	5.007	0.829	2.011	4.722	0.887	2.523	5.294
17	117.200	0.158	0.168	0.427	0.742	0.558	2.141	5.520	0.938	2.211	5.096	1.058	2.815	5.862
18	111.400	0.150	0.171	0.435	0.839	0.535	2.090	5.233	0.877	2.063	4.285	1.018	2.761	5.449
19	113.600	0.153	0.176	0.444	0.806	0.513	2.067	5.574	0.867	2.065	4.544	0.937	2.694	5.876
20	113.200	0.152	0.179	0.468	0.774	0.529	2.132	5.994	0.851	1.952	4.717	0.985	2.774	6.433

Figure A.2: Part of a table with summary statistics of December 1994.

120

Appendix B

Properties of the aggregated process of single site models

The functions that need to be integrated for the calculation of the covariance of the aggregated process are of the following three types:

$$\int_{-h}^{h} (h - |\tau|) \, e^{-m \, (kh+\tau)} d\tau \;\; = \;\; \frac{g_1(m)}{m^2},$$

$$\int_{-h}^{h} (h - |\tau|) \, (kh + \tau) \, e^{-m \, (kh+\tau)} d\tau \;\; = \;\; \frac{g_1(m)}{m^3} \, (mhk + 2) - \frac{e^{-mkh}}{m^2} \, h \, (e^{mh} - e^{-mh}),$$

$$\int_{-h}^{h} (h - |\tau|) \, (kh + \tau)^2 \, e^{-m \, (kh+\tau)} d\tau \;\; = \;\; \frac{g_1(m)}{m^4} \, (m^2 h^2 (k^2 + 1) + 4mhk + 6)$$
$$+ \frac{e^{-mkh}}{m^4} 2mh \, \{mh - (mhk + 2) \, (e^{mh} - e^{-mh})\},$$

where $g_1(m) = (1 - e^{-mh})^2 \, e^{-m(k-1)h}$.

The variance, $\mathrm{var}\{Y_{t_i}^{(h)}\}$, involves the derivation of the following integrals:

$$\int_{0}^{h} (h - \tau) \, e^{-m\tau} d\tau \;\; = \;\; \frac{1}{m^2} \, (e^{-mh} - 1 + mh),$$

$$\int_{0}^{h} (h - \tau) \, \tau \, e^{-m\tau} d\tau \;\; = \;\; \frac{1}{m^3} \, \{mh(e^{-mh} + 1) + 2(e^{-mh} - 1)\},$$

$$\int_{0}^{h} (h - \tau) \, \tau^2 \, e^{-m\tau} d\tau \;\; = \;\; \frac{1}{m^4} \, \{e^{-mh} \, (m^2 h^2 + 4mh + 6) + 2mh - 6\}.$$

We give the second order properties of the aggregated process for the Bartlett-Lewis based rainfall models with dependent cell duration and intensity. When $E(X|L) = f e^{-cL}$ we have:

$$E\{Y_{t_i}^{(h)}\} = h\,E\{Y(t)\} = h\,\lambda\,E(C)\,\frac{f\,\eta}{(c+\eta)^2},$$

$$\mathrm{var}\{Y_{t_i}^{(h)}\} = \frac{4\,\lambda\,\eta\,E(C)\,f^2}{(2c+\eta)^4}\{e^{-(2c+\eta)h} - 1 + (2c+\eta)h\}$$
$$+ \frac{2\,\lambda\,\beta\,\gamma\,\eta^2\,E(C)\,f^2}{(c+\eta)^2\{\gamma^2 - (c+\eta)^2\}}\left\{\frac{e^{-(c+\eta)h} - 1 + (c+\eta)h}{(c+\eta)^3} - \frac{e^{-\gamma h} - 1 + \gamma h}{\gamma^3}\right\},$$

$$\mathrm{cov}(Y_{t_i}^{(h)}, Y_{t_i+k}^{(h)}) = \frac{2\,\lambda\,\eta\,E(C)\,f^2}{(2c+\eta)^4}\{(1 - e^{-(2c+\eta)h})^2\,e^{-(2c+\eta)(k-1)h}\} + \frac{\lambda\,\beta\,\gamma\,\eta^2\,E(C)\,f^2}{(c+\eta)^2\{\beta^2 - (c+\eta)^2\}}$$
$$\times\left\{\frac{(1 - e^{-(c+\eta)h})^2\,e^{-(c+\eta)(k-1)h}}{(c+\eta)^3} - \frac{(1 - e^{-\gamma h})^2\,e^{-\gamma(k-1)h}}{\gamma^3}\right\},$$

where the last equation holds for $k \geq 1$.

When $E(X|L) = f\,L\,e^{-cL}$ the properties of the aggregated process are:

$$E\{Y_{t_i}^{(h)}\} = h\,\lambda\,E(C)\,\frac{2\eta\,f}{(c+\eta)^3},$$

$$\mathrm{var}\{Y_{t_i}^{(h)}\} = 2\,\lambda\,\eta\,E(C)\,f^2\frac{g_3(1) - 4\{5 - 3h(2c+\eta)\}}{(2c+\eta)^6} + \frac{2\,\lambda\,\beta\,\eta^2\,E(C)\,f^2}{(c+\eta)^4\{(c+\eta)^2 - \gamma^2\}}$$
$$\times\left\{\frac{g_2(\gamma)}{\gamma^2}\{4(c+\eta)^2 - \gamma^2\} + \frac{g_2(c+\eta)\,\gamma\,\{5\gamma^2 - 11(c+\eta)^2\}}{2\,(c+\eta)^3\,\{(c+\eta)^2 - \gamma^2\}}\right.$$
$$\left. - \frac{3\,\gamma}{2(c+\eta)^3}\{h(c+\eta)\,(e^{-(c+\eta)h} + 1) + 2\,(e^{-(c+\eta)h} - 1)\}\right\},$$

$$\mathrm{cov}\{Y_{t_i}^{(h)}, Y_{t_i+k}^{(h)}\} = 2\lambda\eta E(C)f^2\frac{\{g_3(k+1) - 2\,g_3(k) + g_3(k-1)\}}{(2c+\eta)^6} + \frac{\lambda\,\beta\,\eta^2\,E(C)f^2}{(c+\eta)^4\,\{(c+\eta)^2 - \gamma^2\}}$$
$$\left\{\frac{g_1(\gamma)}{\gamma^2}\{4(c+\eta)^2 - \gamma^2\} + \frac{g_1(c+\eta)\,\gamma\,\{11\gamma^2 - 17(c+\eta)^2\}}{2(c+\eta)^3\,\{(c+\eta)^2 - \gamma^2\}} - \frac{3\,\gamma\,g_4(c+\eta)}{2\,(c+\eta)^2}\right\},$$

where

$g_2(m) = e^{-mh} - 1 + mh,$

$g_3(k) = \{20 + 8hk(2c+\eta) + h^2k^2(2c+\eta)^2\}e^{-hk(2c+\eta)},$

$g_4(m) = \{h(k+1)e^{-mh} - 2hk + h(k-1)e^{mh}\}\,e^{-mhk},$

and the last expression is valid for $k \geq 1$.

For the Neyman-Scott based rainfall models, the second-order properties are derived from the corresponding expressions of the Bartlett-Lewis based models after replacement of parameter γ by β and of $E(C)$, in the second term of the variance and covariance function, by $E(C^2 - C)/2$.

Appendix C

Cross-correlation function for model with random $\mathcal{W}$

Initially, we define the functions $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ as follows:

$$\mathcal{G}_1(\epsilon_1, \epsilon_2) = \int\limits_{\epsilon_1}^{\epsilon_2} f_{\mathcal{W}}(w)\, dw = \gamma(\nu\epsilon_2, \theta) - \gamma(\nu\epsilon_1, \theta) \tag{C.1}$$

where $\gamma(t, \theta) = \frac{1}{\Gamma(\theta)} \int\limits_0^t u^{\theta-1} e^{-u} du$ is the incomplete gamma function,

$$\mathcal{G}_2(\epsilon_1, \epsilon_2) = \int\limits_{\epsilon_1}^{\epsilon_2} w\, f_{\mathcal{W}}(w)\, dw = \frac{\theta}{\nu} \left\{ \gamma(\nu\epsilon_2, \theta+1) - \gamma(\nu\epsilon_1, \theta+1) \right\}, \tag{C.2}$$

and

$$\mathcal{G}_3(\zeta, \epsilon_1, \epsilon_2) = \int\limits_{\epsilon_1}^{\epsilon_2} e^{-\zeta w}\, f_{\mathcal{W}}(w)\, dw = \left(\frac{\nu}{\nu+\zeta} \right)^\theta \left\{ \gamma(\epsilon_2(\nu+\zeta), \theta) - \gamma(\epsilon_1(\nu+\zeta), \theta) \right\}. \tag{C.3}$$

Then, it is straightforward using Equations (C.1)-(C.3), to integrate with respet to w the functions $W_m(\zeta, d_{ij}w, k)$, $m = 1, 2, 3$, in (4.5)-(4.7), over an interval $[\epsilon_1, \epsilon_2)$, when $d_{ij}\, w + kh$ is positive and negative.

- When $d_{ij} + kh > 0$, then we get

$$
\begin{aligned}
W_{1+}(\zeta, d_{ij}, k, \epsilon_1, \epsilon_2) &= \frac{1}{\zeta^2}(e^{-\zeta h} - 2 + e^{\zeta h}) \int\limits_{\epsilon_1}^{\epsilon_2} e^{-\zeta(d_{ij}\, w + kh)}\, f_{\mathcal{W}}(w)\, dw \\
&= \frac{e^{-\zeta kh}}{\zeta^2}(e^{-\zeta h} - 2 + e^{\zeta h})\, \mathcal{G}_3(\zeta d_{ij}, \epsilon_1, \epsilon_2),
\end{aligned}
$$

$$W_{2+}(\zeta, d_{ij}, k, \epsilon_1, \epsilon_2)$$

$$= \frac{1}{\zeta^2} \int_{\epsilon_1}^{\epsilon_2} \{e^{-\zeta(k-1)h} e^{\zeta d_{ij} w} - 1 + \zeta(k-1)h - \zeta d_{ij} w\} f_{\mathcal{W}}(w) \, dw$$

$$= \frac{1}{\zeta^2} \left\{ e^{-\zeta(k-1)h} \mathcal{G}_3(-\zeta d_{ij}, \epsilon_1, \epsilon_2) + \{\zeta(k-1)h - 1\} \mathcal{G}_1(\epsilon_1, \epsilon_2) - \zeta d_{ij} \mathcal{G}_2(\epsilon_1, \epsilon_2) \right\},$$

and

$$W_{3+}(\zeta, d_{ij}, k, \epsilon_1, \epsilon_2)$$

$$= \frac{1}{\zeta^2} \int_{\epsilon_1}^{\epsilon_2} \{(e^{-\zeta h} - 2) e^{-\zeta k h} e^{-\zeta d_{ij} w} + 1 - \zeta(k-1)h - \zeta d_{ij} w\} f_{\mathcal{W}}(w) \, dw$$

$$= \frac{1}{\zeta^2} \left\{ (e^{-\zeta h} - 2) e^{-\zeta k h} \mathcal{G}_3(\zeta d_{ij}, \epsilon_1, \epsilon_2) + \{1 - \zeta(k-1)h\} \mathcal{G}_1(\epsilon_1, \epsilon_2) - \zeta d_{ij} \mathcal{G}_2(\epsilon_1, \epsilon_2) \right\}.$$

- When $d_{ij} w + kh < 0$, then the three functions become

$$W_{1-}(\zeta, d_{ij}, k, \epsilon_1, \epsilon_2) = \frac{1}{\zeta^2}(e^{-\zeta h} - 2 + e^{\zeta h}) \int_{\epsilon_1}^{\epsilon_2} e^{\zeta(d_{ij} w + kh)} f_{\mathcal{W}}(w) \, dw$$

$$= \frac{e^{\zeta k h}}{\zeta^2}(e^{-\zeta h} - 2 + e^{\zeta h}) \mathcal{G}_3(-\zeta d_{ij}, \epsilon_1, \epsilon_2),$$

$$W_{2-}(\zeta, d_{ij}, k, \epsilon_1, \epsilon_2)$$

$$= \frac{1}{\zeta^2} \int_{\epsilon_1}^{\epsilon_2} \{e^{-\zeta(k+1)h} e^{-\zeta d_{ij} w} - 1 + \zeta(k+1)h + \zeta d_{ij} w\} f_{\mathcal{W}}(w) \, dw$$

$$= \frac{1}{\zeta^2} \left\{ e^{-\zeta(k+1)h} \mathcal{G}_3(\zeta d_{ij}, \epsilon_1, \epsilon_2) + \{\zeta(k+1)h - 1\} \mathcal{G}_1(\epsilon_1, \epsilon_2) + \zeta d_{ij} \mathcal{G}_2(\epsilon_1, \epsilon_2) \right\},$$

and

$$W_{3-}(\zeta, d_{ij}, k, \epsilon_1, \epsilon_2)$$

$$= \frac{1}{\zeta^2} \int_{\epsilon_1}^{\epsilon_2} \{(e^{-\zeta h} - 2) e^{\zeta k h} e^{\zeta d_{ij} w} + 1 + \zeta(k+1)h + \zeta d_{ij} w\} f_{\mathcal{W}}(w) \, dw$$

$$= \frac{1}{\zeta^2} \left\{ (e^{-\zeta h} - 2) e^{\zeta k h} \mathcal{G}_3(-\zeta d_{ij}, \epsilon_1, \epsilon_2) + \{\zeta(k+1)h + 1\} \mathcal{G}_1(\epsilon_1, \epsilon_2) + \zeta d_{ij} \mathcal{G}_2(\epsilon_1, \epsilon_2) \right\}.$$

C.1 $k = 0$

The zero lag cross-covariance of the aggregated process for the model with random $\mathcal{W}$, when $d_{ij} \leq 0$, is obtained by integrating $C_{ij,3}^{(h)}(w d_{ji}, 0)$ over the interval $[0, h/d_{ji})$, where

125

$d_{ji} = -d_{ij}$, and $C_{ij,4}^{(h)}(wd_{ji}, 0)$ over $[h/d_{ji}, \infty)$, which results in the following

$$E_W[C_{ij}^{(h)}(wd_{ij}, 0)] = \int_0^{\frac{h}{d_{ji}}} C_{ij,3}^{(h)}(wd_{ji}, 0)\, dw + \int_{\frac{h}{d_{ji}}}^{\infty} C_{ij,4}^{(h)}(wd_{ji}, 0)\, dw$$

$$= \lambda_{ij} E(C) E(X^2) V_{1+}(d_{ji}, 0) + \frac{\lambda_{ij}\beta}{2} E(C^2 - C) E(X)^2 V_{2+}(d_{ji}, 0),$$

where

$$V_{1+}(d_{ji}, 0) = \frac{1}{\eta_i} W_{3-}(\eta_i, d_{ji}, 0, 0, \frac{h}{d_{ji}}) + \frac{1}{\eta_j} W_{2-}(\eta_j, d_{ji}, 0, 0, \frac{h}{d_{ji}})$$

$$-(\frac{1}{\eta_{ij}})^+ W_{3-}(\eta_{ij}, d_{ji}, 0, 0, \frac{h}{d_{ji}}) - (\frac{1}{\eta_{ji}})^+ W_{2-}(\eta_{ji}, d_{ji}, 0, 0, \frac{h}{d_{ji}})$$

$$+\frac{1}{\eta_i} W_{1-}(\eta_i, d_{ji}, 0, \frac{h}{d_{ji}}, \infty) - (\frac{1}{\eta_{ij}})^+ W_{1-}(\eta_i, d_{ji}, 0, \frac{h}{d_{ji}}, \infty),$$

and

$$V_{2+}(d_{ji}, 0) = \frac{W_{3-}(\beta, d_{ji}, 0, 0, \frac{h}{d_{ji}})}{(\eta_j + \beta)(\eta_i - \beta)} + \frac{W_{2-}(\beta, d_{ji}, 0, 0, \frac{h}{d_{ji}})}{(\eta_i + \beta)(\eta_j - \beta)} + \frac{W_{1-}(\beta, d_{ji}, 0, \frac{h}{d_{ji}}, \infty)}{(\eta_j + \beta)(\eta_i - \beta)}$$

$$- \frac{2\beta}{(\eta_i + \eta_j)} \left\{ \frac{W_{3-}(\eta_i, d_{ji}, 0, 0, \frac{h}{d_{ji}})}{(\eta_i^2 - \beta^2)} + \frac{W_{2-}(\eta_j, d_{ji}, 0, 0, \frac{h}{d_{ji}})}{(\eta_j^2 - \beta^2)} + \frac{W_{1-}(\eta_i, d_{ji}, 0, \frac{h}{d_{ji}}, \infty)}{(\eta_i^2 - \beta^2)} \right\}.$$

When $d_{ij} \geq 0$, then t_{ij} should also be non-negative and in this case the cross-covariance is obtained by integrating $C_{ij,2}^{(h)}(wd_{ji}, 0)$ over the interval $[0, -h/d_{ji})$ and $C_{ij,4}^{(h)}(wd_{ji}, 0)$ over $[-h/d_{ji}, \infty)$, that is,

$$E_W[C_{ij}^{(h)}(wd_{ij}, 0)] = \int_0^{-\frac{h}{d_{ji}}} C_{ij,2}^{(h)}(wd_{ji}, 0)\, dw + \int_{-\frac{h}{d_{ji}}}^{\infty} C_{ij,1}^{(h)}(wd_{ji}, 0)\, dw$$

$$= E_W[C_{ji}^{(h)}(wd_{ji}, 0)],$$

because of the equivalence of $C_{ij,2}^{(h)}(t_{ij}, 0)$ with $C_{ji,3}^{(h)}(t_{ji}, 0)$ and of $C_{ij,1}^{(h)}(t_{ij}, 0)$ with $C_{ji,4}^{(h)}(t_{ji}, 0)$.

C.2 $k = 1$

The cross-covariance of the rainfall intensities at sites i and j, during successive time intervals ($k = 1$), is given in two parts depending on the sign of d_{ij}. So, if $d_{ij} \leq 0$, then

we have

$$
E_W[C_{ij}^{(h)}(wd_{ij}, 1)] = \int\limits_0^{\frac{h}{d_{ji}}} C_{ij,2}^{(h)}(wd_{ji}, 1)\, f_W(w)\, dw + \int\limits_{\frac{h}{d_{ji}}}^{\frac{2h}{d_{ji}}} C_{ij,3}^{(h)}(wd_{ji}, 1)\, f_W(w)\, dw
$$

$$
+ \int\limits_{\frac{2h}{d_{ji}}}^{\infty} C_{ij,4}^{(h)}(wd_{ji}, 1)\, f_W(w)\, dw
$$

$$
= \lambda_{ij}\, E(C)\, E(X^2)\, V_{1+}(d_{ji}, 1) + \frac{\lambda_{ij}\beta}{2}\, E(C^2 - C)\, E(X)^2\, V_{2+}(d_{ji}, 1),
$$

where

$$
V_{1+}(d_{ji}, 1) = \frac{1}{\eta_i}\, W_{2+}(\eta_i, d_{ji}, 1, 0, \frac{h}{d_{ji}}) + \frac{1}{\eta_j}\, W_{3+}(\eta_j, d_{ji}, 1, 0, \frac{h}{d_{ji}})
$$

$$
-(\frac{1}{\eta_{ij}})^+ W_{2+}(\eta_{ij}, d_{ji}, 1, 0, \frac{h}{d_{ji}}) - (\frac{1}{\eta_{ji}})^+ W_{3+}(\eta_{ji}, d_{ji}, 1, 0, \frac{h}{d_{ji}}) + \frac{1}{\eta_i}\, W_{3-}(\eta_i, d_{ji}, 1, \frac{h}{d_{ji}}, \frac{2h}{d_{ji}})
$$

$$
+ \frac{1}{\eta_j}\, W_{2-}(\eta_j, d_{ji}, 1, \frac{h}{d_{ji}}, \frac{2h}{d_{ji}}) - (\frac{1}{\eta_{ij}})^+ W_{3-}(\eta_{ij}, d_{ji}, 1, \frac{h}{d_{ji}}, \frac{2h}{d_{ji}}) - (\frac{1}{\eta_{ji}})^+ W_{2-}(\eta_{ji}, d_{ji}, 1, \frac{h}{d_{ji}}, \frac{2h}{d_{ji}})
$$

$$
+ \frac{1}{\eta_i}\, W_{1-}(\eta_i, d_{ji}, 1, \frac{2h}{d_{ji}}, \infty) - (\frac{1}{\eta_{ij}})^+ W_{1-}(\eta_{ij}, d_{ji}, 1, \frac{2h}{d_{ji}}, \infty),
$$

and

$$
V_{2+}(d_{ji}, 1) = \frac{W_{2+}(\beta, d_{ji}, 1, 0, \frac{h}{d_{ji}})}{(\eta_j + \beta)(\eta_i - \beta)} + \frac{W_{3+}(\beta, d_{ji}, 1, 0, \frac{h}{d_{ji}})}{(\eta_i + \beta)(\eta_j - \beta)} + \frac{W_{3-}(\beta, d_{ji}, 1, \frac{h}{d_{ji}}, \frac{2h}{d_{ji}})}{(\eta_j + \beta)(\eta_i - \beta)}
$$

$$
+ \frac{W_{2-}(\beta, d_{ji}, 1, \frac{h}{d_{ji}}, \frac{2h}{d_{ji}})}{(\eta_i + \beta)(\eta_j - \beta)} + \frac{W_{1-}(\beta, d_{ji}, 1, \frac{2h}{d_{ji}}, \infty)}{(\eta_j + \beta)(\eta_i - \beta)} - \frac{2\beta}{(\eta_i + \eta_j)} \left\{ \frac{W_{2+}(\eta_i, d_{ji}, 1, 0, \frac{h}{d_{ji}})}{(\eta_i^2 - \beta^2)} \right.
$$

$$
+ \frac{W_{3+}(\eta_j, d_{ji}, 1, 0, \frac{h}{d_{ji}})}{(\eta_j^2 - \beta^2)} + \frac{W_{3-}(\eta_i, d_{ji}, 1, \frac{h}{d_{ji}}, \frac{2h}{d_{ji}})}{(\eta_i^2 - \beta^2)} + \frac{W_{2-}(\eta_j, d_{ji}, 1, \frac{h}{d_{ji}}, \frac{2h}{d_{ji}})}{(\eta_j^2 - \beta^2)}
$$

$$
\left. + \frac{W_{1-}(\eta_i, d_{ji}, 1, \frac{2h}{d_{ji}}, \infty)}{(\eta_i^2 - \beta^2)} \right\}.
$$

When $d_{ij} \geq 0$ we get

$$
E_W[C_{ij}^{(h)}(wd_{ij}, 1)] = \int\limits_0^{\infty} C_{Y_{ij,1}^{(h)}}(wd_{ji}, 1)\, dw
$$

$$
= \lambda_{ij} E(C) E(X^2) V_{1-}(d_{ji}, 1) + \frac{\lambda_{ij}\beta}{2} E(C^2 - C) E(X)^2 V_{2-}(d_{ji}, 1),
$$

where

$$
V_{1-}(d_{ji}, 1) = \frac{1}{\eta_j}\, W_{1+}(\eta_j, d_{ji}, 1, 0, \infty) - \frac{1}{\eta_{ji}^+}\, W_{1+}(\eta_{ji}, d_{ji}, 1, 0, \infty),
$$

$$V_{2-}(d_{ji},1) \;=\; \frac{W_{1+}(\beta,d_{ji},1,0,\infty)}{(\eta_i+\beta)\,(\eta_j-\beta)} - \frac{2\beta\,W_{1+}(\eta_j,d_{ji},1,0,\infty)}{(\eta_i+\eta_j)\,(\eta_j^2-\beta^2)}.$$

C.3 $k \geq 2$

Finally, the cross-covariance function for the other time lags greater than one, when $d_{ij} \leq 0$, is

$$
E_W[C_{ij}^{(h)}(wd_{ij},k)] \;=\; \int_0^{\frac{h(k-1)}{d_{ji}}} C_{ij,1}^{(h)}(wd_{ji},k)\,f_W(w)\,dw \;+\; \int_{\frac{h(k-1)}{d_{ji}}}^{\frac{hk}{d_{ji}}} C_{ij,2}^{(h)}(wd_{ji},k)\,f_W(w)\,dw
$$

$$
\;+\; \int_{\frac{hk}{d_{ji}}}^{\frac{h(k+1)}{d_{ji}}} C_{ij,3}^{(h)}(wd_{ji},k)\,f_W(w)\,dw \;+\; \int_{\frac{h(k+1)}{d_{ji}}}^{\infty} C_{ij,4}^{(h)}(wd_{ji},k)\,f_W(w)\,dw
$$

$$
\;=\; \lambda_{ij}E(C)E(X^2)\,V_{1+}(d_{ji},k) + \frac{\lambda_{ij}\beta}{2}E(C^2-C)E(X)^2\,V_{2+}(d_{ji},k),
$$

where

$$
\begin{aligned}
V_{1+}(d_{ji},k) \;=\;& \frac{1}{\eta_j}\,W_{1+}(\eta_j,d_{ji},k,0,\tfrac{h(k-1)}{d_{ji}}) - \frac{1}{\eta_{ji}^+}W_{1+}(\eta_{ji},d_{ji},k,0,\tfrac{h(k-1)}{d_{ji}}) \\
&+\frac{1}{\eta_i}\,W_{2+}(\eta_i,d_{ji},k,\tfrac{h(k-1)}{d_{ji}},\tfrac{hk}{d_{ji}}) + \frac{1}{\eta_j}\,W_{3+}(\eta_j,d_{ji},k,\tfrac{h(k-1)}{d_{ji}},\tfrac{hk}{d_{ji}}) \\
&-\frac{1}{\eta_{ij}^+}\,W_{2+}(\eta_{ij},d_{ji},k,\tfrac{h(k-1)}{d_{ji}},\tfrac{hk}{d_{ji}}) - \frac{1}{\eta_{ji}^+}\,W_{3+}(\eta_{ji},d_{ji},k,\tfrac{h(k-1)}{d_{ji}},\tfrac{hk}{d_{ji}}) \\
&+\frac{1}{\eta_i}\,W_{3-}(\eta_i,d_{ji},k,\tfrac{hk}{d_{ji}},\tfrac{h(k+1)}{d_{ji}}) + \frac{1}{\eta_j}\,W_{2-}(\eta_j,d_{ji},k,\tfrac{hk}{d_{ji}},\tfrac{h(k+1)}{d_{ji}}) \\
&-\frac{1}{\eta_{ij}^+}\,W_{3-}(\eta_{ij},d_{ji},k,\tfrac{hk}{d_{ji}},\tfrac{h(k+1)}{d_{ji}}) - \frac{1}{\eta_{ji}^+}\,W_{2-}(\eta_{ji},d_{ji},k,\tfrac{hk}{d_{ji}},\tfrac{h(k+1)}{d_{ji}}) \\
&+\frac{1}{\eta_i}\,W_{1-}(\eta_i,d_{ji},k,\tfrac{h(k+1)}{d_{ji}},\infty) - \frac{1}{\eta_{ij}^+}W_{1-}(\eta_{ij},d_{ji},k,\tfrac{h(k+1)}{d_{ji}},\infty),
\end{aligned}
$$

$$
\begin{aligned}
V_{2+}(d_{ji},k) \;=\;& \frac{W_{1+}(\beta,d_{ji},k,0,\tfrac{h(k-1)}{d_{ji}})}{(\eta_i+\beta)\,(\eta_j-\beta)} + \frac{W_{2+}(\beta,d_{ji},k,\tfrac{h(k-1)}{d_{ji}},\tfrac{hk}{d_{ji}})}{(\eta_j+\beta)\,(\eta_i-\beta)} \\
&+\frac{W_{3+}(\beta,d_{ji},k,\tfrac{h(k-1)}{d_{ji}},\tfrac{hk}{d_{ji}})}{(\eta_i+\beta)\,(\eta_j-\beta)} + \frac{W_{3-}(\beta,d_{ji},k,\tfrac{hk}{d_{ji}},\tfrac{h(k+1)}{d_{ji}})}{(\eta_j+\beta)\,(\eta_i-\beta)} \\
&+\frac{W_{2-}(\beta,d_{ji},k,\tfrac{hk}{d_{ji}},\tfrac{h(k+1)}{d_{ji}})}{(\eta_i+\beta)\,(\eta_j-\beta)} + \frac{W_{1-}(\beta,d_{ji},k,\tfrac{h(k+1)}{d_{ji}},\infty)}{(\eta_j+\beta)\,(\eta_i-\beta)} \\
&-\frac{2\beta}{(\eta_i+\eta_j)}\left\{ \frac{W_{1+}(\eta_j,d_{ji},k,0,\tfrac{h(k-1)}{d_{ji}})}{(\eta_j^2-\beta^2)} + \frac{W_{2+}(\eta_i,d_{ji},k,\tfrac{h(k-1)}{d_{ji}},\tfrac{hk}{d_{ji}})}{(\eta_i^2-\beta^2)} \right.
\end{aligned}
$$

$$+\frac{W_{3+}(\eta_j, d_{ji}, k, \frac{h(k-1)}{d_{ji}}, \frac{hk}{d_{ji}})}{(\eta_j^2 - \beta^2)} + \frac{W_{3-}(\eta_i, d_{ji}, k, \frac{hk}{d_{ji}}, \frac{h(k+1)}{d_{ji}})}{(\eta_i^2 - \beta^2)}$$

$$+\frac{W_{2-}(\eta_j, d_{ji}, k, \frac{hk}{d_{ji}}, \frac{h(k+1)}{d_{ji}})}{(\eta_j^2 - \beta^2)} + \frac{W_{1-}(\eta_i, d_{ji}, k, \frac{h(k+1)}{d_{ji}}, \infty)}{(\eta_i^2 - \beta^2)} \Bigg\}.$$

For non-negative values of d_{ij}, we have

$$\begin{aligned}
E_W[C_{ij}^{(h)}(wd_{ij}, k)] &= \int_0^\infty C_{ij,1}^{(h)}(wd_{ji}, k)\, f_W(w)\, dw \\[2mm]
&= \lambda_{ij} E(C) E(X^2)\, V_{1-}(d_{ji}, k) + \frac{\lambda_{ij}\beta}{2} E(C^2 - C) E(X)^2\, V_{2-}(d_{ji}, k),
\end{aligned}$$

where

$$V_{1-}(d_{ji}, k) = \frac{1}{\eta_j} W_{1+}(\eta_j, d_{ji}, k, 0, \infty) - \frac{1}{\eta_{ji}^+} W_{1+}(\eta_{ji}, d_{ji}, k, 0, \infty),$$

$$V_{2-}(d_{ji}, k) = \frac{W_{1+}(\beta, d_{ji}, k, 0, \infty)}{(\eta_i + \beta)(\eta_j - \beta)} - \frac{2\beta\, W_{1+}(\eta_j, d_{ji}, k, 0, \infty)}{(\eta_i + \eta_j)(\eta_j^2 - \beta^2)}.$$

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

Bartlett, M. S. (1963). The spectral analysis of point processes. *J. R. Stat. Soc. B*, **25**, 264–296.

Chandler, R. E. (1997). A spectral method for estimating parameters in rainfall models. *Bernoulli*, **3**(1), 1–22.

Cowpertwait, P. S. P., O'Connell, P. E., Metcalfe, A. V., and Mawdsley, J. A. (1996). Stochastic point process modelling of rainfall. I. Single-site fitting and validation. *J. Hydrology*, **175**, 17–46.

Cowpertwait, P. S. P. (1991). Further developments of the Neyman-Scott clustered point process for modelling rainfall. *Water Resour. Res.*, **27**, 1431–1438.

Cowpertwait, P. S. P. (1994). A generalized process model for rainfall. *Proc. R. Soc. Lond. A*, **447**, 23–37.

Cox, D. R. and Isham, V. S. (1980). *Point Processes*. Chapman and Hall, London.

Cox, D. R. and Isham, V. S. (1988). A simple spatial-temporal model of rainfall. *Proc. R. Soc. Lond. A*, **415**, 317–328.

Cox, D. R. and Isham, V. S. (1994). *Statistics for the Environment 2: Water Related Issues.*, chap. Stochastic models of precipitation. John Wiley & Sons Ltd., London.

Eagleson, P., Fennessey, N. M., Qinliang, W., and Rodriguez-Iturbe, I. (1987). Application of spatial poisson models to air mass thunderstorm rainfall. *J. Geophys. Res.*, **92**, 9661–9678.

Entekhabi, D., Rodriguez-Iturbe, I., and Eagleson, P. S. (1989). Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation. *Water Resour. Res.*, **25**, 295–302.

Georgiou, A. and Vassiliou, P. C. G. (1993). *Μη Γραμμικες Μεθοδοι Βελτιστοποιησης (Non-linear methods of optimisation)*. Ziti Press, Thessaloniki.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University press.

Hutchinson, M. F. (1990). A point rainfall model based on a three-state continuous markov occurrence process. *J. Hydrology*, **114**, 125–148.

Kavvas, M. L. and Delleur, J. W. (1981). A stochastic cluster model of daily rainfall sequences. *Water Resour. Res.*, **17**, 1151–1160.

Kendall, M. G. and Moran, P. A. P. (1963). *Geometrical probability*. Charles Griffin, London.

Le Cam, L. (1961). A stochastic description of precipitation. In Neyman, J. (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **3**, pp. 165–186, Berkeley, California.

Lewis, P. A. W. (1964). A branching poisson process model for the analysis of computer failure patterns. *J. R. Stat. Soc. B*, **26**, 398–456.

Mardia, K. V. (1972). *Statistics of directional data*. Academic Press, London.

Mason, J. (1986). Numerical weather prediction. *Proc. R. Soc. Lond. A*, **407**, 51–60.

Neyman, J. and Scott, E. L. (1958). Statistical approach to problems of cosmology. *J. R. Stat. Soc. B*, **20**, 1–43.

Northrop, P. J. (1996). *Modelling and statistical analysis of spatial-temporal rainfall fields*. Ph.D. thesis, University of London.

Onof, C. J. and Wheater, H. S. (1993). Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model. *J. Hydrology*, **149**, 67–95.

Onof, C. J., Wheater, H. S., and Isham, V. S. (1994). Note on the analytical expression of the inter-event time characteristics for Bartlett-Lewis type rainfall models. *J. Hydrology*, **157**, 197–210.

Onof, C. J. (1992). *Stochastic modelling of British rainfall data using Poisson processes.* Ph.D. thesis, University of London.

Phelan, M. J. and Goodall, C. R. (1990). An assessment of a generalized Waymire-Gupta-Rodriguez-Iturbe model for GARP atlantic tropical experiment rainfall. *J. Geophys. Res.*, **95**, 7603–7615.

Rodriguez-Iturbe, I., Cox, D. R., and Eagleson, P. S. (1986). Spatial modeling of total rainfall. *Proc. R. Soc. Lond. A*, **403**, 27–50.

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

Rodriguez-Iturbe, I., Febres de Power, B., and Valdes, J. B. (1987b). Rectangular pulses point process models for rainfall: Analysis of empirical data. *J. Geophys. Res.*, **92**, 9645–9656.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. S. (1988). A point process model for rainfall : further developments. *Proc. R. Soc. Lond. A*, **417**, 283–298.

Smith, J. A. and Karr, A. F. (1985). Parameter estimation for a model of space-time rainfall. *Water Resour. Res.*, **21**, 1251–1257.

Stein, M. (1986). A simple model for spatial-temporal processes. *Water Resour. Res.*, **22**, 2107–2110.

Stern, R. D. and Coe, R. (1984). A model fitting analysis of daily rainfall data. *Proc. R. Soc. Lond. A*, **147**, 1–34.

Valdes, J. B., Rodriguez-Iturbe, I., and Gupta, V. K. (1985). Approximations of a temporal rainfall from a multidimensional model. *Water Resour. Res.*, **21**, 1259–1270.

Velghe, T., Troch, P. A., De Troch, F. P., and Van de Velde, J. (1994). Evaluation of cluster-based rectangular pulses point process models for rainfall. *Water Resour. Res.*, **30**, 2847–2857.

Waymire, E. and Gupta, V. K. (1981a). The mathematical structure of rainfall representations 1. a review of the stochastic rainfall models. *Water Resour. Res.*, **17**, 1261–1272.

Waymire, E. and Gupta, V. K. (1981b). The mathematical structure of rainfall representations 3. some applications of the point process theory to rainfall processes. *Water Resour. Res.*, **17**, 1287–1294.

Waymire, E., Gupta, V. K., and Rodriguez-Iturbe, I. (1984). A spectral theory of rainfall intensity at a meso-β scale. *Water Resour. Res.*, **20**, 1453–1465.