Single-site point process-based rainfall models in a nonstationary climate

by

Joanna Maria Kaczmarska

A Thesis submitted for the degree of
Doctor of Philosophy

in the

Faculty of Mathematical & Physical Sciences

Department of Statistical Science

University College London

March 2013

Declaration of Authorship

I, Joanna Maria Kaczmarska, declare that this dissertation represents my own work, except where due acknowledgement is made.

Signed: _____

Date: _____

Abstract

Long series of simulated rainfall are required at point locations for a range of applications, including hydrological studies. Clustered point process-based rainfall models have been used for generating such simulations for many decades. One of their main advantages is the fact that they generate simulations in continuous time, allowing aggregation to different timescales in a consistent way, and such models generally perform well in representing rainfall at hourly to daily timescales. An important disadvantage, however, is their stationarity. Although seasonality can be allowed for by fitting separate models for each calendar month or season, the models are unsuitable in their basic form for climate impact studies.

In this thesis we develop new methodology to address this limitation. We extend the current fitting approach by replacing the discrete covariate, calendar month, with continuous covariates which are more directly related to the incidence and nature of rainfall. The covariate-dependent model parameters are estimated for each time interval using a kernel-based nonparametric approach within a Generalised Method of Moments framework.

An empirical study using the new methodology is undertaken using a time series of five-minute rainfall data. In addition to addressing the need for temporal non-stationarity, which is our main focus, we also carry out a systematic comparison of a number of key variants of the basic model, in order to identify which features are required for an optimal fit at sub-hourly resolution. This generates some new insights into the models, leading to the development of a new model extension, which introduces dependence between rainfall intensity and duration in a simple way. The new model retains the 'rectangular pulses' (i.e. rain cells with a constant intensity) of the original clustered point process model, which had previously been considered inappropriate for fine-scale data, obviating the need for a computationally more intensive 'instantaneous pulse' model.

Acknowledgements

Becoming a student again over twenty years after leaving university was a significant change for me, and I missed many aspects of my previous office life, especially my former colleagues. However, this was more than made up for by the joy of exchanging routine for days full of learning and research, and the excitement of facing new challenges. I am glad that I chose to return to my studies at UCL, and that I had Professor Valerie Isham and Dr. Paul Northrop as my supervisors. I would like to thank Valerie and Paul for their support and guidance during the last three years. As well as providing invaluable feedback on my work and advice on statistical theory, they were always good company and generous with their time. I would also like to thank Dr. Richard Chandler for his helpful ideas at my MPhil Upgrade, and for always giving detailed and considered answers to any questions I asked him. Thanks also to Dr. Christian Onof from Imperial College who provided helpful input from a hydrological perspective.

The Engineering and Physical Sciences Research Council is gratefully acknowledged for their financial support, and Deutsche Montan Technologie and Emschergenossenschaft/Lippeverband in Germany and the National Institute of Water and Atmospheric Research in New Zealand for providing the data.

Finally, thanks are due to my family, especially my husband, Andrzej, for their love and support, and to my friends and fellow PhD students for their companionship.

Contents

List of Figures

List of Tables

Abbreviations

Models

BLIP	Bartlett Lewis Instantaneous Pulse model
BLIP2	Bartlett-Lewis Instantaneous Pulse model with two superposed processes
BLIPR	Random Parameter Bartlett-Lewis Instantaneous Pulse model
BLRP	Bartlett-Lewis Rectangular Pulse model
BLRP2	Bartlett-Lewis Rectangular Pulse model with two superposed processes
BLRPR	Random Parameter Bartlett-Lewis Rectangular Pulse model
BLRPR_X	Random Parameter Bartlett-Lewis Rectangular Pulse model with dependent intensity-duration ($\mu_X \propto \eta$)
BLRPR_{X_c}	Random Parameter Bartlett-Lewis Rectangular Pulse model with dependent intensity-duration ($\mu_X \propto \eta^c$)

Covariates

geo200	Geo-potential height at 200 hPa, m (similarly, geo500 & geo700 are geo-potential heights at 500 hPa and700 hPa respectively)
rhum	Relative surface humidity, %
rhum700	Relative humidity at 700 hPa, %
shum	Specific surface humidiy, g/kg
shum700	Specific humidiy at 700 hPa, g/kg
slp	Sea-level pressure, mb
temp	Surface temperature, °C
thick	Thickness of the atmosphere between 500hPa and 1000hPa
uwind	Zonal (west-east) wind component, m/s
vwind	Meridional (north-south) wind component, m/s

Other Acronyms

AIC	Akaike's information criterion
AOGCM	Atmosphere-Ocean General Circulation Model
BIC	Bayes' information criterion

GAM	Generalised Additive Model
GCM	Generalised Circulation Model
GEV	Generalised Extreme Value
GLM	Generalised Linear Model
IPCC	Intergovernmental Panel on Climate Change
MOS	Model Output Statistics
NAO	North Atlantic Oscillation
NCEP	National Centre for Environmental Prediction
RCM	Regional Climate Model
SRES	Special Report on Emission Scenarios
UKCIP02	UK Climate Impacts Programme, 2002
UKCP09	UK Climate Projections, 2009

Chapter 1

Introduction

Long series of simulated rainfall are required by hydrologists, telecommunications engineers and those involved in the modelling of climate impacts on agriculture and the environment. The timescales at which these are required depend on the particular application, and range from monthly down to a few minutes. Applications for which very fine scale data are required at point locations include urban drainage design and radio telecommunications networks. Observed data series are generally too short (and may suffer from quality issues), particularly at these finer timescales. Over the last few decades this has led to the development of stochastic rainfall models, from which series of any desired length can be simulated. These simulated series are required to resemble real precipitation in terms of reproducing key characteristics such as variability, correlations between successive intervals, extremes and wet/dry sequences. A number of different approaches exist and practical applications are numerous.

More recently, however, the issue of climate change has become an increasing concern. Warming of the climate system is now "unequivocal", according to the Intergovernmental Panel on Climate Change (IPCC), whose latest report (the Fourth Assessment report (IPCC 2007)), warns that, expressed as a global average, surface temperatures have increased by about 0.74°C over the past hundred years, and are projected to increase by about 0.2°C per decade over the next two decades. Further projections increasingly depend on specific SRES (Special Report on Emission Scenarios, Nakićenović & Swart (2000)) emission scenarios. Observed changes to precipitation have also been noted, including changes to the amount, intensity, frequency and type of precipitation. There is more uncertainty here, due to the difficulties of measurement, regional influence of aerosols, gaps in our understanding of certain precipitation feedbacks, and high natural variability. Precipitation is influenced by El Niño and atmospheric circulation patterns such as the North Atlantic Oscillation, which are themselves influenced by climate change (Trenberth et al. (2007)). Despite these uncertainties, it is considered very likely that hydrological characteristics will change in the future, including in particular an increase in the frequency of

heavy precipitation events (Bates et al. 2008).

Although various different approaches have been proposed in order to generate artificial rainfall series that incorporate climate change scenarios at a point location, none has so far done so in a way that addresses all the requirements in a straightforward way, and this remains an ongoing challenge. The simpler approaches tend not to be able to reproduce the structure and characteristics of rainfall (such as the correlations between successive intervals, extremes, variability etc.). Another common limitation of many models is that they are only applicable at a single timescale, usually daily. More sophisticated models address these issues, but typically at the expense of a great deal of complexity and computational burden, often requiring several components, particularly where sub-hourly series are required.

Generalised Circulation Models (GCMs) are the main tool for predicting future climate impacts resulting from the increase of greenhouse gases in the atmosphere (see Barry & Chorley (2003) for an introduction). They model large-scale movements over the entire globe, which is divided into a coarse grid of boxes, over tens to hundreds of years using a set of physical equations. Typical surface resolutions are 1–2.5° × 1.5–4°, with of the order of ten to twenty atmospheric levels. The equations link the five key physical quantities of the atmosphere — wind velocity, air temperature, pressure, density and humidity, and are numerically integrated forward in small time-steps (of the order of a few minutes to tens of minutes), starting from a set of initial conditions, and respecting a set of boundary conditions (e.g. between the atmosphere and the Earth's land or ocean surface). In order to solve the coupled equations, the models also need to keep track of processes within the grid boxes (known as 'parameterisations') such as absorption and scattering of radiation, cloud formation, precipitation etc. Typically details of coastlines, mountains, surface vegetation and soil are incorporated, but much of the regional detail is lost in the averaging over grid squares. According to the Fourth Assessment report, there is considerable confidence that Atmosphere-Ocean General Circulation Models (AOGCMs) provide credible quantitative estimates of future climate change, although confidence in these estimates is higher for some climate variables e.g. temperature, than for others, e.g. precipitation (Randall et al. 2007). However, AOGCMs cannot provide information at scales finer than their computational grid, typically of the order of 200 km. (Christensen et al. 2007).

The techniques for generating data at a finer temporal and/or spatial resolution from climate model output are referred to as 'downscaling', and are classified as either dynamical or statistical. The former are physically based, such as Regional Climate Models (RCMs). These cover smaller areas with a finer resolution, typically with grid cells of approximately 50km by 50km (although a few recent models have had grid sizes below 20km), allowing more local features to be represented. RCMs are often embedded into a GCM in a region of interest and use the global model information as boundary conditions. However, even

at the finer resolution level of RCMs, the models' ability to produce realistic future rainfall projections at sub-daily resolution at point locations is questionable. This type of physical model is also computationally very intensive, making it infeasible for impact studies which require the analysis of a large number of possible scenarios.

Statistical downscaling techniques have the important advantage of computational speed. They exploit observed relationships between the large-scale climate variables of the GCM (or RCM) and local climate. Since modelled values of future atmospheric variables are considered fairly reliable, future rainfall may be derived by relating rainfall to these atmospheric variables e.g. using regression techniques.

There are numerous different approaches to statistical downscaling. These are difficult to categorise as many use combinations of a number of different elements and techniques. Useful recent reviews which focus specifically on precipitation and hydrological modeling include Fowler et al. (2007) and Maraun et al. (2010). One of the most popular techniques is regression, which can involve different levels of complexity, and form part of more sophisticated methods. At the simplest level, a multiple regression may be used to estimate mean monthly rainfall amounts conditionally on a set of suitable large-scale atmospheric predictor variables. At the daily timescale, this is likely to be problematic, due to the highly non-normal distribution of rainfall. Transformations are commonly used to address both non-normality and non-linear relationships between the predictor and predictand.

A common solution in respect of daily rainfall amounts is to use a power-transformed and truncated normal distribution (e.g. Bardossy & Plate (1992)). An alternative approach is to use separate models for daily rainfall occurrence and daily rainfall amounts, conditionally on a wet day. For example, Kilsby et al. (1998) take the log of the rainfall amount and use the logit function $\log(\frac{p}{1-p})$ for daily rainfall occurrence, p. However these transformations introduce bias when applying the backward transformation, which needs to be adjusted for. A better approach is to use generalised linear models (GLMs), which can respect the actual distribution of the data. Thus rainfall occurrence, a binary variable, may be modelled using logistic regression, whereas for wet-day amounts, the normal distribution assumption of the linear regression may be replaced by a more suitable alternative, say gamma, within the GLM (Chandler & Wheater 2002, Furrer & Katz 2007).

Non-parametric regression techniques are a more flexible way to address the issue of non-linear relationships between the predictors and predictand e.g. Generalised Additive Models (GAMs) (Beckmann & Buishand 2002) which replace the linear functions of GLMs with nonparametric ones. Another nonlinear regression technique is the artificial neural network (e.g. Haylock et al. (2006), Cavazos & Hewitson (2005)). Techniques to reduce the dimensionality of the covariates include the use of principal components (Cavazos & Hewitson 2005) or canonical correlation analysis (von Storch & Zwiers 1999).

Another popular statistical downscaling approach involves the use of weather typing

schemes (Jones et al. 1993, Buishand & Brandsma 1997). This approach categorises weather into discrete types, and associates each type with either a rainfall mean, or some more extensive properties of rainfall (such as the parameters of a rainfall model). The weather types may be defined using subjective judgement, or using an objective approach, such as cluster analysis. Climate change is then estimated by evaluating the change in frequency of the weather classes simulated by the GCM. An advantage of weather types is that that are a straightforward way to allow for nonlinear relationships between the underlying raw climate variables and the response (Maraun et al. 2010). However, a disadvantage of discrete weather types, highlighted by Wilby (1997), is the difficulty of maintaining the uniqueness of rare events at the same time as grouping sufficient data to ensure statistical integrity. The weather types are also potentially local in character, and not readily used across locations.

Random cascade models represent another methodology that is sometimes included under the umbrella of statistical downscaling. These are models that use multifractal techniques to subdivide rainfall at a coarse resolution repeatedly to successively smaller subintervals. These models cannot (at least currently) be used as a stand-alone method of statistical downscaling, since the precipitation output from GCMs is not considered reliable, and the random cascade models have no parameters that relate to climate variables. However, they are often used in conjunction with other methods, and discussed briefly in this context in Section 1.3.

Usually, statistical downscaling models are fitted to (or calibrated against) historical observed or reanalysis data, and it is assumed that the predictors are perfectly modeled by the GCM or RCM, in which case they are categorised as 'perfect prognosis' models. An alternative category of models, termed MOS (Model Output Statistics) models, are instead calibrated against simulated predictors. Typically this involves using simulated precipitation from an RCM as a predictor, in which case MOS can effectively be thought of as 'correcting' the precipitation of the RCM, although usually spatial downscaling to a more local scale is also involved. MOS was originally developed in weather forecasting (Klein & Glahn 1974), and is only recently being applied in this way (Maraun et al. 2010).

Some of the simpler downscaling approaches do not allow a random element, and thus tend to underestimate local-scale variability. Many do not explicitly impose an appropriate temporal structure (although there will be an implicit assumption within the predictors), and this may lead to performance issues in terms of reproducing the characteristics of real weather series. Weather generators are stochastic models which produce local synthetic time series of weather data, and which address both of these issues. Often daily precipitation is modelled first because of its importance and relative complexity, with other variables of interest at the local scale, such as daily maximum and minimum temperature, solar radiation, humidity and windspeed then modelled conditionally on precipitation occurrence. Weather generators can be used for statistical downscaling in a number of different ways,

with varying levels of sophistication, and may involve both regression and weather typing. A simple example of an unconditional weather generator (i.e. one which does not involve large-scale covariates) is just a two-state (wet/dry) first order Markov chain with daily transition probabilities, and a distribution (typically gamma) for precipitation amounts on wet days (Katz 1977, Richardson 1981). Second or third order Markov chains may be used instead in order to reflect the temporal structure more accurately (Stern & Coe 1984). The GLMs described above may also be categorised as weather generators, and are essentially extensions of the Markov chain approach. This type of model can be categorised as purely statistical, and makes no attempt to represent the physical rainfall process itself in any way. A consequent limitation is that the model provides estimates only at the timescale at which it has been fitted, which is typically daily. GLMs and similar models are not feasible at subdaily resolution, as the complicated dependency structure of rainfall would require an excessive number of parameters at these scales (Chandler et al. 2007a). An example of a recent, more sophisticated (multi-site) model of this type used in the context of statistical downscaling is the extended nonhomogeneous hidden Markov model of Charles et al. (2004), which combines an underlying unobserved (hidden) stochastic process of weather states with an observed process of daily rainfall occurrence, which is assumed to be temporally independent given the hidden process. The hidden process of weather states is assumed to follow a first order Markov chain, with transition probabilities conditional on atmospheric predictors.

One of the most important categories of weather generator, particularly in the field of hydrology, consists of clustered point process-based rainfall models, which are the focus of this thesis. These models differ from the purely statistical ones described above, in that they attempt to model the physical rainfall process, albeit in a very simplified way. The deterministic differential equations of numerical weather models are replaced by a stochastic process with a small number of physically interpretable parameters. Because the underlying process runs in continuous time, simulations can be generated which can be aggregated to different timescales in a consistent way, and this is one of the principal advantages of this type of model.

Clustered point process-based models do not in themselves constitute a statistical downscaling method, since in their basic form they are simply fitted to historical observed rain-gauge data and used to simulate a stationary rainfall series. In this form they are classed as unconditional weather generators. However, the recent literature includes many approaches where these have been used in conjunction with large-scale atmospheric variables, as part of a statistical downscaling methodology. In the next sections we introduce the basic models and then discuss how these have been used within a statistical downscaling context. We conclude this introduction with a summary of the research question that is addressed in this thesis, and an outline of the thesis structure.

1.1 Clustered point process-based rainfall models

Empirical evidence has shown that rainfall exhibits clustering in both space and time (Austin & Houze 1972). Clustered point process-based rainfall models have been used for over twenty-five years, since a seminal paper by Rodriguez-Iturbe et al. (1987). Examples of practical application are numerous (Onof et al. 2000, Wheater et al. 2005, Cowpertwait 2006, Kilsby et al. 2007, Burton et al. 2008), with generally very good performance. A clustered point process rainfall model is also used within the Weather Generator tool of the UK Climate Projections (UKCP09) project to downscale climate projections (Jones et al. 2009). The original single-site models have been extended to the spatial-temporal domain, but here we focus on point locations.

The models assume that rain-events or 'storms' arrive in a Poisson process. Each rain event consists of a cluster of rain cells, with the temporal location of cells relative to the event origin specified by one of two clustering mechanisms — Bartlett-Lewis or Neyman-Scott. Each cell is assumed to have a random duration and intensity. In the most commonly used models, the intensity remains constant over the duration of the cell, giving rise to their description as 'rectangular pulse models'. The models are fitted to discrete data from rain-gauges, typically using the generalised method of moments. This is a fairly subjective method, for which there is considerable flexibility, particularly in terms of the number and types of properties chosen for fitting. The dependence structure of the models, particularly when aggregated, makes a maximum likelihood method impracticable.

Since their introduction, many refinements have been introduced in order to improve the fit of the basic models to specific rainfall properties. Key amongst these have been those which have allowed for different types of rainfall. These include models with multiple cell-types (Cowpertwait 1994), or multiple superposed processes (Cowpertwait 2004, Cowpertwait et al. 2007). In order to keep parameter numbers manageable, these methods have generally limited the number of cell types or processes to just two, which can be thought of as representing heavy, short-duration convective and lighter, long-duration stratiform types of rainfall. An alternative modification to enable variation between storms is the randomisation of the cell duration parameter between storms (Rodriguez-Iturbe et al. (1988), Entekhabi et al. (1989)). In effect this allows a continuous range of storm types.

While much of the application of the models has been at hourly or longer timescales, the requirement for sub-hourly resolution, in particular for the design of stormwater sewerage systems, was the motivation for the development of the Bartlett-Lewis Pulse model (Cowpertwait et al. 2007), which replaces the rectangular rain cells of the original Bartlett-Lewis model with a Poisson process of instantaneous pulses (thus incorporating two levels of clustering, and allowing greater variability in rain intensity at short timescale). However, no study has as yet investigated the performance of the rectangular-pulse models at sub-hourly timescales, nor compared these against the new instantaneous pulse version. In

order for the models to be able to reproduce rainfall properties at a sub-hourly timescale, properties at this timescale must be included within the fit.

Apart from issues relating to a sub-hourly timescale, the most commonly noted shortcomings of the models, with respect to their ability to simulate realistic rainfall series, relate to the reproduction of wet/dry properties and to extremes. The former was addressed to some extent by randomising the cell duration parameter, as discussed, the latter by the introduction of the skewness coefficient as one of the fitting properties (Cowpertwait 1998). Parameter identifiability can also be a problem, particularly with the model variants with relatively high numbers of parameters, such as those with multiple cell types or superposed processes.

The most significant shortcoming of these models, however, is that they are stationary. Historically, as for most unconditional weather generators, seasonal variation in rainfall characteristics has been addressed by fitting a separate model for each calendar month or season. However, this approach is no longer viable in a world of changing climate and seasons. In the next section we consider approaches to this problem in the existing literature.

1.2 Downscaling approaches using point process models

As mentioned above, the point process rainfall models are fitted to a selected set of rainfall statistics. A simple approach re-scales some or all of the historical statistics to allow for future climate change, with separate models then fitted to each such re-scaled set of statistics. Examples of this approach include Kilsby et al. (1998), Kilsby et al. (2007) and Burton et al. (2010). Kilsby et al. (1998) fit separate models for each month, updating just two rainfall statistics (the mean daily rainfall amount and the proportion of dry days), and allowing two of the parameters, the storm arrival rate and the mean cell intensity to vary, keeping the other model parameters (which are based on the present climate) constant over the whole period. The two statistics are regressed on large-scale atmospheric variables each month. Rather than using regression to predict future values of rainfall statistics, Kilsby et al. (2007) derive these simply by multiplying current observed values (or appropriately transformed values) by 'change factors', based on UKCIP02 future climate scenarios (Hulme et al. 2002). A 'change factor' is just the ratio of the appropriate rainfall statistic's value from the given climate model in respect of a future 'time slice' divided by its value for the control period. Separate models are fitted for three time slices — 2020s, 2050s and 2080s, and four IPCC SRES emissions scenarios. Note that this scaling approach can be thought of as a simple MOS (Maraun et al. 2010). Burton et al. (2010) extend this work to allow for a greater level of non-stationarity, interpolating the 'change factors' to provide a monthly series of statistics. While these approaches are conceptually simple, they are computationally burdensome and rather inflexible. Typically

constraints need to be imposed in order to ensure smoothness (e.g. by keeping some parameters constant). Parameter uncertainty has not been considered, and its estimation is likely to be difficult, given the fitting to 'quasi-observed' data. The change factor approach also makes the implicit assumption that the observations represent a stationary climate (Burton et al. 2010).

Fowler et al. (2000) present a rather different model which combines a semi-Markov chain with a clustered point process model. The former generates a sequence of daily weather states. A point process model is fitted for each of the (six) weather states, and the combined model is used to provide synthetic rainfall series at a single site. Climate change scenarios can be accommodated by changing the frequencies of the different weather types. There are some issues with this model, in terms of the behaviour of rainfall at the transition points between states, as storm durations may exceed the daily time scale at which weather states are modelled. It is thus not straightforward to determine the rainfall belonging to each state in the observed data.

Typically these downscaling approaches have been used only for downscaling to a daily timescale. A second step is then required, if rainfall at a sub-daily timescale is required. Here again, clustered point processes form a part of many techniques. A brief description of some of the key methods is given below, which may be applied to the daily data from any of the downscaling techniques.

1.3 Generating rainfall at finer resolution

One of the simplest methods of generating rainfall at a sub-daily scale from a coarser resolution is by using scaling relationships (Kilsby et al. 2007, Chandler et al. 2007b). Sub-daily statistics are derived using observed relationships between sub-daily and daily statistics, and then these are included as fitting statistics within the point process model. Very broadly the relationship between log timescale and log statistic is close to linear for the variance and proportion of intervals which are wet, although more complex relationships are required in practice to avoid some odd results. However, scaling relationships to derive sub-hourly statistics from the daily values have not been developed as yet.

Various alternative methods exist, some of which are suitable for generating rainfall down to a sub-hourly resolution. If the method attempts to match actual daily totals, rather than just the distribution or moments, then it is termed 'disaggregation', rather than 'downscaling'. Several disaggregation approaches are based on point process models (Glasbey et al. 1995, Koutsoyiannis & Onof 2000, 2001). The broad idea behind these is to simulate long records at the appropriate timescales from the point process model, and then find short sequences which provide good matches for the daily level series (or more specifically for each wet-day cluster), within some level of tolerance. A different class of disaggrega-

tors is based on the fact that rainfall intensities exhibit simple scaling behaviour (Lovejoy & Schertzer 1995, Gupta & Waymire 1993) and uses multifractal techniques. The cascade model of Onof et al. (2005) subdivides rainfall at a coarser scale iteratively until the required resolution is achieved.

If historical data at an appropriate resolution are available, it would clearly be preferable to fit a model that is suitable at the desired timescale directly, without the need for further disaggregation or downscaling.

1.4 Research objectives and outline of thesis

The performance of both dynamical and statistical downscaling methods is continually improving, with developments in technology allowing ever finer resolution, and numerous research initiatives identifying and refining useful methodologies. However, a number of gaps remain between currently available approaches, and end-user requirements. Here, we focus on the ability to generate subdaily rainfall, allowing for temporal non-stationarity, with relatively low computational demands.

Our proposed approach is based on the Bartlett-Lewis model (although it could equally well be applied to the Neyman-Scott models), and thus shares the advantages of other clustered point process-based approaches, discussed in the previous sections. It has similarities with those of Kilsby et al. (1998) and Burton et al. (2010), in that we assume that future climate is stationary over monthly periods. Similarly to Kilsby et al. (1998) also, we use regression techniques to relate local rainfall to large-scale atmospheric variables. However, Kilsby et al. (1998)'s approach requires a two stage process — firstly the regression models are used to derive a series of statistics, then a series of point process based models must be fitted to each set of statistics.

We propose instead to remove a stage of the process by relating atmospheric variables directly to the model parameters within the fitting process, using a kernel-based nonparametric approach. It will be shown that the method proposed is a natural extension of the current approach, replacing the discrete covariate, calendar month, with various continuous covariates within a Generalised Method of Moments framework. We will also show how a tuning constant ('bandwidth') can be adjusted in order to ensure optimal use of the available data, such that an appropriate effective sample size is used at each covariate point, with parameters changing smoothly with covariate values.

A key advantage of our method over many other approaches is that, by using a range of statistics over a set of different timescales, we relate not just one or two key statistics, such as the mean daily rainfall and proportion of dry days, to large scale atmospheric parameters, but also other important properties of rainfall, such as variability and skewness. This means that our method allows great flexibility in allowing the structure of rainfall to

vary if the relationships and future values of our predictands indicate that this should be so. In addition, no constraints are imposed on the form of the relationships between the fitted parameters and the covariates, and valuable intuition is gained by examining these relationships. As well as allowing the clustered point process-based models to incorporate potential climate change scenarios, relating the model parameters to suitable large scale atmospheric covariates should also improve the interannual variability, which is generally under-estimated, even in a stable climate.

Although the main new development of this thesis is the non-stationary version of the Bartlett-Lewis clustered rainfall model, we are interested also in the feasibility of a model with just one fitting stage, capable of providing a rainfall series down to a five minute resolution (assuming that appropriate historical data are available for fitting), without the need for downscaling or disaggregation. We therefore first carry out a review of the key existing variants of the Bartlett-Lewis based model, developing two new alternatives as part of this review. Our aim is to clarify the drivers behind model performance at a sub-hourly timescale, and to identify the optimal choice for fine-scale data.

The structure of the rest of this thesis is as follows. First, in Chapter 2, we discuss the fitting methodology of the point process-based models, looking at new developments in the theory, including the issue of parameter uncertainty and identifiability. In Chapter 3 we carry out a review and introduce two new model variants, as discussed, using a time series of 69 years of five minute data from Bochum in Germany. We focus solely on the Bartlett-Lewis suite of models here, rather than on models based on the Neyman-Scott clustering mechanism, principally because methodology for an additional level of clustering has been developed for the former, but the two mechanisms generally exhibit similar performance (Rodriguez-Iturbe et al. 1987). Chapter 4 introduces the methodology for the proposed nonparametric models, with Chapters 5 and 6 providing a practical study, including the selection of suitable covariates and detailed results of the fitting respectively. The focus of this practical work is to demonstrate the viability of the methodology using reanalysis data, and we do not produce any simulations in respect of future climate scenarios here. This next stage would require consideration of the most appropriate source of large-scale atmospheric variables, including issues such as model biases, model uncertainty and the use of ensembles, different grid scales etc. Chapter 7 concludes the thesis with some thoughts on this and other future work.

The software: *momfit* (Chandler et al. 2010) has been used for fitting the global models in Chapter 3, and individual functions from this suite have also been used within Chapter 6. All the programs developed within this thesis, have been written in R (R Development Core Team 2010). R packages used include:

- *RNetCDF* (Michna 2012) for interfacing with the NCEP NetCDF datasets,
- *Lattice* (Sarkar 2010) and *Fields* (Furrer et al. 2011) within some of the plotting

routines, and

- *mvtnorm* (Genz et al. 2012, Genz & Bretz 2009) for sampling from the Multivariate Normal distribution.

The methodology developed here is intended for practical application, and so computation times are of interest. Any such times given throughout the thesis are based on runs carried out on a home computer.

Chapter 2

Methodology in respect of existing point process-based models

In this chapter, we consider the fitting methodology in respect of the point process-based rainfall models, focusing on the simplest of the clustered Bartlett-Lewis models, the Barlett-Lewis Rectangular Pulse (BLRP) model, as an example. We start by examining the structure of this model, then give a description of the fitting method and associated issues. We decide on the fitting properties that are used in all the empirical work of this thesis, describe our numerical optimisation approach, and consider the uncertainty of the parameter estimates. The chapter concludes with some thoughts on the assessment of model performance, and on model selection. More complex variants of the basic model are considered in Chapter 3.

2.1 Description of the BLRP model

The specification of the basic Bartlett-Lewis Rectangular Pulse (BLRP) model (Rodriguez-Iturbe et al. 1987) is as follows:

- Rain-events (or 'storms') arrive in a Poisson process of rate λ.
- Each event generates a cluster of cell arrivals, with a cell at the storm origin itself. The Bartlett-Lewis clustering mechanism assumes that the time intervals between successive cells are independent, identically distributed random variables. It is normally assumed that the intervals between cells are exponentially distributed, so that the cell arrivals constitute a secondary Poisson process of rate β.
- Each cell is associated with a rectangular pulse of rain, of random duration, L, and with random intensity, X. In the simplest version of the model, these are both

assumed to be exponentially distributed with parameters η and $1/\mu_X$ respectively, and are independent of each other.

- The cell origin process terminates after a time that is exponentially distributed with rate γ.

This basic version thus has five parameters in total. Both storms and cells may overlap, and the total intensity of rain at any point in time, $Y(t)$ is given by the sum of all pulses 'active' at time t. Note that there may be periods of zero rain within a storm. The process in respect of a single storm is illustrated in Figure 2.1.

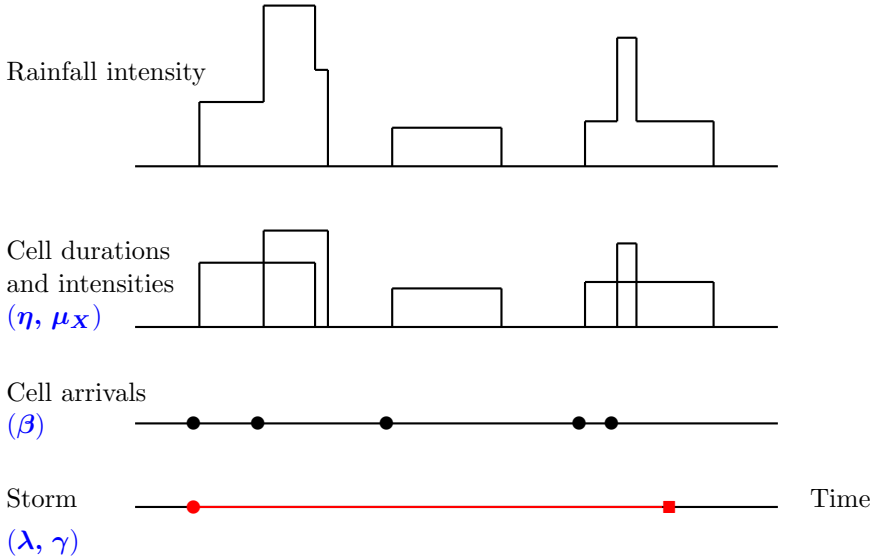


Figure 2.1: Illustration of a single storm within the Bartlett-Lewis Rectangular Pulse model, with parameters shown in blue.

Additional flexibility can be added by allowing for a distribution with more parameters for pulse intensities. In particular, a distribution with a longer tail may help with the fit of extreme values, and popular variants include the Gamma and Weibull distributions. One additional parameter is required for either of these. Moving away from the exponential distribution for cell durations would be less straightforward, however, as the Markov structure of the model would be lost, and hence its mathematical tractability.

2.2 Fitting methodology

As mentioned in Section 1.1, the models are fitted to discrete data from rain-gauges, and their dependence structure, particularly when aggregated, makes a full maximum likelihood method impracticable. Some attempts have been made to use likelihood based

methods, however, and we consider these very briefly. Chandler (1997) uses an approximate likelihood function based on the Fourier transform of the data. This simplifies the likelihood construction, because the sample Fourier coefficients are asymptotically independent and normally distributed. However, the method's reliance on only the mean and second-order properties of the data, results in poor performance in respect of wet and dry interval properties. Salim & Pawitan (2003) create a 'quasi-likelihood' by assuming that vectors of rainfall data over disjoint seventy-two hour intervals are independently drawn from a 72-dimensional multivariate normal distribution. The dependence of the rainfall process is thus characterised by the variance-covariance matrix of the 72 hourly totals. This choice is made primarily for computational convenience, and clearly does not reflect the highly skewed nature of hourly rainfall. The fit to the proportion of intervals that are dry is shown to be poor, and it is unlikely that such a method could accurately reflect properties other than those of first and second-order.

A different approach is taken by Northrop (2006), who constructs a marginal likelihood for the binary sequence of rainfall occurrences over consecutive time intervals for the unclustered Poisson rainfall model. The temporal parameters of the Poisson model are estimated from this marginal likelihood, with the intensity parameters estimated using the method of moments. However, for the more realistic, clustered models, direct maximisation of the marginal likelihood is computationally intensive, so strategies to alleviate the computational burden would be necessary. Rodriguez-Iturbe et al. (1988) warn that in any case, it is not at all clear that a method such as maximum likelihood could be the optimal method because of the idealised nature of the model, particularly the rectangular pulses. They argue that a likelihood function would tend to give undue emphasis to this unrealistic feature.

Due to the difficulties inherent in a likelihood-based approach, fitting of the point process-based rainfall models was originally carried out by the method of moments (Rodriguez-Iturbe et al. 1987). In this method, sample moments (or indeed other sample statistics or 'properties') are equated with their expectations according to the model. The number of statistics selected exactly matches the number of parameters, resulting in a set of simultaneous equations. For example, Rodriguez-Iturbe et al. (1987) apply the five-parameter BLRP model to data from Denver, Colorado with a fit based on the hourly mean, and the variance and lag-1 autocorrelations at aggregation levels of six and twelve hours. Later, it was deemed preferable to fit a larger set of sample moments approximately, rather than a smaller set exactly (Cowpertwait et al. 1996, Wheater et al. 2005), and models are now generally fitted using the generalised method of moments (GMM).

Some of the drive to use likelihood-based methods stems from a desire to compare different models using measures such as the Akaike information criterion (see, for example, Section 4.7 of Davison (2003)). The alternative generalised method of moments fitting approach, is also perceived as being overly subjective, due to the choice of weights and fitting properties.

However, as we highlight in this chapter, there is a theoretically optimal choice in respect of the weights, and methods for model and moment selection in the GMM framework also exist, albeit they are not widely used or developed.

We will use the generalised method of moments methodology to fit the models in this thesis. The theory is introduced in the next section, and we consider its application to the point process-based models in more detail in Section 2.4.

2.3 Generalised method of moments

The generalised method of moments is an extension of the method of moments, in which the number of fitting properties exceeds the number of unknown parameters. The GMM estimator is very popular in econometrics, dating in particular from the seminal paper of Hansen (1982). Advantages include the fact that it is computationally convenient, and, assuming certain conditions, gives consistent, asymptotically normally distributed estimators of the parameters. Hall (2010) provides a useful summary of GMM estimation, and Hall (2005) a comprehensive treatment. In this chapter the key results are presented without proofs. The methodology described is the starting point from which the new local approach of Chapter 4 is developed.

GMM estimation requires an equation of the form:

$$\mathrm{E}[g(Y, \theta_0)] = 0, \tag{2.1}$$

where θ_0 is the unknown parameter vector, of dimension q say, Y is a vector of random variables, and $g(\cdot)$ is a vector of functions. Such an equation is known as a set of 'population moment conditions'. In order for the system to be 'identified', this equation must have a unique solution, so that $\mathrm{E}[g(Y, \theta_0)]$ does not equal zero for $\theta \neq \theta_0$.

Commonly, and for the rainfall models fitted here, g takes the form:

$$g(Y, \theta) = T(Y) - \tau(\theta),$$

where $T(Y)$ is a vector of summary statistics, and $\tau(\theta)$ is the vector of their expected values under the model.

If $T(Y)$ also has dimension q, then the estimator $\hat{\theta}$, obtained by replacing the population moment condition with its sample equivalent, is the usual 'method of moments' estimator. If, however, $T(Y)$ has dimension $k > q$, then generally, due to sampling variation, there will be no unique solution. The model is then said to be 'over-identified'. In that case, the GMM estimator is given by the value of θ that minimises the distance between $T(Y)$

and $\tau(\theta)$, with distance measured by the following quadratic form:

$$S_n(Y, \theta) = [T(Y) - \tau(\theta)]^{\mathrm{T}} W_n [T(Y) - \tau(\theta)], \tag{2.2}$$

for some positive definite weighting matrix W_n, which ensures that we put a positive, non-zero weight on all the moment conditions. The subscript n here indicates the sample size, and S_n is referred to as the 'objective function'. W_n may depend on the data, but in order for the estimator to be well-defined it is required to converge in probability to W, a positive definite matrix of constants. The theoretically optimal approach (in terms of minimising the variance of the parameter estimates) of Hansen (1982), takes W_n as proportional to the inverse of the covariance matrix of the statistics. This is intuitively appealing, since a statistic with a relatively small variance is more informative than one with a large variance, and therefore should be given a greater weight. Also, it seems sensible to adjust for the correlations between the statistics, which are unlikely to be entirely uncorrelated in practice.

The estimators are consistent, provided certain regularity conditions are met. As already discussed, a valid set of population moment conditions is required which identifies θ_0. Identification requires both an appropriate model structure, and sufficient information in the observed data so that θ_0 can be determined. In practice, identification problems are often addressed by restricting the possible parameter space. The primary purpose of other conditions is to ensure that the Law of Large Numbers can be applied to the sample statistics, $T(Y)$, so that these tend in probability to their expectations $\tau(\theta)$. This is true for the statistics typically selected in rainfall model fitting.

2.4 Fitting approach: the rainfall application

In this section, we consider the GMM fitting approach in more detail, in the context of our practical application. Note that GMM may be thought of as a special case of an estimating function, and the theory and methodology are reviewed in this context by Jesus & Chandler (2011), who derive asymptotic results, and take the Poisson rainfall model as an example. This is the simplest of the point process-based models, with no clustering. The fitting methodology used for the models in this thesis broadly follows that described by Jesus & Chandler (2011).

In order to apply the methodology in practice, a number of decisions are required in respect of various components, key amongst which are the following:

- derivation of the weighting matrix,
- selection of the fitting properties,

- minimisation of the objective function, which typically requires numerical optimisation techniques.

These will be considered in turn in this section. In addition, it is important to consider the level of uncertainty involved in the parameter estimates, as well as validation of the model, and selection between alternative models. These are covered in Sections 2.5 and 2.6 respectively.

2.4.1 The weighting matrix

Historically, a diagonal weighting matrix has generally been assumed, so that the objective function becomes:

$$S_n(Y,\theta) = \sum_{i=1}^{k} w_i \, [T_i(y) - \tau_i(\theta)]^2,$$

where the weight w_i is applied to statistic i. Usually, the weights have been chosen so that each property contributes broadly equally to the fit. Jesus & Chandler (2011) carry out a simulation study involving the Poisson rainfall model, and compare results using these traditional approaches, against those obtained with the theoretically optimal inverse covariance matrix. The analytical expressions in respect of our fitting statistics are complex, and deriving theoretical expressions for the covariances between the statistics is not possible. However, empirical estimates can be used instead. Jesus & Chandler (2011) confirm that variance-based weighting schemes deliver much less variable estimators than the popular fixed weighting schemes. However, they find that a two-step approach is required in order to derive a reliable sample estimate of the full covariance matrix. Using just a single step with the sample covariance matrix, the diagonal matrix of inverse variances performs as well as the full covariance matrix, and is close to optimal. In order to reduce the computational burden (which becomes more onerous in the local estimation of Chapter 4), we have therefore chosen to use the simpler diagonal matrix in the practical applications in this thesis.

In order to estimate the diagonal matrix of inverse variances, we follow the approach originally suggested by Rodriguez-Iturbe et al. (1988), which involves first calculating the fitting properties separately for each observation month, to give $T(Y_t),\ t = 1, \ldots n$, where Y_t denotes the data for observation month t, and we have n months of data. Separate models are fitted in respect of each calendar month to allow for seasonality, with monthly vectors of statistics assumed independent across years. The vector of mean summary statistics in respect of calendar month m is then given by:

$$T_m = \frac{1}{n_m} \sum_{t=1}^{n} I(m_t = m) \, T(Y_t),$$

where m_t is the calendar month of the tth observation month, and $n_m = \sum_{t=1}^{n} I(m_t = m)$.

I is the indicator function, such that $I(x) = 1$ if x is true, and 0 otherwise. The sample covariance matrix of the mean of the statistics in respect of calendar month m can then be calculated as:

$$\mathrm{Var}(T_m) = \frac{1}{n_m(n_m - 1)} \sum_{t=1}^{n} I(m_t = m)\ (T(Y_t) - T_m)(T(Y_t) - T_m)^{\mathrm{T}}. \qquad (2.3)$$

We take the inverse of the diagonal elements as our weights. (Note that we have used the variance of the mean statistics here, but we could equally well have used the variance of the statistics themselves.)

There is an issue in terms of finite sample bias with this approach, since components of the statistics vector $T(Y_t)$ that are at the daily timescale will be based on only around 30 observations. However, in their simulation study Jesus & Chandler (2011) found that any bias in the summary statistics was negligible compared with sampling variability, and this is not thought to be a significant issue from a practical perspective.

Bootstrapping is an alternative approach to the calculation of the covariance matrix of fitting statistics. In order to preserve the dependencies in the rainfall time-series, some form of block-bootstrapping is required. The main advantage of such an approach is the reduction of the finite-sample biases discussed above, since statistics can be calculated over the entire re-sampled data sets. However, the method adds a substantial additional computational burden, and was found not to yield any clear improvement (in terms of the coverage of confidence intervals for the parameters) by Jesus (2012), in the context of his simulation study of the Poisson rainfall model.

An additional appeal of the proposed approach, whereby we treat the data for each model as a sample of n_m replicates of the summary statistics, is that it extends neatly to the local modelling framework.

2.4.2 Fitting properties

The choice of which, and how many, statistics to include, and at which aggregation levels is a fairly subjective one. Rodriguez-Iturbe et al. (1988) suggest that the properties chosen should be likely to have relatively small sampling errors, and not to be highly mutually correlated. Chandler (2003) recommend choosing statistics that have variances that are as small as possible, and that vary rapidly with respect to the model parameters. In practice, the choice of properties is rather restricted if it is deemed essential that analytical expressions be available for $\tau(\theta)$. Although simulations can be used, they substantially increase the computational burden. Thus most studies have focused on the mean, variance, autocorrelations and the proportion of dry intervals at a range of aggregation levels. Cowpertwait (1998) argues for the inclusion of the skewness coefficient in order to get a better fit to the tail of the empirical distribution. In order to reflect the observed structure of the

rainfall process, there need to be properties at a timescale which will provide information on the behaviour of single rain cells, and at least one property should correspond to the timescale of the interval between storms.

Cowpertwait et al. (2002) favours dimensionless properties, so, for example, the coefficient of variation would be included instead of the variance. These have a particular advantage when fitting multi-site models, when most parameters can be found using dimensionless properties across all sites, with the scale parameter for cell intensity calculated separately for each site, based on its sample mean.

Here we follow Cowpertwait et al. (2007) in our choice of fitting properties and use the hourly mean, plus the coefficient of variation, lag-1 autocorrelation and skewness at timescales of 5 minutes, 1 hour, 6 hours and 24 hours. Thus T and τ are vectors of dimension 13.

The derivations of the mean and second order properties of the aggregated rainfall series for the BLRP model are shown in some detail in Appendix A.2 as an example of the methodology required for these derivations. Appendix A.3 gives the analytical expressions for the fitting properties for the BLRP model.

2.4.3 Numerical Optimisation

Equation (2.2) can be expressed in the form of an estimating equation by differentiating with respect to θ and setting equal to zero, but in practice the estimator is found by minimising the objective function itself using numerical optimisation techniques. The approach followed here is that of Wheater et al. (2005), and we have used the optimisation routines developed for that project. Firstly, a set number of optimisations are carried out using the Nelder-Mead simplex method, each starting with a different initial value for the set of parameters. This set of initial values is generated by random perturbation about a single user-supplied value. The best parameter set is then used as a new starting value for a further set of optimisations, which now use a Newton-type algorithm. The reason for the use of two different optimisation routines is that the first is more robust and thus well suited to identifying promising regions of the parameter space, whereas the second is more powerful if given good starting values.

If different starting values give radically different final parameters, then it suggests that the model is over-parameterised. If we do not wish to reduce the number of parameters, it may be necessary to introduce constraints on the parameters, as discussed in Section 2.3 to keep them within physically realistic bounds (see for example Cowpertwait (1998)), and to try to ensure smoothly changing parameters month on month.

An alternative approach to encourage smoothly changing parameters is to carry out this two-stage optimisation only for the first fit, for the month of January, say. The fitted

parameter set from January is then taken as the initial value for the fit in February, and so on. Each of these subsequent fits is based on only a single Newton-type optimisation. This helps to ensure both a quick calculation time, and smooth changes in the parameters between months, since the fits in respect of neighbouring months should be close. These issues are far more important when we consider continuous covariates in Chapters 4 and 6, for which we have many more evaluation points. For the initial review of Chapter 3, however, models only include calendar month as a single, discrete covariate. Here we will take the first approach, i.e. that of Wheater et al. (2005). A brief comparison of the two approaches is undertaken in Section 4.6.1.

In practice, for the purposes of the numerical optimisation, we parameterise the objective function with the logarithms of the rainfall model parameters. This ensures that the fitted parameters are positive, and has also been found to improve the stability of the numerical optimisation. Thus for the five-parameter BLRP model, the parameter vector is given by $\theta = (\log\lambda, \log\mu_X, \log\beta, \log\gamma, \log\eta)^{\mathrm{T}}$. If we want to replace the exponential intensity distribution with the Weibull, for example, we need to add a shape parameter, $\log\varrho$, say.

2.5 Estimation of uncertainty

Different approaches to estimating parameter uncertainty have been taken in the literature. Rodriguez-Iturbe et al. (1988) look at parameter stability for the random parameter Bartlett-Lewis model by perturbing the input statistics by small amounts ($\pm$ 2%) and looking at the impact on the resulting parameter estimates. Cowpertwait (1998) uses a bootstrap approach, obtaining 100 sets of parameter estimates by fitting a Neyman-Scott model 100 times, each time using whole years sampled with replacement from the series of observed data. Sampling whole years (separately for each calendar month) ensures that the dependencies in the rainfall series are captured.

Approximate standard errors can be calculated using the theory of estimating functions. This broadly involves taking a Taylor series expansion of the estimating equation and considering the limiting behaviour as the sample size increases. The theory gives an asymptotic distribution for the estimated parameter vector $\hat{\theta}_m$ for calendar month m (where θ_m is the true value of the parameter vector and W_m is the limiting weighting matrix), of:

$$n_m^{1/2}\{\hat{\theta}_m - \theta_m\} \sim N_q\Bigg(0, \left\{\left[\frac{\partial\tau(\theta_m)}{\partial\theta}\right]^{\mathrm{T}} W_m \frac{\partial\tau(\theta_m)}{\partial\theta}\right\}^{-1} \left[\frac{\partial\tau(\theta_m)}{\partial\theta}\right]^{\mathrm{T}} W_m \mathrm{Var}[T(Y)|M=m]\, W_m \frac{\partial\tau(\theta_m)}{\partial\theta} \left\{\left[\frac{\partial\tau(\theta_m)}{\partial\theta}\right]^{\mathrm{T}} W_m \frac{\partial\tau(\theta_m)}{\partial\theta}\right\}^{-1}\Bigg), \quad (2.4)$$

where $\mathrm{Var}[T(Y)|M=m]$ denotes the covariance matrix of the statistics conditional on the calendar month being m, and the notation $\frac{\partial\tau(\theta_m)}{\partial\theta}$ is used to represent $\frac{\partial\tau}{\partial\theta}\big|_{\theta=\theta_m}$.

The standard errors of the parameter estimates can then be approximated by taking the square root of the diagonal elements of the covariance matrix, divided by $n_m^{1/2}$. We replace θ_m by $\hat{\theta}_m$, the limiting weights matrix, W_m with that based on our sample, W_{n_m}, and estimate $n_m^{-1}\text{Var}[T(Y)|M = m]$ by the sample covariance matrix of the mean statistics, given by Equation (2.3). The derivatives are calculated numerically. We can use the asymptotic distribution to construct approximate confidence intervals for the individual parameters, or a confidence region for the whole parameter set. Examination of the correlation matrix for the parameter set can identify where a model is over-parameterised. Bootstrapping may also be useful here, allowing the relationships between pairs of fitted parameters to be plotted and examined.

If we choose W_{n_m} to be the optimal weights matrix given by the inverse of $\text{Var}[T(Y)|M = m]$ as explained in Section 2.3, then the variance estimate simplifies to:

$$\text{Var}(\hat{\theta}_m) \approx \left\{ \left[\frac{\partial \tau(\theta_m)}{\partial \theta} \right]^{\mathrm{T}} \{\text{Var}[T_m]\}^{-1} \frac{\partial \tau(\theta_m)}{\partial \theta} \right\}^{-1}. \tag{2.5}$$

Another approach, used by Chandler (2003), is the examination of profile objective functions. Each parameter in turn is fixed at each of a set of values, and the objective function is optimised over the remaining parameters. The resulting plot for each parameter showing the set of parameter values against the optimised objective function provides a useful means for assessing the identifiability of the parameter — for example, a very flat objective function indicates a wide range of plausible values. Approximate confidence regions can also be calculated using the objective function $S_n(Y, \theta)$ itself, defined as the set of parameters for which $S_n(Y, \theta)$ is less than some threshold. In order to determine the appropriate threshold for a given level of confidence, Chandler (2003) takes a second order Taylor series expansion of $S_n(Y, \theta)$ about $\hat{\theta}$, to get:

$$2[S_n(Y, \theta_m) - S_n(Y, \hat{\theta}_m)] \approx (\theta_m - \hat{\theta}_m)^{\mathrm{T}} \, H(\theta_m)(\theta_m - \hat{\theta}_m), \tag{2.6}$$

which is a quadratic form in normally distributed random variables. Here $H(\theta_m)$ is the expected value of the second derivative of the objective function, given by $-\left[\frac{\partial \tau(\theta_m)}{\partial \theta}\right]^{\mathrm{T}} W_m \frac{\partial \tau(\theta_m)}{\partial \theta}$. Chandler (2003) then follows an idea of Bowman & Azzalini (1997) (Chapter 5) whereby the real distribution of this quadratic form (which is intractable unless the optimal weighting matrix is used) is replaced by a more convenient distribution with the same first three moments: a scaled and shifted χ^2 distribution.

The rth cumulant of the quadratic form is given (Kuonen 1999) by:

$$\kappa_r = 2^{r-1} \, \Gamma(r) \, \text{tr}\{[\text{Var}(\hat{\theta}_m) H(\theta_m)]^r\}, \tag{2.7}$$

where tr() denotes the trace operator. In practice we replace θ_m with $\hat{\theta}_m$ as before.

The distribution of the quadratic form is then approximated by that of $aX + c$, where $X \sim \chi^2_b$ and:

$$a = \frac{|\kappa_3|}{4\kappa_2} \qquad b = \frac{8\kappa_2^3}{\kappa_3^2} \qquad c = \kappa_1 - ab. \tag{2.8}$$

A $100(1-\alpha)\%$ confidence region can then be taken as the set of parameter values for which the objective function is below the threshold given by:

$$S_n(Y, \hat{\theta}_m) + \frac{1}{2}\left[a\,\chi^2_{b,(1-\alpha)} + c\right],$$

where $\chi^2_{b,(1-\alpha)}$ is the upper $100\alpha\%$ percentile of the χ^2 distribution with b degrees of freedom.

Confidence intervals for an individual parameter, ρ, say, can be derived using similar methodology. In that case, Chandler (2003) shows that $H(\theta_m)$ in Equation (2.7) should be replaced by the inverse of the element of H^{-1} that corresponds to ρ, and $\mathrm{Var}(\hat{\theta}_m)$ should be replaced with the variance of ρ. Denoting these two scalars respectively by h and v, the constants of Equation (2.8) are then given by $a = hv$, $b = 1$ and $c = 0$.

Generally, throughout this thesis approximate 95% confidence intervals are derived, where given, by adding ± 2 standard errors to the parameter estimates, with the standard errors calculated using Equation (2.4). Since we are fitting logged parameters, these are confidence intervals about the geometric, rather than the arithmetic, mean parameter estimates. Other ideas discussed here will also be illustrated for certain models. Care must be taken with all inference, given that results are asymptotic, whereas we have a finite sample, and further noting that the asymptotic results rely on Equation (2.1) being true. This is discussed in a bit more detail in the next section.

2.6 Model selection, assessment of model performance and model misspecification

In the literature there are numerous variants of the basic BLRP model, some of which have a very high number of parameters. Although generally increasing the complexity of the models improves the fit, as expected, it comes with a number of disadvantages, including identifiability issues, and consequent variation in the fitted parameter sets from month to month. Since one of the claimed advantages of this type of model is the fact that the parameters have physical meaning, this is rather undesirable. Further, while different parameter sets may give similar estimates in respect of the fitting properties, it is possible that they generate very different behaviour in respect of other features of the rainfall process, which again is unlikely to be desirable. Some form of objective approach to model selection would be helpful.

Jesus & Chandler (2011) consider formal tests that allow us to compare two competing models, where one model can be defined as a constrained version of the other. A set of $r < q$ linear constraints may be expressed in the form $C\theta_m = c_0$ where C is an $r \times q$ matrix, and c_0 an $r \times 1$ vector. For example, we can test whether a Weibull intensity distribution is preferable to an exponential by setting: $\theta_m = (\log \lambda_m, \log \mu_{X_m}, \log \varrho_m, \log \beta_m, \log \gamma_m, \log \eta_m)^{\mathrm{T}}$, $C = (0, 0, 1, 0, 0, 0)$ and $c_0 = 0$.

Using the asymptotic results from Section 2.5, we have that $\hat{c}_m$, the estimator of c_0 for month m, has the approximate distribution:

$$\hat{c}_m \sim N_r(C\,\theta_m, C\,\mathrm{Var}(\hat{\theta}_m)\,C^{\mathrm{T}}). \tag{2.9}$$

Then

$$(\hat{c}_m - c_0)^{\mathrm{T}}[C\,\mathrm{Var}(\hat{\theta}_m)\,C^{\mathrm{T}}]^{-1}(\hat{c}_m - c_0) \tag{2.10}$$

has approximately a χ^2 distribution with r degrees of freedom, and so we reject the null hypothesis $H_0 : C\theta_m = c_0$ if this statistic is greater than, say, the 95% quantile of the χ^2_r distribution. This hypothesis would need to be tested separately in respect of each calendar month, m. This is known as a 'quasi-Wald' test. This test may also be extended to the case where the restrictions are not linear: see Hall (2005), Section 5.3 for this and alternative nested hypothesis tests in respect of the parameter vector.

These types of methods are only useful in certain limited situations. For likelihood-based fitting, more general techniques are available for selection between non-nested models, including for example Akaike's information criterion (AIC) and Bayes' information criterion (BIC). Both of these methods include penalties for additional parameters. Similar approaches are currently under-developed, however, for models fitted using the generalised method of moments, although some ideas have been proposed. Andrews & Lu (2001) suggest selection criteria for both model and moment selection (MMSC), which are based on the so-called 'J-statistic' for testing over-identifying restrictions (i.e. the validity of the model), with a penalty term for the use of more parameters, and a reward term for the use of more moment conditions. The J-statistic is analogous to (minus) the log-likelihood of methods such as AIC and BIC. It is given by the value of the sample size times the objective function evaluated at $\hat{\theta}$, with the weights matrix given by the inverse of the covariance matrix of the statistics, using a two-step estimation procedure, so for our model by:

$$J_{n_m} = n_m\,[T_m - \tau(\hat{\theta}_m)]^{\mathrm{T}}\,(\mathrm{Var}[T(Y)|M = m])^{-1}\,[T_m - \tau(\hat{\theta}_m)]. \tag{2.11}$$

J_{n_m} converges to the χ^2_{k-q} distribution under the null hypothesis that the moment conditions given by Equation (2.1) hold, and so may be used to test whether these are supported by the data. Andrews & Lu (2001)'s MMSC-BIC criterion selects the model with q pa-

rameters, and k moment conditions, such that $J_{n_m} - (k-q)\log n_m$ is minimised.

It is important to note, however, that all these tests, and the asymptotic theory in general, only hold if our model is correctly specified, i.e. if the moment conditions given by Equation (2.1) hold. Where the model is misspecified, $(n_m)^{1/2}[T_m - \tau(\hat{\theta}_m)]$ diverges, and $J_{n_m} \to \infty$ as $n_m \to \infty$. Hall (2005) (Section 5.3) therefore stresses the importance of using a model specification test (such as the J-test using the J-statistic described above) first before undertaking inference about the parameters.

In practice, however, we may want to find the model that gives the best prediction, without assuming that any of the models are correct. As in the likelihood setting, an alternative approach to model selection uses cross-validation or some other form of out-of-sample validation. We use this type of approach in Section 6.4 in the context of covariate selection for the local models. However, this method (discussed in more detail in Section 6.4) is arguably less appropriate in the case where the selection is between competing models with different numbers of parameters. This is because it effectively minimises the approximate weighted sum of mean squared errors of the estimated statistics, not of the estimated parameters. There is thus no effective penalty applied in respect of badly identified models, with markedly different parameter sets in different months, as long as the fitted statistics are stable. As discussed earlier, however, such variation is considered undesirable.

Another alternative is to use a form of selection based on the AIC. In the likelihood setting this is based on the minimisation of the expected (over the distribution of $\hat{\theta}$) Kullback-Leibler discrepancy. After some approximation, this leads to the minimisation of the statistic:

$$2\{-l(\hat{\theta}) + \text{trace}(\hat{H}^{-1}\hat{K})\} \tag{2.12}$$

where H is minus the expected second derivative of the log-likelihood, and K is the covariance matrix of the score function. The factor of 2 is there for historical reasons, and could be omitted. For a correctly specified model, $K = H^{-1}$, so that $\text{trace}(H^{-1}K) = \text{trace}(I_q) = q$, where q is the dimension of the parameter vector, θ.

In the GMM framework, Andrews & Lu (2001) replace minimisation of the negative log-likelihood by minimisation of the J-statistic, as discussed earlier in the case of the MMSC-BIC statistic. Taking H as the expected second derivative of the objective function, and K as the covariance matrix of its gradient vector, we have:

$$H = 2\left[\frac{\partial\tau(\theta_m)}{\partial\theta}\right]^{\mathrm{T}} W_m \frac{\partial\tau(\theta_m)}{\partial\theta} \tag{2.13}$$

and

$$K = 4\left[\frac{\partial\tau(\theta_m)}{\partial\theta}\right]^{\mathrm{T}} W_m \,\mathrm{Var}[T(Y)|M=m]\, W_m \,\frac{\partial\tau(\theta_m)}{\partial\theta}. \tag{2.14}$$

With the optimal weighting matrix, given by the inverse of the covariance matrix of the fitting statistics, it can be seen that $\mathrm{trace}(H^{-1}K) = 2\,\mathrm{trace}(I_q) = 2q$. So the statistic to be minimised is given by $J_{n_m} + 2q$. If we also reward additional moment conditions by subtracting $2k$, then this gives Andrews & Lu (2001)'s MMSC-AIC criterion.

In the case of a sub-optimal weighting matrix, however, such as the diagonal matrix of inverse variances used in the empirical work here, the penalty would not reduce in this way, and would be given by $\mathrm{trace}(H^{-1}K)$. It is not immediately clear, however, whether the model selection criterion based on minimisation of $J_{n_m} + \mathrm{trace}(H^{-1}K)$ would give a proper analogue of the usual (likelihood-based) selection procedures. In particular, further analysis is required in order to understand the implication of the J-statistic no longer having a χ^2 distribution. Since the main focus of this thesis is to develop and fit point process-based rainfall models that allow for temporal non-stationarity, we do not consider the issues of moment selection or model misspecification and its impact on inference further here. However, the latter is undoubtedly a topic of interest for further research in respect of the point process-based rainfall models, and some form of Information Criterion selection procedure looks promising.

It could be argued in any case that, given that none of the models are considered to be 'correct'(even if the moment conditions hold), it is sensible to consider model selection in the context of how well they perform in respect of features that are important to practitioners, rather than in terms of global measures of fit (Hansen 2005). Since the number of properties included in the objective function exceeds the number of parameters and we have a finite sample, there is no guarantee that there will be a good fit to all the fitting properties. In Chapter 3, the adequacy of the fit is thus assessed by considering properties used in the fitting procedure, as well as others that are of interest in hydrological applications. Some properties will need to be assessed using simulations, for example, extreme values (which are considered in Section 2.6.1). Assessing the output from some hydrological model for which our simulations form an input, such as a rainfall-runoff model, may also be a good way to investigate lack of fit, although not considered here.

2.6.1 Extreme value performance

Finally in this chapter we consider model performance in respect of rainfall extremes. These are very important for hydrological design, but unfortunately the point process-based models often do not perform well in this respect. This is a feature of many stochastic weather generators, and successfully modelling both typical and extreme rainfall with a

parsimonious number of parameters is a challenging undertaking. It is particularly so here, where we try and achieve a good fit simultaneously at timescales ranging from five minutes, to daily.

In order to assess the fit to extremes, we compare observed maxima at each timescale (five minutes, and one, six and twenty-four hours) against 100 simulations from each model, where each simulation is of the same length as the observed data. In order to reflect parameter uncertainty, the parameter set θ_m for each of these simulations is sampled from the $MVN_q(\hat{\theta}_m, \mathrm{Var}[\hat{\theta}_m])$ distribution (described in Section 2.5). Note that, since θ_m represents the logarithms of the underlying rainfall model parameters, the latter are assumed to be distributed as multivariate lognormal, and so are right-skewed.

It is normally assumed that the yearly rainfall maxima at each required timescale are distributed according to a type I Generalised Extreme Value (GEV) distribution, otherwise known as a Gumbel distribution, which has the cumulative distribution function:

$$F(z) = \exp\left\{-\exp\left(-\frac{z-\mu}{\sigma}\right)\right\}, \quad -\infty < z < \infty, \tag{2.15}$$

where μ and σ are location and scale parameters respectively. Re-arranging to find the upper pth quantile, z_p, we get:

$$z_p = \mu - \sigma \log\{-\log(1-p)\}. \tag{2.16}$$

The upper pth quantile, z_p, is known as the 'return level' associated with the 'return period' $1/p$ i.e. the approximate average time period within which rainfall of the specified magnitude can be expected to occur once (see Coles (2001), Chapter 3 for a description of the GEV models and related inference). The quantity $y(p) = -\log\{-\log(1-p)\}$ is known as the 'Gumbel reduced variate', and it can be seen that if z_p is plotted against $y(p)$, then we should get an approximately straight line. This representation is potentially useful for extrapolating past the range of the observed data. Effectively, $F(z)$ is transformed onto a more convenient scale.

This type of plot, which is known as a 'return level plot', is commonly used in the hydrological literature to examine extreme value performance. Both the observed maxima, and those from the model simulations can be plotted in this way, and the plots compared. In practice (Shaw et al. 2011), to eliminate bias, p is often calculated according to the Gringorten plotting position formula, given by:

$$p = \frac{k - 0.44}{n + 0.12}, \tag{2.17}$$

where k is the rank of the observation in the ordered set of maxima, and n is the number of years of data.

The fit to extremes should be examined separately for each calendar month in order to assess the individual models, and also (and more onerously) annually.

In the next chapter we carry out an empirical study, fitting the BLRP model and a number of variants to a time-series of five-minute rainfall data from Bochum in Germany, using the GMM methodology discussed here.

Chapter 3

Review of Bartlett-Lewis models, and further development

3.1 Motivation

So far, we have considered the simplest of the clustered Bartlett-Lewis rainfall models — the Bartlett Lewis Rectangular Pulse (BLRP) model. However, a variant of this, the Random Parameter Bartlett-Lewis Rectangular Pulse (BLRPR) model, in which the cell duration parameter is allowed to vary randomly between storms, is usually preferred (Rodriguez-Iturbe et al. 1988, Onof & Wheater 1993, Wheater et al. 2005). The BLRPR model effectively allows for a continuous range of storm types, with the addition of one further parameter. Both of these models have been considered unsuitable at sub-hourly resolution, however, because of the rectangular structure of the rain cells, and this led to the development of the Bartlett Lewis Instantaneous Pulse (BLIP) model (Cowpertwait et al. 2007). In this model, the rectangular cells with constant intensities of the original BLRP model are replaced by a Poisson process of instantaneous pulses, in order to allow greater variability in rainfall intensity over small time intervals. A version with two superposed processes provided a good fit to five-minute data from New Zealand, but required an excessive number of parameters (11 or 12).

The reason for using two superposed processes is to allow for the two main types of rainfall: convective and stratiform. The former is thus named as it arises from the convective clouds formed from rising thermals. These are more common in the summer months, when land temperatures heat up during the day, and the warm air close to the surface starts to rise. As condensation occurs, latent heat is released, and the rising air can continue to remain warmer than its surroundings. In an unstable atmosphere it will develop vertically into deep cumulus or even cumulonimbus clouds. These deep clouds, with their vigorous updraughts, give plenty of time for large raindrops to develop, and rain from these types of

clouds is typically of short duration, but heavy. In the extreme, these are thunderstorms. Stratiform rain is less intense, but with a longer duration, and arises from clouds that form in sheets or strata as a result of a mass of air being lifted, either due to a physical barrier such as a mountain, or due to the meeting of two distinct masses of air at different temperatures. Drizzle is produced by very shallow stratus clouds, where the raindrops have little time to develop.

Weather in the mid-latitudes is characterised by frontal systems, which are the areas where large air masses meet. Cold fronts tend to bring primarily convective rain, whereas warm fronts bring stratiform, although in practice these systems can produce many different kinds of weather, often in quick succession.

The BLRPR model addresses the issue of different types of storm in a far more parsimonious way than the approach with two superposed processes. The original motivation behind the work in this chapter was therefore to develop a random parameter version of the new BLIP model, and to undertake a systematic comparison of the performance of the three versions of the two model types when fitted to a time series of rainfall at five minute intervals. The versions here refer to: fixed parameter/random parameter/two superposed processes, and the types to: rectangular pulse/cluster of instantaneous pulses. Fitting follows the methodology described in Chapter 2, using the generalised method of moments, with the thirteen fitting properties selected in Section 2.4.2. This is not intended as a comprehensive review, rather a limited review of a few selected variants of the Bartlett-Lewis based models in order to find the one most suitable where fine-scale data is available.

In this chapter, we first introduce our dataset. The existing models of interest are then described, followed by the development of the new Random Parameter Bartlett-Lewis Instantaneous Pulse (BLIPR) model. All the models are fitted, and preliminary comparisons made using minimum objective function values. Conclusions drawn from these results lead us to the development of a new model variant, and selected models are then compared in more detail, using simple plots of observed versus fitted properties. Parameter uncertainty is considered using a number of methods from Chapter 2 in respect of our chosen model. Finally, we conclude with some thoughts for further improvements, and a summary of the key findings.

3.2 Bochum data

The data to which we fit the models is from a single site in Bochum in Germany, and includes sixty-nine years of five minute rainfall data, from January 1931 to December 1999. Bochum is a city in North Rhine-Westphalia, in the urban Ruhr area of Western Germany, with a temperate climate. Most of the rainfall measurements were obtained

using a Hellmann rain gauge, in which rain displaces a float and a marking pen attached to the float makes a continuous trace on a recording chart.

The percentage of days with missing data was highest for the month of December, at 3.4%, and lowest for September, at 0.9%. Inspection of the data identified three (non-consecutive) days where the date field had clearly been mis-coded. This was clear since the file contains duplicates of all the intervals within these days, with different data in each duplicated interval, and in each of the three cases, there is no record coded as belonging to the previous day. We have therefore recoded the dates of the first of each of these duplicated records to the previous days' dates. No other quality issues were found.

Samples of the Bochum rainfall series are shown in Figure 3.1, the first over the whole month of January 1981, and the second zooming in on a single day. A key characteristic of

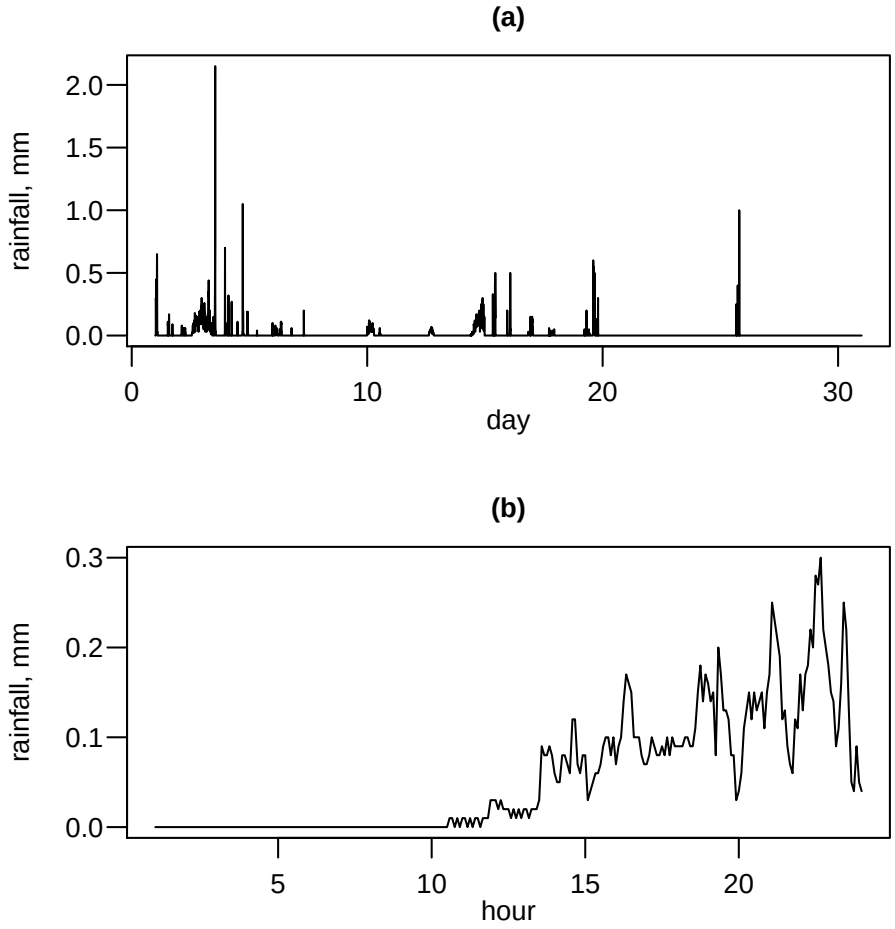


Figure 3.1: Samples of the Bochum series of 5 minute rainfall totals. (a) January 1981 (b) 14th January 1981

these data, clear in the first plot, is the fact that the totals are very often exactly zero, and this is what makes rainfall more difficult to model than other weather variables. The second

plot shows the high variability in rainfall intensities over five minute intervals, which, as discussed, was the motivation behind the new BLIP model. Two further exploratory plots are given in Figure 3.2. These examine the mean hourly rainfall by year and by calendar month. The high interannual variability of rainfall is evident, as are the seasonal differences, with no trend discernible over the period. As discussed in Section 2.6, model performance will be assessed by comparing moments and other properties of the observed and fitted series, and therefore further graphs summarising various features of the data will be given throughout the thesis, as appropriate.

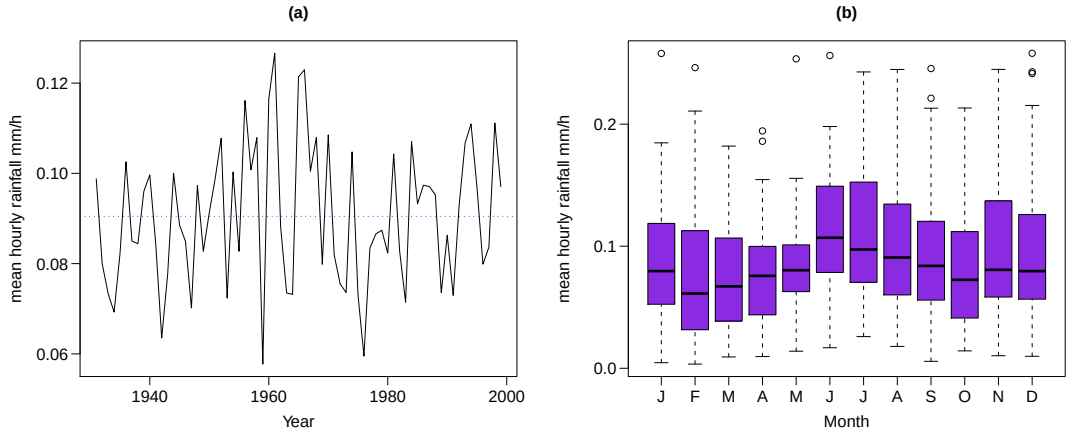


Figure 3.2: Mean hourly rainfall, mm, in Bochum (a) by year, with the overall mean of the series given in blue (b) by calendar month.

3.3 Existing Model descriptions

The basic BLRP model has been described in Section 2.1. The other two models of interest are summarised in this section. Both of these are based on the Bartlett-Lewis clustering mechanism, with an exponential distribution for inter-cell arrivals, so that cells within storms arrive according to a secondary Poisson process. Analytical expressions for the thirteen fitting properties for each of the models are given in Appendix A.

3.3.1 Random Parameter Bartlett-Lewis Rectangular Pulse (BLRPR) model

This randomised version of the BLRP model was introduced by Rodriguez-Iturbe et al. (1988), and the analytical expressions for the fitting properties are given in Appendix A.4. In this model the parameter η, that specifies the duration of cells, is allowed to vary randomly between storms. This is achieved by assuming that the η values for distinct storms are independent, identically distributed random variables from a gamma distribu-

tion with index α and scale parameter ν. The model is re-parameterised so that, rather than keeping the cell arrival rate, β, and the storm termination rate, γ constant for each storm, it is the ratio of both of these parameters to η that is kept constant. Thus, for a higher η (i.e. typically shorter cell durations), we have correspondingly shorter storm durations, and shorter cell interarrival times. Essentially the effect is that all storms have a common structure, but distinct storms occur on different (random) timescales.

The durations of cells and storms (more precisely cell origin processes) are both exponentially distributed, conditional on the cell duration parameter, η. Their unconditional distributions are Pareto type II. For example, the unconditional density of the cell duration, L, is given by:

$$\begin{aligned} f(l) &= \int_0^\infty f(\eta) f(l|\eta) \mathrm{d}\eta \\ &= \frac{\nu^\alpha}{\Gamma(\alpha)} \int_0^\infty \eta^\alpha e^{-\eta(\nu+l)} \mathrm{d}\eta \\ &= \frac{\alpha\, \nu^\alpha}{(\nu+l)^{\alpha+1}}, \qquad l \geq 0. \end{aligned}$$

This heavy-tailed distribution has an infinite mean if α is less than 1, and an infinite variance if α is less than 2. Also, in terms of the aggregated rainfall process, it turns out that, for values of α smaller than 3, the variance is infinite, and for values smaller than 4, the skewness is infinite. This is potentially problematic. For example, in practice it has been found that simulations with unconstrained values of α occasionally generate unrealistically long periods of rainfall (Onof & Arnbjerg-Nielsen 2009, Verhoest et al. 2010). Potential solutions to this issue include:

- setting constraints on α,
- rejecting storms or cells beyond a certain length or cells with an excessive intensity within any simulations,
- truncating the Gamma distribution for the cell duration parameter, η, with support (ε, ∞) (Onof & Arnbjerg-Nielsen 2009). The lower limit, ε, for the integrals over η can be pre-specified, or alternatively, can constitute a further parameter to be determined. This approach requires integrals in the analytical expressions to be evaluated using approximations.

Initial fits using the truncated Gamma approach were not successful here, with very long computation times, and this method was not pursued. Further investigation would be required to identify which features of our fit were causing the problem, which was not encountered by Onof & Arnbjerg-Nielsen (2009). This might, for example, relate to the inclusion of properties at five minutes, the particular fitting properties selected, the numerical optimisation routine used etc. In any case, for the fits using the BLRPR model

carried out here, α was not found to be small.

3.3.2 Bartlett-Lewis Instantaneous Pulse (BLIP) model

The Bartlett Lewis Instantaneous Pulse model (Cowpertwait et al. 2007), intended for fitting to fine-scale (of the order of five to fifteen minute) data, has a minimum of six parameters (one more than the original Bartlett-Lewis model), and is defined and parameterised as follows:

- Storm origins arrive in a Poisson process of rate λ.
- Each storm origin initiates a Poisson process of cell origins of rate β; in contrast to the basic Bartlett-Lewis model, it is not assumed that there is a cell at the storm origin itself, so a storm may have no rainfall. This is purely for mathematical convenience and does not lead to any loss of generality.
- Each cell origin initiates a further Poisson process of rainfall pulses of rate ξ. Again, it is not assumed that there is a pulse at the cell origin, so a cell may have no rainfall. Note that the pulses are instantaneous — they have a depth, but no duration. This Poisson process of instantaneous pulses replaces the rectangular pulse assumption of the original Bartlett-Lewis model.
- Both the duration of the cell origin process, and the cell duration are assumed to be exponentially distributed, the former with rate γ, and the latter with rate η. The process of pulses terminates with the cell or storm lifetime, whichever is the sooner.
- Associated with each pulse is a depth, X, so the pulse process is a marked point process (Cox & Isham (1980)). The model developed by Cowpertwait et al. (2007) allows pulse depths from a single cell to be dependent, but those from distinct cells are assumed independent. No specific dependence structure is specified, and the model fitted in the paper assumed independent, exponentially distributed pulse depths, with mean depth μ_X.

Analytical expressions for the fitting properties are given in Appendix A.6. As discussed, the model actually fitted by Cowpertwait et al. (2007) assumed two superposed processes. A common depth parameter was assumed across the two storm types, giving a total of eleven parameters.

3.4 Development of the Random Parameter Bartlett-Lewis Instantaneous Pulse (BLIPR) model

For the randomisation of η in the BLIP model, we take the same approach as for the original Bartlett-Lewis model, but now with the additional assumption that the ratio, $(\omega = \xi/\eta)$, of the pulse arrival rate to the cell duration parameter is kept constant.

In order to calculate the moments, it is helpful to think of the random parameter model as the superposition of a continuum of independent processes with random cell duration parameter, η, and storm origin rate, $\lambda f(\eta)$, where $f(\eta)$ is the density function of η. Now, the rth cumulant of a sum of independent random variables is the sum of their rth cumulants. Therefore the mean, variance and 3rd central moment (which are the first three cumulants) can simply be obtained by replacing λ with $\lambda\, f(\eta)$ in their original equations, and integrating over possible values of η.

The integration approach described requires some expectations of functions of η. In particular, we need $E_\eta\left[\left(\frac{1}{\eta}\right)^k e^{-\eta x}\right]$ for $k = 1$ and various values of x, given by:

$$\begin{aligned} E_\eta\left[\left(\frac{1}{\eta}\right)^k e^{-\eta x}\right] &= \frac{\nu^\alpha}{\Gamma(\alpha)}\int_0^\infty \eta^{\alpha-1-k} e^{-(\nu+x)\eta} d\eta \\ &= \frac{\nu^\alpha}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha-k)}{(\nu+x)^{\alpha-k}}. \end{aligned}$$

Note that, in order for the integral not to diverge at zero, we require $\alpha > k$. This proved to be an issue for the original Bartlett-Lewis model, as discussed in Section 3.3.1, where the skewness integral included elements with $k = 4$. For the Bartlett-Lewis Instantaneous Pulse model, we only require $\alpha > 1$ in order for the integrals for the variance and skewness of the aggregated rainfall not to diverge. However, we may still want the constraint $\alpha > 2$ in order to prevent the simulation of unrealistically long rain events, as discussed in Section 3.3.1

The moments are derived from the original equations of Cowpertwait et al. (2007), by taking expectations over η and using the formula above. All the moments can be expressed exactly, which is an advantage for this type of model where numerical integrations can lead to slow computational speeds. The moments for the new model are given in Appendix A.7. As in the original fixed parameter BLIP model, this retains the flexibility to allow pulse depths to be dependent within cells. In their empirical fits, Cowpertwait et al. (2007) assumed these to be independent, but intuitively, dependent pulse depths should allow higher values of extremes at short timescales. This is desirable since the fits understated five-minute extreme values.

3.5 Models Fitted

The models were fitted, using the methodology and fitting properties discussed in Chapter 2, to the Bochum data. A separate fit was produced for each month, to allow for seasonality. In each case, we assume that $\sigma_X/\mu_X = 1$, and that $\mathrm{E}[X^3] = 6\mu_X^2$ (consistent with X being exponentially distributed). Initially, no further constraints were imposed on the parameters, other than that they should be greater than zero. The six models initially fitted were:

Rectangular Pulse Models

1. the Bartlett-Lewis Rectangular Pulse model (BLRP)
2. the Random Parameter Bartlett-Lewis Rectangular Pulse Model (BLRPR)
3. the Bartlett-Lewis Rectangular Pulse model with two superposed processes (BLRP2);

Instantaneous Pulse Models

1. the Bartlett-Lewis Instantaneous Pulse Model (BLIP)
2. the Random Parameter Bartlett-Lewis Instantaneous Pulse model, introduced in Section 3.4 (BLIPR)
3. the Bartlett-Lewis Instantaneous Pulse model with two superposed processes (BLIP2).

For the Bartlett-Lewis Rectangular Pulse model, on randomising the cell duration parameter, η, the fitted solution gave such a high precision to the mean cell duration, that it effectively replicated the non-random solution. Thus, the fitted parameter set for the BLRPR model is simply a re-parameterised version of the set of BLRP parameters, and there is thus no improvement in the fit compared with the fixed η version. This appears to contradict examples in the literature where the randomised η version had shown an improved fit compared to the fixed η model (Rodriguez-Iturbe et al. 1988, Wheater et al. 2005). On further investigation, we concluded that the improvement in the fit to proportion dry that had previously been found by randomising η was at the expense of a deterioration in the fit to the skewness, which had not been included as a fitting property in these earlier analyses. In particular, if skewness is not included in the fit, it is highly overestimated in the summer months at timescales of six and twenty-four hours.

Fitting the models with two superposed processes proved problematic. Although the BLRP2 model with no parameter constraints gave a very good fit in terms of a low minimum objective function value, the parameters thus obtained were highly unstable, unrealistic and inconsistent from month to month, and no standard errors could be found. It was clear that there was insufficient information in our observed data to identify the large

number of required parameters. Introducing constraints for the parameters increased the minimum objective function values, and did not resolve the situation, with resulting solutions having many parameters lying on the constraint boundaries. We therefore concluded that ensuring realistic and reasonably smooth parameters across months would require constraints on the relationships between parameters, rather than just setting bounds on individual parameters. There were similar issues with the BLIP2 model. Ultimately we decided that both of these models' parameter identifiability issues made them unsuitable for practical application.

Given the above findings, we present further results here for the following three models only: BLRP, BLIP, BLIPR.

For the BLIP and BLIPR models we initially followed Cowpertwait et al. (2007) and assumed that pulses within a single cell had independent depths. However, for the BLIPR model an alternative assumption was also considered, whereby pulses within a single cell have a common depth (the most extreme form of dependance). The latter achieved a lower minimum objective function value in all months, and a better fit in respect of properties not included in the fitting process, such as wet/dry properties. For both of these options, the unconstrained solution gave an extremely high number of pulses per hour (of the order of 10^5–10^6), so for practical reasons, μ_X was constrained to be 0.001, reducing the number of parameters by one. All other fitted parameters were broadly as before, except for a corresponding change in ω. The quality of the fit was unchanged with this constraint, as the product term $\mu_X\,\omega$ effectively forms a single composite parameter over most of the possible parameter space. We also considered two alternative constraints on α: $\alpha > 1$ or $\alpha > 2$, as discussed in Section 3.4. The former only affects July, whereas the latter affects all the summer months.

A comparison of the performance of the three fitted models, together with the findings discussed earlier in this section and consideration of the fitted parameter sets (shown in Tables B.1 to B.3 of Appendix B) led us to a hypothesis, which we present in the next section.

3.6 Initial performance comparison of the fitted models

Table 3.1 shows the minimum objective function value for each of the models that we have successfully fitted, for each month. Since the same set of moments and weights were used for each model, these are directly comparable.

Key findings from the results are summarised below:

- The BLRP model outperforms the BLIP model, with a lower minimum objective

	BLRP	BLIP[1] independent pulse depths	BLIPR[1], independent pulse depths	BLIPR[1], common pulse depths	BLIPR[2], common pulse depths
Jan	83	67	45	40	40
Feb	38	56	30	24	24
Mar	100	113	58	48	48
Apr	110	168	85	66	66
May	141	239	93	76	78
Jun	152	275	92	72	80
Jul	162	345	110	95	97
Aug	140	268	86	76	79
Sep	149	271	87	65	72
Oct	92	150	71	50	50
Nov	68	76	30	25	25
Dec	68	67	32	28	28

Table 3.1: Comparison of minimum objective function values; [1]: $\alpha > 1$; [2]: $\alpha > 2$.

.

function value in all months except January and December. The model with rectangular pulses has generally been considered unsuitable for timescales shorter than the mean cell duration, due to the unrealistic intensity shape. However, when fine-scale data are available for fitting, the fitted model tends to have shorter, more frequent cells than if only hourly data are available (of the order of 5–10 minutes, compared with 20–40 minutes for most months), which are still within a realistic range. With these shorter cells, and given also the potential for cells to overlap, repetition of the same rainfall totals over consecutive five minute intervals is relatively infrequent. The fitted parameters are shown in Table B.1 of Appendix B, along with some key properties such as mean storm and cell inter-arrival times and durations.

- When skewness is included in the fit, there is no benefit to randomising the cell duration parameter in respect of the BLRP model, as discussed in Section 3.5. However, there is a clear benefit in respect of the BLIP model, with the randomised version showing the best performance of all the models.

- The fitted BLIPR model has a very high number of pulses per cell (particularly if we do not apply constraints, as discussed), with very short inter-arrival times, and the better performing version has common within-cell pulse depths. Effectively then, the cells are 'rectangular'.

These results imply that it is not the replacement of rectangular pulses by clusters of instantaneous ones that leads to the improved performance of the BLIPR model, compared with the BLRPR model. Instead, the improved performance can be attributed to the fact that the BLIPR model allows rainfall intensity to vary with cell duration, since the pulse rate effectively drives the intensity and is proportional to the cell duration parameter, η. Our new model variant thus gives a simple, but effective way of introducing dependence

between cell duration and intensity.

This suggests that the same effect could be achieved by amending the BLRPR model, so that the mean cell intensity parameter, μ_X is also varied in proportion to the cell duration parameter, η. This is preferable from a computational point of view, eliminating the need for simulation of a vast number of instantaneous pulses.

3.7 Testing our hypothesis

Extending the BLRPR model to allow μ_X to vary in proportion to the cell duration parameter, η, is straightforward, and follows the methodology discussed in Section 3.4. We re-parameterise the BLRPR model so that the ratio, $\iota = \mu_X/\eta$ is now kept constant, and express $E(X^2)$ and $E(X^3)$ in terms of ι also (for which the formulae depend on the choice of distribution for the rainfall intensity). We then take expectations over η as before. The analytical expressions for this new model, which we denote the BLRPR$_X$ model, are given in Appendix A.5, with the required moments of the rainfall intensity distribution, X specified in Appendix A.1.

Unlike the cell and storm durations considered in Section 3.3.1, which have an unconditional Pareto Type II distribution when conditionally exponentially distributed, the unconditional distribution of X does not have a tractable form (again assuming X is conditionally exponentially distributed). Using simulations for a typical parameter set, we find that the unconditional distribution has a higher skewness, and fatter tails than the exponential distribution.

The fitted parameter set, assuming an exponential distribution for cell intensities as before, is given in Table B.4 of Appendix B. Comparing this with the fitted parameters of the Random Parameter Bartlett-Lewis Instantaneous Pulse (BLIPR) model with common within-cell pulse depths, shown in Table B.3, the strong similarity between the two models is evident. In particular, the new parameter, ι, of the BLRPR$_X$ model broadly equates to $\mu_X\,\omega$ of the BLIPR model (noting that μ_X represents an intensity in the rectangular pulse models, but a depth in those with instantaneous pulses). Values of the minimum objective function (see Table 3.2) and plots of the two fits are also found to match, thus supporting our hypothesis.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
BLIPR	40	24	48	66	79	80	97	79	72	50	25	28
BLRPR$_X$	39	22	46	63	74	76	92	74	68	47	23	26

Table 3.2: Comparison of minimum objective function value; α constrained to be at least 2.

We have therefore established that the new rectangular pulse model variant is effectively

equivalent to the BLIPR model with common within-cell pulse depths, and that there is therefore no need to replace the rectangular pulses with a process of instantaneous pulses for fine-scale data. This is the optimal model, at least in terms of the minimum objective function values. In the next section we examine the performance of the three models (BLRP, BLIP, BLRPR$_X$) in more detail, firstly in terms of the fitted moments, and then by considering wet/dry properties, which were not included within the objective function, and extreme value performance.

3.8 Performance comparison of the fitted models

3.8.1 Fitted Moments

Plots of the fits of the models (BLRP, BLIP, BLRPR$_X$) against the observed data for each month in respect of the mean, coefficient of variation, lag-1 autocorrelation and skewness coefficient are shown in Figures 3.3-3.6. The y-axes for these and other similar plots in this thesis have been selected automatically such that, for each individual plot, the axis spans the range covered by the observed and fitted values. This means that the fit in respect of an individual model tends to look worse if all models fit well, than if at least one of the other plotted models has a poor fit (since in the latter case the scale will be wider). Care should therefore be taken to consider also the scale when examining such plots. All the models generally perform well with respect to the properties included in the fitting. They reproduce the mean exactly (this is not a given, since the number of properties fitted exceeds the number of parameters), and fit the coefficient of variation well at all timescales. All tend to underestimate the lag-1 autocorrelation at longer timescales. All also tend to underestimate the skewness at the shorter timescales, with the BLRPR$_X$ model showing the best fit in respect of 5 minute skewness, and the BLIP model the worst.

3.8.2 Wet/dry properties

The proportion of dry intervals is a very important property for hydrological applications. Although this could have been included as one of the fitting properties, it is useful to reserve an important feature for subsequent model validation, as this gives an independent test of the appropriateness of the model structure. Plots of the fits of the models against the observed data for each month in respect of the proportion dry are shown in Figure 3.7. The BLRPR$_X$ model can be seen to outperform the other models with respect to the fit to proportion dry, across all timescales. It is also of interest to consider the wet and dry

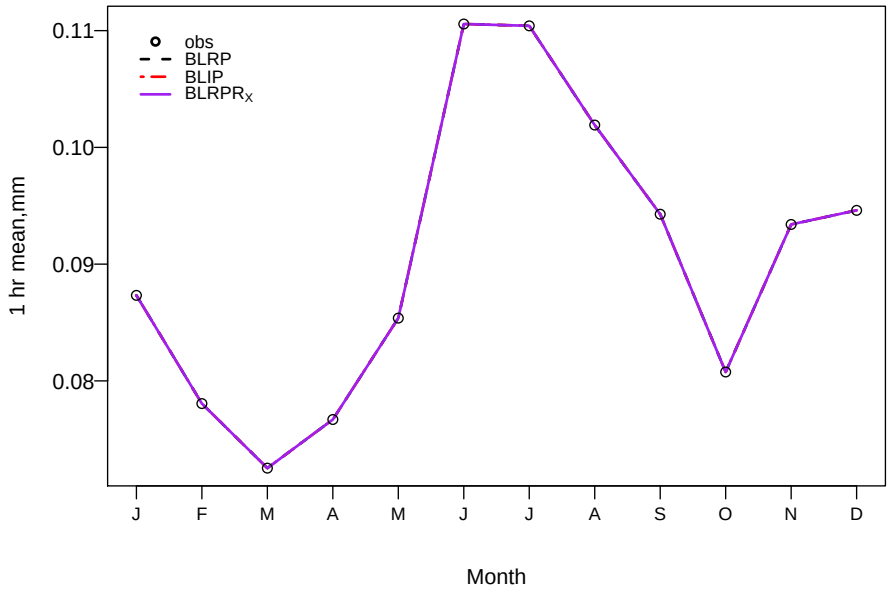


Figure 3.3: Mean 1-hour rainfall by month, fitted v observed.

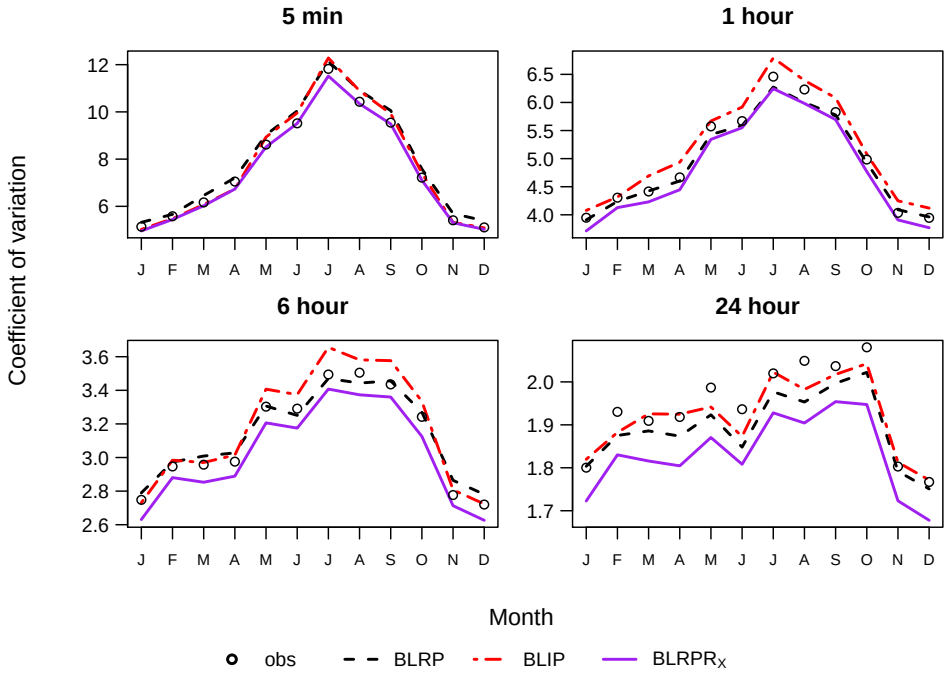


Figure 3.4: Coefficient of variation by month, fitted v observed.

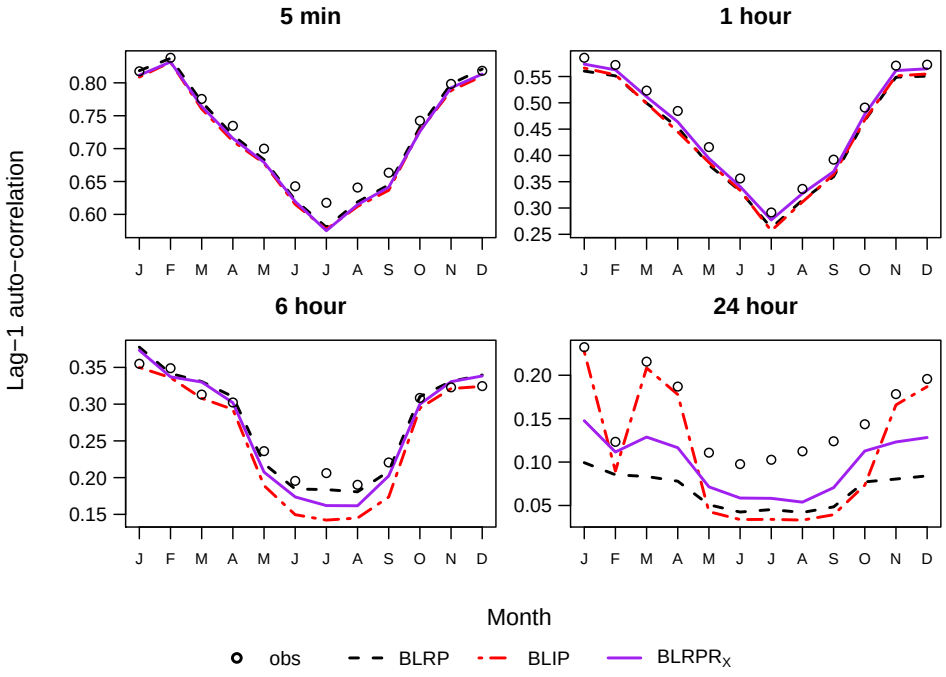


Figure 3.5: Lag-1 autocorrelation by month, fitted v observed.

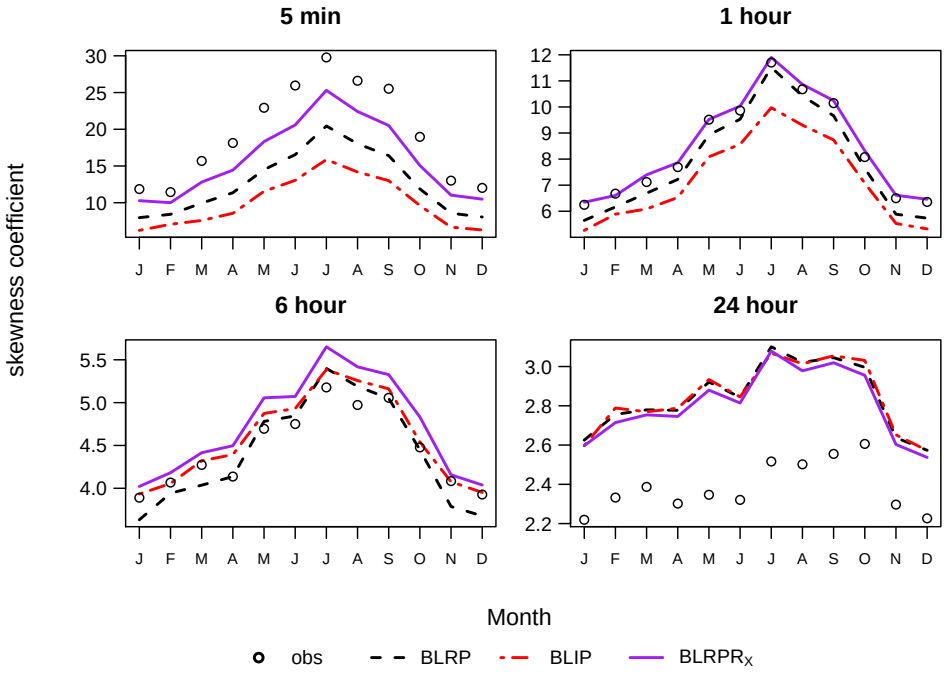


Figure 3.6: Coefficient of skewness by month, fitted v observed.

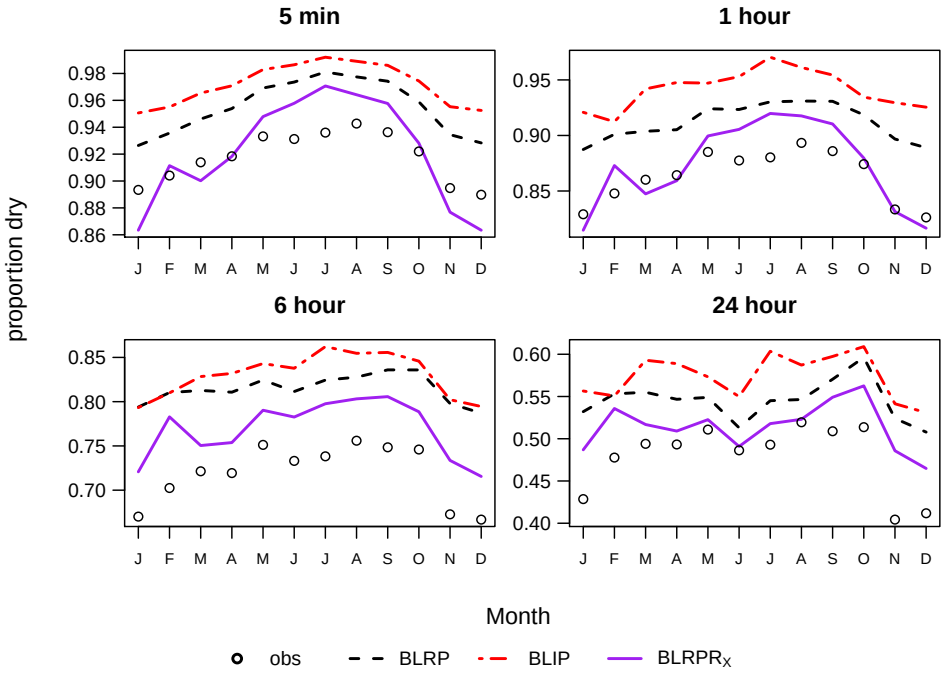


Figure 3.7: Proportion dry by month, fitted v observed.

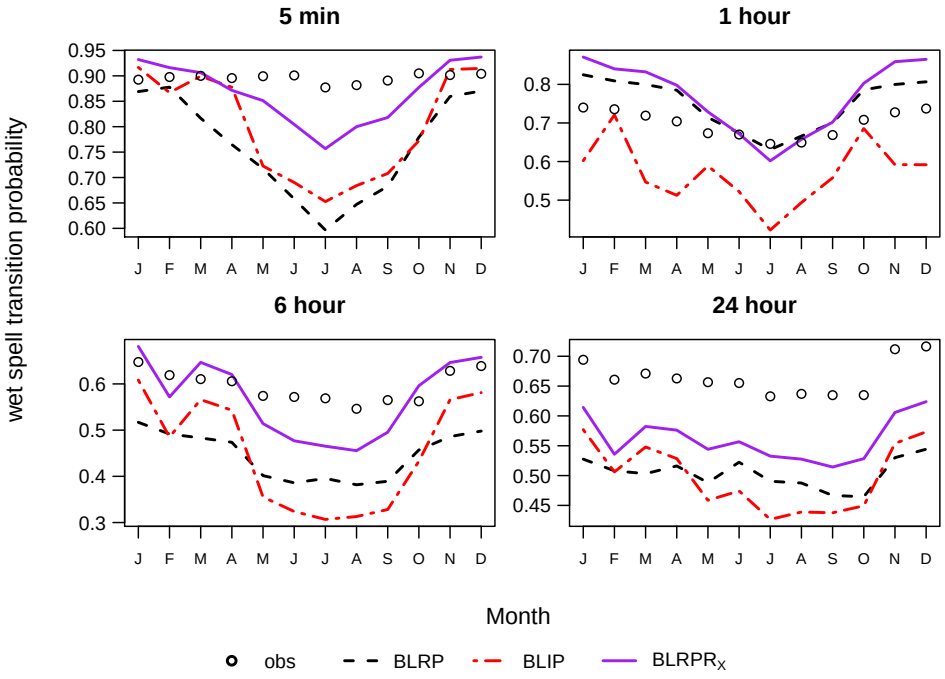


Figure 3.8: Transition probability of a wet interval being followed by another wet interval, by month, fitted v observed.

spell transition probabilities (i.e the probability that a wet interval is followed by another wet interval, or a dry by another dry), which are important for the accurate modelling of antecedent conditions. Figure 3.8 shows that the BLRPR$_X$ model again outperforms the other models with respect to the wet spell transition probability. While the BLRP model has a good fit at the hourly timescale, it performs poorly at other timescales, with only the BLRPR$_X$ model showing consistency of performance across timescales. There is less difference between models for the dry spell transition probabilities, with all models providing a reasonable fit at all timescales. The fit to the wet/dry properties in respect of the summer months would be further improved if we did not impose the constraint that $\alpha > 2$. However, this would be at the expense of allowing storms and cells of unrealistic durations in the simulations (as discussed in Section 3.4), and also a slight deterioration of the fit to the 24 hour variance, and 6 hour lag-1 autocorrelation.

3.8.3 Extreme value performance

For our data, the months with the highest rainfall, rainfall variability and skewness are the summer months, and these are also the months with the highest extremes. A comparison of the fit of extremes for July for the BLRPR$_X$ model is given in Figure 3.9, using Gumbel plots, following the methodology of Section 2.6.1. We use 100 simulations of 69 years each, sampling from the distribution of the parameter set for each simulation. The graphs for July show that the model has a tendency to underestimate extremes, as has been noted before for this type of model. Results for other months give a fairly similar picture. A comparison for the BLRPR$_X$ model showing annual extremes is shown in Figure 3.10. Again, we use 100 simulations of 69 years here, sampling from the distribution of the appropriate parameter set each month. A good fit here is a more onerous requirement than that for a single month, and the effect of a slight understatement in respect of most individual months can be seen to lead to a more substantial understatement in the annual comparison. A comparison showing mean annual extremes (averaged over fifty simulations) for a number of alternative models at the five minute and hourly timescales is also shown in Figure 3.11. At the five minute timescale, the BLRPR$_X$ model gives the best performance, although all the models underestimate the extremes. Results are closer at the one-hour timescale, and for longer timescales, there is essentially no difference between models. Based on our analysis, the BLRPR$_X$ is shown to be the best performing of the models compared, both in terms of the moments fitted, and more importantly, in respect of the wet/dry properties and extreme values, neither of which is included in the fit. It is also intuitively appealing, since we know that the intensity of rainfall does vary inversely with the duration of the rain event. Further, this dependence has been introduced to the BLRPR model without the need for any additional parameters or complexity. Considering

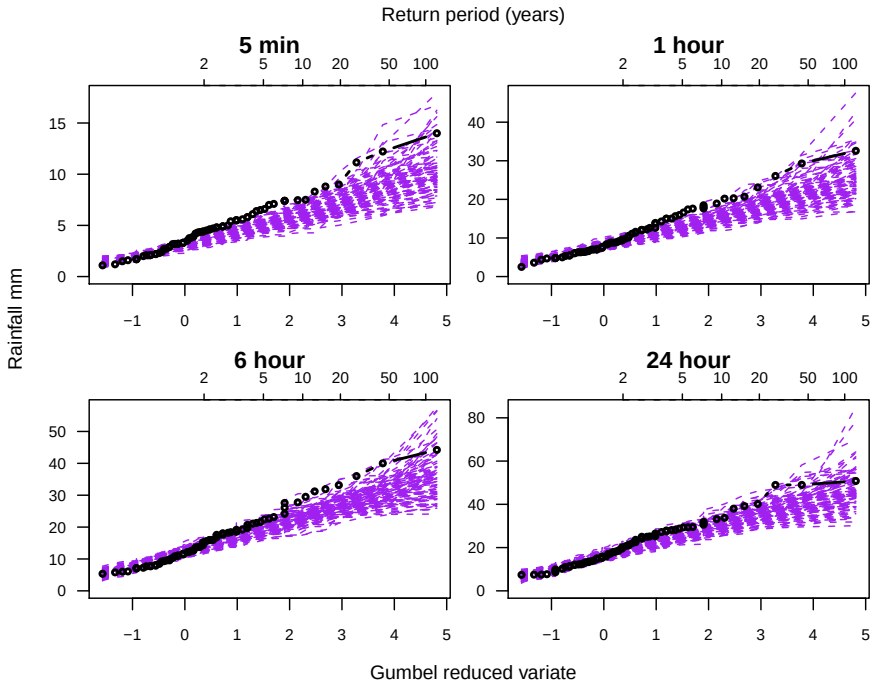


Figure 3.9: Gumbel plots of observed (black) v simulated (purple) extremes for July, using the BLRPR_X model and 100 simulations, each of 69 years; rainfall intensity assumed to follow an exponential distribution; α constrained to be greater than 2.

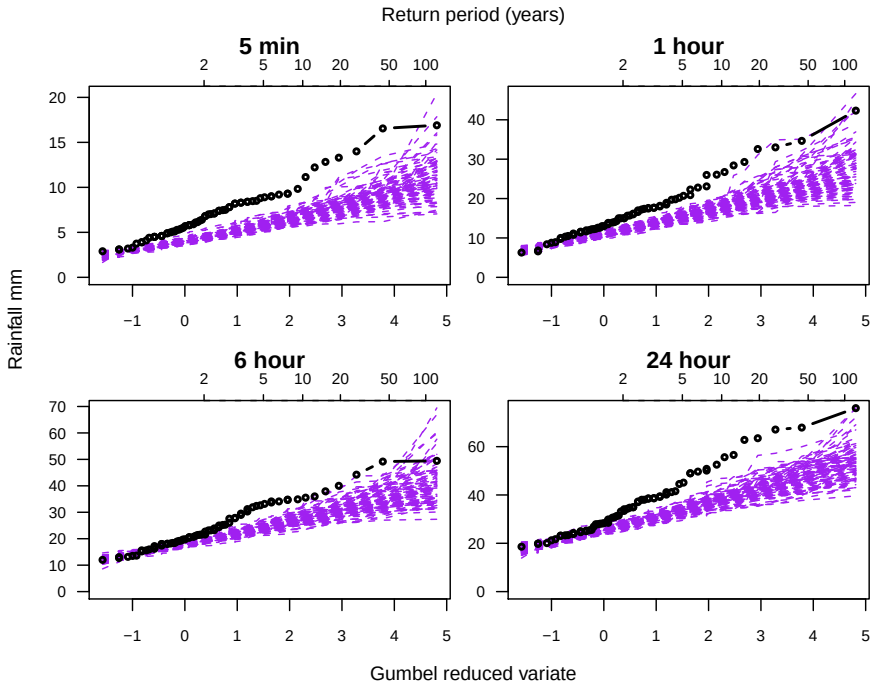
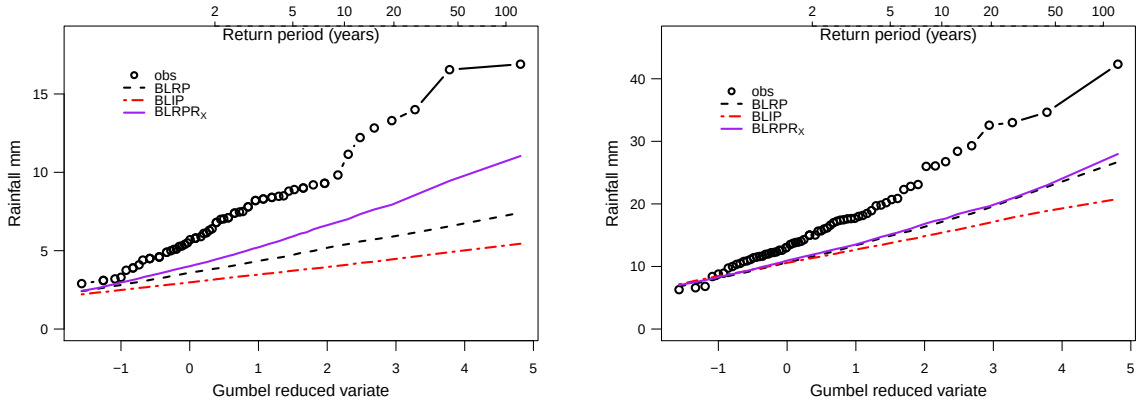


Figure 3.10: Gumbel plots of observed (black) v simulated (purple) annual extremes, using the BLRPR_X model and 100 simulations, each of 69 years; rainfall intensity assumed to follow an exponential distribution; α constrained to be greater than 2.



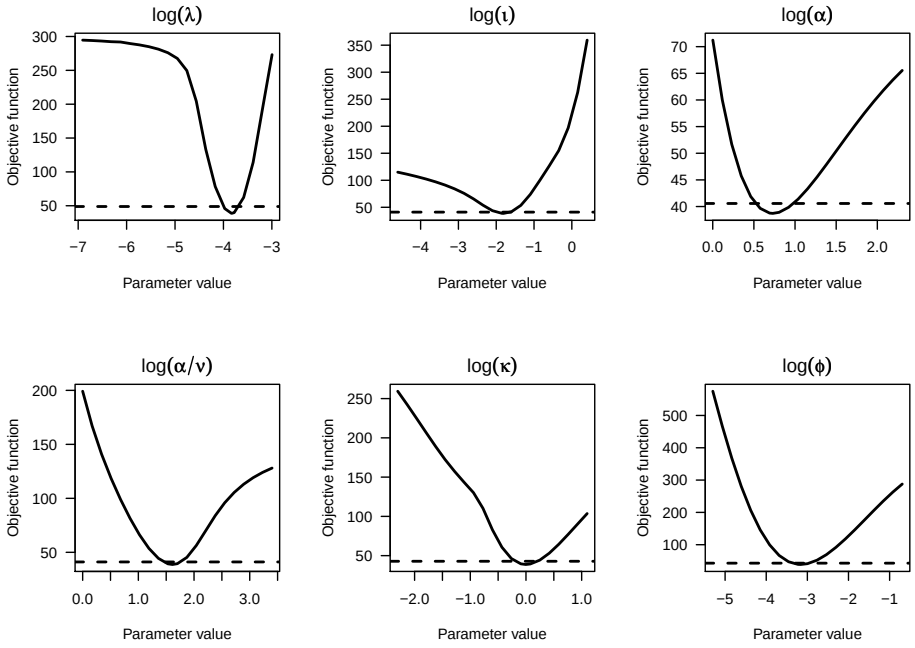
(a) 5 minute **(b)** 1 hour

Figure 3.11: Annual Gumbel plots of observed v simulated extremes for variants of the Bartlett-Lewis model; rainfall intensity assumed to follow an exponential distribution.

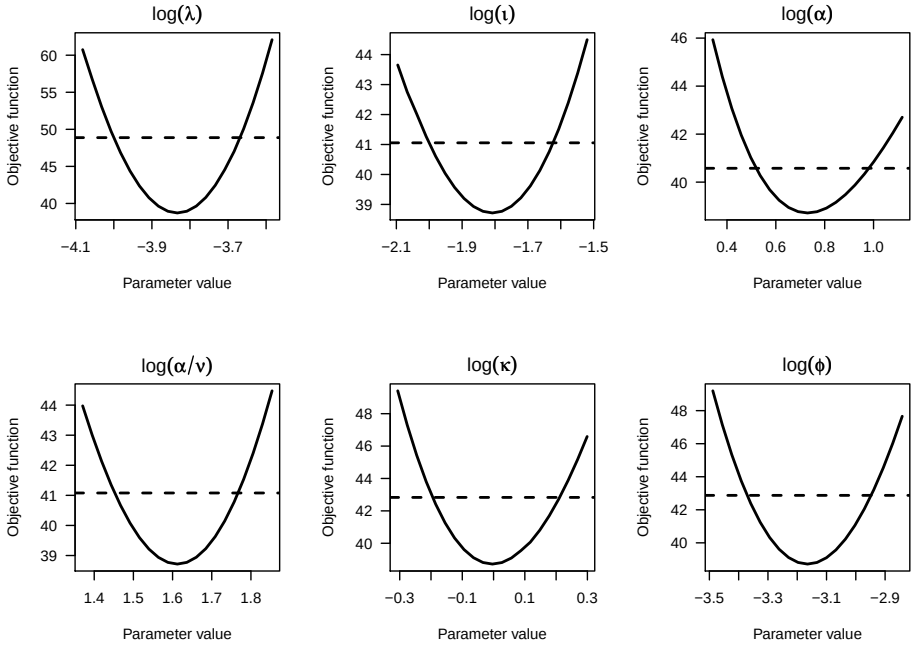
the fitted parameter set, shown in Table B.4 of Appendix B, the parameter values change fairly smoothly from month to month. Comparing with empirical observations from Houze & Hobbs (1982), the parameter values seem reasonable. Winter storms last several hours, have around 20 cells, which each last on average around 22 minutes. In summer, storms and cells are shorter, and have around 8 cells. However, these have a correspondingly much higher intensity, giving broadly the same amount of rainfall per storm over all months. In Section 3.10 we consider how performance might be further improved, particularly in respect of extreme values. First though, we take a look at parameter uncertainty, and identifiability in respect of the new model.

3.9 Parameter Identifiability and Confidence Intervals

Parameter identifiability with this type of model is a known problem, particularly for the variants with a large number of parameters, as noted in Section 2.6. This can be explored using plots of profile objective functions, as described in Section 2.5. Profile objective functions for the logarithm of the parameters of our optimal model, the BLRPR$_X$ model, are shown in Figure 3.12 for the month of January. As before, we have constrained the value of α to be greater than 2 (except in the plot for α itself). The first set of plots shows a wide range of possible parameter values, allowing us to check whether there are multiple local minima, for example, or extensive regions where the objective function is flat. We have then reduced the parameter range so that the approximate 95% confidence intervals can be seen more clearly. These are based on the objective function, again as described in Section 2.5. The plots show that the parameters are fairly well identified



(a) wide parameter range



(b) reduced parameter range

Figure 3.12: Profile objective function plots for the Random Parameter Bartlett-Lewis Rectangular Pulse model with dependent intensity-duration; rainfall intensity is assumed to follow an exponential distribution, conditional on η. The plots are for January.

in January. The confidence intervals are found to be in agreement with those based on ± 2 asymptotic standard errors about the mean parameter values. Results for July (not shown) again indicate good parameter identification (although now a lower limit for α cannot be found). Profile objective functions have not been considered for other months, but intervals have been calculated instead using asymptotic standard errors.

Figure 3.13 shows pointwise confidence intervals based on asymptotic standard errors for all months, and compares these against approximate bootstrap intervals. The latter have been obtained by resampling whole years with replacement, and re-fitting the model 500 times. The interval limits are then taken as the 2.5th and 97.5th percentiles over the 500 samples. The asymptotic intervals are generally close to the bootstrap intervals, which indicates that parameters are fairly well-identified in all months. (For models where this is not the case, the bootstrap intervals are much wider.) There is slightly more parameter uncertainty in the summer months, which is fairly typical.

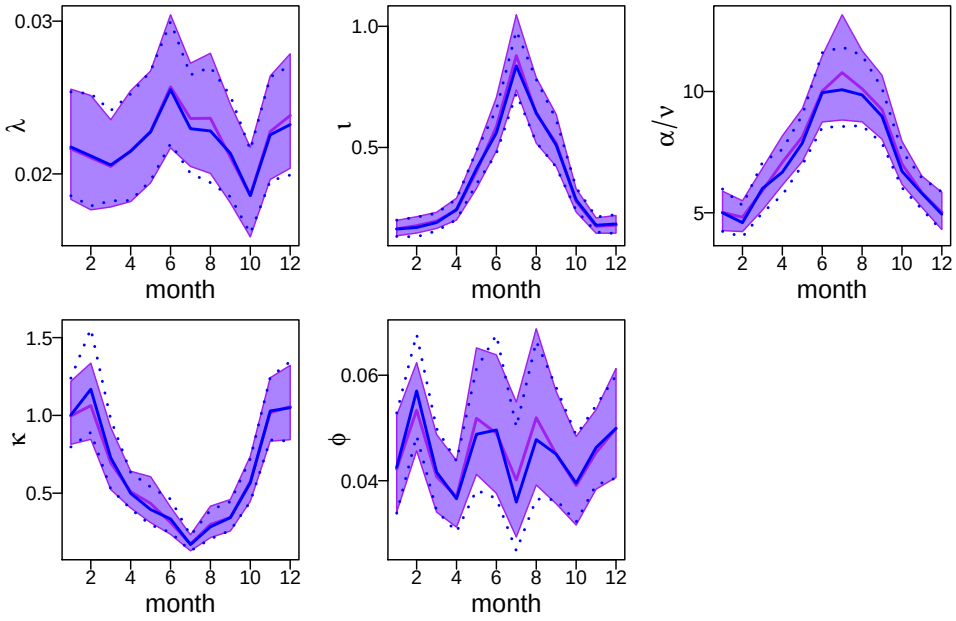
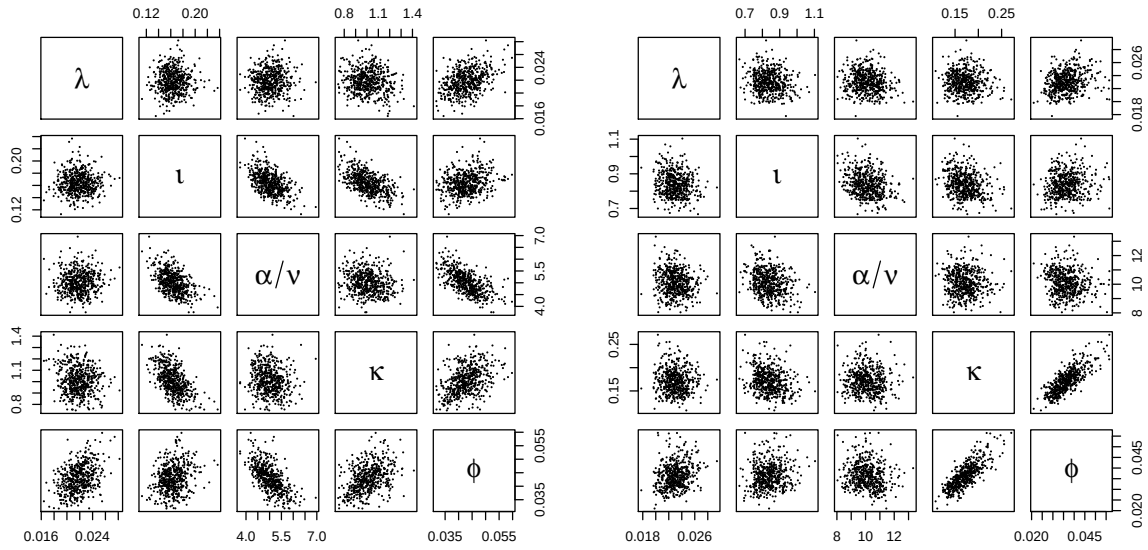


Figure 3.13: Mean parameters and 95% pointwise confidence intervals for the BLRPR$_X$ model: asymptotic results (purple) v bootstrap estimates (blue) based on 500 samples; the asymptotic intervals are based on the mean $\pm$ 2 standard errors, and marked with blue dotted lines. The bootstrap intervals are based on the 2.5th and 97.5th percentiles, and indicated with purple shading ($\alpha > 2$).

Plotting pairs of parameter estimates against each other in respect of the 500 bootstrap fits gives an indication of the relationships between them, and again can highlight problems with parameter identifiability. Scatterplots are shown for January and July in Figure 3.14. The strongest relationships in January include those between $\hat{\alpha/\nu}$ (which is the mean value of the cell duration parameter η) and $\hat{\iota}$, between $\hat{\alpha/\nu}$ and $\hat{\phi}$, and between $\hat{\iota}$ and $\hat{\kappa}$. The last

of these shows that, to some extent, a particular rain event may be generated by fewer, more intense cells or by a greater number of lighter ones. These relationships are far weaker in July, perhaps due to a greater proportion of the rain being convective, with the highest correlation in the summer months being between $\hat{\kappa}$ and $\hat{\phi}$. This indicates that the mean number of cells per storm, μ_C, which is equal to $1 + \kappa/\phi$ remains broadly constant over the different parameter sets. The asymptotic correlation matrix indicates the same relationships.



(a) January **(b)** July

Figure 3.14: Scatterplots of parameter estimates for the BLRPR$_X$ model from 500 bootstrap fits.

3.10 Potential further improvements

Although the new model fits fairly well, there are areas for improvement, the most important of which is the fit to extreme values, which are understated. This is perhaps surprising, as intuitively the inclusion of the skewness coefficient as one of the fitting properties should lead to an improved fit in respect of extremes. On investigation, we found that our approach of averaging the skewness over 69 separate observation months, rather than calculating a single statistic over the whole of the data, tends to understate the skewness coefficient itself, particularly at the 5 minute timescale. This is found to be related to the effective weights that are applied to periods of high skewness under the two alternative approaches, rather than to sampling variation, or to the choice of mean (local or global) about which the moments are centred. The 'separate observation months'

approach was originally introduced purely as a means of deriving a covariance matrix for the fitting properties, but it is integral to the fitting approach when we introduce large scale atmospheric covariates in Chapter 4. We considered exchanging the dimensionless properties of coefficient of variation and skewness coefficient for variance and the third central moment respectively, since these are much more robust in respect of the choice of calculation methodology. However, at least in the context of the Bochum data, this was not found to be effective, leading to a substantial deterioration to the fit of the mean rainfall and proportion of intervals that are dry (although extremes were then well fitted). For the rest of the thesis, we maintain our original set of fitting properties. Note that the problem with averaging the dimensionless properties should in any case be reduced if we can find suitable covariates which explain some of the interannual variability, and this is addressed in the rest of this thesis.

It is generally thought that a distribution for the rainfall intensity with 'fatter tails' should improve the fit to extremes. We investigated this by replacing the exponential distribution with the gamma and Weibull distributions, both of which have the exponential as a special case, and which can therefore be tested against the exponential using the quasi-Wald test of Section 2.6. The Weibull distribution gave the better performance here. However, we found that the addition of a further parameter caused problems in terms of parameter identifiability, with less consistency from month to month. Also the asymptotic results showed a very high correlation (between 0.92 and 0.98) between the estimated shape parameter, ϱ and the intensity parameter, ι, suggesting that an additional parameter is not justified. Table B.5 of Appendix B shows that the fitted shape parameter is close to 0.6 in most months. We therefore chose to consider the two alternative constraints of a shape parameter of 1 (our original exponential) and a shape parameter of 0.6. The objective function values may be compared in Table 3.4.

Applying the quasi-Wald test, we calculate the statistic $\log(\hat{\varrho})^2/\mathrm{Var}(\hat{\varrho})$ to test H_0: intensity is exponential, and $[\log(\hat{\varrho}) - \log(0.6)]^2/\mathrm{Var}(\hat{\varrho})$ to test H_0: shape parameter is 0.6. In each case, we reject the null hypothesis at a significance level of α if the statistic is greater than the $100(1-\alpha)\%$ quantile of the χ^2 distribution with one degree of freedom. The statistics and p-values in respect of the two tests are given in Table 3.3. The 95% quantile of the χ^2_1 distribution is equal to 3.84, and the 99% quantile to 6.63. The tests suggest that a shape parameter of 0.6 is reasonable, since we would not reject the hypothesis in any month at the 99% level, and only in July at the 95% level (assuming that we have a correctly specified model).

Another alternative considered for the BLRPR_X model involved allowing a more flexible intensity/duration relationship, by letting the mean intensity be proportional to the cell duration parameter, raised to some fixed power, the level of which is to be determined i.e. to have $\iota = \mu_X/\eta^c$ for some additional parameter, c. Fitting properties can be calculated

ϱ		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	s	2.81	9.72	4.49	45.36	41.36	17.61	0.59	3.04	14.06	71.89	14.09	2.78
	p	0.094	0.002	0.034	0.000	0.000	0.000	0.442	0.081	0.000	0.000	0.000	0.095
0.6	s	0.39	0.00	0.25	0.65	1.56	0.71	6.19	0.41	0.01	3.59	0.00	0.28
	p	0.532	1.000	0.617	0.420	0.212	0.399	0.013	0.522	0.920	0.058	1.000	0.597

Table 3.3: Statistics (s) and p-values (p) in respect of hypothesis tests for the shape parameter, ϱ of the rainfall intensity distribution; BLRPR_X model.

following the approach of Section 3.7, but now with $\mu_X = \iota\eta^c$, $\text{E}(X^2) = f_1\iota^2\eta^{2c}$ and $\text{E}(X^3) = f_2\iota^3\eta^{3c}$, with $f_1 = \text{E}(X^2)/\mu_X^2$ and $f_2 = \text{E}(X^3)/\mu_X^3$. The fitted parameter set for this model, which we denote the BLRPR_{X_c} model, is shown in Table B.6 of Appendix B. It can be seen that the fitted values of c are fairly close to 1, and we could test whether $c = 1$ as we did for the shape parameter. The minimum objective function values may be compared against our original model, and the variants with Weibull intensity distributions in Table 3.4. This model has the lowest minimum objective function value,

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
BLRPR_X, Exp	39	22	46	63	74	76	92	74	68	47	23	26
BLRPR_X, Wei[1]	38	20	43	54	61	59	90	69	53	35	21	25
BLRPR_X, Wei[2]	38	20	44	54	61	60	95	69	53	36	21	26
BLRPR_{X_c} Exp	26	4	36	45	59	66	69	54	61	32	17	18

Table 3.4: Comparison of minimum objective function values for variants of the BLRPR_X model (Exp = exponential intensity distribution, Wei = Weibull); α constrained to be at least 2; [1]: unconstrained shape parameter; [2]: shape parameter of 0.6

although the differences between the variants tend to be rather small in most months. Plots in respect of selected properties comparing the new model variants are shown in Appendix C. These show that there is little difference in the fits of the three models, and that an improvement in one property is generally at the expense of some deterioration in another. Using a Weibull conditional intensity distribution rather than an exponential improves the fits to 5 minute skewness and to extremes at short timescales, but with some deterioration in the lag-1 autocorrelation at longer timescales. Allowing more flexibility in the intensity/duration relationship gives the best minimum objective function values, but the improvements are primarily to the longer term coefficient of variation and skewness, and cause a deterioration in the fit to extremes at all timescales, probably due to the problems with using these dimensionless properties, discussed earlier.

Our preferred model is therefore the BLRPR_X model with a Weibull intensity, conditional on η, with a fixed shape parameter of 0.6, but there is little to choose between this and the exponential intensity distribution.

3.11 Conclusions: optimal model for fine-scale data?

As has been evident in this review, there are numerous different elements involved in the fitting of point process-based models, including the choice of fitting properties and weights applied to these, the specific model variant selected, the location and resolution of the data, and the numerical optimisation method. These make comparisons between different studies and recommendations based on a single study rather difficult.

Our review has highlighted some limitations in all the models, notably an inability to achieve a good fit to all properties in the summer months of a temperate climate, when rainfall exhibits particularly high variability and skewness. However, such limitations are not surprising when we consider the simplicity of these models compared with the highly complex (and as yet not even fully understood) real physical rainfall process. The challenge of achieving a good fit at timescales that cover the wide range from five minutes up to daily, is particularly demanding, and the performance of the original rectangular pulse model has far exceeded prior expectations.

The key findings of this review (which have been validated using an alternative series of 5-minute rainfall from Kelburn in Wellington, New Zealand, from 1940 to 2004) include the following:

- The BLRP model, originally considered unsuitable for fine-scale data due to the unrealistic rectangular pulses, achieves a reasonable fit across the range of properties (when fitted to 5 minute data), and outperforms the model with instantaneous pulses (the BLIP model).
- Although a random parameter version of the BLIP model with dependent pulse depths shows an improved performance, the same effect can be achieved more easily by amending the Random Parameter Bartlett-Lewis Rectangular Pulse (BLRPR) model, so that the mean rainfall intensity, μ_X, is also randomised. Compared to the non-random version, the fit is improved in respect of short term skewness and extremes, and wet/dry properties. This dependency is also intuitively desirable, and adds no additional complexity or parameters.
- The BLRPR$_X$ model, with six parameters, provides a robust solution, with consistent parameters from month to month. Adding further parameters adds little to the fit, since typically improvements in some properties cause degradation in others, and more parameters bring issues of parameter identifiability and consistency. Replacing the exponential intensity distribution with a Weibull with a fixed shape parameter, however, may be desirable.

In theory, properties such as variance and the third central moment are preferred to the dimensionless coefficient of variation and skewness coefficient when using average prop-

erties over separate observation months, since their estimation is more robust. However, where, as here, a desirable fit cannot be achieved in all months, these may not be desirable in practice. As discussed in Section 3.10, if we can identify covariates that can explain some of the interannual variability in our fitting properties, then the problems with the dimensionless properties will in any case be reduced.

There were generally fewer issues with the Kelburn data, where there is much less variability across months in respect of most of the statistics, and where the skewness levels at short timescales are much lower. We found, for example, that the estimated α in an unconstrained fit was only below 2 for a single month. Fitting with the 3rd central moment rather than the skewness coefficient did not cause the same deterioration to the mean rainfall (although some slight deterioration did occur), and the differences between the skewness coefficient, if calculated as the mean over separate observation months rather than over the whole period, were also much smaller. The fitted values of the shape parameter of the Weibull distribution covered a similar range, with an average of 0.66.

In Chapter 4 we introduce the theory in respect of continuous covariates, and consider the selection of suitable variables in Chapter 5. For the practical application of Chapter 6 we start with the simplest of the clustered models — the BLRP model with an exponential intensity distribution. Once the ideas have been proven viable, our preferred model is also considered.

Chapter 4

New approach to address non-stationarity

In the previous chapter we compared different versions of the Bartlett-Lewis point process-based rainfall model. All of the models reviewed suffer from the limitation that they are stationary. As discussed, to accommodate seasonal variation in rainfall characteristics, current practice is to fit a separate model for each calendar month or season. This allows for seasonality given a stable climate. However, under climate change scenarios, calendar month is likely to become increasingly less reliable as an indicator of rainfall behaviour, and is in any case just a proxy for the real drivers, atmospheric variables such as temperature and pressure. In this chapter, we therefore extend the current approach by replacing the discrete covariate, calendar month, with continuous covariates which are more directly related to the incidence and nature of rainfall. We find that a natural extension to current practice is to use a kernel-based nonparametric approach within a generalised method of moments framework (local GMM). Discussion of the selection of suitable covariates is covered in Chapter 5. First though, in this chapter, we develop the methodology.

4.1 Motivation for local modelling

4.1.1 Kernel Smoothing

Recall our original (global) fitting approach. Here we minimise the following objective function with respect to θ:

$$S_n(Y,\theta) = [T(Y) - \tau(\theta)]^{\mathrm{T}} W_n [T(Y) - \tau(\theta)],$$

where $T(Y)$ is a vector of statistics, and $\tau(\theta)$ is the corresponding vector of expected values, determined from our chosen model (e.g. the Bartlett-Lewis Rectangular Pulse

model, or a more complex variant of this). θ is a q-dimensional parameter vector, and we fit to k properties, where $k > q$ so that T and τ are $k \times 1$, and W_n, the weighting matrix, is $k \times k$. (If we have $k = q$ then a weighting matrix is not required, and we solve $T(Y) - \tau(\theta) = 0$.)

In practice, $T(Y)$ is taken as the mean of statistics calculated over each month of the observation period. This approach is taken in order to be able to calculate a sample covariance matrix, which is required for the weighting matrix. As discussed above, it is common practice to allow for seasonality by fitting a separate model for each calendar month. We can express this approach by writing the equation for the estimator of θ for calendar month m as:

$$\hat{\theta}_m = \text{argmin}_\theta \left[\left\{ \frac{1}{\sum_{t=1}^{n} I(m_t=m)} \sum_{t=1}^{n} I(m_t = m)[T(Y_t) - \tau(\theta_m)] \right\}^{\mathrm{T}} \times W_{n_m} \left\{ \frac{1}{\sum_{t=1}^{n} I(m_t=m)} \sum_{t=1}^{n} I(m_t = m)[T(Y_t) - \tau(\theta_m)] \right\} \right], \tag{4.1}$$

where Y_t is the vector of all the rainfall data in observation month t, m_t is the calendar month of the tth observation month, and $T(Y_t)$ is the vector of statistics for that month, with $\tau(\theta_m)$ the vector of expected values for calendar month m. W_{n_m} is the weighting matrix for month m (for which we have n_m observation months of data), and I is the indicator function, such that $I(x) = 1$ is x is true, and 0 otherwise.

Now we want to replace the discrete covariate, calendar month, with one or more continuous covariates. Firstly, we need to decide on a suitable time interval over which to measure the covariates. We already calculate statistics separately for each observation month (denoted by t in the equations), as discussed above, and this would seem a natural choice. Although detail of individual weather systems is clearly lost at this timescale, a monthly interval length has many advantages. Firstly it is short enough such that it is reasonable to treat the series within each interval as stationary. On the other hand, sample autocorrelations for monthly rainfall series tend to be very small, so a month is long enough to permit treatment of the data as independent between intervals. It is also long enough for any small sample biases in the statistics to be negligible, which is an issue primarily in respect of statistics at the daily timescale. Additionally, mean monthly values of many atmospheric variables are readily available. Of course, a key requirement for the modelling to be successful is that there must exist reasonably strong relationships between rainfall behaviour over a month, and suitable covariates, aggregated over the month. Evidence from existing literature suggests that this should be so (e.g. Kilsby et al. (1998)). This is discussed further in Chapter 5, and investigated in respect of the Bochum data. For the rest of this chapter, we assume that such covariates can indeed be found, and that the distribution of the monthly rainfall statistics, conditional on these covariates, is stationary.

A very simple way of proceeding might be to partition the monthly continuous covariate into a number of discrete ordered bins, then fit a separate model for each bin, as per the existing method. However, this type of approach, while easy, is very crude and unlikely to be helpful in terms of allowing for a gradually changing climate over time.

As a simple motivating example, we consider the same issue in respect of the relationship between two random variables X and Y, which is examined by means of a scatterplot. In

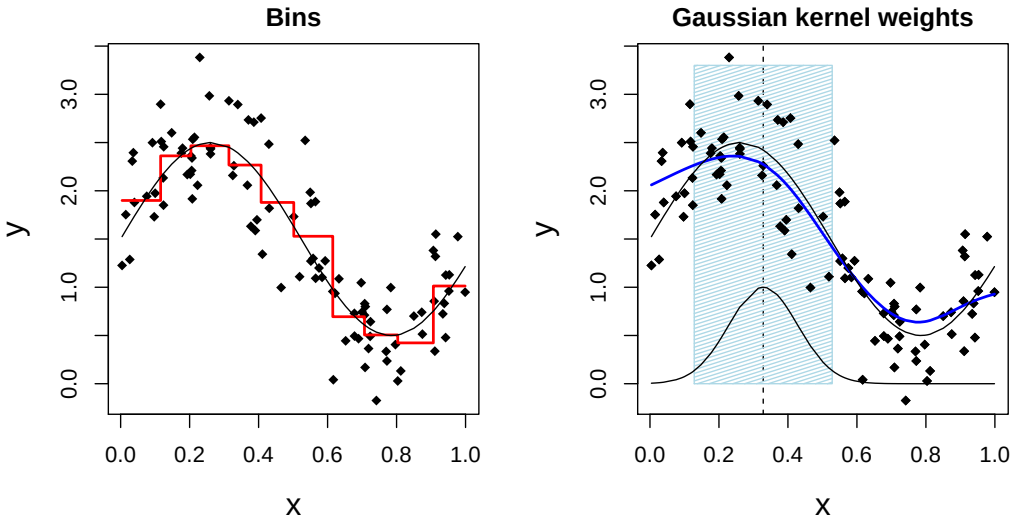


Figure 4.1: Scatterplots showing the relationship between x and y, generated from the equation $Y = \sin(6 * X) + 1.5 + \varepsilon$, $\varepsilon \sim N(0, 0.14)$, with two different smoothing approaches. The lower curve in the righthand plot shows the relative weights used in fitting $\hat{\theta}(x_0)$, the point on the dashed line. The blue rectangle shows the range over which observations contribute non-negligible weight to the fit at this point. The true curve is shown by the black line.

the first plot of Figure 4.1, we have grouped the data into 10 bins, calculated the average in each bin, and plotted the resulting step function. We can think of this as fitting a value y at x_0 that is a weighted average of the observed y, where the weights are either 1 if within our defined 'neighbourhood' (here bin), or 0 if not. However, in estimating y at x_0, it is clearly desirable to give greater weight to observations that are closest to x_0. A better approach therefore is to use weighted averages, where the weights are dependent on the distance away. This gives a smooth curve, as shown in the graph on the right. Here, we have used a 'local mean' or 'Nadaraya-Watson' (Nadaraya 1964, Watson 1964) estimate of $\theta(x_0) = \mathrm{E}[Y|X = x_0]$, given by:

$$\hat{\theta}(x_0) = \frac{\sum_{i=1}^{n} w(X_i - x_0)\, y_i}{\sum_{i=1}^{n} w(X_i - x_0)},$$

which must be calculated for each required value of x_0. This is clearly just a weighted average. The weight function, $w(z)$, peaks at zero, is symmetric, and decreases as $|z|$ increases. The weights in our example have been calculated using the Normal density function with mean x_0 and standard deviation 0.1. Altering the standard deviation would

alter the size of the neighbourhood (the blue rectangle on the graph), and hence the smoothness of the curve. Weight functions used in this context are known as 'kernel' weights. The weights depend on a parameter which determines the amount of smoothing, which here is given by the standard deviation of the normal density function. Such a parameter may be referred to as a 'tuning constant', and, depending on the smoothing approach used, may take the form of a 'bandwidth' (as here) or a 'span' (a fixed percentage of the data that contributes a positive weight to each local fit). A kernel function, denoted by $K(\cdot)$, is usually chosen to be a symmetric density function which integrates to 1, and is scaled to have a variance of 1. Assuming a bandwidth of h, the weights are then given by:

$$w(X_i - x_0) = K_h(X_i - x_0) = \frac{1}{h}K\left(\frac{X_i - x_0}{h}\right).$$

In practice, the choice of kernel function is relatively unimportant compared to the choice of h (Wand & Jones 1995), and the Gaussian kernel function is often used for convenience. It is given by $K(t) = (\sqrt{2\pi})^{-1}\exp(-t^2/2)$. We can see that:

$$K_h(X_i - x_0) = \frac{1}{\sqrt{2\pi h^2}}\exp\left(-\tfrac{1}{2h^2}(X_i - x_0)^2\right)$$

is the normal density function with mean x_0 and standard deviation h. (Note that the constant term $\frac{1}{\sqrt{2\pi h^2}}$ does not need to be included in the fitting.)

This simple example has illustrated the idea of replacing the indicator weighting functions of a discrete covariate or binning approach, with kernel weighting functions which provide a smooth fit in respect of a continuous covariate. In the local mean example, as the width of the neighbourhood is increased, the solution tends to the horizontal line given by $\bar{y}$, hence the description of this method as 'local mean' or 'local constant'. This approach can be extended, so that the solution tends instead to a best fit straight line through the points ('local linear regression'), or indeed to any order of polynomial. Now we approximate $\mathrm{E}[Y_i|X_i] = \theta(X_i)$ locally by the polynomial $\psi(X_i)$ of order p such that:

$$\begin{aligned}\psi(X_i) &= \theta(x_0) + (X_i - x_0)\theta^{(1)}(x_0) + \frac{(X_i - x_0)^2}{2}\theta^{(2)}(x_0) + ... + \frac{(X_i - x_0)^p}{p!}\theta^{(p)}(x_0) \\ &\equiv \sum_{j=0}^{p} b_j(x_0)(X_i - x_0)^j, \end{aligned} \tag{4.2}$$

where $\theta^{(j)}$ denotes the jth derivative of θ with respect to x, and $b_j(x_0) = \theta^{(j)}(x_0)/j!$. Estimation of $\hat{\theta}(x_0)$ involves solving for the set of parameters $\hat{b}(x_0) = (\hat{b}_0(x_0), \ldots, \hat{b}_p(x_0))^{\mathrm{T}}$ to minimise:

$$\sum_{i=1}^{n} K_h(X_i - x_0)\left\{Y_i - \sum_{j=0}^{p} b_j(x_0)(X_i - x_0)^j\right\}^2. \tag{4.3}$$

Differentiating with respect to the b_j and setting to zero gives the following $p+1$ equations:

$$\sum_{i=1}^{n} K_h(X_i - x_0)(X_i - x_0)^k \left\{ Y_i - \sum_{j=0}^{p} \hat{b}_j(x_0)(X_i - x_0)^j \right\} = 0, \quad k = 0, \ldots p. \tag{4.4}$$

The estimate at x_0 is given by $\hat{\theta} = \hat{b}_0(x_0)$, and the whole curve is obtained by running the local polynomial regression across the set of required covariate values, as for the local mean example. We can see that taking $p = 0$ gives the local mean equation. Estimators for the derivatives of θ up to the pth derivative are also available.

The complexity of the model is determined both by the bandwidth (the smaller the neighbourhood, the greater the effective number of parameters in the model), and by the order of the polynomial chosen. A very small value of h effectively interpolates the data, whereas as h increases, the curve becomes smoother, but important features may be lost. As h increases to ∞, the solution becomes equivalent to that of the global polynomial regression of order p. Ultimately an appropriate compromise must be reached, and much of the literature addresses the issues of selection of the optimal order and bandwidth, which we will consider in Sections 4.3 and 4.4 respectively. These sections also include a brief overview of other weighting options.

The idea of local polynomial regression has been around for a long time, proposed originally by Cleveland (1979) and other authors, and there is a wealth of literature in the field. Useful general references include Fan & Gijbels (1996), Wand & Jones (1995), Bowman & Azzalini (1997) and Wasserman (2006).

These ideas have been extended to local likelihood-based methods (see Tibshirani & Hastie (1987), Section 4.9 of Fan & Gijbels (1996) and Section 3.4 of Bowman & Azzalini (1997)). Carroll et al. (1998) combine the theory of estimating equations with local polynomial regression to give an approach with very wide applicability. If we let $g(Y, v) = Y - v$, then the set of $p + 1$ equations (4.4) can be written as:

$$\sum_{i=1}^{n} K_h(X_i - x_0)(X_i - x_0)^k \, g\left\{ Y_i, \sum_{j=0}^{p} \hat{b}_j(x_0)(X_i - x_0)^j \right\} = 0, \qquad k = 0, \ldots p. \tag{4.5}$$

Carroll et al. (1998) extend the methodology by allowing θ to be vector valued, and the function g to be any estimating function (which must have the same dimension as θ). The Y_i are independent, and possibly vector-valued. Taking g to be the derivative of the log-likelihood function, for example, would give local maximum likelihood estimation. The principal example of the paper has g as the usual method of moments function. The unknown parameter set was taken to be the vector of conditional moments of Y given $X = x$, in which case $g(Y, v) = M(Y) - v$, with $M(\cdot)$ the vector of sample moments.

Another interesting development is that of Gozalo & Linton (2000) who present a local

estimation method which allows one to shrink to a favourite non-linear shape, rather than just towards a constant or polynomial. The rationale behind this idea is that, if the form of the local function is close to the true function, then the bias of the estimator will be very small. This is considered particularly relevant for binary data, and in nonlinear time series estimation. The empirical example is of a binary response variable, with the cumulative distribution function of the standard normal distribution as the local function. There are many similarities and areas of overlap between these different nonparametric estimation methods. For example, this empirical example could be framed as a local estimating equation, and local polynomial regression is a special case of both Gozalo & Linton's method, and Carroll et al.'s local estimating equations

Lewbel (2007) develops a local GMM method which uses a local mean estimator in a GMM framework. A number of authors in the econometrics literature combine local fitting with GMM estimation, although generally only in cases where the equations can be explicitly solved (regression-type models with instrumental variables); see for example Cai & Li (2008) and Tran & Tsionas (2010), both of whom consider local GMM in the context of panel data models with varying coefficients.

4.1.2 Splines

All the approaches discussed so far in this section have been based on kernel smoothing. A rather different alternative would be to regress the response on a set of basis functions. Regression splines take these basis functions as polynomials (typically cubics). Splines are piecewise polynomials, joined at certain values of the covariate, x, known as the 'knots'. At the knots there are constraints to ensure smoothness: these require that the 1st to $(k-1)$th derivatives, where k is the degree of the polynomials, are equal at the joins. In order to fit a regression spline, one needs to fix the degree of the spline, the number of knots and their position. The choice of basis set (i.e. the polynomial building blocks) for the splines is key for efficient computation, and a popular choice is the B-spline basis (de Boor 2001). For a cubic spline with k fixed internal knots, a basis set of dimension $k+4$ is required, and the regression involves solving for the coefficients a_j to minimise $\sum_{i=1}^{n}\left\{y_i - \sum_{j=1}^{k+4} a_j B_j(x_i)\right\}^2$, where $B_j(x_i)$ denotes the value of the jth B-spline at x_i. The difficulty here is the choice of the number and position of the knots. One approach, 'smoothing splines', is to fix these at the observed data points, and then to reduce the dimensionality by adding a term to the sum of squares that penalises roughness, based on the second derivative of the regression function. Penalised splines (or 'P-splines') use a similar idea, but here the number of knots is large, but not fixed, and the penalty is based on the differences between the coefficients, a_j, in adjacent segments (see Eilers & Marx (1996), who also give an overview of B-splines). In either case, the effect of the roughness penalty is controlled by a smoothing parameter, which acts in a similar way to the bandwidth in kernel smoothing.

Both kernel-based smoothers and splines are widely used, and both are potential options for relating the parameters of the point-process based models to continuous covariates. Within this thesis we have decided to focus on kernel-based smoothing, since it represents a very natural and intuitive extension of the existing fitting approach. Splines are briefly considered in the context of future work in Chapter 7. In the next section we return to the point process-based rainfall model and consider how these local modelling ideas can be applied, starting with the local mean approach. Although there are potential advantages to assuming a higher order polynomial, in particular a local linear approach, this is a sensible starting point, given the additional complexities that are involved in our models.

4.2 Local mean approach to point process models

Recall the formula with the discrete covariate, calendar month of our current approach, as expressed in Equation (4.1), and re-stated below:

$$\hat{\theta}_m = \text{argmin}_\theta \left[\left\{ \frac{1}{\sum_{t=1}^{n} I(m_t=m)} \sum_{t=1}^{n} I(m_t = m)[T(Y_t) - \tau(\theta_m)] \right\}^{\mathrm{T}} \times W_{n_m} \left\{ \frac{1}{\sum_{t=1}^{n} I(m_t=m)} \sum_{t=1}^{n} I(m_t = m)[T(Y_t) - \tau(\theta_m)] \right\} \right].$$

Now applying the logic from the scatterplot example above to our rainfall model, we can simply replace the indicator functions with kernel summation to solve for parameters at a given covariate value $X = x_0$:

$$\hat{\theta}(x_0) = \text{argmin}_{\theta_{x_0}} \left[\left\{ \frac{1}{n} \sum_{t=1}^{n} K_h(X_t - x_0) \left[T(Y_t) - \tau(\theta(x_0))\right] \right\}^{\mathrm{T}} \times W_n(x_0) \left\{ \frac{1}{n} \sum_{t=1}^{n} K_h(X_t - x_0) \left[T(Y_t) - \tau(\theta(x_0))\right] \right\} \right], \tag{4.6}$$

where $W_n(x_0)$ is a weighting matrix, which will be discussed further in Section 4.2.6. At the moment we need assume only that it depends on the covariate value $X = x_0$, and converges to a positive-definite matrix of constants, which we will denote by W_{x_0}.

This is effectively the local mean GMM approach of Lewbel (2007). The evaluation points (i.e. the values of x_0 at which we choose to solve the equation) can coincide with the observed covariate values, x_i, which will permit an assessment of the goodness of fit. Alternatively an arbitrary set of points can be used, for example a regularly spaced grid, or a set of future values projected by a GCM or other climate model. In the latter case, though, care should be taken over any parts of the range which are sparsely represented in the observed data.

We now consider the methodology for the local mean model, including the derivation of the asymptotic variance and bias of the parameter estimates. In Section 4.3, we extend the equations to a local linear framework, and consider the issues involved in choosing the optimal order of local polynomial. Note that for the local linear case, we refer to the true function at x as $\theta(x)$, and the polynomial approximation at x close to a neighbouring point x_0 as $\psi(x)$, where $\psi(x) = \theta(x_0) + (x - x_0)\theta'(x_0)$. For the local mean case, however, we just use θ throughout, since it is clear whether we are referring to the true value of the function at x given by $\theta(x)$ or the local mean approximation given by $\theta(x_0)$. Finally, we will conclude this chapter with an investigation of methods for determining a suitable bandwidth, and a recommended approach for our model.

At this stage, we do not need to specify any particular point process model, and assume only that our parameter vector, θ has q components, and that we fit to k properties, with $q < k$.

4.2.1 Background to the asymptotic derivations, and key assumptions

The principal aim of this thesis is to develop a local fitting approach to the point process-based rainfall models that can be applied in practice, and the asymptotic expressions are not directly used in such an application. The primary purpose of the derivations is rather to gain an understanding of the key drivers behind the behaviour of the local estimators, and to inform our choice of bandwidth. In addition, a 'quasi-asymptotic' expression for the variance, together with the asymptotic normality of the estimators, will allow us to determine approximate uncertainty levels associated with our estimators.

Throughout the derivations, we assume that the regularity conditions required for standard GMM (discussed briefly in Section 2.3) hold. The moment conditions and further conditions required for identifiability are discussed in Section 4.2.2, and broadly follow Jesus & Chandler (2011), except that now we condition on the covariate, X. We assume that the density at any required evaluation point, x_0 is greater than zero (i.e. $(f(x_0) > 0)$.

Various smoothness conditions are required in order to allow local averaging. These ensure that Taylor series expansions about x_0 can be taken to the required order, such that remainder terms tend to zero as the sample size increases to infinity and the bandwidth goes to zero. Here we assume that the following functions are sufficiently smooth in a neighbourhood of x_0 to permit differentiation as required, and that the functions and derivatives are finite at x_0: the parameter vector function, $\theta(x)$, the design density, $f(x)$, the conditional variance $\mathrm{Var}[T(Y)|X = x]$, and the composite function $\tau(\theta(x))$. The last of these requires similar smoothness for τ in θ.

Finally we assume that the kernel function is a continuous, symmetric density function

with the following properties:

$$\begin{gathered}\int K(z)\,\mathrm{d}z = 1, \qquad \int zK(z)\,\mathrm{d}z = 0 \qquad \int z^2K(z)\,\mathrm{d}z = k_2 \neq 0. \\ \int z^{2r}K(z)\,\mathrm{d}z < \infty, \quad r = 1, 2, \ldots.\end{gathered} \tag{4.7}$$

Although the kernel function does not have to be compactly supported, it should decay fast enough to eliminate the impact of a remote data point (Fan & Gijbels 1996). Recall also, that we assume that the pairs (X_t, Y_t) are independent across months.

4.2.2 Target of the estimation

As for standard GMM, we want to establish whether the proposed equation, Equation (4.6), has a 'target' value, i.e. a unique value $\theta_0(x_0)$ in respect of each covariate value x_0 to which the estimator $\hat{\theta}(x_0)$ converges in probability. Jesus & Chandler (2011) have shown this to be the case for the global GMM case in respect of the point process-based rainfall models (assuming appropriate choice of moment conditions). Here we extend their results to kernel-based estimation, and then consider the asymptotic bias, variance and distribution of the parameter estimators. Conditions on τ are as for standard GMM: we require that $\tau(\cdot)$ is twice differentiable with respect to θ, the parameter vector, and that the derivatives are bounded, for all θ such that $\theta(x) \in \Theta(x)$, a compact subset of $\mathbb{R}^q$. It is also required that $\partial\tau/\partial\theta$ be of full rank.

Then, for a particular rainfall process, for example in Bochum, we assume that there exists a unique, true value of the parameter vector, $\theta_0(x) \in \Theta(x)$, such that:

$$\mathrm{E}[T(Y)|X = x] = \tau(\theta_0(x)), \tag{4.8}$$

where x is fixed, but arbitrary. This is our moment condition.

For the asymptotic derivation, for simplicity, we assume that X is scalar, although the theory can be extended to a vector covariate, and we will be considering multivariate predictors in our empirical study. We assume also that we are considering an 'internal point' i.e. a point x which does not lie near the boundary of the design region.

We define:

$$G_n(\theta(x)) = \frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x)\,[T(Y_t) - \tau(\theta(x))], \tag{4.9}$$

where t is the observation month, and the pairs (X_t, Y_t) are i.i.d., as (X, Y). The bandwidth is assumed to be a function of the sample size, n, although for notational simplicity we write h rather than h_n. Note that G_n is a vector of dimension k. Then the local GMM

estimator at $x = x_0$ is given by:

$$\hat{\theta}(x_0) = \text{argmin}_{\{\theta(x_0)\}} \, G_n(\theta(x_0))^{\mathrm{T}} \, W_n(x_0) \, G_n(\theta(x_0)), \tag{4.10}$$

where $W_n(x_0)$ is a $k \times k$ weighting matrix, which may depend on the data (X_t, Y_t), but which converges in probability to a positive-definite weighting matrix: $W_n(x_0) \to^p W_{x_0}$. This is an example of an extremum estimator, sometimes referred to as an 'M-estimator' (see Chapter 5 of van der Vaart (1998) or Newey & McFadden (1986)). Note that, although the theory generally refers to estimators which *maximise* an objective function, it is equally applicable here, where we solve for a minimum, since this could be considered as the maximum of the negative objective function. At $\theta(x_0) = \hat{\theta}(x_0)$ the derivative of the minimand in Equation (4.10) is equal to zero. In this form, the equation is an example of an estimating equation, and is given by:

$$\begin{aligned} 0 &= \left[\frac{\partial G_n(\hat{\theta}(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_n(x_0) \, G_n(\hat{\theta}(x_0)) \qquad (4.11) \\ &= \left\{\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)\left[\frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}\right]\right\}^{\mathrm{T}} W_n(x_0) \left\{\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)[T(Y_t) - \tau(\hat{\theta}(x_0))]\right\}, \end{aligned}$$

where the notation $\frac{\partial G_n(\hat{\theta}(x_0))}{\partial \theta}$ and $\frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}$ is used to represent $\frac{\partial G_n}{\partial \theta}\big|_{\theta=\hat{\theta}(x_0)}$ and $\frac{\partial \tau}{\partial \theta}\big|_{\theta=\hat{\theta}(x_0)}$ respectively (i.e. the Jacobian matrices of G_n and τ, evaluated at $\theta = \hat{\theta}(x_0)$).

In order to demonstrate consistency of the estimator $\hat{\theta}(x_0)$ for $\theta_0(x_0)$, Newey & McFadden (1986) argue that it is preferable to consider the estimator as the global maximum (here, minimum) of an objective function, rather than as a solution to first order conditions i.e. in the form of Equation (4.10), rather than Equation (4.11). This is because the first-order condition can have multiple roots even when the objective function has a unique maximum. The former is the equation that we actually solve numerically, and is in any case arguably the more straightforward, so this is the approach that we take here.

In order for there to exist a unique value to which the estimator converges as the sample size increases, we require that $G_n(\theta(x_0))^{\mathrm{T}} \, W_n(x_0) \, G_n(\theta(x_0))$ converges uniformly in probability to a non-random function which has a unique minimum at the true value $\theta(x_0) = \theta_0(x_0)$. It should also be the case that only parameters in the neighbourhood of θ_0 give values of the objective function close to this minimum value. The convergence is required to be uniform (and the parameter space compact, as already assumed), in order to allow minimising and limiting operations to be exchanged i.e. for the limit of the minimum to be equal to the minimum of the limit as the sample size goes to infinity.

Consider first the asymptotic behaviour of $G_n(\theta(x_0))$. We have:

$$\begin{aligned}
\mathrm{E}[G_n(\theta(x_0))] &= \mathrm{E}\Big\{\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)\,[T(Y_t) - \tau(\theta(x_0))]\Big\} \\
&= \mathrm{E}\left\{K_h(X - x_0)\,[T(Y) - \tau(\theta(x_0))]\right\} \\
&= \int\int \frac{1}{h}K\left(\frac{x - x_0}{h}\right)\,[T(y) - \tau(\theta(x_0))]\,f(y|x)\,f(x)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int \frac{1}{h}K\left(\frac{x - x_0}{h}\right)\mathrm{E}[T(Y) - \tau(\theta(x_0))|X = x]\,f(x)\,\mathrm{d}x. \qquad (4.12)
\end{aligned}$$

Letting $R_\theta(x) = \mathrm{E}[T(Y) - \tau(\theta(x_0))|X = x]$ and making the substitution $z = (x - x_0)/h$, we have:

$$\begin{aligned}
\mathrm{E}[G_n(\theta(x_0))] &= \int K(z)\,R_\theta(x_0 + zh)\,f(x_0 + zh)\mathrm{d}z \\
&= \int K(z)\Big\{R_\theta(x_0)f(x_0) + zh\big[R'_\theta(x_0)f(x_0) + R_\theta(x_0)f'(x_0)\big] \\
&\quad + \frac{(zh)^2}{2}\big[R''_\theta(x_0)f(x_0) + 2R'_\theta(x_0)f'(x_0) + R_\theta(x_0)f''(x_0)\big] + o(h^2)\Big\}\mathrm{d}z,
\end{aligned}$$

where we have taken a Taylor series expansion of the product $R_\theta(x_0 + zh)\,f(x_0 + zh)$ about x_0, keeping terms up to order h^2. Now $\int K(z)\,\mathrm{d}z = 1$ and $\int K(z)z\,\mathrm{d}z = 0$ from the kernel properties defined in (4.7), so we are left with:

$$\begin{aligned}
\mathrm{E}[G_n(\theta(x_0))] &= R_\theta(x_0)f(x_0) \;+\; h^2\int K(z)z^2\,\mathrm{d}z\,\Big\{\frac{1}{2}R''_\theta(x_0)f(x_0) + R'_\theta(x_0)f'(x_0) \\
&\quad + \frac{1}{2}R_\theta(x_0)f''(x_0)\Big\} + o(h^2). \qquad (4.13)
\end{aligned}$$

Now consider the variance of $G_n(\theta(x_0))$.

$$\begin{aligned}
&\mathrm{Var}[G_n(\theta(x_0))] \\
&= \mathrm{Var}\Big\{\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)\,[T(Y_t) - \tau(\theta(x_0))]\Big\} \\
&= \frac{1}{n}\mathrm{Var}\left\{K_h(X - x_0)\,[T(Y) - \tau(\theta(x_0))]\right\} \\
&= \frac{1}{n}\mathrm{E}\big\{K_h^2(X - x_0)[T(Y) - \tau(\theta(x_0))]\,[T(Y) - \tau(\theta(x_0))]^{\mathrm{T}}\big\} \\
&- \frac{1}{n}\mathrm{E}\big\{K_h(X - x_0)[T(Y) - \tau(\theta(x_0))]\big\}\,\mathrm{E}\big\{K_h(X - x_0)[T(Y) - \tau(\theta(x_0))]\big\}^{\mathrm{T}} \\
&= \frac{1}{n}\int\int K_h^2(x - x_0)\,[T(y) - \tau(\theta(x_0))]\,[T(y) - \tau(\theta(x_0))]^{\mathrm{T}}\,f(y|x)\,f(x)\,\mathrm{d}y\mathrm{d}x + O\Big(\frac{1}{n}\Big) \\
&= \frac{1}{n}\int \frac{1}{h^2}K^2\left(\frac{x - x_0}{h}\right) \\
&\times\; \mathrm{E}\{[T(Y) - \tau(\theta(x_0))]\,[T(Y) - \tau(\theta(x_0))]^{\mathrm{T}}\,|\,X = x\}\,f(x)\,\mathrm{d}x + O\Big(\frac{1}{n}\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\int \frac{1}{h^2}K^2\left(\frac{x-x_0}{h}\right)\left[\mathrm{Var}[T(Y)|X=x] + R_\theta(x)R_\theta(x)^{\mathrm{T}}\right] f(x)\,\mathrm{d}x + O\Big(\frac{1}{n}\Big) \\
&= \frac{1}{n}\int \frac{1}{h}K^2(z)\left[\mathrm{Var}[T(Y)|X=x_0+zh] + R_\theta(x_0+zh)R_\theta(x_0+zh)^{\mathrm{T}}\right] f(x_0+zh)\,\mathrm{d}z \\
&+ O\Big(\frac{1}{n}\Big) \\
&= \frac{f(x_0)}{nh}\int K^2(z)\,\mathrm{d}z\left[\mathrm{Var}[T(Y)|X=x_0] + R_\theta(x_0)R_\theta(x_0)^{\mathrm{T}}\right] + O\Big(\frac{1}{n}\Big), \qquad (4.14)
\end{aligned}
$$

where we have made the substitution $z = (x - x_0)/h$ in the second from last line, and the last line follows by taking a Taylor series expansion of $\left[\mathrm{Var}[T(Y)|X = x_0 + zh] + R_\theta(x_0 + zh)R_\theta(x_0 + zh)^{\mathrm{T}}\right] f(x_0 + zh)$ of order zero about x_0. Note that the same approach and definition of z as used for the derivations of $\mathrm{E}[G_n(\theta(x_0))]$ and $\mathrm{Var}[G_n(\theta(x_0))]$ will be used in the derivation of subsequent results involving kernel summations, which will therefore be shown in less detail.

By the weak law of large numbers, $G_n(\theta(x_0))$ converges in probability to its expectation, provided its variance converges to zero, which requires that $nh \to \infty$ as $n \to \infty$. Further, the expectation converges to $R_\theta(x_0)f(x_0)$ provided $h \to 0$. Therefore, for $G_n(\theta(x_0))$ to be a consistent estimator of $R_\theta(x_0)f(x_0)$, we must have the bandwidth tending to zero, but at a slower rate than $1/n$.

If we choose the weighting matrix such that it converges to a matrix of constants, then, by Slutsky's theorem, $G_n(\theta(x_0))^{\mathrm{T}}\, W_n(x_0)\, G_n(\theta(x_0))$ tends to a non-random function of θ, given by:

$$
f^2(x_0)R_\theta(x_0)^{\mathrm{T}}\, W_{x_0} R_\theta(x_0). \qquad (4.15)
$$

We have shown pointwise convergence and assumed compactness of Θ. Stochastic equicontinuity is then a sufficient condition for the convergence to be uniform (Newey 1991). In order to demonstrate this, we first consider the asymptotic behaviour of $\frac{\partial G_n(\theta(x_0))}{\partial\theta}$, given by:

$$
-\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)\left[\frac{\partial\tau(\theta(x_0))}{\partial\theta}\right].
$$

Since $\frac{\partial\tau(\theta(x_0))}{\partial\theta}$ is a matrix of constants, we need only consider $\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)$, which

is just the standard density estimator of $f(x_0)$, with:

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0) &= \mathrm{E}[K_h(X_t - x_0)] + O_P\left\{\sqrt{\mathrm{Var}\left[\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)\right]}\right\} \\ &= \int \frac{1}{h} K\left(\frac{x - x_0}{h}\right) f(x)\,\mathrm{d}x + O_P\left\{\frac{1}{\sqrt{n}}\sqrt{\mathrm{Var}[K_h(X - x_0)]}\right\} \\ &= \int K(z)\, f(x_0 + zh)\,\mathrm{d}z + O_P\left(\frac{1}{\sqrt{nh}}\right) \\ &= f(x_0)\int K(z)\mathrm{d}z + hf'(x_0)\int zK(z)\mathrm{d}z + o(h) + O_P\left(\frac{1}{\sqrt{nh}}\right) \\ &= f(x_0) + O_P\left(h^2 + \frac{1}{\sqrt{nh}}\right), \end{aligned} \tag{4.16}$$

where, as before, we have made the substitution $z = (x - x_0)/h$, and noted that $\int K(z)\,\mathrm{d}z = 1$ and $\int zK(z) = 0$. Provided that $h \to 0$ and $nh \to \infty$, and given that the first order derivative of τ is bounded, we have shown that $\frac{\partial G_n(\theta(x_0))}{\partial \theta}$ is $O_p(1)$, and converges to $-f(x_0)\frac{\partial \tau(\theta(x_0))}{\partial \theta}$.

Now, by the Mean Value Theorem for vector valued functions of several variables (see, for example, Apelian & Surace (2009)), we have:

$$\begin{aligned} G_n(\tilde{\theta}(x_0)) - G_n(\theta(x_0)) &= \begin{bmatrix} \frac{\partial G_{n_1}(\vartheta_1(x_0))}{\partial \theta} \\ \frac{\partial G_{n_2}(\vartheta_2(x_0))}{\partial \theta} \\ \vdots \\ \frac{\partial G_{n_k}(\vartheta_k(x_0))}{\partial \theta} \end{bmatrix} [\tilde{\theta}(x_0) - \theta(x_0)] \\ &= D_G(\vartheta_1, \vartheta_2, \ldots \vartheta_k)\, [\tilde{\theta}(x_0) - \theta(x_0)], \end{aligned} \tag{4.17}$$

say, where $\frac{\partial G_{n_i}(\vartheta_i(x_0))}{\partial \theta}$ represents the ith row of the matrix $\frac{\partial G_n}{\partial \theta}$ evaluated at the point ϑ_i, which lies on the segment $(\tilde{\theta}, \theta)$ (which is assumed to be entirely contained within Θ). So:

$$\begin{aligned} ||G_n(\tilde{\theta}(x_0)) - G_n(\theta(x_0))|| &= \left|\left| D_G(\vartheta_1, \vartheta_2, \ldots \vartheta_k)[\tilde{\theta}(x_0) - \theta(x_0)] \right|\right| \\ &\leq ||D_G(\vartheta_1, \vartheta_2, \ldots \vartheta_k)||\, ||\tilde{\theta}(x_0) - \theta(x_0)|| \\ &\leq M\, ||\tilde{\theta}(x_0) - \theta(x_0)||, \end{aligned} \tag{4.18}$$

where $M = \max_{\vartheta_1 \ldots \vartheta_k \in \Theta} ||D_G(\vartheta_1, \vartheta_2, \ldots \vartheta_k)||$ is $O_p(1)$, as demonstrated earlier, and $||A|| = \sqrt{\sum_{i,j} |a_{ij}|^2}$ (i.e. $||\cdot||$ represents the Euclidean matrix norm). This Lipschitz condition is sufficient for stochastic equicontinuity (Newey 1991), and therefore implies uniform convergence, as required.

The limiting function (4.15) (which only takes values greater than or equal to zero) has a unique minimum at the true value $\theta = \theta_0$, since $R_{\theta_0}(x_0) = 0$, and $R_\theta(x_0) \neq 0$ for $\theta \neq \theta_0$ (by our initial moment condition). These conditions mean that as $n \to \infty$ (with the

additional proviso that $h \to 0$ and $nh \to \infty$), our estimating equation defines a unique estimator $\hat{\theta}$ that is consistent for θ.

4.2.3 Asymptotic Variance

Now we consider the asymptotic variance of the estimator. We apply the mean value theorem again, now for the line segment $(\hat{\theta}, \theta_0)$, so we have:

$$\begin{aligned} G_n(\hat{\theta}(x_0)) &= G_n(\theta_0(x_0)) + \begin{bmatrix} \frac{\partial G_{n_1}(\check{\theta}_1(x_0))}{\partial \theta} \\ \frac{\partial G_{n_2}(\check{\theta}_2(x_0))}{\partial \theta} \\ \vdots \\ \frac{\partial G_{n_k}(\check{\theta}_k(x_0))}{\partial \theta} \end{bmatrix} [\hat{\theta}(x_0) - \theta_0(x_0)] \\ &= G_n(\theta_0(x_0)) + D_G(\check{\theta}_1, \check{\theta}_2, \dots \check{\theta}_k)\, [\hat{\theta}(x_0) - \theta_0(x_0)], \end{aligned} \tag{4.19}$$

where the points $\check{\theta}_1, \check{\theta}_2, \dots \check{\theta}_k$ lie on the line segment $(\hat{\theta}, \theta_0)$. Substituting this into Equation (4.11) gives:

$$0 = \left[\frac{\partial G_n(\hat{\theta}(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_n(x_0) \ \left[G_n(\theta_0(x_0)) + D_G(\check{\theta}_1, \check{\theta}_2, \dots \check{\theta}_k)\, [\hat{\theta}(x_0) - \theta_0(x_0)]\right]. \tag{4.20}$$

Since $\hat{\theta}$ is consistent and converges to θ_0, then so do $\check{\theta}_1, \check{\theta}_2, \dots \check{\theta}_k$, as they lie on the segment $(\hat{\theta}, \theta_0)$. We have also shown that $\frac{\partial G_n(\theta(x_0))}{\partial \theta}$ converges in probability to the non-random function $-\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} f(x_0)$, and we have $W_n(x_0) \to^p W_{x_0}$ (by appropriate selection) and so this may be restated as:

$$-f(x_0)\left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_{x_0} \ \left[G_n(\theta_0(x_0)) - f(x_0)\left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right] [\hat{\theta}(x_0) - \theta_0(x_0)]\right] = o_p(1). \tag{4.21}$$

Since $\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}$ is of full rank, and W_{x_0} is positive-definite, then $\left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta}$ is invertible, and we have:

$$\begin{aligned} \hat{\theta}(x_0) - \theta_0(x_0) \approx \frac{1}{f(x_0)} &\left\{\left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right\}^{-1} \\ &\times \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_{x_0} G_n(\theta_0(x_0)). \end{aligned} \tag{4.22}$$

So:

$$\begin{aligned}\mathrm{Var}[\hat{\theta}(x_0)] \approx \frac{1}{f^2(x_0)} \Bigg\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \Bigg\}^{-1} \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} \\ W_{x_0}\, \mathrm{Var}[G_n(\theta_0(x_0))]\, W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \left\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right\}^{-1}.\end{aligned} \tag{4.23}$$

From Equation (4.14) above, noting that $R_{\theta_0}(x_0) = 0$ by the moment condition, we have:

$$\mathrm{Var}[G_n(\theta_0(x_0))] \approx \frac{f(x_0)}{n\,h} \mathrm{Var}[T(Y)|X = x_0] \int K^2(z) \mathrm{d}z.$$

So finally, the asymptotic expression for the variance of our estimator is given by:

$$\begin{aligned}\mathrm{Var}[\hat{\theta}(x_0)] \approx \frac{1}{n\,h f(x_0)} \int K^2(z)\, \mathrm{d}z \\ \times \left\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right\}^{-1} \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_{x_0} \\ \times \mathrm{Var}[T(Y)|X = x_0]\, W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \left\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right\}^{-1}.\end{aligned} \tag{4.24}$$

It can be seen that the asymptotic variance decreases as the bandwidth h increases and as the density $f(x_0)$ increases. This is as expected, as the factor $nhf(x_0)$ can be thought of as controlling the effective local sample size. As h increases, we increase the size of the local neighbourhood, and, where $f(x_0)$ is relatively higher, the density of points in the neighbourhood is greater.

Alternatively, we can take a 'quasi-asymptotic' approach, whereby we replace expressions in Equation (4.24) with appropriate sample summations. This approach is advocated by both Fan & Gijbels (1995) (in the context of local polynomial regression) and Carroll et al. (1998) (in the context of local estimating equations), for making no more use of asymptotics than needed. Replacing $\theta_0(x_0)$ with $\hat{\theta}(x_0)$, and $W(x_0)$ with $W_n(x_0)$ we get:

$$\begin{aligned}\mathrm{Var}[\hat{\theta}(x_0)] \approx \frac{\sum_{t=1}^{n} K_h^2(X_t - x_0)}{\{\sum_{t=1}^{n} K_h(X_t - x_0)\}^2} \\ \times \left[\left[\frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_n(x_0) \frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta} \right]^{-1} \left[\frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_n(x_0)\, \mathrm{Var}[T(Y)|\, x_0] \\ \times W_n(x_0) \frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta} \left[\left[\frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_n(x_0) \frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta} \right]^{-1}.\end{aligned} \tag{4.25}$$

The conditional variance of $T(Y)$ is unknown, and will have to be estimated at each required value of X. We can use a local mean estimator with some fixed bandwidth h^* as

follows:

$$\mathrm{Var}[T(Y)|\ X = x_0] \approx \frac{\sum_{t=1}^{n} K_{h^*}(X_t - x_0)[T(y_t) - \tau(\breve{\theta}(x_0))][T(y_t) - \tau(\breve{\theta}(x_0))]^{\mathrm{T}}}{\sum_{t=1}^{n} K_{h^*}(X_t - x_0)}. \quad (4.26)$$

where $\breve{\theta}$ denotes the parameter vector fitted with the bandwidth, h^*. Carroll et al. (1998) take a slightly simpler approach, which here would involve estimating the variance of $G_n[\theta_0(x_0)]$ directly as:

$$\mathrm{Var}(G_n[\theta_0(x_0)]) \approx \frac{1}{n^2} \sum_{t=1}^{n} K_h^2(X_t - x_0)\,[T(y_t) - \tau(\hat{\theta}(x_0))][T(y_t) - \tau(\hat{\theta}(x_0))]^{\mathrm{T}}. \quad (4.27)$$

We will consider the relative merits of these two approaches in the context of our practical application in Section 6.1.

4.2.4 Asymptotic bias

Next we consider the bias, returning to Equation (4.22) and taking the expectation of both sides to get:

$$\begin{aligned}\mathrm{Bias}[\hat{\theta}(x_0)] \approx \frac{1}{f(x_0)} &\left\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right\}^{-1} \\ &\times \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0}\, \mathrm{E}[G_n(\theta_0(x_0))]. \end{aligned} \quad (4.28)$$

From Equation (4.13), now putting $R_{\theta_0}(x_0) = 0$:

$$\mathrm{E}[G_n(\theta_0(x_0))] \approx h^2 \int K(z) z^2\, \mathrm{d}z \left\{ \frac{1}{2} R''_{\theta_0}(x_0) f(x_0) + R'_{\theta_0}(x_0) f'(x_0) \right\}. \quad (4.29)$$

We have:

$$\begin{aligned} R_{\theta_0}(x) &= \mathrm{E}[T(Y) - \tau(\theta_0(x_0)) \,|\, X = x] \\ &= \mathrm{E}[T(Y) \,|\, X = x] - \tau(\theta_0(x_0)) \\ &= \tau(\theta_0(x)) - \tau(\theta_0(x_0)). \end{aligned} \quad (4.30)$$

Also:

$$R'_{\theta_0}(x) = \frac{\partial \tau(\theta_0(x))}{\partial \theta}\ \theta'_0(x) \quad (4.31)$$

and

$$R''_{\theta_0}(x) = \frac{\mathrm{d}}{\mathrm{d}x} \left[\frac{\partial \tau(\theta_0(x))}{\partial \theta} \right] \theta'_0(x) + \frac{\partial \tau(\theta_0(x))}{\partial \theta}\ \theta''_0(x). \quad (4.32)$$

Substituting back:

$$\begin{aligned}\mathrm{E}[G_n(\theta_0(x_0))] \approx h^2 \int K(z) z^2 \,\mathrm{d}z \\ \times \left\{ \left[\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}x} \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right] f(x_0) + \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} f'(x_0) \right] \theta_0'(x_0) \right. \\ \left. + \frac{1}{2} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} f(x_0)\, \theta_0''(x_0) \right\}. \end{aligned} \tag{4.33}$$

We leave the first term in this form for ease of notation, since the second differential of $\tau(\theta)$ with respect to θ would give a three-dimensional array and consequent notational complexity. So finally, the asymptotic expression for the bias is given by:

$$\begin{aligned}\mathrm{Bias}[\hat{\theta}(x_0)] \approx h^2 \int K(z) z^2 \,\mathrm{d}z \; \left\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right\}^{-1} \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \\ \times \left\{ \left[\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}x} \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right] + \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \frac{f'(x_0)}{f(x_0)} \right] \theta_0'(x_0) + \frac{1}{2} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \theta_0''(x_0) \right\}. \end{aligned} \tag{4.34}$$

As for the variance, we can gain some useful insights by considering this asymptotic result. It can be seen that the absolute bias increases as the size of the local neighbourhood, which is controlled by the bandwidth h, is increased. This is intuitive, since it means that the estimator at a given value x_0 is based on observations which include values increasingly dissimilar to x_0. The bias also depends on both the gradient and the curvature of the true curves (given by the first and second derivatives $\theta'(\cdot)$ and $\theta''(\cdot)$ respectively). The problem is that the expression involves a number of unknown terms, including the design density and its derivative, and the first and second derivatives of θ. In practice, it may be possible to estimate these using 'plug-in estimators'. This is discussed further in Section 4.4.

An alternative approach to the asymptotic calculation of $\mathrm{E}[G_n(\theta_0(x_0))]$ follows Carroll et al. (1998), and uses the fact that $\mathrm{E}[T(Y) - \tau(\theta_0(x)) \,|\, X = x] = 0$. This allows us to re-write Equation (4.12) (now writing θ for θ_0 for notational simplicity), as:

$$\begin{aligned}\mathrm{E}[G_n(\theta(x_0))] &= \int\int \left[T(Y) - \tau(\theta(x_0)) - \{T(Y) - \tau(\theta(x))\} \right] \\ &\quad \times K_h(x - x_0)\, f(y|x)\, f(x) \,\mathrm{d}y \,\mathrm{d}x \\ &= \int \left[\tau(\theta(x)) - \tau(\theta(x_0)) \right] K_h(x - x_0)\, f(x) \,\mathrm{d}x. \end{aligned} \tag{4.35}$$

Carroll et al. (1998)'s approach next involves taking a first order Taylor series expansion

of $\tau(\theta(x_0))$ about $\theta(x)$, which for our equation gives:

$$\mathrm{E}[G_n(\theta(x_0))] \approx \int \left[\frac{\partial\tau(\theta(x))}{\partial\theta}\right][\theta(x)-\theta(x_0)]\; K_h(x-x_0)\; f(x)\,\mathrm{d}x. \tag{4.36}$$

Then $\theta(x)$ is estimated by taking terms up to second order in a Taylor series expansion about x_0, such that $[\theta(x)-\theta(x_0)]$ is given by:

$$\theta'(x_0)\,(x-x_0) + \theta''(x_0)\,(x-x_0)^2/2.$$

Thus:

$$\begin{aligned}\mathrm{E}[G_n(\theta(x_0))] &\approx \int \left[\frac{\partial\tau(\theta(x))}{\partial\theta}\right] K_h(x-x_0) \times \left[\theta'(x_0)\,(x-x_0) + \theta''(x_0)\,\frac{(x-x_0)^2}{2}\right] f(x)\,\mathrm{d}x \\ &= \int K(z)\left[\frac{\partial\tau(\theta(x_0+zh))}{\partial\theta}\right]\left\{hz\,\theta'(x_0) + \frac{(hz)^2}{2}\theta''(x_0)\right\} f(x_0+zh)\,\mathrm{d}z.\end{aligned} \tag{4.37}$$

Next we take a Taylor series expansion of $\left[\frac{\partial\tau(\theta(x_0+zh))}{\partial\theta}\right] f(x_0+zh)$ about x_0, keeping terms up to order h^2 to give:

$$\begin{aligned}&\mathrm{E}[G_n(\theta(x_0))] \\ &\approx h^2 \int K(z)z^2\,\mathrm{d}z \\ &\times \left\{\left[\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial\tau(\theta(x_0))}{\partial\theta}\right] f(x_0) + \frac{\partial\tau(\theta(x_0))}{\partial\theta} f'(x_0)\right]\theta'(x_0) + \frac{1}{2}\frac{\partial\tau(\theta(x_0))}{\partial\theta} f(x_0)\,\theta''(x_0)\right\},\end{aligned} \tag{4.38}$$

where terms in $\int K(z)z\,\mathrm{d}z$ have been eliminated as they are equal to zero.

In this alternative approach, the idea of going up to order $p+2$ in the polynomial expansion of $\theta(x)$ (so that the bias is based on the $(p+1)$th and $(p+2)$th terms) is in line with that followed by Fan & Gijbels (1996) and is particularly appealing when considering higher order local polynomial regression i.e. when $p > 1$.

If we compare Equation (4.38) with our earlier Equation (4.33) then we see that the first term in the latter is half of that in the former. This difference arises because the calculation as given by Carroll et al. (1998) omits a term in h^2, for which we need to include a second order term in the approximation of $\tau(\theta(x_0))$ in Equation (4.36). We demonstrate this below, where we take the case where θ is scalar for simplicity, and now have:

$$\tau(\theta(x_0)) \approx \tau(\theta(x)) + \left[\frac{\mathrm{d}\tau(\theta(x))}{\mathrm{d}\theta}\right][\theta(x_0)-\theta(x)] + \frac{1}{2}\left[\frac{\mathrm{d}^2\tau(\theta(x))}{\mathrm{d}\theta^2}\right][\theta(x_0)-\theta(x)]^2. \tag{4.39}$$

Approximating $[\theta(x)-\theta(x_0)]$ by $\theta'(x_0)\,(x-x_0) + \theta''(x_0)\,(x-x_0)^2/2$ and making the

substitution $z = (x - x_0)/h$ as before, the new quadratic term gives an additional element in the estimation of $\mathrm{E}[G_n(\theta(x))]$ of:

$$\begin{aligned}
&- \int K(z)\, \frac{1}{2}\left[\frac{\mathrm{d}^2\tau(\theta(x_0 + zh))}{\mathrm{d}\theta^2}\right] (hz)^2 \left[\theta'(x_0)\right]^2 \, f(x_0 + zh)\, \mathrm{d}z + o(h^2) \\
&= -\frac{1}{2}h^2 \int K(z) z^2 \mathrm{d}z \left[\frac{\mathrm{d}^2\tau(\theta(x_0))}{\mathrm{d}\theta^2}\right] \left[\theta'(x_0)\right]^2 \, f(x_0) \, + o(h^2) \\
&= -\frac{1}{2}h^2 \int K(z) z^2 \mathrm{d}z \left\{ \left[\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial\tau(\theta(x_0))}{\partial\theta}\right]\right. f(x_0)\, \theta'(x_0) \, + o(h^2). \qquad (4.40)
\end{aligned}$$

Adding this term to Equation (4.38), we now have agreement with our original Equation (4.33).

We can replace expressions with appropriate sample summations, as we did for the asymptotic variance calculation, setting θ to $\hat{\theta}$, and replacing W_{x_0} with $W_n(x_0)$ to give the bias estimate:

$$\begin{aligned}
\mathrm{Bias}[\hat{\theta}(x_0)] \approx &\left[\left[\frac{\partial\tau(\hat{\theta}(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_n(x_0) \frac{\partial\tau(\hat{\theta}(x_0))}{\partial\theta}\right]^{-1} \left[\frac{\partial\tau(\hat{\theta}(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_n(x_0) \\
&\times \frac{\sum_{t=1}^n K_h(X_t - x_0) \, \left[\tau(\hat{\theta}(x_t)) - \tau(\hat{\theta}(x_0))\right]}{\sum_{t=1}^n K_h(X_t - x_0)}. \qquad (4.41)
\end{aligned}$$

This fairly crude approach to the calculation of bias is similar to that used in deriving the bias-corrected 'twicing estimator' of Stuetzle & Mittal (1979) for ordinary kernel regression and of Kauermann et al. (1998) for local estimating equations.

4.2.5 Asymptotic distribution

Finally, we consider the asymptotic distribution of our estimator, where we can appeal to a form of the Central Limit Theorem to demonstrate the approximate large sample normality of $\hat{\theta}(x_0)$ (Schuster 1972). We have already specified that we must have $h \to 0$ and $nh \to \infty$ in the derivations above. We have also seen that there is a bias-variance trade-off in selecting h: as h increases, the variance decreases, but the bias increases and vice-versa. If we want to choose the mean-square optimal bandwidth, then we require $h = h^*$ to equate the rates of convergence to zero of the squared bias and variance. Thus we need h^* to satisfy:

$$O((h^*)^4) = O\left(\frac{1}{nh^*}\right) \qquad (4.42)$$

i.e. we want $h^4 \propto \frac{1}{nh}$, so we pick h such that $nh^5 \to c$, a constant. Note that, given this choice of h, the limiting distribution will be biased, although the bias term disappears asymptotically as $h \to 0$.

We define the bias at x_0 as $h^2 B(x_0)$, where (from Equation (4.34)):

$$B(x_0) = \int K(z) z^2 \, \mathrm{d}z \, \left\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right\}^{-1} \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \times \left\{ \left[\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}x} \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right] + \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \frac{f'(x_0)}{f(x_0)} \right] \theta_0'(x_0) + \frac{1}{2} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \theta_0''(x_0) \right\}. \tag{4.43}$$

Then, returning to Equation (4.22), subtracting the bias, multiplying both sides by $(nh)^{1/2}$, and using the variance result from Equation (4.24), we have:

$$\begin{aligned} &(nh)^{1/2} \{ \hat{\theta}(x_0) - \theta_0(x_0) - h^2 B(x_0) \} \to^D \\ &\quad N(0, \frac{1}{f(x_0)} \int K^2(z) \, \mathrm{d}z \, \left\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right\}^{-1} \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \\ &\quad \times V(x_0) \, W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \left\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right\}^{-1}, \end{aligned} \tag{4.44}$$

where:

$$V(x_0) = \mathrm{Var}[T(Y) | X = x_0]. \tag{4.45}$$

Note that, using the mean square optimal bandwidth, which is proportional to $n^{-1/5}$, the estimator converges at rate $n^{-2/5}$, compared with a rate of $n^{-1/2}$ for a parametric fit.

An alternative approach, taken by Lewbel (2007) and many other authors, assumes $nh^5 \to 0$. This assumption for the bandwidth makes the bias shrink faster than the variance ('undersmoothing'), so is not mean square optimal, but simplifies the limiting distribution since no bias term is then required. It is not clear, however, how one chooses such an estimator in practice. This approach is also criticised by Hansen (2012) for giving an inefficient and misleading estimator, which misses the bias-variance trade-off that is an inherent part of nonparametric estimation. We prefer the mean-square optimal approach.

The limiting distribution (given by Equation (4.44)) can be used to calculate approximate pointwise confidence intervals, although note that these will be for $\mathrm{E}(\hat{\theta}(x_0))$ rather than $\theta(x_0)$ itself, due to the bias. Although it is possible to estimate the bias term, this would itself involve terms in $\hat{\theta}$ (its first and second derivatives), and so would increase the variance of the estimator from that given in Equation (4.44). In practice it is usually deemed sufficient to indicate a level of variability without adjusting for bias. In order to avoid confusion such intervals are referred to as 'variability bands', rather than confidence intervals (Bowman & Azzalini 1997).

4.2.6 The weighting matrix

So far we have just assumed that the weighting matrix $W_n(x_0)$ may depend on the data, and that it converges to the positive definite non-random matrix, W_{x_0}. Now we see that, if we take the weighting matrix to be the inverse of the conditional covariance matrix of the statistics, such that $W_{x_0} = V(x_0)^{-1}$, then the expression for the variance simplifies and we have:

$$(nh)^{1/2}\{\hat{\theta}(x_0) - \theta_0(x_0) - h^2 B(x_0)\} \to^D$$
$$N\left(0,\ \frac{1}{f(x_0)} \int K^2(z)\,\mathrm{d}z \left\{ \left[\frac{\partial \tau(\theta_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} V(x_0)^{-1} \frac{\partial \tau(\theta_0(x_0))}{\partial \theta} \right\}^{-1}\right). \tag{4.46}$$

This choice for the weighting matrix gives optimal efficiency, and a two-step procedure can be used, as described in Section 2.3 in respect of standard GMM. However, computation time is now an even more onerous constraint, since this procedure would be required in respect of each evaluation point. As before we use the conclusions of Jesus & Chandler (2011)'s simulation study to justify using just a diagonal matrix of inverse variances. With a continuous covariate, however, we only have a single observation at each evaluation point, so we cannot take a straightforward sample estimate, as we did when modelling conditional on month. A practical approach may be to group the data into bins based on the value of the covariate, calculating sample variances within each bin, and then potentially smoothing across individual observations. This is considered in more detail in Section 4.6.2, which describes our practical approach.

4.3 Extending to higher order polynomials

The Nadaraya-Watson or local mean approach suffers from some limitations. Consider a covariate value, x_0, at which the true curve has a positive gradient, and where there are more observation points in the local neighbourhood with a lower covariate value, then a higher one. The local mean weighting approach will then give a fitted value at x_0 which is biased downwards. The converse is true if there are more points with a higher covariate value. This sort of bias is known as design bias, and arises when the design is not equi-spaced. A similar problem arises at (or near) the boundaries where the observation points in the local neighbourhood lie only (or primarily) to one side, and the fitted curve will tend to be too flat as a result. This is known as boundary bias. Various approaches exist in the literature aimed at addressing these limitations. These include boundary kernel methods (Müller 1991, Jones 1993), reflection methods and 'pseudo-data'. The methods are reviewed by Dai & Sperlich (2010), who note that boundary correction methods are not much used in practice, perhaps because they are not allowed for in most standard software, and are also seen as complex. These methods are in any case primarily aimed at

density estimation, rather than regression, since in the latter, extending to a local linear approach is generally considered to be a more appealing way to address boundary issues.

In Figure 4.2 we return to the simple motivating example of Section 4.1 to illustrate the difference between a local mean and local linear fit. The local linear fit's correction for the boundary bias is evident. In fact Fan & Gijbels (1996) (Section 3.3) show that taking the

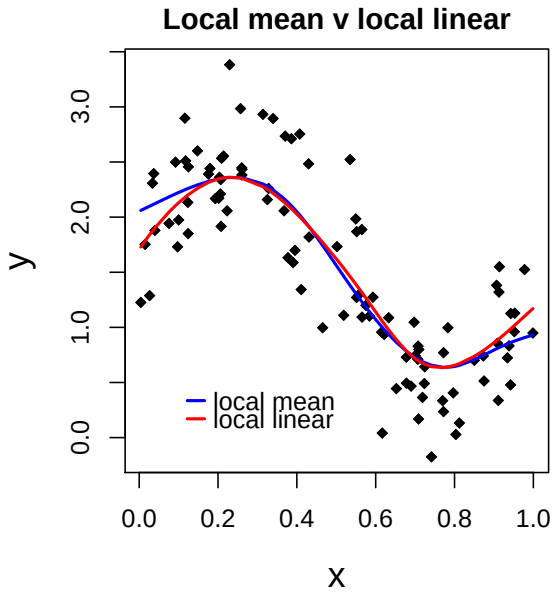


Figure 4.2: Scatterplots showing the relationship between x and y, generated from the equation $Y = \sin(6X) + 1.5 + \varepsilon$, $\varepsilon \sim N(0, 0.14)$, showing local mean (blue) and local linear (red) fits.

order of the polynomial to be odd in local polynomial regression reduces bias compared with the preceding even order estimate, without increasing asymptotic variance. So it appears that odd order fits are (at least asymptotically) preferable to those using even orders. There is no clear-cut comparison between two odd order fits however — increasing the order (from 1 to 3, say) reduces bias, but at the expense of increased variance.

Order $p = 1$ is therefore a very popular choice, although a higher value of p may be required if estimates of the derivatives of θ are also required (in which case, one might choose the order of derivative required plus 1). Ruppert & Wand (1994) give a word of caution, however, against taking the local linear estimator as the automatic benchmark. They acknowledge the importance of reducing boundary bias, but they point out that, while for interior points, the ratio of the asymptotic variance of the local linear estimator to the variance of the local mean estimator is 1, at the boundaries it can be considerably higher, and in some circumstances, near the boundary, the latter may provide a more accurate estimate. Note also that while the two methods have identical asymptotic variances at interior points, in finite samples the Nadaraya-Watson estimator tends to have a smaller variance. Thus in cases where the regression function is fairly flat, the Nadaraya-Watson estimator may have the advantage. We now consider the estimating equation, variance

and bias for the local linear fitting approach in respect of our model.

4.3.1 Derivations for the local linear model

Now, when solving for the parameters at the evaluation point x_0, we have $\psi(X_t) = b_0(x_0) + b_1(x_0)(X_t - x_0)$, where $b_0(x_0) = \theta(x_0)$ and $b_1(x_0) = \theta'(x_0)$. The fitted value of θ at x_0 is therefore equal to $\hat{b}_0(x_0)$, with $\hat{b}_1(x_0)$ providing an estimator for the gradient of the curve at x_0. The parameters $b_0(\cdot)$ and $b_1(\cdot)$ will need to be found for each required evaluation point, as before. For ease of manipulation we will write the parameter set in a vector of length $2q$ as $b(x) = (b_0(x)^{\mathrm{T}}, b_1(x)^{\mathrm{T}})^{\mathrm{T}}$, and define the vector X_{x_0} as $(1, X_t - x_0)^{\mathrm{T}}$. We no longer write a subscript of 0 to denote the 'true' values of parameters, with the subscripts here simply indicating the appropriate coefficient.

We develop the method for the local linear case using a mixture of practical and theoretical investigation. We will describe the route by which we reach our conclusions here, including our original approach, although this was ultimately proven to be invalid. We define G_n broadly as before, but now it is a function of the vector $b(x)$, which includes components in respect of θ and θ', and we write $\psi(X_t)$ in the summation rather than $\theta(x_0)$, since θ is no longer assumed to be constant in the neighbourhood of x_0:

$$G_n(b(x)) = \frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x)\,[T(Y_t) - \tau(\psi(X_t))], \tag{4.47}$$

where $\psi(X_t) = b_0(x) + b_1(x)(X_t - x)$. We originally attempted to solve for $\hat{b}$ at the point x_0, by minimising $G_n(b(x_0))^{\mathrm{T}}\, W_n(x_0)\, G_n(b(x_0))$ numerically, as before. This practical approach failed, with estimates of b_0 and b_1 varying wildly with x. Returning to the theory, it is clear that the problem is that the moment condition implicit in this approach is just:

$$\mathrm{E}[T(Y)|X = x] = \tau(b_0(x)), \tag{4.48}$$

which only defines b_0, and gives no information on b_1. Further, the matrix $\frac{\partial G_n(b(x))}{\partial b}$ is (asymptotically) not of full rank, since:

$$\begin{aligned}
\mathrm{E}\left[\frac{\partial G_n(b(x_0))}{\partial b}\right]^{\mathrm{T}} &= \mathrm{E}\left[-\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)\; X_{x_0} \otimes \left[\frac{\partial \tau(\psi(X_t))}{\partial \theta}\right]^{\mathrm{T}}\right] \\
&= -\int \begin{pmatrix} 1 \\ zh \end{pmatrix} K(z) \otimes \left[\frac{\partial \tau(\psi(x_0 + zh))}{\partial \theta}\right]^{\mathrm{T}} f(x_0 + zh)\mathrm{d}z \\
&= -\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \left[\frac{\partial \tau(b_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} f(x_0) + o(h^2),
\end{aligned} \tag{4.49}$$

where the notation $\frac{\partial \tau(\psi(X_t))}{\partial \theta}$ represents $\frac{\partial \tau}{\partial \theta}$ evaluated at $\theta = \psi(X_t) = b_0(x_0) + b_1(x_0)(X_t -$

x_0), and the $\otimes$ symbol denotes the Kronecker product, which takes two matrices, and multiplies each element of the first matrix by the entire second matrix, putting each result in place of the element of the first matrix. Then, if we let $M_{x_0}^{\mathrm{T}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \left[\frac{\partial \tau(b_0(x_0))}{\partial \theta}\right]^{\mathrm{T}} f(x_0)$, the calculations for the asymptotic bias and variance require us to invert the matrix $M_{x_0}^{\mathrm{T}} W_{x_0} M_{x_0}$, but this matrix is not invertible. (For example, in the case where $q = 1$, $k = 2$, only the $(1,1)$th entry in this 2×2 matrix is non-zero.)

It is clear then, that an alternative fitting equation is required, which allows the gradients of the components of θ to be identified. We show in the next section that this can be achieved by applying the kernel weights to the quadratic form in $[T(Y_t) - \tau(\psi(X_t))]$, rather than to $[T(Y_t) - \tau(\psi(X_t))]$ itself.

4.3.2 Alternative local approach

Recall that we have so far been considering a local GMM approach where we solve for $\hat{\theta}(x_0)$ using the equation:

$$\hat{\theta}(x_0) = \mathrm{argmin}_{\theta(x_0)} \left\{ \frac{1}{n} \sum_{t=1}^{n} K_h(X_t - x_0)[T(Y_t) - \tau(\psi(X_t))] \right\}^{\mathrm{T}} W_n(x_0) \times \left\{ \frac{1}{n} \sum_{t=1}^{n} K_h(X_t - x_0)[T(Y_t) - \tau(\psi(X_t))] \right\}, \tag{4.50}$$

where $\psi(X_t)$ is equal to $\theta(x_0)$ for the local mean case, and $b_0(x_0) + b_1(x_0)(X_t - x_0)$ (i.e. $\theta(x_0) + \theta'(x_0)(X_t - x_0)$) for the local linear case. Our alternative approach applies the kernel weights to the quadratic form instead, to give:

$$\hat{\theta}(x_0) = \mathrm{argmin}_{\theta(x_0)} \left\{ \frac{1}{n} \sum_{t=1}^{n} K_h(X_t - x_0) \left[T(Y_t) - \tau(\psi(X_t))\right]^{\mathrm{T}} W_n(x_0) \left[T(Y_t) - \tau(\psi(X_t))\right] \right\}. \tag{4.51}$$

The determination of the appropriate weighting matrices to use here is less clear, but we have chosen to use $W_n(x_0)$ rather than $W_n(x_t)$ in the equation since this will prove to be more convenient, and in any case we are assuming local homoscedasticity. For simplicity, we will also treat the weighting matrix in the local linear derivations as it were known, and so will refer to it as W_{x_0}. In practice, this is not necessarily the case (e.g. if it is based on the variance of the statistics), in which case the results will include an additional element of approximation. Differentiating this equation with respect to b will give us the

following two sets of equations:

$$0 = \frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)\left[\frac{\partial\tau(\psi(X_t))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0}\,[T(Y_t) - \tau(\psi(X_t))]. \tag{4.52}$$

$$0 = \frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)\,(X_t - x_0)\left[\frac{\partial\tau(\psi(X_t))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0}\,[T(Y_t) - \tau(\psi(X_t))]. \tag{4.53}$$

These equations exactly identify the $2q$ parameters, and are effectively the sample equivalents of the required moment conditions.

Now we consider the asymptotic variance and bias of the estimators using this alternative approach, for the local linear case. We follow Carroll et al. (1998) by letting the unknown parameters be $a_0 = b_0$ and $a_1 = h\,b_1$, which simplifies the expressions. Now $\psi(X_t) = a_0(x) + a_1(x)\,(X_t - x)/h$.

We differentiate the right-hand side of Equation (4.51) with respect to $a = (a_0^{\mathrm{T}}, a_1^{\mathrm{T}})^{\mathrm{T}}$, ignoring the constant multiplier of 2 (which does not affect the location of the minimum), and define:

$$L_n(a(x_0)) = -\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0)\,X_{x_0}^h \otimes \left[\frac{\partial\tau(\psi(X_t))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0}\,[T(Y_t) - \tau(\psi(X_t))], \tag{4.54}$$

with $X_{x_0}^h$ denoting the vector $(1, (X_t - x_0)/h)^{\mathrm{T}}$, and $L_n(\hat{a}(x_0) = 0$. Note that $L_n(\hat{a}(x_0))$ is a vector of length $2q$.

Then:

$$\begin{aligned}\frac{\partial}{\partial(a_0^{\mathrm{T}}, a_1^{\mathrm{T}})} L_n(a(x_0)) = \frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x_0) X_{x_0}^h {X_{x_0}^h}^{\mathrm{T}} \\ \otimes \left\{\left[\frac{\partial\tau(\psi(X_t))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial\tau(\psi(X_t))}{\partial\theta} - \Phi(X_t, Y_t)\right\}\end{aligned}$$

where $\Phi(X_t, Y_t)$ is the $q \times q$ matrix with (i,j)th term given by:

$$\Phi(X_t, Y_t)_{(i,j)} = \sum_{l=1}^{k}\sum_{m=1}^{k} \frac{\partial^2\tau_l(\psi(X_t))}{\partial\theta_i\partial\theta_j}\, W_{(l,m)_{x_0}}\,[T_m(Y_t) - \tau_m(\psi(X_t))]$$

So,

$$\begin{aligned}
&\frac{\partial}{\partial(a_0^{\mathrm{T}}, a_1^{\mathrm{T}})} L_n(a(x_0)) \\
&= \mathrm{E}\left[K_h(X - x_0) X_{x_0}^h X_{x_0}^{h\ \mathrm{T}} \otimes \left\{ \left[\frac{\partial \tau(\psi(X))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\psi(X))}{\partial \theta} - \Phi(X, Y) \right\} \right] + O_p(1/\sqrt{nh}) \\
&= \int K_h(x - x_0) \begin{pmatrix} 1 & (x - x_0)/h \\ (x - x_0)/h & (x - x_0)^2/h^2 \end{pmatrix} \otimes \left\{ \left[\frac{\partial \tau(\psi(x))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\psi(x))}{\partial \theta} \right. \\
&\left. - \mathrm{E}[\Phi(X, Y) | X = x] \right\} f(x)\, \mathrm{d}x \ + \ O_p(1/\sqrt{nh}) \\
&= \int K(z) \begin{pmatrix} 1 & z \\ z & z^2 \end{pmatrix} \otimes \\
&\quad \left\{ \left[\frac{\partial \tau(\psi(x_0 + zh))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\psi(x_0 + zh)}{\partial \theta} \right\} f(x_0 + zh)\, \mathrm{d}z + O(h^2) + O_p(1/\sqrt{nh}) \\
&= f(x_0) \begin{pmatrix} 1 & 0 \\ 0 & \int K(z) z^2 \mathrm{d}z \end{pmatrix} \otimes \left[\frac{\partial \tau(a_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(a_0(x_0))}{\partial \theta} \ + O_p(h + 1/\sqrt{nh}). \quad (4.55)
\end{aligned}$$

Note that this is a matrix of dimension $2q \times 2q$. The term in $\mathrm{E}[\Phi(X, Y)|X = x]$ in the first line is $O(h^2)$ since, if we treat it as a function of x and take a Taylor series expansion about x_0, the first two terms are equal to zero. This is because $\mathrm{E}[T_m(Y) - \tau_m(\psi(X))]|X = x_0]$ and its first derivative with respect to x are both zero in the local linear case. (A similar result in respect of $\mathrm{E}[L_n(a(x_0))]$ which is discussed below in more detail, should make this clearer.)

In order to derive the asymptotic variance and bias of the estimators we use a first order Taylor series expansion of $L_n(\hat{a}(x_0))$, which is equal to zero, about $a(x_0)$, and rearrange to get:

$$\begin{aligned}
&\begin{pmatrix} \hat{a}_0(x_0) - a_0(x_0) \\ \hat{a}_1(x_0) - a_1(x_0) \end{pmatrix} \\
&\approx -\frac{1}{f(x_0)} \left\{ \begin{pmatrix} 1 & 0 \\ 0 & \int K(z) z^2 \mathrm{d}z \end{pmatrix} \otimes \left[\frac{\partial \tau(a_0(x_0))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(a_0(x_0))}{\partial \theta} \right\}^{-1} \ L_n(a(x_0)). \quad (4.56)
\end{aligned}$$

Next we derive $\mathrm{E}[L_n(a(x_0)]$ and $\mathrm{Var}[L_n(a(x_0)]$:

$$\begin{aligned}
&\mathrm{E}[L_n(a(x_0))] \\
&= -\int K_h(x - x_0) \begin{pmatrix} 1 \\ (x - x_0)/h \end{pmatrix} \otimes \left[\frac{\partial \tau(\psi(x))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} [T(y) - \tau(\psi(x))]\ f(y|x)\, f(x)\, \mathrm{d}y \mathrm{d}x \\
&= -\int K_h(x - x_0) \begin{pmatrix} 1 \\ (x - x_0)/h \end{pmatrix} \otimes \chi[\psi(x)]^{\mathrm{T}} W_{x_0} R(x)\ f(x)\, \mathrm{d}x,
\end{aligned}$$

where we write $\chi[\psi(x)]$ for $\frac{\partial\tau}{\partial\theta}$ evaluated at $\psi(x) = a_0(x_0) + a_1(x)(x - x_0)/h$ for notational convenience, and

$$\begin{aligned} R(x) &= \mathrm{E}[T(Y) - \tau(\psi(x))|X = x] \\ &= \tau(\theta(x)) - \tau(\psi(x)), \end{aligned} \tag{4.57}$$

noting that $R(x_0)$ and $R'(x_0)$ are both zero, since $\theta(x_0) = a_0(x_0)$ and $\theta'(x_0) = a_1(x_0)/h$. Also:

$$R''(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial\tau(\theta(x))}{\partial\theta}\right)\theta'(x) + \frac{\partial\tau(\theta(x))}{\partial\theta}\theta''(x) + \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial\tau(\psi(x))}{\partial\theta}\right)\frac{a_1}{h},$$

and so:

$$R''(x_0) = \frac{\partial\tau(\theta(x))}{\partial\theta}\theta''(x). \tag{4.58}$$

Continuing by letting $z = (x - x_0)/h$ as before:

$$\begin{aligned} \mathrm{E}[L_n(a(x_0))] &= -\int K(z)\begin{pmatrix}1\\ z\end{pmatrix} \otimes \chi[\psi(x_0 + zh)]^{\mathrm{T}} W_{x_0} R(x_0 + zh)\ f(x_0 + zh)\,\mathrm{d}z \\ &= -\frac{1}{2}\begin{pmatrix}h^2\int K(z)z^2\mathrm{d}z\\ 0\end{pmatrix} \otimes \chi[a_0(x_0)]^{\mathrm{T}} W_{x_0} R''(x_0)\ f(x_0) \\ &\quad - \frac{1}{6}\begin{pmatrix}0\\ h^3\int K(z)z^4\mathrm{d}z\end{pmatrix} \otimes \Big[3\chi'[a_0(x_0)]^{\mathrm{T}} W_{x_0} R''(x_0)\ f(x_0) \\ &\quad + \chi[a_0(x_0)]^{\mathrm{T}} W_{x_0} R^{(3)}(x_0) f(x_0) + 3\,\chi[a_0(x_0)]^{\mathrm{T}} W_{x_0} R''(x_0) f'(x_0)\Big] + o(h^3) \end{aligned} \tag{4.59}$$

and:

$$\begin{aligned} \mathrm{Var}[L_n(a(x_0)] &= \frac{1}{n}\mathrm{Var}\left[K_h(X - x_0)\ X_{x_0}^h \otimes \chi[\psi(X)]^{\mathrm{T}} W_{x_0}\ [T(Y) - \tau(\psi(X))]\right] \\ &= \frac{1}{n}\int K_h^2(x - x_0) X_{x_0}^h {X_{x_0}^h}^{\mathrm{T}} \otimes\ \chi[\psi(x)]^{\mathrm{T}} W_{x_0} \\ &\quad \times\ [T(y) - \tau(\psi(x))]\,[T(y) - \tau(\psi(x))]^{\mathrm{T}} W_{x_0}\ \chi[\psi(x)]\ f(y|x) f(x)\ \mathrm{d}y\,\mathrm{d}x + O\left(\frac{1}{n}\right) \\ &= \frac{1}{n}\int K_h^2(x - x_0) X_{x_0}^h {X_{x_0}^h}^{\mathrm{T}} \otimes\ \chi[\psi(x)]^{\mathrm{T}} W_{x_0} \\ &\quad \times\ \left[\mathrm{Var}[T(Y)|X = x] + R(x)R(x)^{\mathrm{T}}\right] W_{x_0}\ \chi[\psi(x)]\ f(x)\ \mathrm{d}x + O\left(\frac{1}{n}\right) \\ &= \frac{f(x_0)}{nh}\int K^2(z)\begin{pmatrix}1 & z\\ z & z^2\end{pmatrix}\mathrm{d}z \otimes \chi[a_0(x_0)]^{\mathrm{T}} W_{x_0} \\ &\quad \times\ \mathrm{Var}[T(Y)|X = x_0]\, W_{x_0}\ \chi[a_0(x_0)]\ +\ O\left(\frac{1}{n}\right), \end{aligned}$$

where the last line follows since $R(x_0) = 0$. Now we use the approximations of $\mathrm{E}[L_n(a(x_0))]$ and $\mathrm{Var}[L_n(a(x_0)]$ and Equation (4.56) to derive the variance and bias of the parameter estimates:

$$\begin{aligned}\mathrm{Var}\begin{pmatrix}\hat{a}_0(x_0)\\ \hat{a}_1(x_0)\end{pmatrix} \approx \frac{1}{nh\,f(x_0)}\Bigg\{ \begin{pmatrix}1 & 0\\ 0 & \int K(z)z^2\mathrm{d}z\end{pmatrix} \otimes \left[\frac{\partial\tau(a_0(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial\tau(a_0(x_0))}{\partial\theta}\Bigg\}^{-1}\\ \times \int K^2(z)\begin{pmatrix}1 & z\\ z & z^2\end{pmatrix}\mathrm{d}z\\ \otimes \left[\frac{\partial\tau(a_0(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0}\,\mathrm{Var}[T(Y)|X=x_0]\,W_{x_0}\,\frac{\partial\tau(a_0(x_0))}{\partial\theta}\\ \times \Bigg\{ \begin{pmatrix}1 & 0\\ 0 & \int K(z)z^2\mathrm{d}z\end{pmatrix} \otimes \left[\frac{\partial\tau(a_0(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial\tau(a_0(x_0))}{\partial\theta}\Bigg\}^{-\mathrm{T}} \qquad (4.60)\end{aligned}$$

and:

$$\begin{aligned}\mathrm{Bias}\begin{pmatrix}\hat{a}_0(x_0)\\ \hat{a}_1(x_0)\end{pmatrix} \approx -\frac{1}{f(x_0)}\Bigg\{ \begin{pmatrix}1 & 0\\ 0 & \int K(z)z^2\mathrm{d}z\end{pmatrix}\\ \otimes \left[\frac{\partial\tau(a_0(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial\tau(a_0(x_0))}{\partial\theta}\Bigg\}^{-1} \mathrm{E}[L_n(a(x_0)]. \qquad (4.61)\end{aligned}$$

We are interested primarily in the bias of $\hat{a}_0(x_0) = \hat{\theta}(x_0)$, which we now write in full as:

$$\begin{aligned}\mathrm{Bias}[\hat{a}_0(x_0)] \approx \Bigg\{ \left[\frac{\partial\tau(a_0(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0} \frac{\partial\tau(a_0(x_0))}{\partial\theta}\Bigg\}^{-1}\\ \times\ \frac{1}{2}h^2 \int K(z)z^2\mathrm{d}z \left[\frac{\partial\tau(a_0(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_{x_0} \left[\ \frac{\partial\tau(a_0(x_0))}{\partial\theta}\ \ \theta''(x_0)\right], \qquad (4.62)\end{aligned}$$

where we have substituted for $R''(x_0)$ using Equation (4.58). We can see that the asymptotic variance of $\hat{b}_0$ $(= \hat{a}_0)$ of the local linear estimator, given by Equation (4.60), is equivalent to that of the local mean estimator in Equation (4.24), whereas the asymptotic bias is reduced (comparing Equations (4.62) and (4.34)). The bias is no longer dependent on the gradient of θ, nor on the design density. This is similar to results in respect of local regression, and in line with expectations.

As for the local mean estimator, the asymptotic expressions include unknown terms, and we can instead take a quasi-asymptotic approach, replacing expressions with sample sum-

mations. For example, the expression for the variance of $\hat{a}(x_0)$ is given by:

$$\begin{aligned}
&\mathrm{Var}[\hat{a}(x_0)] \\
&\approx \left\{ \sum_{t=1}^{n} K_h(X_t - x_0)\; X_{x_0}^h X_{x_0}^{h^{\mathrm{T}}} \otimes \left[\frac{\partial \tau(\hat{\psi}(X_t))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\hat{\psi}(X_t))}{\partial \theta} \right\}^{-1} \\
&\times \sum_{t=1}^{n} K_h^2(X_t - x_0)\;\; X_{x_0}^h X_{x_0}^{h^{\mathrm{T}}} \otimes \left[\frac{\partial \tau(\hat{\psi}(X_t))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0}\, \mathrm{Var}[T(Y)|X = x_0]\, W_{x_0} \frac{\partial \tau(\hat{\psi}(X_t))}{\partial \theta} \\
&\times \left\{ \sum_{t=1}^{n} K_h(X_t - x_0)\;\; X_{x_0}^h X_{x_0}^{h^{\mathrm{T}}} \otimes \left[\frac{\partial \tau(\hat{\psi}(X_t))}{\partial \theta} \right]^{\mathrm{T}} W_{x_0} \frac{\partial \tau(\hat{\psi}(X_t))}{\partial \theta} \right\}^{-1}, \qquad (4.63)
\end{aligned}$$

with the variance for $\hat{a}_0(x_0)$ given by the (1,1)th term.

Note that if we used this alternative approach for the local mean estimator (i.e. applying the kernel weights to the quadratic form), then we would find the same solution as before, and the asymptotic bias and variance would also be unchanged. The advantage of our original method in the local mean case is purely practical. Averaging the statistics, rather than the quadratic form, allows practitioners to use existing software for the optimisation routine, simply replacing the average over calendar month, with the appropriate kernel based average before calling the routine.

We will not consider polynomials of order greater than 1, since it is clear that these are neither desirable nor likely to be viable for the rainfall models, requiring the estimation of too many parameters relative to the amount of available information.

So far we have been assuming that the bandwidth h is a function of the sample size, but we have not considered how it may be determined in practice. We look at this in the next section.

4.4 Choosing a bandwidth

We use a global bandwidth, which means that h is constant across the whole data range, but will briefly consider the alternatives. A local bandwidth allows h to be different at each evaluation point. This gives more flexibility, for example allowing a smaller value of h over areas with high curvature and/or where the density of observation points is high. Another approach allows a fixed percentage of the data (termed the 'span') to contribute a positive weight to each local fit. This is the idea behind 'nearest-neighbour' methods such as Loess (Cleveland 1979, Cleveland & Devlin 1988), which aim to give a broadly constant variance for the estimator across all covariate values (assuming the variance of the response is itself broadly constant). However, as already discussed, choosing a tuning constant is a question of balancing the bias and variance, which move in opposite directions as the tuning constant is varied. Thus the stability of variance that the nearest neighbour

method brings is at the expense of more variable bias. This type of method is likely to be desirable for curves with a complex structure, whereas a constant global bandwidth should be adequate for relatively smooth curves, or where the amount of data is not sufficient to justify a local approach.

Now we consider how a suitable global bandwidth may be chosen. A popular practical approach in the case of local regression is simply to choose this subjectively after a visual inspection of the fits with different values of h. In the case of a scatterplot smooth, the fitted line can be compared to the actual data points, and a practitioner can take an informed view as to whether particular features identified in the fits are 'real' or just due to randomness. A slightly more complex, but still subjective, idea is to decide on an appropriate complexity of the model in terms of 'effective degrees of freedom'. Thus for example, if it is felt that a function should be equivalent in complexity to a cubic polynomial, then a fit would be desirable which gives an effective degree of freedom of around 4. This approach requires the approximation of the effective degrees of freedom for a given bandwidth. For local regression an explicit solution exists which is linear in the response, so that the estimator can be expressed in the form $\hat{\theta} = Sy$. This means that the approximate degrees of freedom can be determined by analogy with normal linear regression as the trace of the smoothing matrix, S. (Recall that in normal linear regression, the degrees of freedom are given by the trace of the 'hat' matrix, $H = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$, where $\hat{Y} = HY$.)

For our model, however, there is no explicit solution, and it is not clear how one might estimate the effective degrees of freedom. Subjective assessment is also more difficult because we can only compare the statistics $T(Y)$ against their expectations $\tau(\theta)$ (of which we have k components), and cannot directly compare our fitted parameters against observed responses.

Alongside these subjective approaches, automatic bandwidth selection methods generally aim to minimise the integrated mean squared error or a proxy for this. Fan & Gijbels (1996) define the IMSE as:

$$\mathrm{IMSE} = \int [\mathrm{Bias}(\hat{\theta}(x))^2 + \mathrm{Var}(\hat{\theta}(x))]\, w(x)\mathrm{d}x,$$

for some weight function w. Bowman & Azzalini (1997) have $w(x) = f(x)$, the design density.

4.4.1 Plug-in method

One of the main approaches to automatic bandwidth selection is called the 'plug-in' method, and involves the minimisation of the asymptotic expression for the IMSE (or MSE for local h). The asymptotic expression is thus differentiated with respect to h and

set to zero to get an expression for the optimal h. This gives a bandwidth that is $O(n^{-1/5})$ as we saw in Section 4.2.5. The problem is that this depends on a number of unknown quantities, so these then need to be estimated or 'plugged-in'. Returning to the asymptotic equations for the variance and bias of the local mean estimator given by Equations (4.24) and (4.34) respectively, we see that estimates would be required of $f(x)$, $f'(x)$, $\theta'(x)$, $\theta''(x)$ and $\mathrm{Var}(T(Y)|X = x)$ (or appropriate functions of these).

Fan & Gijbels (1996) suggest a relatively simple approach to estimating the unknown derivatives of θ in the context of local polynomial regression which they call the 'rule of thumb' method. This involves fitting a polynomial of order $p+3$ (e.g. of order 4 for local linear regression) globally to $\theta(x)$, so that we get a parametric fit given by:

$$\breve{\theta} = \breve{\alpha}_0 + ... + \breve{\alpha}_4 x^4.$$

Then an estimator for $\theta''(x)$ is given by the quadratic $2\breve{\alpha}_2 + 6\breve{\alpha}_3 x + 12\breve{\alpha}_4 x^2$, which allows a reasonable flexibility in estimating the curvature. An estimator for θ' is also clearly available (although not required in the case of local linear regression). The conditional variance is estimated by the standardised residual sum of squares from this parametric fit, and the need to estimate $f(x)$ is avoided by replacement of an integral with a summation over observed data. Fan & Gijbels (1996) suggest that this method is in a sense similar to the normal reference method used for density estimation (where unknown values in the calculation for the optimal bandwidth are estimated by simply assuming the normal density). The added complexity of the regression estimation problem means that there is no straightforward equivalent.

This method is not viable for the local rainfall model, since, even for the local mean estimator, it would require of the order of 20 parameters in the global fit, and it is highly unlikely that enough independent moment conditions could be found for a numerically stable solution. Other approaches to finding higher order derivatives of θ involve carrying out local fits of the required order using a 'pilot bandwidth'. Since we require the second derivative of θ, we would need at least a local quadratic fit at the pilot bandwidth, which again would give rise to identification problems. Overall, this would appear to be a general problem with plug-in methods, making them inappropriate for our models.

The main alternative to the plug-in method is some form of cross-validation approach, which we will consider in the next section.

4.4.2 Cross-validation techniques

An ideal way to select an optimal smoothing parameter would be to compare the prediction performance of the model with different values of h on a new set of test data. Each model would be estimated from the original 'training' data set, and the models' prediction

errors could be compared on the new 'test' data set. Such new test data are not generally available, however, and a simple alternative approach is just to split the data randomly into two sets — estimating the model from one, and testing it on the other, which is known as a 'hold-out' sample.

Cross-validation techniques are more sophisticated versions of the simple 'hold-out' sample, artificially splitting the data into several 'training' and 'test' data sets, using various different approaches (see Arlot & Celisse (2011) for a comprehensive survey).

Leave-one-out cross-validation is an example of exhaustive data splitting, which means that all possible combinations of a given split ratio are considered: in this case the split ratio is $1 : n - 1$. We will consider this first in the context of a scatterplot smoother. The idea is to predict a response value for each data point, (Y_i, X_i) from the remainder of the data i.e. excluding the ith observation itself. The estimator calculated this way is denoted $\hat{\theta}_{h,-i}$. An estimate of the performance of the estimation is then obtained from the least squares loss function:

$$CV(h) = n^{-1} \sum_{i=1}^{n} \{Y_i - \hat{\theta}_{h,-i}\}^2.$$

The expectation of this is given by:

$$E[CV(h)] = n^{-1} \sum_{i=1}^{n} E[\{\hat{\theta}_{h,-i} - \theta(X_i)\}^2] + n^{-1} \sum_{i=1}^{n} \sigma^2(X_i),$$

where $\sigma^2(X_i)$ is the conditional variance of Y given $X = x_i$. The first term is a discrete approximation to the IMSE, and the second term is independent of h, so we can see that minimising $CV(h)$ with respect to h will minimise the IMSE.

However, as already discussed, our local rainfall models have a multivariate response, which is the set of parameters of the underlying point process model. This can be addressed by taking as our CV statistic (assuming a diagonal weights matrix):

$$CV(h) = n^{-1} \sum_{t=1}^{n} \sum_{i=1}^{k} [T_i(Y_t) - \tau_i(\hat{\theta}_{h,-t})]^2 w_{t_i},$$

where the w_{t_i} are the k diagonal components of the matrix W_{x_t}. The expectation of this is given by:

$$E[CV(h)] = n^{-1} \sum_{t=1}^{n} \sum_{i=1}^{k} E\{[\tau_i(\hat{\theta}_{h,-t}) - \tau_i(\theta(X_t))]\}^2 w_{t_i} + n^{-1} \sum_{t=1}^{n} \sum_{i=1}^{k} \mathrm{Var}(T_i(Y_t)|X_t) w_{t_i}$$

and so minimising this cross-validation criterion would effectively minimise the integrated weighted sum of the mean squared errors of the components of τ. This is arguably preferable to consideration of the mean squared errors of the individual parameter estimates in any case, since practitioners are more concerned with the simulations that are subsequently

generated from the models, than with the values of the individual parameters. However, it is worth noting that, although there should be no problem with a well-identified model, this would potentially allow very variable parameters for some of the more complex variants. The cross-validation (CV) approach is appealing as it is straightforward to apply, and does not require estimation of any additional parameters. The computation time for leave-one-out cross-validation is likely to be prohibitive, however. Also, while it is approximately unbiased for the true prediction error, it can have high variance as the n 'training' data sets are so similar to each other. A similar, but computationally cheaper approach is K-fold cross-validation, where the data set is randomly partitioned into K equally sized subsets. The same approach is taken as for leave-one out CV, but now we leave out the kth data set, and fit to the remaining $K-1$ sets, doing this for $k = 1, 2, ...K$ and summing the prediction errors of the K sets. If $K = n$, this is equivalent to leave-one-out CV. The lower the value of K, the lower the variance, and the quicker the calculation time, but potentially at the expense of some upward bias i.e. overestimation of the true prediction error (Hastie et al. 2001). Values of $K = 5$ or $K = 10$ are recommended as a good compromise. Hastie et al. (2001) note that in practice the model chosen is often that which is the most parsimonious of models whose prediction error is no more than one standard error above the error of the best model.

Another alternative, where calculation times are a constraint, is repeated random sub-sampling, which improves on the simple hold-out sample by averaging over several such random splits of the data. K-fold cross-validation is often preferred to repeated random sub-sampling, on the basis that in the former all observations are used for both training and validation, and all are used for validation only once, whereas in the latter there may be overlaps. However, it could be argued that its reliance on a single random permutation may result in an 'unfortunate' split, which is not representative. It also has the disadvantage that the choice of K specifies both the number of repetitions, and the size of the validation set. There are no such restrictions with repeated random sub-sampling, where there is flexibility over the training/validation split, and no limit to the number of repetitions. In the context of the rainfall models, this is our preferred approach. Hengartner & Wegkamp (2002) carry out a simulation study where such a method performs well against the Akaike information criterion (AIC) and Generalised Cross Validation (which is an approximation to leave-one-out CV which is generally faster, but only applicable where a solution can be expressed as a linear function of the response). The test/training split is derived by taking the size of the testing sample as n^{β} where n is the size of the training sample and β is in the range 0.8–0.95. The optimal bandwidth is calculated for each such split — a histogram or density plot of the results across all the splits then also provides useful insight. Either the mean or the median may be selected as the optimal bandwidth.

4.4.3 Other approaches

Other approaches aim to minimise the IMSE using alternative methods to estimate the variance and bias at the observation points over an appropriate grid of bandwidths, from which the optimal choice can be made. These may involve jacknife or bootstrap estimation, although computation time is likely to be a serious constraint, given the requirement for a large number of bootstrap samples at each required covariate point at each potential bandwidth. There are also empirical approaches which rely on the estimation of the variance and bias using 'quasi-asymptotic' methods. For example, the asymptotic variance can be estimated by the sandwich method (see Equation 4.25). Ruppert (1997) also proposes an empirical method of bias estimation, where the bias is approximated by a polynomial function $f(h, \gamma)$, the parameters of which are found using least-squares estimation. This idea is based on asymptotic theory which shows that the asymptotic bias has an expansion in powers of h, beginning with power $p + 1$. Carroll et al. (1998) deal with a multivariate response by minimising the mean squared error of a scalar function of all the responses. The variance of this derived scalar is estimated using the delta method, and the bias is estimated directly using the empirical bias approach.

4.4.4 Suggested approach

Key constraints in the context of our model include computation time and the identifiability of parameters, as already discussed. Also, as mentioned, we have a multivariate response. It could be argued that each element of the response (i.e. each of the rainfall model parameters) should have a different bandwidth, since the relationships between the parameters and the covariates may differ in terms of their relative complexity. However, it is not clear how such an approach could be implemented in the local GMM framework, and in any case it is unlikely to be viable from a practical perspective.

It is often said that plug-in methods perform better than cross-validation (Gasser et al. 1991, Ruppert et al. 1995), which is criticised for variability and a tendency to lead to under-smoothing, with the main advantage of cross-validation cited as its universal applicability. However, Loader (1999) challenges this opinion, finding that plug-in methods are highly dependent on the pilot bandwidth, and are prone to over-smoothing complex functions. He argues that the variability of cross-validation is not a problem, but a symptom of the difficulty of bandwidth selection, where the selector has to make a decision purely from the data. Where performance of CV is poor, this may be caused by other problems, and might be better addressed in different ways, for example by allowing the bandwidth to vary, or by the use of robust techniques. Blind reliance on any automatic method of bandwidth selection should in any case be avoided.

In the context of the point process-based rainfall models, we need a pragmatic approach.

We do not expect the relationships to be unduly complex, and are limited by issues of computation time and numerical stability of solutions. We therefore look for a method that is straightforward to apply, but aim to use it in conjunction with a subjective assessment, as recommended by Loader (1999).

We have discounted plug-in methods due to the difficulties of identification of the large number of required parameters. Either the empirical estimation methods or cross-validation are feasible, although on balance, a cross-validation-type approach is preferred. Due to computational time constraints, rather than leave-one-out cross-validation, we have decided to use repeated random sub-sampling, which involves randomly splitting the data into test and training sets a number of times. We have based the proportions on the suggestions of Hengartner & Wegkamp (2002) (described in Section 4.4.2). For each such split the model is fitted to the test data points using just the training observations, and the bandwidth is identified that gives the lowest prediction error over the test data. The prediction error is taken as the mean weighted sum of squared residuals i.e. as:

$$n_{ts}^{-1} \sum_{t=1}^{n_{ts}} \sum_{i=1}^{k} [T_i(Y_t) - \tau_i(\hat{\theta}_{t_{h,tr}})]^2 w_{t_i} \tag{4.64}$$

where the summation is over observations in the test set (denoted ts), with $\hat{\theta}$ based on observations in the training set (denoted tr), and with the weights based on the smoothed sample variances of the statistics as before. This method is appealing not only in terms of computation time, but also because it gives some insight into the variability of the optimal bandwidth across samples. Further details are given in Section 6.3 which describes practical results.

4.5 Multiple covariates

So far we have assumed a univariate covariate. In theory at least, generalisation to multi-dimesional X is straightforward. Examples of practical application (using loess) are given by Cleveland & Devlin (1988), while the theoretical properties, including the asymptotic bias and variance of local linear regression are discussed by Ruppert & Wand (1994).

The local mean model can be expressed as before, but now X is a d-dimensional vector, i.e. $X_t = (X_{t1}, ...X_{td})^{\mathrm{T}}$. In order to define neighbourhoods in d dimensions, we need a d-dimensional kernel function, K, and a $d \times d$ symmetric positive definite smoothing matrix, H. As for the univariate case, K is assumed to be a probability density function that integrates to 1, and has zero mean and other odd-order moments. We assume also that:

$$\int z_i z_j K(z) dz = \delta_{ij}\, \mu_2(K), \tag{4.65}$$

where δ_{ij} is the Kronecker delta, and $\mu_2(K)$ is a scalar greater than 0. We define:

$$K_H(z) = \frac{1}{|H|^{1/2}} K\big(H^{-1/2}\, z\big), \tag{4.66}$$

where z is a d-dimensional vector. Taking K as the $N(0, I_d)$ density, for example, would give:

$$K_H(X_t - x_0) = \frac{1}{(2\pi)^{d/2}|H|^{1/2}} \exp\left(-\tfrac{1}{2}(X_t - x_0)^{\mathrm{T}} H^{-1}(X_t - x_0)\right), \tag{4.67}$$

which is the multivariate normal density with variance matrix H, and mean x_0. The bandwidth matrix, H controls both the size and the direction of smoothing.

Wand & Jones (1995) describe various levels of sophistication when specifying the bandwidth matrix H. The simplest approach is to take $H = h^2 I_d$, where I_d is the d-dimensional identity matrix. (Note that we have h^2 here because we have defined H to be the covariance matrix of the kernel density. Some authors define $K_H(z) = \frac{1}{|H|} K\big(H^{-1}\, z\big)$ instead, in which case here we would have $H = hI_d$.) This means that the same amount of smoothing is carried out along each of the coordinate axes of the covariates (so in the bivariate case, the contour plots of the kernel are circles). For this to be sensible, it would be necessary to standardise the covariates first, so that they are on the same scale. Using this approach (which is taken by Cleveland & Devlin (1988) for their loess procedure), we would have:

$$K_H(X_t - x_0) = h^{-d} K\left(\Big[\sum_{j=1}^{d} (X_{tj} - x_{0j})^2\Big]^{\frac{1}{2}} / h\right), \tag{4.68}$$

where K is a one-dimensional kernel function, and we use Euclidean distance.

A more flexible, but still straightforward, approach is to take H to be diagonal, but allow the diagonal elements to be different, so that we can apply different degrees of smoothing to the different covariates. The kernel contours are now ellipses with axes corresponding to the coordinate directions. We can then write $K_H(X_t - x_0)$ as a product of the univariate kernels (i.e. we use a 'product' kernel):

$$K_H(X_t - x_0) = K_{h_1}(X_{t1} - x_{01})\, K_{h_2}(X_{t2} - x_{02}) \ldots K_{h_d}(X_{td} - x_{0d}). \tag{4.69}$$

The most flexible option is to specify a full bandwidth matrix, which allows smoothing in directions different to those of the coordinate axes. This requires $d(d-1)$ additional smoothing parameters. Wand & Jones (1995) acknowledge that in certain circumstances such a choice may be warranted, but suggest that the diagonal option is generally sufficient. The three options in respect of a bivariate covariate are illustrated in Figure 4.3.

To find an optimal bandwidth matrix, cross-validation techniques could be used. Alternatively one could develop multivariate versions of the plug-in methods (see for example

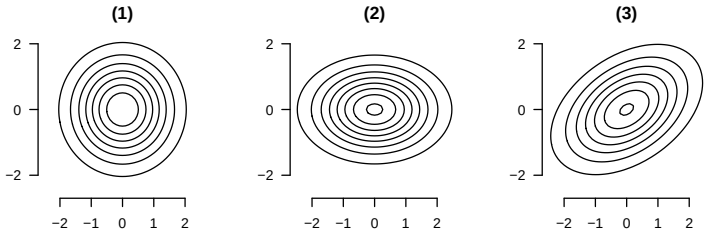


Figure 4.3: Contours of a Gaussian Kernel illustrating different choices of bandwidth matrix for a bivariate covariate: (1) $H = h^2 I_2$; (2) H diagonal; (3) fully flexible H.

Fan & Gijbels (1996) for a brief description of a possible method).

The main problem with multiple polynomial regression is what has been termed the 'curse of dimensionality' (Bellman 1961) which simply means that as the dimension of the covariate vector increases, so the data become increasingly sparse. This means that there are either very few points in the local neighborhood, or the neighborhood ceases to be very 'local". You therefore either need to increase the size of your data exponentially with the dimension d, to maintain a given degree of accuracy, or, for the same sized data set, suffer highly increased uncertainty in your estimates. Hansen (2012) refers to the curse of dimensionality as "the phenomenon that the rate of convergence of nonparametric estimation decreases as the dimension increases". Assuming a common bandwidth, h, for all dimensions, then the variance in the multivariate case (with d-dimensional covariate) will be of order $1/(nh^d)$, with bias of order h^2, as for the single covariate case. Thus the mean square optimal value of h is $O\left(n^{-1/(4+d)}\right)$, and the estimator converges at rate $n^{-2/(4+d)}$.

The dimension therefore either needs to be kept appropriately low, or some kind of constraints need to be introduced in the model, which effectively reduce the dimensionality. Thus, the fully flexible model, as defined above tends to be limited in practice to two or three variables.

In our practical investigation we consider up to three covariates for the local mean estimator, and assume a diagonal bandwidth matrix. The asymptotic expressions for the variance and bias could in principle be derived, but would include further unknowns, since we would now also need the first and second partial derivatives of $\tau(\theta)$ with respect to the vector X, as well as joint density functions. Rather than using asymptotic expressions, we will estimate the variance using the sandwich method and, as discussed, will use cross-validation type techniques in order to determine the optimal bandwidth. No additional theory is then required, and we simply replace the single Gaussian kernel in our equations with the product of individual kernels in respect of each required covariate, as in Equation (4.69). While this is conceptually very simple, there are practical limitations and even with a diagonal bandwidth matrix, selecting the optimal values in the multidimensional setting is computationally intensive. This is considered further in Section 6.4, where we take a pragmatic approach and impose some constraints on the possible bandwidths in or-

der to reduce the computational burden. Extension to the local linear case is also possible in theory, in which case θ would be given by $b_0 + b_{11}(X_{t1} - x_{01}) + \ldots b_{1d}(X_{td} - x_{0d})$.

4.6 Methodology in Practice

In the next two chapters we carry out a practical study, using the techniques and equations discussed in this chapter to fit local point process-based rainfall models to the rainfall time series from Bochum. We primarily focus on the local mean approach, which is considered to be more appropriate given constraints in respect of computational time and parameter identification. However, the impact of extending to a local linear estimator will also be examined. We will start with the simplest of the Bartlett-Lewis clustered point process-based models, the Bartlett-Lewis Rectangular Pulse model with a single-parameter distribution for the rainfall intensity, so that θ has five components (i.e. $q = 5$). The same set of fitting properties as in Chapter 3 will be used, so that $k = 13$. Key elements of our proposed methodology are summarised below.

4.6.1 Fitting methodology

The evaluation points (i.e. the values of x_0 at which we choose to solve the equation) will be taken as the observed covariate values, which will allow us to test the goodness of fit of the models. As for the global model, numerical optimisation is used to solve Equation (4.50) for the local mean estimator, or Equation (4.51) for the local linear estimator, and we solve for the logarithms of the rainfall parameters, as before. Now, however, we have to carry out a fit at each required evaluation point, which gives of the order of several hundred fits, compared with the 12 for the seasonal models. Recall that when fitting the monthly models we followed the approach of Wheater et al. (2005), using the optimisation routines developed for that project. Firstly, a set number of optimisations are carried out using the Nelder-Mead method, each starting with a different initial value for the set of parameters. The best parameter set is then used as a new starting value for a further set of optimisations, which now use a Newton-type algorithm.

For the local fits carried out here, this two-stage optimisation is carried out only once for a selected observation with a large number of neighbours. Subsequent fits are carried out in covariate order, stepping from the selected observation down to that with the smallest value of the covariate, then returning to the selected observation and stepping up to that with the highest value. Each fitted parameter set in turn is then taken as the initial value for the next fit. Each of these subsequent fits is based on only a single Newton-type optimisation. This ensures both a quick calculation time, and a smooth curve, since the fits in respect of neighbouring covariates should be close.

The impact of this approach was tested for the discrete covariate, calendar month, where we fitted a model for January using the two stage method, and then the other months in turn using just a single optimisation. Results were then compared with those of Chapter 3. For the BLRP model, there were no significant differences found in the parameters for the other months. Computation time for January was of the order of 0.3 minutes, with a total computation time for all 12 months of 0.4 minutes (including calculation of standard errors), giving a reduction in computation time of around 90% using our method, compared with using the two-stage optimisation for every month. Time savings will be even more substantial for the continuous covariates. In the case of multivariate predictors, we take a pragmatic approach and sort by an arbitrarily chosen covariate.

4.6.2 The weighting matrix

As discussed in Section 4.2.6, we will use a single step GMM fit, with a diagonal matrix of weights of inverse conditional variances. The practical approach we take here involves first binning the observations into 12 equally sized groups, based on the value of the covariate. The sample variance is then calculated for each bin, and this is treated as the variance conditional on the covariate value given by the midpoint of the bin. Variances conditional on other values of the covariate are then derived from these 12 midpoints using a Nadaraya-Watson estimator. It was found that the fits were not overly sensitive to the bandwidth used in this smoothing, which was thus selected subjectively. In the case of multiple covariates, a single covariate is picked for the calculation of the weights, rather than grouping observations into bins based on all the covariates. Altering the choice of covariate was not found to have a significant difference on the fitted parameter sets.

4.6.3 Estimation of the variance of the estimators

The variance of the estimators is calculated using the sandwich approach of Equation (4.25), and will be used to determine approximate variability bands. A difference from the approach used in Chapter 3 is that in the calculation of the conditional covariance matrix of the statistics, $\mathrm{Var}[T|X]$, the conditional expected values of the statistics will be estimated by their fitted values, $\tau(\theta(X))$, rather than by sample means, since the latter are not available for a continuous covariate. An alternative approach would be to estimate the variances by the smoothed sample values used for the weights matrix.

We now move on to a practical study, where we will consider the application of the theory discussed in this chapter to our Bochum database.

Chapter 5

Choice of suitable covariates

In this chapter we identify a suitable set of covariates for further investigation in the practical application of Chapter 6. First we consider selected studies in the existing literature, in order to identify a range of potential predictors of interest. Note, though, that there is a key difference between our statistical downscaling methodology and the majority of applications in the literature. The latter generally only consider rainfall occurrence or rainfall amounts as predictands, whereas we are also interested in identifying predictors that explain the variability of other rainfall properties, such as autocorrelations, and skewness. In the second part of this chapter, therefore, we look more specifically at relationships between various potential predictors and the rainfall properties of Bochum.

Within this thesis, we only consider historical observed data (or surrogate observed data from reanalysis initiatives). This allows the methodology to be developed, and relationships between the fitted parameters of the clustered point process-based rainfall models, and components of the climate system to be examined. In order for the methodology to be of practical use, however, it is important that any predictors selected will be well represented by climate models. As discussed in Section 4.1, the methodology is based on monthly time intervals, so the predictors should be aggregated to this timescale. Monthly mean values of the sort of large scale atmospheric variables that we consider here are expected to meet this requirement. There are numerous issues in sourcing suitable future predictor values, however. For example, it is now generally accepted that output from an 'ensemble' of climate models is required, rather than from a single model, and various techniques exist for combining such outputs and allowing for the uncertainty within them. These include Bayesian hierarchical models, as well as much less sophisticated methods, such as simple weighting schemes (see Knutti et al. (2010) for a discussion of the issues, and Chandler (2011) for a recent example of the Bayesian approach, which also includes a simple 'poor man's version'). Biases in the climate model outputs are another problem, and it is common to work with standardised, rather than absolute, values of the predictors to allow for this. Standardisation involves subtraction of the mean and division by the

standard deviation of the predictor for a baseline period, which is often taken as 1961–1990. Averaging may be carried out separately for each calendar month, if seasonality is to be allowed for explicitly. There may also be differences between the grid resolution of historical and model data, so that data need to be re-gridded. These considerations are common to all statistical downscaling techniques, and constitute an active area of research. Wilby et al. (2004) provide some useful guidelines, which cover standardisation and re-gridding. For the purposes of developing and testing the methodology in this thesis, we use absolute values of the predictors rather than standardised values, for ease of interpretation. We do not allow for seasonality explicitly, but assume it is implicitly included within the predictors (which importantly allows seasonal behaviour to change under the impact of climate change).

For our modelling methodology to be successful, there must exist reasonably strong relationships between at least some of our fitting properties and the large-scale atmospheric predictors. An implicit assumption is that these observed, empirical relationships remain valid under future climate conditions i.e. that the relationships are stationary. Again, this assumption is common to statistical downscaling techniques in general. Although it is impossible to ascertain in respect of future climate conditions that are outside the historical observed experience, this does not seem unreasonable.

5.1 Variables used in the existing literature on rainfall downscaling

Numerous studies in the existing literature covering different regions, and taking different approaches to downscaling, have identified a range of suitable predictors, with no clear consensus as to which are optimal. In part this clearly depends on the exact definition, spatial resolution and timescale of the predictand e.g. rainfall occurrence or amounts, at a point location, or averaged over a grid cell, monthly or daily, etc. The strong correlations between many of the atmospheric variables themselves is another factor, so that essentially the same information can be obtained from different combinations of variables.

Consideration of the precipitation process, in order to identify physically meaningful predictors, is a natural starting point. Obvious candidates are the atmospheric circulation variables (such as sea-level pressure, geopotential heights, zonal and meridional wind velocities etc.), as well as temperature and humidity-related variables. Availability of quality data over a sufficiently long timescale is clearly a requirement, and many recent studies have used NCEP (National Centre for Environmental Prediction) reanalysis data (Kalnay et al. 1996, Kistler et al. 2001), which uses an analysis/forecast system to perform data assimilation using historical data from 1948 to the present. This includes an extensive set of weather variables with 4-times daily, daily and monthly values at a range of heights in

the atmosphere on a 2.5° × 2.5° grid covering the whole globe.

Most commonly, predictor variables are selected from the grid overlying the target location, but it has been shown that often this is not optimal (Wilby & Wigley 2000), and the optimal location and spatial extent of each predictor should also be considered.

5.1.1 Studies with predictors at a daily level

A very comprehensive study of predictor variables has been carried out by Cavazos & Hewitson (2005), who examined the performance of twenty-nine potential atmospheric variables from the NCEP reanalysis dataset as predictors of grid cell area-averaged daily precipitation in fifteen locations using artificial neural networks. The results were then tested also at individual stations and at regional scales. Not surprisingly, the performance was found to be poorer at the local scale due to the importance of elements such as convection, local topography and sea breezes. However, there was generally good agreement between modelled synoptic events at local and regional scales. The potential predictors included sea-level pressure plus the following seven variables at different pressure level in the atmosphere — zonal (west-east) and meridional (north-south) wind components, divergence, vorticity, geopotential height, specific and relative humidity, as well as the atmospheric thickness of two pressure bands. The thickness of an atmospheric layer is effectively a measure of temperature, since the height of the troposphere fluctuates as a result of the different heating of the Earth's surface — the stronger the heating the thicker the tropospheric layer. Geopotential height is the height necessary to reach a given pressure level, and this gives similar information to surface pressure charts, but for the upper air. Vorticity is a measure of atmospheric rotation, with positive values corresponding to cyclonic flow (low pressure), and negative values corresponding to anticyclonic flow (high pressure). Divergence measures the horizontal outflow of air from a region. Divergence at the upper levels of the atmosphere is associated with low pressure systems and storms. Cavazos & Hewitson (2005) found that mid-tropospheric geopotential heights (at 500 and 700 hPa) and mid-tropospheric specific humidity (at 700 hPa) were the most important predictors at all the locations and both the seasons analysed (Dec–Feb and Jun–Aug). The tropospheric thickness (500–1000 hPa) and surface and mid-tropospheric meridional wind components were also important, but they were regionally and seasonally dependent. Generally performance was better in the winter than the summer, and for mid-latitude locations rather than the tropics or equatorial regions. This is likely to be due at least in part to the greater difficulties of predicting convective processes, particularly given the spatial resolution at which predictors are available.

An earlier extensive study carried out by Wilby & Wigley (2000) looked at fifteen potential predictor variables for six US regions. The predictands were daily precipitation occurrence, and wet-day amounts. They also found that in general the highest explained

variances for observed data are in winter and that they tend to be higher for rainfall occurrence, than for wet-day amounts. The largest percentages of variance in observed rainfall occurrence are explained by: specific humidity, maximum (summer) or minimum (winter) temperature, mean sea level pressure, zonal velocity component, 500hPa geopotential height and divergence at 500hPa. For observed wet-day amounts, these were: specific humidity (winter only), maximum (summer) or minimum (winter) temperature, divergence, and 500hPa geopotential height.

5.1.2 Studies with predictors at a monthly level

At the monthly timescale, detail of individual weather systems is lost, but a monthly timescale has a number of advantages, as discussed in Section 4.1 and noted also by Kilsby et al. (1998). Kilsby et al. (1998)'s study into 67 sites across England and Wales, at the monthly timescale, was carried out before the development of the reanalysis data, and hence restricted by the availability of data. The mean daily rainfall amount and the proportion of dry days were originally regressed on four atmospheric circulation variables: mean sea-level pressure, zonal and meridional flow and vorticity, as well as a number of geographic variables. Temperature was not found to improve performance for this study, although the authors acknowledge its importance for future impact studies. Sea-level pressure was preferred as an alternative to vorticity.

Chandler et al. (2007b) used mean monthly sea-level pressure, relative humidity and temperature as predictors for daily rainfall occurrence and wet-day rainfall amounts. The motivation for using covariates at the monthly level was partly one of convenience, because the number of days in a climate model year does not match those in actual calendar years. However, the improved reliability of climate model outputs at the monthly timescale compared with daily was also noted as a positive factor. Further, the nature of the models (GLMs with covariates relating also to previous days' rainfall) were such that weather systems on smaller temporal scales would still be well represented by the simulations. This should also be the case for our approach — we are fixing parameters at a monthly level, but the continuous and stochastic nature of the model should ensure realistic rainfall simulations down to sub-daily timescales.

5.2 Selection of covariates for further research

We return now to our rainfall time-series from Bochum, Germany for a practical study. Recall that we have 5-minute rainfall observations from 1931 to 1999. We fit the BLRP model to 13 selected statistics as before: mean hourly rainfall, plus the coefficient of variation, skewness and lag-1 auto-correlation at timescales of 5 minutes, and 1, 6 and 24 hours.

We have decided to focus on the NCEP reanalysis data, given their accessibility and quality. These are available from January 1948, giving a total of 624 monthly observations (52 years) when combined with the time span over which we hold rainfall data. We start by selecting a number of variables based on the studies above, taking the time series of their monthly mean values. The correlations between each of these and our statistical properties are considered to get an idea of their relevance. We found that these correlations were not necessarily strongest for the grid point nearest to Bochum's location. However, the optimal point varied with the predictor and property considered, so for simplicity a single location was selected for all predictors, based on an overall impression of these results. This is the grid point with latitude 52.5N and longitude 7.5E. The variables considered included:

- sea-level pressure (slp),
- geo-potential heights at 200, 500 and 700 hPa (geo200, geo500, geo700)
- surface temperature (temp),
- thickness of the atmosphere between 500hPa and 1000hPa (thick),
- relative humidity at the surface, and at 700hPa (rhum, rhum700),
- specific humidity at the surface, and at 700hPa (shum, shum700),
- zonal (west-east) and meridional (north-south) wind components (uwind, vwind).

In addition to monthly mean values, we include monthly minima and maxima of temperature. These have been derived from NCEP series of daily minima and maxima, based on six-hour hindcasts at 2m. We also consider the North Atlantic Oscillation (NAO Index), which tracks the normalised pressure difference between the permanent low in Iceland and the permanent high in the Azores. These pressure differences, which vary over time, control the strength and direction of westerly winds and storm tracks across the North Atlantic. A large difference leads to increased westerlies and generally higher rainfall in Central Europe, with the strongest impact on the weather in the winter months. There are various versions of the index, with slightly different definitions and different southern stations. The index used here is the monthly index based on stations in Gibraltar and Iceland, which has been extended back to 1821 (Jones & Thornton 1997). We consider both the overall correlation, and that taking only the winter months — December, January and February. Table 5.1 shows the correlations between these potential predictor variables and selected Bochum statistics. Correlations between the potential predictors themselves are given in Table 5.2.

The choice of which covariates to include for further investigation is influenced by the following points:

	mean	coeff of var		skewness		ac lag 1	
		5 min	6 hour	5 min	6 hour	5 min	6 hour
slp	-0.538	0.304	0.526	0.121	0.385	0.046	-0.013
geo200	0.030	0.608	0.414	0.525	0.364	-0.470	-0.305
geo500	-0.096	0.629	0.499	0.512	0.420	-0.422	-0.280
geo700	-0.195	0.627	0.550	0.485	0.451	-0.372	-0.258
temp (mean)	0.170	0.577	0.275	0.546	0.265	-0.542	-0.344
temp (min)	0.217	0.493	0.196	0.488	0.208	-0.512	-0.318
temp (max)	0.102	0.579	0.307	0.539	0.276	-0.530	-0.339
thick	0.085	0.592	0.370	0.527	0.331	-0.484	-0.307
rhum	0.140	-0.586	-0.417	-0.465	-0.318	0.413	0.240
rhum700	0.169	-0.580	-0.434	-0.448	-0.328	0.385	0.229
shum	0.243	0.538	0.224	0.528	0.238	-0.534	-0.343
shum700	0.289	0.479	0.195	0.467	0.217	-0.478	-0.295
uwind	0.374	-0.204	-0.384	-0.092	-0.275	-0.068	-0.022
vwind	0.213	-0.356	-0.368	-0.232	-0.286	0.124	0.084
nao	0.069	-0.045	-0.094	-0.030	-0.087	-0.036	-0.012
nao(winter)	0.372	-0.182	-0.235	0.070	-0.207	-0.164	0.016

Table 5.1: Correlations between selected potential covariates and selected Bochum rainfall statistics.

- For the purposes of this investigation, we narrow down the selection of variables early, leaving only a small number for further investigation. This is because of the twin constraints of calculation times and the 'curse of dimensionality'. For computationally simpler models, and where more data are available, such an initial filter may not be required.

- Sea-level pressure is the variable with the strongest (linear) relationship with mean rainfall, which is arguably the most important property, so this is a strong contender.

- All three predictors relating to temperature have a very low correlation with sea-level pressure and fairly strong relationships with all the other properties (coefficient of variation, skewness and lag-1 autocorrelation), particularly at the shorter durations. Temperature is also intuitively appealing, given its significance in future climate change scenarios. There is little to choose between the three options, which are very highly correlated with each other, as expected, and we have selected the mean monthly series.

- Geo-potential heights at the lower pressure levels are very highly correlated with temperature, whereas at higher pressure levels, they are highly correlated with sea-level pressure. Therefore, they are not expected to contribute much further useful information, given the inclusion of these two variables. Thickness is also highly correlated with temperature, and therefore excluded.

- Although the wind components generally exhibit lower correlations with the properties of interest than some of the other variables, they do benefit from low correlation with the other potential predictor variables and are therefore likely to add useful

	slp	geo200	geo500	geo700	temp (mean)	temp (min)	temp (max)	thick	rhum	rhum700	shum	shum700	uwind	vwind	nao
slp	1.0	0.30	0.47	0.62	0.04	-0.02	0.07	0.17	-0.24	-0.30	-0.02	-0.08	-0.12	-0.34	0.16
geo200	0.30	1	0.98	0.92	0.93	0.85	0.87	0.98	-0.61	-0.59	0.91	0.88	0.00	-0.26	0.05
geo500	0.47	0.98	1	0.98	0.87	0.77	0.81	0.95	-0.61	-0.61	0.83	0.80	-0.05	-0.30	0.08
geo700	0.62	0.92	0.98	1	0.79	0.69	0.75	0.87	-0.61	-0.62	0.74	0.69	-0.08	-0.35	0.11
temp (mean)	0.04	0.93	0.87	0.79	1	0.93	0.95	0.95	-0.70	-0.66	0.97	0.91	0.03	-0.30	0.02
temp (min)	-0.02	0.85	0.77	0.69	0.93	1	0.88	0.87	-0.61	-0.56	0.93	0.87	0.06	-0.27	0.04
temp (max)	0.07	0.87	0.81	0.75	0.95	0.88	1	0.88	-0.77	-0.74	0.91	0.84	-0.06	-0.40	-0.03
thick	0.17	0.98	0.95	0.87	0.95	0.87	0.88	1	-0.59	-0.56	0.94	0.92	-0.01	-0.20	0.03
rhum	-0.24	-0.61	-0.61	-0.61	-0.70	-0.61	- 0.77	-0.59	1	0.99	-0.56	-0.46	0.21	0.61	0.11
rhum700	-0.30	-0.59	-0.61	-0.62	-0.66	-0.56	-0.74	-0.56	0.99	1	-0.53	-0.42	0.22	0.61	0.10
shum	-0.02	0.91	0.83	0.74	0.97	0.93	0.91	0.94	-0.56	-0.53	1	0.95	0.04	-0.22	0.01
shum700	-0.08	0.88	0.80	0.69	0.91	0.87	0.84	0.92	-0.46	-0.42	0.95	1	0.01	-0.16	-0.03
uwind	-0.12	0.00	-0.05	-0.08	0.03	0.06	-0.06	-0.01	0.21	0.22	0.04	0.01	1	0.39	0.47
vwind	-0.34	-0.26	-0.30	-0.35	-0.30	-0.27	-0.4	-0.20	0.61	0.61	-0.22	-0.16	0.39	1	0.39
nao	0.16	0.05	0.08	0.11	0.02	0.04	-0.03	0.03	0.11	0.10	0.01	-0.03	0.47	0.39	1

Table 5.2: Correlations between potential predictor variables.

information. However, they are also likely to be less universally applicable i.e. they apply at local levels. We have decided to include the zonal component (uwind) only at this stage, due to its higher correlation with mean rainfall.

- On first investigation, relationships with the NAO index look weak. However, the correlation with mean rainfall in the winter is actually fairly high at 0.37, with lower correlations in the other seasons (negative in summer). As for the wind components (with which it shows moderate positive correlations), the relationship will be location-dependent, rather than having universal application. We have not investigated it further here.

- Relative humidity is intuitively appealing, with similar levels of correlation to temperature. Although the relationship with temperature is clearly strong (correlation of -0.7), this may still add useful information. There is little to choose between the surface and the 700hPa levels

- Finally, the specific humidity is included as a potential alternative to temperature. Here, we take the value at 700hPa, primarily because the surface level data are not available for the same grid point as our other data.

The following five potential covariates have been selected for further investigation: sea-level pressure, temperature, relative humidity, specific humidity at 700hPa, and the zonal wind component. All are monthly mean values.

5.2.1 Seasonality

As discussed at the start of this chapter, seasonality is not explicitly allowed for, and we assume that it is effectively captured by the covariates. This is a deliberate choice, since seasonal behaviour may change under the impact of climate change. In Figure 5.1 we examine the relationship between calendar month and selected covariates at Bochum, over the period of our data (1948–1999). It is clear that there is a very strong relationship between calendar month and temperature, which is expected to capture much of the seasonality effect. Strong relationships are also seen between calendar month and specific humidity, which is highly positively correlated with temperature, and between calendar month and relative humidity, which has a negative correlation with temperature. The median sea-level pressure is fairly constant throughout the year, but variability is much higher in the winter months. The zonal wind component also shows more variability in the winter. Variability of minimum temperatures is higher in the winter, whereas variability of maximum temperatures is higher in the summer.

A potential concern is that, while the seasonal behaviour in a future climate might change, there are nevertheless seasonal effects that will not be captured by temperature. For

example, while summer and winter are clearly differentiated by mean temperature, spring and autumn have similar temperature levels, but experience different weather. This is due to various factors, including, for example, differences in ocean temperatures between these seasons (since these lag behind the changes to surface temperatures). These differences should be captured, at least to some extent, by the other selected covariates. The zonal wind component can be seen to be higher in autumn then in spring (a mean of 1.5 m/s over September to November, compared with 0.7 m/s over March to May), in line with a greater frequency of Westerlies, and relative humidity also tends to be higher.

Another concern, raised at a presentation of this research, is that temperature is effectively being used here for two purposes, both to indicate seasonality, and anomalies (e.g. Indian summers etc.). A suggested solution to this issue is to represent seasonality by mean temperature averaged across all the grid squares at our selected latitude, and local relative differences from seasonal averages by the anomalies at Bochum. We have not pursued this idea here, but it may be of interest for future work, although the value of adding another covariate needs to be considered carefully.

A compromise solution might be to include seasonality explicitly, but rather than taking month as a discrete covariate, it could be smoothed using a fairly large bandwidth, so that broad seasonal effects could be allowed for, without using too many degrees of freedom. In order to allow for the periodic nature of month, the Gaussian kernel weights should be replaced by a periodic kernel function such as the von Mises distribution.

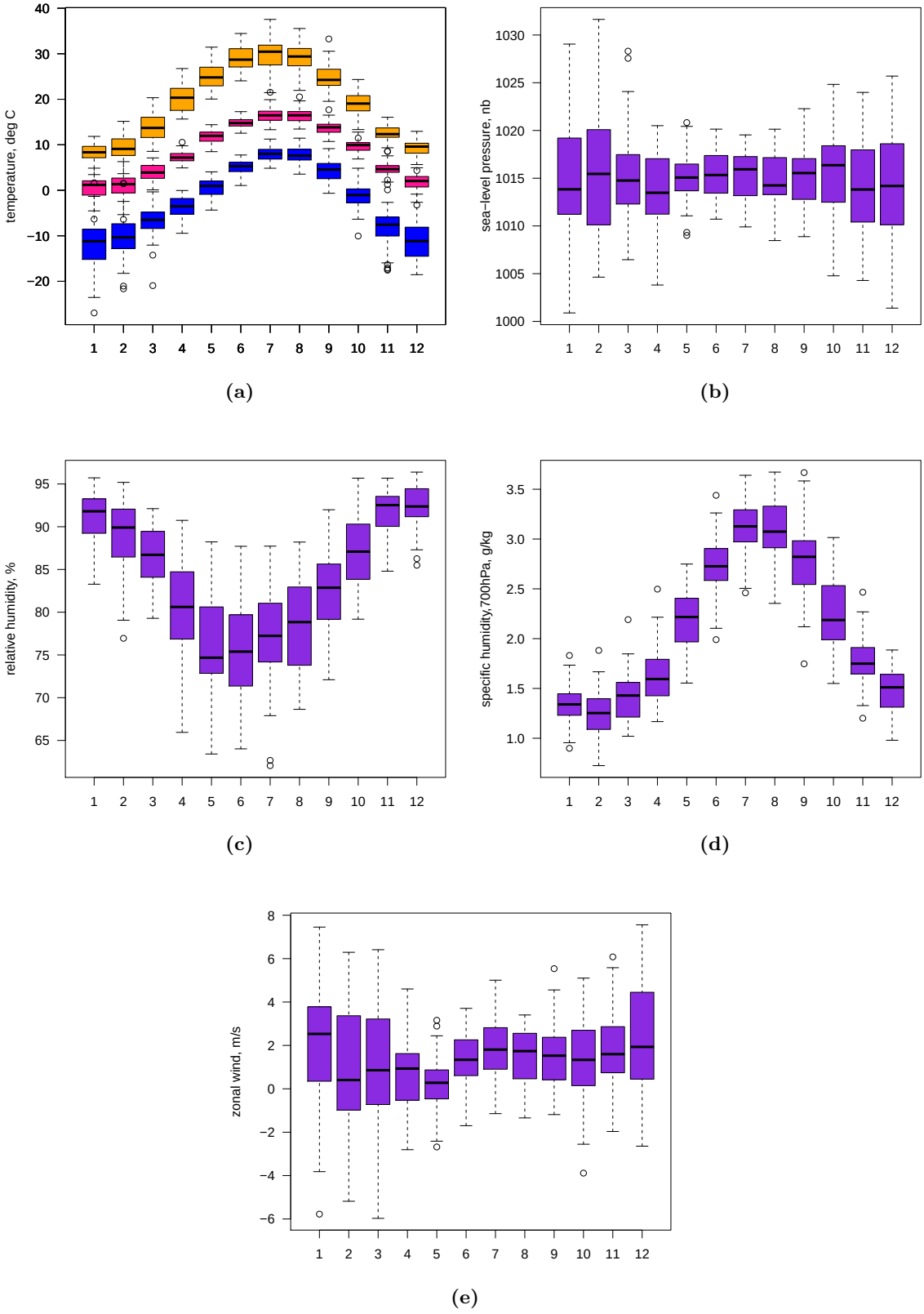


Figure 5.1: Relationship between calendar month and selected covariates. In (a), mean temperatures are shown in pink, minimum temperatures in blue, and maxima in orange.

Chapter 6

Practical results

In this chapter we apply the methodology of Chapter 4 to fit a local BLRP model (with an exponential intensity distribution) to the Bochum rainfall time series. This model is selected as the simplest of the clustered point process-based models. The key components of our study, which have been discussed in the earlier sections, are summarised below for convenience:

- The BLRP model, assuming a one-parameter rainfall intensity distribution, has five free parameters, so θ is a vector of length 5, with elements: λ, the storm arrival rate; γ, the storm termination rate; β, the cell arrival rate; η, the cell termination rate; μ_X, the mean rainfall intensity.
- We fit the model to the following 13 statistics as before: mean hourly rainfall, plus the coefficient of variation, skewness and lag-1 auto-correlation at timescales of 5 minutes, and 1, 6 and 24 hours.
- The required sample statistics are calculated for each observation month between January 1948 and December 1999, giving a total of 624 monthly observations (52 years). Potential continuous covariates investigated are monthly averages in respect of sea-level pressure (slp), temperature (temp), relative humidity (rhum), specific humidity at 700hPa (shum), and the zonal wind component (uwind).

Initially, we consider a single covariate in order to carry out some exploratory analysis, including examination of the impact of the bandwidth, and to examine the suitability of the methodology in practice. We focus on local mean estimation, but briefly consider local linear fitting in Section 6.2. Section 6.3 covers automatic bandwidth selection. Multiple covariates are introduced in Section 6.4, where we also consider model selection, in respect of the optimal choice of covariates. A small number of preferred models are analysed in more detail in Section 6.5. Finally, alternative datasets and more complex model variants (such as the BLRPR$_X$ model), are briefly considered in Sections 6.6 and 6.7 respectively.

6.1 Initial exploration using a single covariate

6.1.1 Impact of the bandwidth

At this stage we are primarily interested in the methodology rather than interpretation of the results, and we consider the impact of the bandwidth (which we take to be global, as discussed in Section 4.4), with just the single continuous covariate: temperature. When a very small bandwidth is used, fitted parameters (and more particularly, their variability bands) can cover a very wide range. For this reason, we plot the natural logarithm of parameters (indeed, this is what has been fitted). In Section 6.3 we choose the optimal bandwidth for all the univariate options, and plots with these bandwidths will show the actual (i.e. not logged) parameters, for ease of interpretation.

Figure 6.1 shows graphs of the fitted model parameters with temperature as the single covariate, using three different bandwidths. Fitted values have been calculated and plotted at all 624 observed data points. The low numbers of observations at the two ends of the range, which are particularly clear in the graphs with the smallest bandwidth, should be noted. In a straightforward local polynomial regression, as discussed in Section 4.4, the fit can be subjectively assessed by comparison with the scatterplot of observed points. Here, no such visualisation is possible in respect of the fitted parameters, given their indirect relationship with the covariates. However, it still seems intuitively clear that the smallest bandwidth shown of 0.5 is over-fitting with an unreasonably 'wiggly' curve, whereas the largest bandwidth shown of 5 is potentially smoothing out genuine features. It can also be seen that as the bandwidth increases, the curves get smoother and flatter, as expected. A visual comparison of fitted versus observed results *is* possible, if we consider the statistics, rather than the parameters, as shown in Figures 6.2 and 6.3. The former compares the fitted mean against the observed points with the three different bandwidths, whereas the latter shows the other twelve statistics, at the middle bandwidth of 1.5 only. These graphs broadly confirm the relationships that we saw in the correlations of Section 5.2, and a fit with a bandwidth of around 1.5 seems to be of the right order, although the sparsity of points at the boundaries is problematic.

6.1.2 Variance and bias

As discussed in Section 4.4, automatic bandwidth selection is generally based on some means of balancing the bias-variance trade-off. Here, an indication of variance can be given by calculating and plotting variability bands of $\pm$ 2 standard errors, as described in Section 4.2.5. Recall that these are not confidence intervals, but pointwise bands for the expected values of the parameters. Plots showing the fit for the parameter μ_X including variability

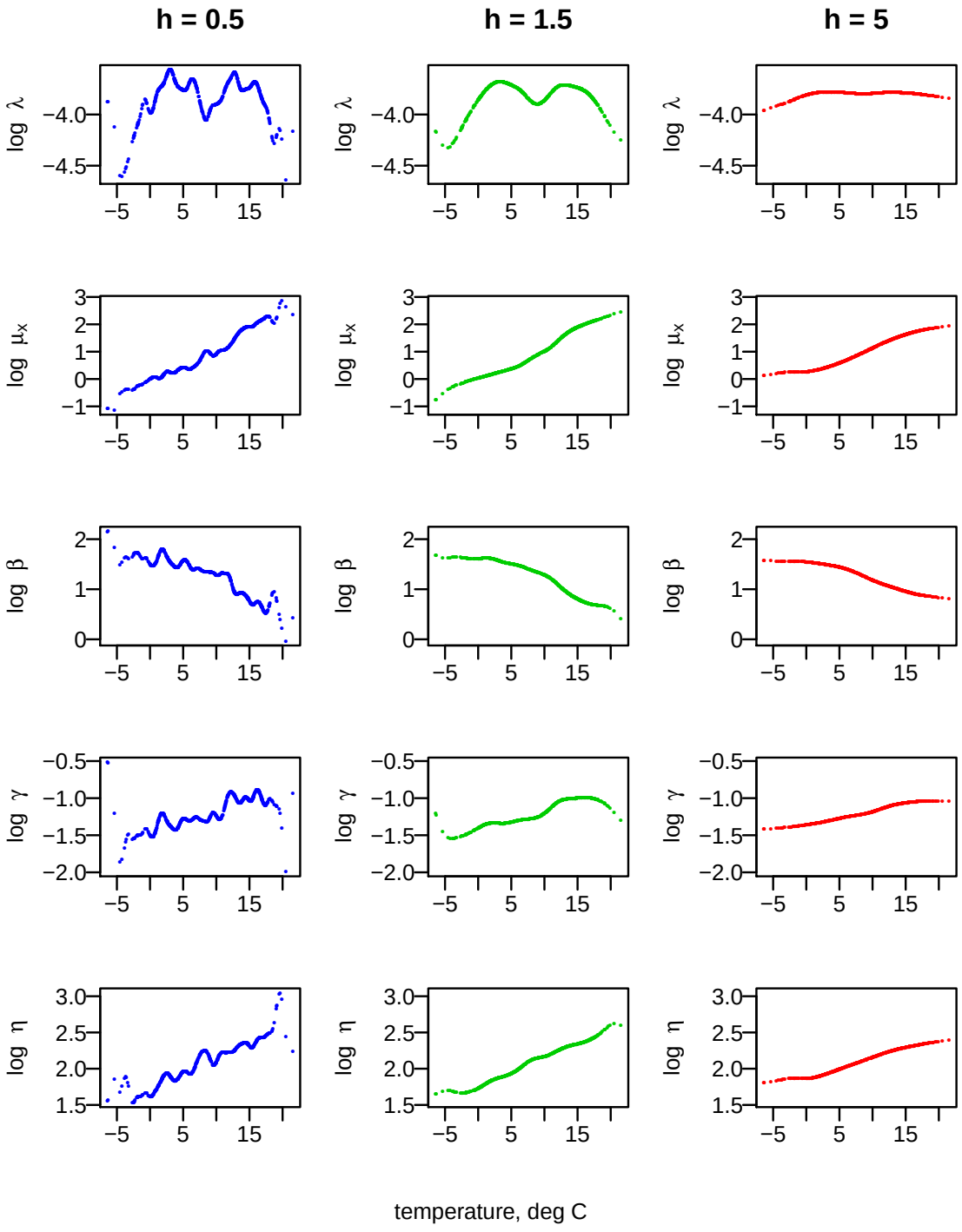


Figure 6.1: Fitted parameters with temperature as a covariate and three different bandwidths (denoted by h).

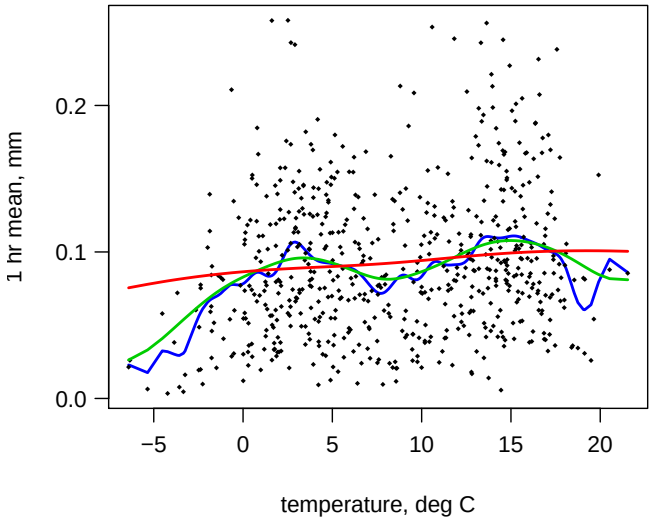


Figure 6.2: Fitted v observed mean with temperature as a covariate and three different bandwidths: h = 0.5 (blue), h = 1.5 (green), h = 5 (red).

bands for the three selected bandwidths are shown in Figure 6.4. Here we have taken the simple approach of Carroll et al. (1998), in which the calculation of the conditional variance of the statistics, T, is an integral part of the calculation of $\mathrm{Var}[G_n(\theta_0(x_0)]$, and hence is based on the same bandwidth as the fit (see Section 4.2.3 and Equation (4.27)). It is clear how the variability decreases as the bandwidth gets larger, and more observations are included in each local fit, particularly at the ends of the range. However, there is an undesirable feature apparent in these plots. With the smallest bandwidth, the estimate of μ_X at the highest value of temperature is rather surprisingly shown as having a very small variance — this is because very few points have been given non-negligible weight here, and the fit at this point itself is very good. The problem lies in the fact that the conditional variance of T should not depend on the selected value of the bandwidth, but in this estimate, it does. Potentially we could improve our calculations by taking the variance out of the summation and evaluating it separately at a fixed bandwidth, h^*, in line with the alternative approach of Equation (4.26), and shown again below:

$$\mathrm{Var}[T(Y)|\, x_0] \approx \frac{\sum_{t=1}^{n} K_{h^*}(X_t - x_0)[T(y_t) - \tau(\breve{\theta}(x_0))][T(y_t) - \tau(\breve{\theta}(x_0))]^{\mathrm{T}}}{\sum_{t=1}^{n} K_{h^*}(X_t - x_0)}. \tag{6.1}$$

where $\breve{\theta}$ denotes the parameter vector fitted with the bandwidth, h^*.

Another alternative would be to use the smoothed sample variance values that we calculated for the weights. Recall from Section 4.6.2 that these are based on a 'double-smoothing' approach, whereby we first group observations into 12 bins, by the covariate value, and then smooth the 12 sample variances over the individual observations, using a local mean approach. (We used a bandwidth of 2 in respect of temperature.) Figure

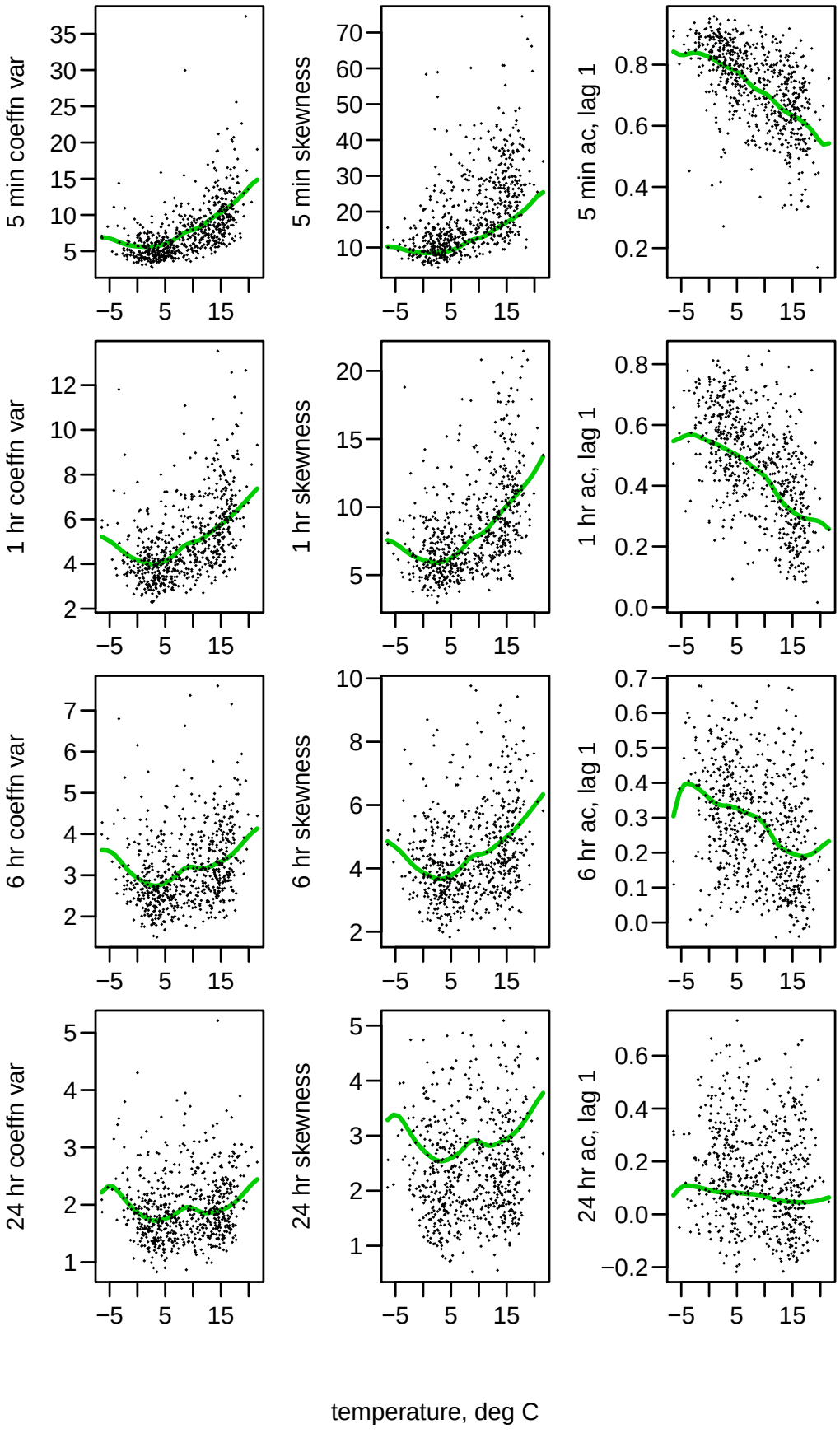


Figure 6.3: Fitted v observed coefficient of variation, skewness and lag-1 auto-correlation at selected timescales, with temperature as a covariate, h = 1.5.

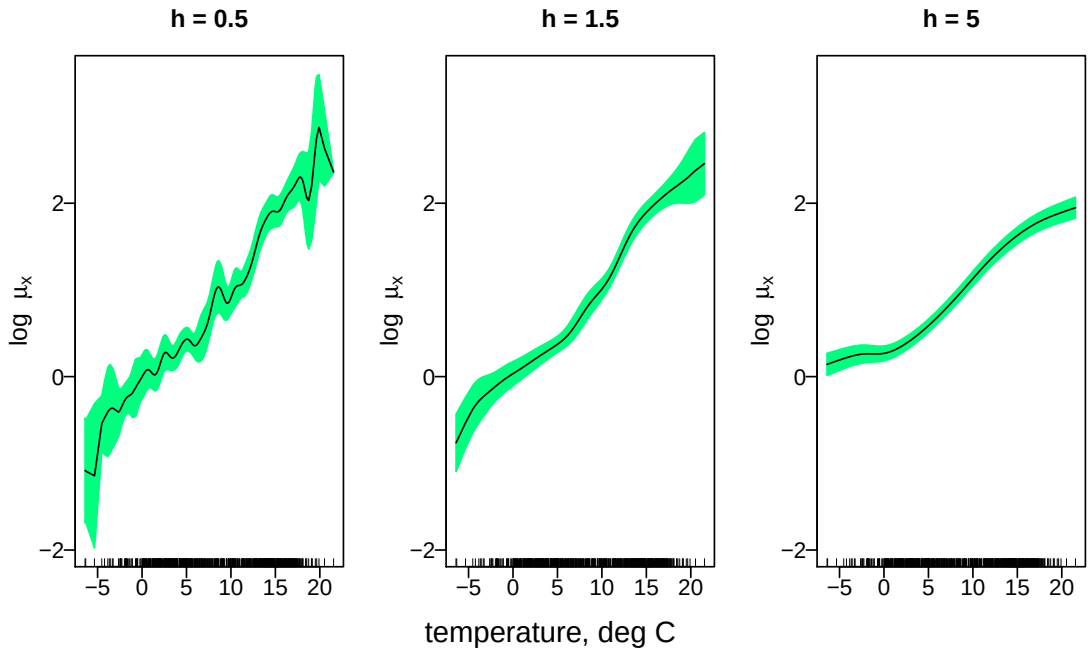


Figure 6.4: Variability bands for μ_X with three different bandwidths; the 'rug' plot along the x-axis indicates the density of the covariate, temperature.

6.5 shows estimated variances of each of the fitting properties at the one hour timescale. The graphs include three curves based on Equation (6.1) with the three bandwidths used in Figure 6.1, as well as the curve based on the smoothed sample variances used for the weighting matrices. Ultimately, we have chosen to use Equation (6.1), which is more convenient than the fully sample-based 'double-smoothing' approach in the case of multiple covariates. We set h^* to 2, which gives a fairly smooth variance curve, while being small enough such that the local variances are not unduly overestimated.

Using the variance estimate for $T|X$ given by Equation (6.1), we have recalculated the three variability bands of Figure 6.4, and these are shown in Figure 6.6 (using a bandwidth of 2). The main differences are at the extremes of the range, particularly at the smallest bandwidth. Further variability bands will be based on this alternative method, which is preferred.

Now we take a look at the bias using the approach described in Section 4.2.4 and Equation

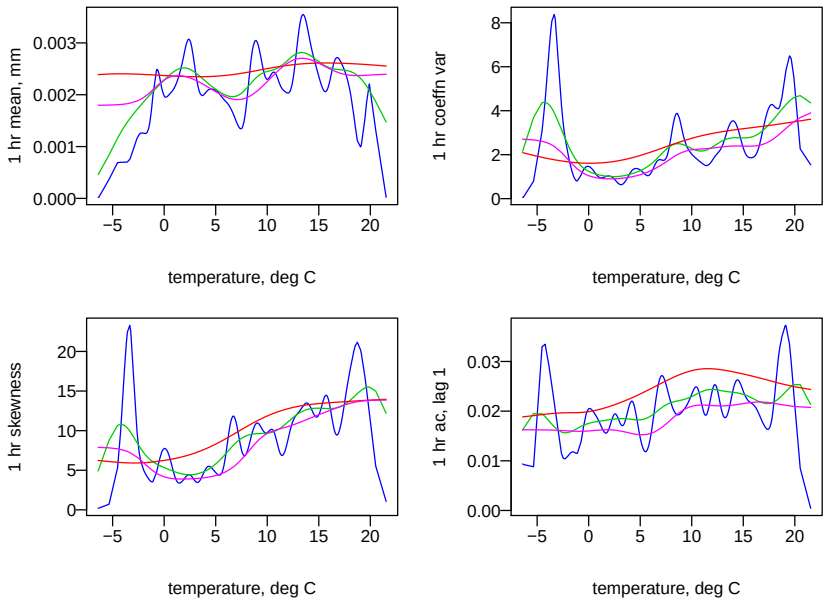


Figure 6.5: Estimated variance of the statistics, T, at the one hour timescale using three different bandwidths (0.5 blue, 1.5 green, 5 red) and Equation (4.26). The smoothed sample variance used for the weighting matrices is also shown in pink.

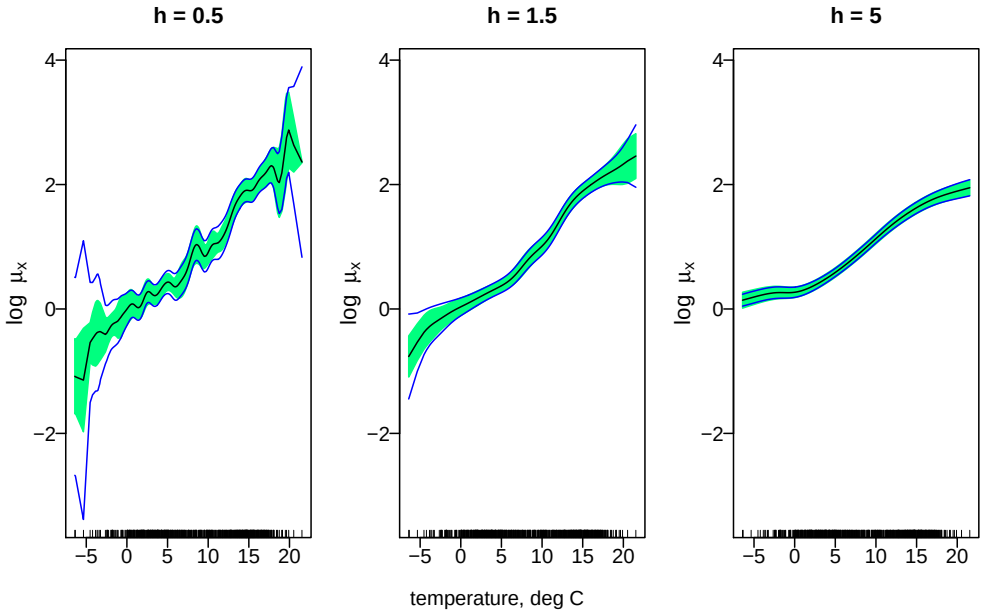


Figure 6.6: Revised variability bands for μ_X with three different bandwidths: the variance of $T|X$ is estimated as in Equation (4.26) and indicated by the blue lines; the original bands are shown in green.

(4.41), restated below:

$$\text{Bias}[\hat{\theta}(x_0)] \approx \left[\left[\frac{\partial\tau(\hat{\theta}(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_n(x_0) \frac{\partial\tau(\hat{\theta}(x_0))}{\partial\theta}\right]^{-1} \left[\frac{\partial\tau(\hat{\theta}(x_0))}{\partial\theta}\right]^{\mathrm{T}} W_n(x_0) \times \frac{\sum_{t=1}^{n} K_h(X_t - x_0)\ [\tau(\hat{\theta}(x_t)) - \tau(\hat{\theta}(x_0))]}{\sum_{t=1}^{n} K_h(X_t - x_0)}. \tag{6.2}$$

Figure 6.7 illustrates this calculation for two of the parameters, λ and μ_X, at the three bandwidths, where the original estimates have been plotted along with 'corrected' estimates from which the above bias estimate has been deducted. The estimated bias increases with the bandwidth, as expected, although is clearly understated at the highest bandwidth of 5. Attempting to estimate the bias with such a high level of h is clearly nonsensical (in the extreme the estimated curve will be a horizontal line, and the bias will be estimated as zero!). At the other bandwidths, the estimated bias reflects the findings from the asymptotic calculations: we have higher bias where the curve is steeper, in areas of high curvature and at the boundaries. Note that the reduction in bias of the bias-corrected estimator comes at the expense of an increase in variance (estimated at approximately 40% by Kauermann et al. (1998) in the context of local estimating equations). While of interest as part of the analysis, we will not be pursuing this undersmoothing approach further, but will instead try to derive the mean square optimal bandwidth, as described in the next section.

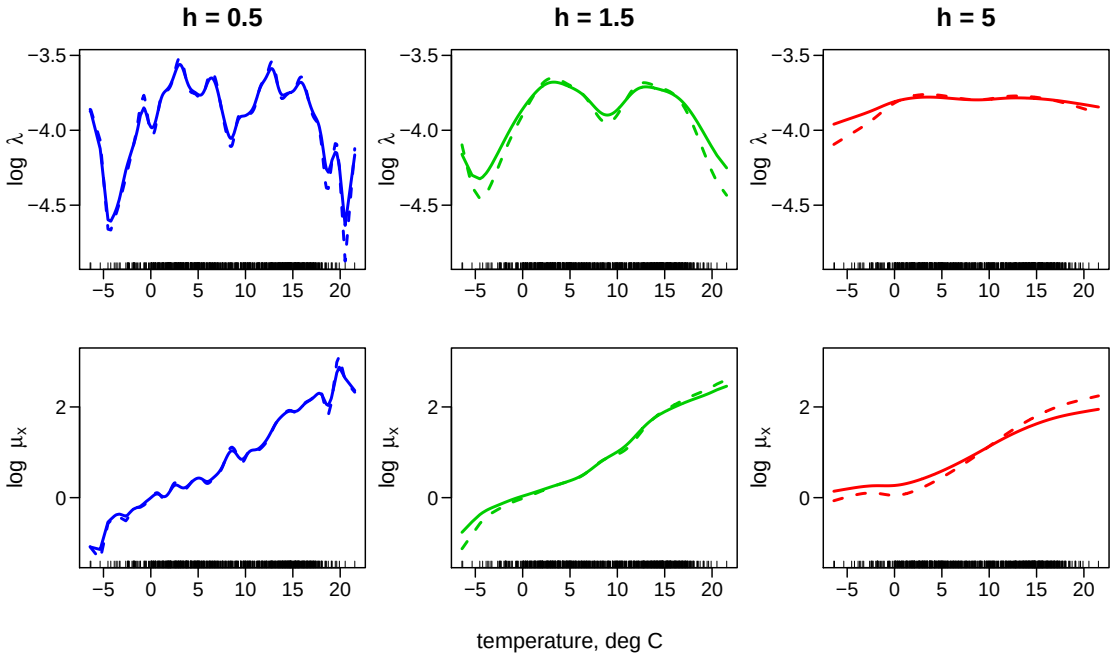


Figure 6.7: Bias-adjusted estimates for λ and μ_X with three different bandwidths: the original estimates are shown as continuous lines, the bias-adjusted estimates as dotted lines.

6.1.3 Fitting points and binning

So far, and indeed for the rest of our investigation, parameters have been estimated at the observation points, which permits an assessment of the goodness of fit. One could instead evaluate parameters at an equally spaced grid of covariate points, or indeed at any required set of points (e.g. the covariate values obtained from a climate model output), provided care is taken not to extrapolate too far beyond the range for which we have information.

For large data sets, 'binning' is sometimes used in combination with local polynomial regression, as a means of speeding up calculation times, and we consider the technique briefly here. Figure 6.8 illustrates the fitted parameters with temperature as a covariate, where we have binned the observations into equally spaced grids of 30 and 100 points. The resulting fits may be compared with those obtained using the original data. The technique which we have used here is referred to as 'linear binning' and is a refinement of simple binning. It involves splitting each observed point $(X_t, T(Y_t))$ into two fractional points assigned to the grid points on either side of X_t, with the fractions reflecting the relative nearness of the two points. The statistic T is then calculated at each grid point as a weighted average of all the allocations to that point, with the fractions as weights. Once the data have been binned, we proceed as normal with the fitting, except that we multiply each point in the kernel summation by the sum of the fractions allocated to that point. Binning is less useful in our situation than in local regression. This is because in the latter, the kernel evaluations are the most computationally expensive element and further time reductions are possible by recognising that, with an equally spaced grid, many of these are the same. For our (non-linear) model, however, it is the numerical optimisations that are computationally expensive, and the number of kernel evaluations is not in any case high. For the local mean case, calculation times are not prohibitive (a few minutes for all evaluation points). This approach also makes the degree of smoothing less transparent, since binning is itself a smoothing technique. Given the relative sparsity of points over parts of our range, binning is not recommended, and we will not consider it further here.

6.1.4 Residuals

Finally, as part of our initial exploration of local mean estimation with a single covariate, we consider the residuals. Our methodology assumes that the vectors of statistics $T(\cdot)$ are independent across months, conditional on the covariate. In order to check the validity of this assumption, we plot correlograms of the residuals in respect of a number of the statistics. Here we are again using a bandwidth of 1.5. The plots are given in Figure 6.9. It can be seen that most of the residual autocorrelations fall within the 95% confidence band around zero, demonstrating that our assumption of independence across months is

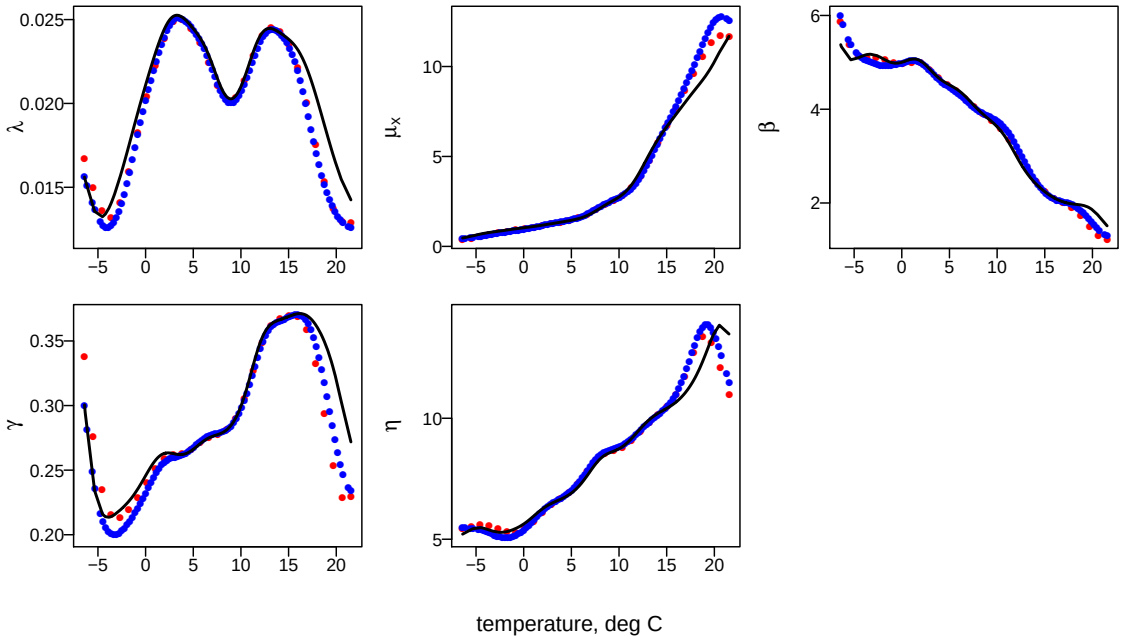


Figure 6.8: Parameter estimates with temperature as a covariate, using 'linear binning' approximation, with 30 points (red) and 100 points (blue). The original fit is indicated by the black line. The bandwidth is 1.5.

not unreasonable. Residuals plotted against the fitted statistics and against the covariate, temperature, may also be of interest. Note, though, that we are not assuming constant variance, and that the mean residuals are also not expected to be zero, given that we have an over-identified model. Examination of such plots did not indicate any concerns, with all mean residuals close to zero, and no strong non-random behaviour.

6.2 Local linear estimation

As discussed in Section 4.3, local linear estimation has many advantages over local mean, particularly in respect of design bias, and boundary bias. We have seen also that the asymptotic variance is unchanged from that of the local mean estimator (at least for interior points), despite the increase in the number of parameters. However, this may not be the case for finite samples or at the boundaries. In the context of our point process-based models, numerical stability and computation time are also potential issues.

In this section we fit a local linear model with the single covariate, temperature, using a global bandwidth of 1.5, and compare the results to the local mean fit. Computation times are found to be of the order of 200 times slower. This is primarily due to the increased number of calculations required in respect of a single evaluation of the objective

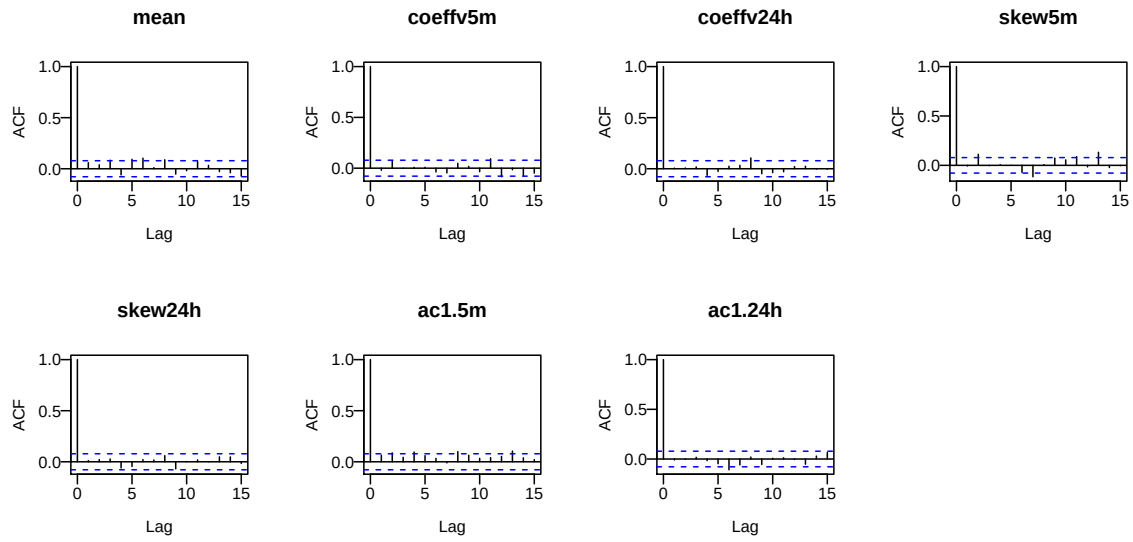


Figure 6.9: Autocorrelation of residuals in respect of selected fitting statistics. The covariate is temperature, with a bandwidth of 1.5.

function. For example, at a given value of the covariate x_0, the vector of expected values of the statistics given by $\tau(\theta(x_t))$ is the same for all t in the local mean case, since $\theta(x_t) = \theta(x_0)$, a constant. In the local linear case, $\theta(x_t) = b_0(x_0) + b_1(x_0)(x_t - x_0)$, and so a separate calculation of τ is required for each t. Another important factor, is the number of iterations required by the numerical optimisation routine in order to find the solution. This is expected to increase with the number of parameters: here we have ten, compared with five for the local mean fit.

In order to keep run times relatively short, the comparison is carried out over an equally spaced grid of sixty temperature values, rather than over all 624 observed data points. This leads to a run time for the local linear fit in respect of all sixty points of around an hour. With sixty equally spaced points, the mean number of iterations required for a local mean fit is 19 (with a range of 11 to 22). For the local linear fit, the mean number of iterations is approximately doubled to 39, with a range of 19 to 63. The higher numbers tend to correspond to points nearer to the boundaries. Note that we are still using the full observed data set within the fitting (i.e. we are not binning).

Figure 6.10 shows the fitted parameters with the two different orders of fit. The results are broadly in line with expectations, with the local linear fitted curve similar for interior points, but generally steeper near the boundaries. While we have tried to ensure that the code is as efficient as possible (for example, by the use of vectors and matrices rather than 'for loops'), it is possible that further analysis might yield additional time savings. The use of a compactly supported kernel, such as the Epanechnikov kernel, instead of the Gaussian kernel, should also improve the speed, since calculations at each fitting point

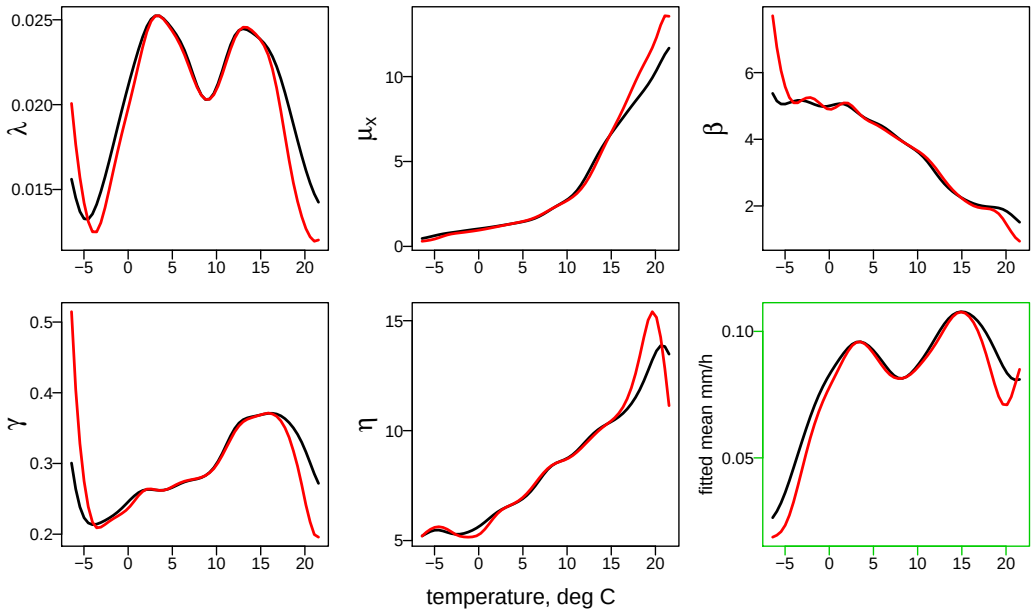


Figure 6.10: Fitted parameters with local mean (black) and local linear (red) fits. The fitted mean is also shown. The covariate is temperature, with a bandwidth of 1.5.

would then require $\theta(x_t)$ to be evaluated over a smaller range of t. It is not obvious, however, that the local linear fit is preferable here, even if the computation speed were not an issue. This is because of the extreme sparsity of observation points at very low and very high mean temperatures, coupled with the high variability of the statistics themselves, so our confidence that the observed statistics at the boundaries are representative is fairly low. In that case, assuming a flatter curve may actually be preferred. Of course, this decision may vary given a particular location and climate. However, it is also likely that computation time and difficulties with numerical optimisation will increase with the number of covariates, and/or with a more complex model. For these reasons, we have decided to focus on the local mean approach for the rest of this thesis.

6.3 Choice of bandwidth

A number of methods for deriving an optimal bandwidth have been discussed in Section 4.4. We have chosen to use a global bandwidth as initial investigations indicated our functions were not 'too wiggly', and our data are relatively limited, given the number of parameters to be estimated. As discussed, we use a form of cross-validation, involving repeated random sub-sampling. This involves randomly splitting the data into test and training sets a number of times. Initially we consider univariate predictors i.e. we search for an optimal bandwidth for each of our set of one-dimensional covariates.

We split our 624 observations into 225 in the test data set, and 399 in the training set, i.e. a split of 36%/64%, which is broadly in line with the recommendation of Hengartner & Wegkamp (2002), who suggest taking the size of the test data-set as the size of the training set to a power between 0.8 and 0.95. We run the exercise twenty-five times, using the same random split for each of the covariates within a run, but a different random split for each of the twenty-five runs. For a given run, and for a given covariate, X, the model is fitted at a grid of potential bandwidths using the data in the training sample, at the covariate values corresponding to the points in the test sample. The optimal bandwidth for each run is then taken as the one that minimises the mean weighted sum of squared errors of the statistics within the test sample (as described in Section 4.4.4 and given by Equation (4.64)). We then calculate the median and mean of the results from the twenty-five runs, and choose our bandwidth based on a combination of these, the distribution of results over the twenty-five runs, and a subjective assessment of graphs at various fixed bandwidths.

Within a given run, and for a given covariate, X, we need to select the grid of potential values of h, the bandwidth. Such grids typically have a geometric progression, with $h_j = C^j h_{min}$, for some factor, C. Here, we first take a wide, but relatively coarse grid over a single sample to get an idea of the approximate location of the optimal bandwidth and the shape and steepness of the curve. We then carry out more runs over a finer, but narrower grid. For the coarse grid we initially had $h_{min} = (X_{(n)} - X_{(1)})/n$, $h_{max} = (X_{(n)} - X_{(1)})$, and $C = 1.4$, which gave a grid of 20 points. However, we found that the smallest bandwidth gave very unstable results for the observations with the most extreme covariate values, so we removed this point. The resulting graphs are shown in Figure 6.11. The finer grid has 27 points, with $h_{min} = 10 \times (X_{(n)} - X_{(1)})/n$, $h_{max} = (X_{(n)} - X_{(1)})/5$, and $C = 1.1$, and the resulting graphs are shown in Figure 6.12. The computation time for a full set of 25 runs across the 27 grid points in respect of a single covariate is of the order of 7 hours.

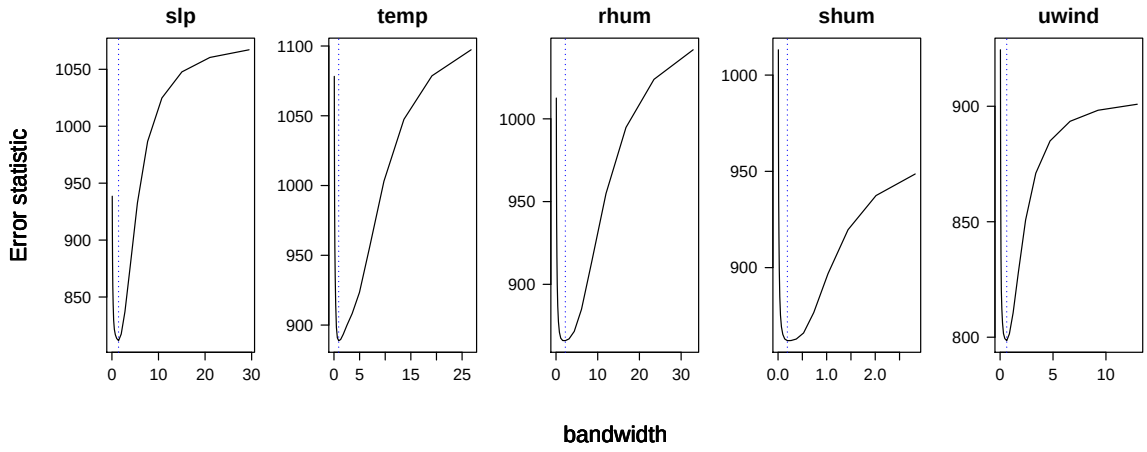


Figure 6.11: Mean weighted sum of squared errors of the fitting properties v bandwidth for a single hold-out sample of 225 observations.

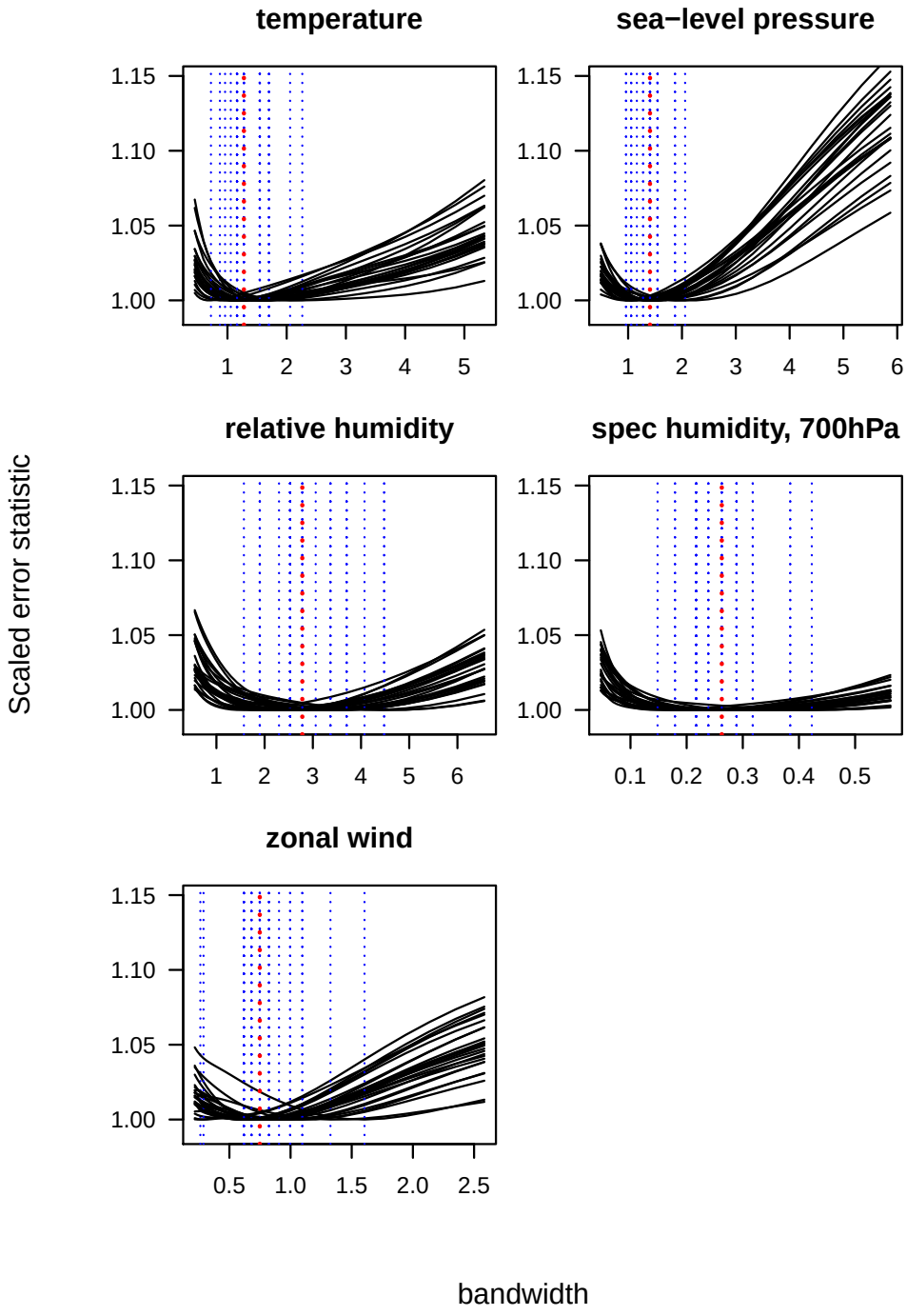


Figure 6.12: Mean weighted sum of squared errors v bandwidth for 25 hold-out samples, each of 225 observations; the dotted lines show the positions of the minima, with the red line indicating the median of these.

Figure 6.11 shows that there is a global minimum for the bandwidth in respect of all the univariate predictors. The curve is steepest for sea-level pressure, and rather flat for specific humidity. This can be seen even more clearly in Figure 6.12. Here we show the curves for all twenty-five test (or 'hold-out') samples over the narrower grid of bandwidths, and scale the error statistics so that the minimum for each run is set to 1. The positions of the minima are also shown. A histogram and density plot of the optimal bandwidth based on the 25 hold-out samples is given in Figure 6.13. Finally, Table 6.1 shows the mean and median optimal bandwidths over the 25 samples. Given the high level of noise in the observed data (which is clear in the graphs of Figures 6.2 and 6.3), it is to be expected that there will be a certain amount of variability in the 'optimal' bandwidth, derived over different subsets of the data. Nevertheless, the appropriate ballpark levels are reasonably clear, and the selected bandwidths, also included in the table, are based on these results, combined with a subjective view of the resulting curves.

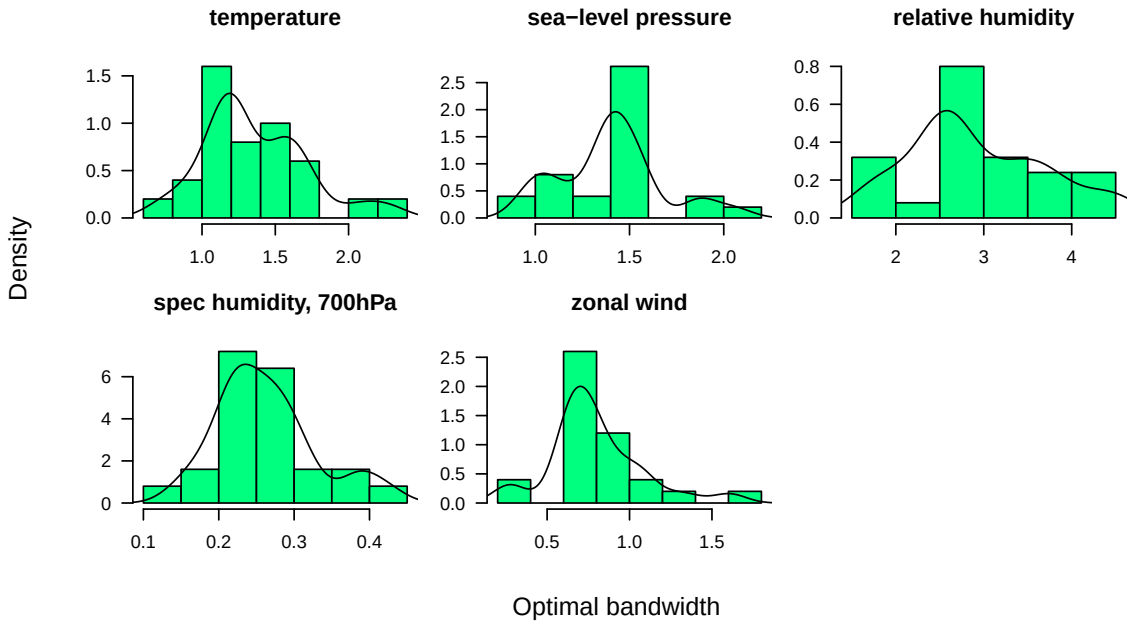


Figure 6.13: Density of the optimal bandwith for each univariate predictor, based on 25 hold-out samples, each of 225 observations.

covariate	Mean optimal bandwidth	Median optimal bandwidth	Selected bandwidth
temp	1.36	1.28	1.3
slp	1.40	1.41	1.5
rhum	2.92	2.78	2.75
shum	0.26	0.26	0.25
uwind	0.79	0.75	0.75

Table 6.1: Optimal bandwidths based on 25 hold-out samples, single covariates.

6.4 Multiple covariates and model comparison

So far we have just considered single covariates, but models with two or three covariates (the maximum likely to be viable given the curse of dimensionality) are also of interest. In this section, we carry out a preliminary comparison of all possible combinations of one, two or three of our selected predictors. A number of pragmatic decisions have been made in respect of the fits with two or three covariates, given the limited amount of data, and desire to avoid unnecessary computational burden. Firstly, in terms of both the order in which the fits are carried out, and the calculation of the weighting matrix, a single component of the covariate has been used, as discussed in Section 4.6. Secondly, initially we use the bandwidths derived in Section 6.3 for each of the single covariates within the product kernels. While these are unlikely to be optimal, this is just intended as a first step, based on the belief that the choice of predictors is more important than using the optimal bandwidth matrix, provided that the bandwidths are in the right ballpark. A small number of models are then selected for more detailed analysis, and these bandwidths are reviewed.

We include also a fit with no covariates, and one with calendar month as the covariate, which is current practice. The model selection approach broadly follows that used in the selection of the optimal bandwidth i.e. we compare the mean weighted sum of squared errors of the thirteen estimated statistics, over the twenty-five hold-out samples. Now, however, the weights used in the calculation of the prediction errors are based on the unconditional empirical variance matrix of the statistics, since a fair comparison requires the same weights to be used for all covariate options.

This preliminary comparison is shown in Figure 6.14, where the errors have been scaled so that the median for no covariates (i.e. a global fit) is 100. Clearly there are various combinations of covariates that give similar results in terms of the levels of error, and our analysis is not sufficient to indicate definitively that one particular set is optimal. In practice, such a decision might depend on a number of factors of which such a comparison would be one, but issues such as consideration of climate physics, and the availability and reliability of the covariate values might be others. Based on these results, we investigate

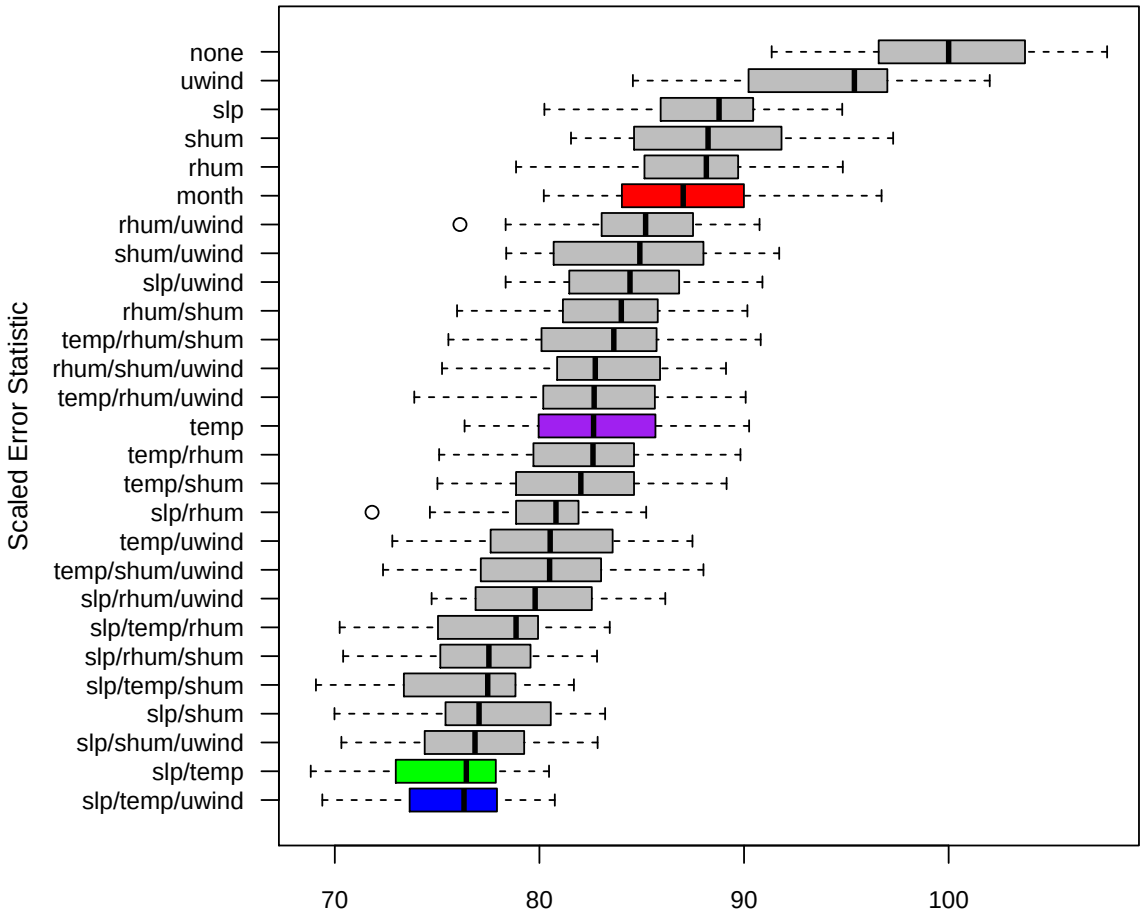


Figure 6.14: Comparison of the mean weighted sum of squared errors of the 13 estimated fitting statistics, for various combinations of covariates over 25 hold-out samples of 225 observations (based on training samples of the remaining 399 observations); product kernels use optimal bandwidths derived from the single covariate fits. The best models with one, two and three covariates, as well as the current default model with covariate month, are shown in colour.

three models in more detail, first refining the selection of the bandwidths for those with multiple covariates. The three models are the best with one, two and three covariates, given by the nested sets below:

1. temperature
2. temperature and sea-level pressure
3. temperature, sea-level pressure and zonal wind component

In order to get closer to the mean square optimal bandwidth for the multivariate predictors, we compare the mean weighted sum of squared errors across a set of potential bandwidth matrices. Full flexibility here would be very onerous, so we limit the investigation to a simple re-scaling of the original diagonal bandwidth matrix (where the diagonal elements comprised the optimal bandwidths for the univariate predictors). Our results give the revised bandwidths shown in Table 6.2. We repeat the model comparison for the

Covariate	As single covariate	With 2 covariates	With 3 covariates
temp	1.3	1.75	2.1
slp	1.5	2.00	2.5
uwind	0.75		1.25

Table 6.2: Estimated mean-square optimal bandwidths for multiple covariates, based on twenty-five hold-out samples.

selected models, using these revised bandwidths, and the results are shown in Figure 6.15, again including also a fit with no covariates, and one with calendar month. The model comparison shows that, using month as a covariate, as is common practice, reduces the median prediction error by around 13% compared with just fitting a global model. Temperature, the best performing single covariate, gives an improved reduction of around 17% compared with no covariates. Results for the other single covariates are similar to month, with the exception of the zonal wind component which performs less well, as expected. The addition of a second covariate, sea-level pressure, to the optimal single covariate, temperature, gives a substantial further improvement, with an overall reduction of 25% compared to the global model, based on the revised bandwidths. Further improvement from the addition of a third covariate, is more limited, reducing the error by a further 2%. The optimal choice, the zonal wind component, reflects its low correlation with the other covariates. These are encouraging results. Replacing month with other covariates potentially has value even if the level of prediction error is broadly the same, if we believe that climate change will lead to different seasonal patterns. Here we have shown that this approach can in fact also lead to a notable improvement in fit.

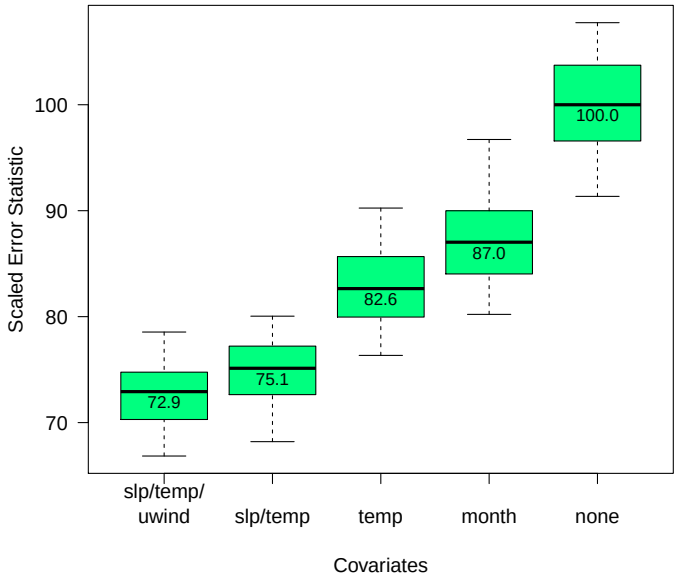


Figure 6.15: Comparison of the mean weighted sum of squared errors for selected covariates over 25 hold-out samples of 225 observations (based on training samples of the remaining 399 observations); figures shown are median errors. The bandwidths are as in Table 6.2.

In the next section we look at the fitted values of the parameters against the covariates for our three selected nested models, and consider the potential interpretation of these in the context of physical weather processes. We consider also the uncertainty involved in these estimates. We then examine the fit of the models.

6.5 Further analysis of selected models

6.5.1 Interpretation of results

The fitted parameters with the single covariate, temperature, are shown in Figure 6.16, including variability bands of $\pm$ 2 standard errors as before. We have also shown the fitted mean hourly rainfall, for which the standard errors have been calculated using the delta method approximation:

$$\mathrm{Var}[h(\theta(x))] \approx h'[\hat{\theta}]^{\mathrm{T}} \,\mathrm{Var}(\hat{\theta})\, h'[\hat{\theta}], \tag{6.3}$$

where $h(\theta)$ denotes the required function of θ (here the expected mean rainfall), with $h'(\theta)$ the first derivative with respect to θ.

The derivative of the mean rainfall with respect to θ is already available from the numerical calculations required for the calculation of the variance of θ. A similar calculation could be undertaken to assess the uncertainty in the estimation of any of the other statistics

of interest. Similar graphs in respect of the other single covariates: sea-level pressure,

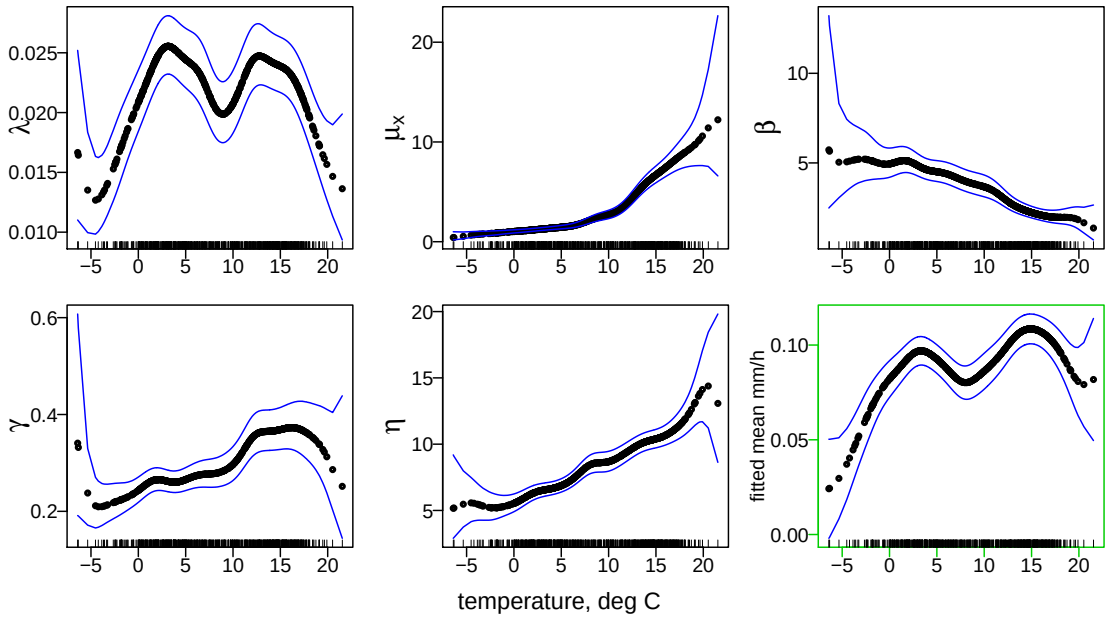
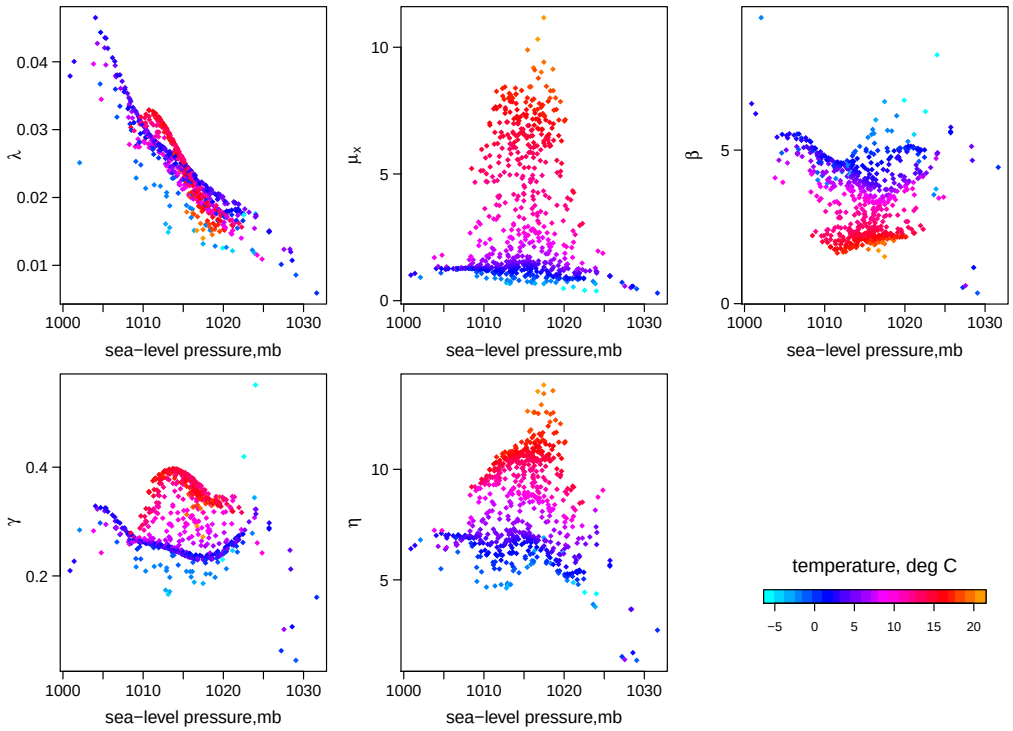


Figure 6.16: Fitted parameters plus mean rainfall v temperature, $h = 1.3$. Variability bands plotted at $\pm$ 2 standard errors.

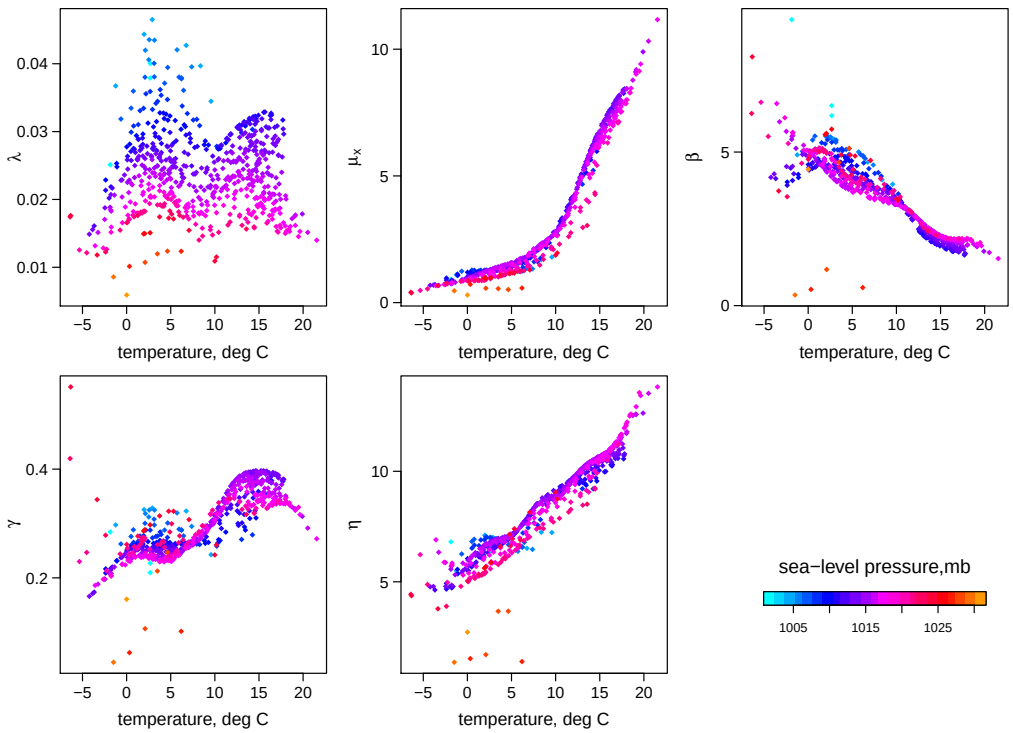
relative humidity, specific humidity and zonal wind component are shown in Figure D.1 of Appendix D.

Fitted parameters with the best combination of two covariates, sea-level pressure and temperature, are shown in Figure 6.17. In order to make it easier to identify interesting relationships, we have plotted two sets of graphs. These both show exactly the same results, but the axes have been exchanged in the second set. The fitted parameters with the optimal set of three parameters: sea-level pressure, temperature, and the zonal wind component, are shown in Figure D.2 of Appendix D, using panel plots. In higher dimensions, it is harder to show the uncertainty in a way that is readily interpretable, and this will be examined in the next subsection. Examining the results from the plotted fits in the context of physical weather processes, we recall that the covariates are monthly means, and so we are not capturing details of individual weather events. Plotting the relationships as points rather than lines or curves allows us to see where data are sparse, and relationships here need to be treated with care. We observe the following:

- Rainfall intensity, μ_X, can be seen to increase with temperature. This is in line with intuition, since increased temperatures lead to greater moisture content in the atmosphere, and increased convective activity. The marked increase in steepness at around 10 degrees could represent the point at which convective rainfall becomes



(a) sea-level pressure on x-axis



(b) temperature on x-axis

Figure 6.17: Scatterplots of fitted parameters vs sea-level pressure and temperature; bandwidths: sea-level pressure: 2.0; temperature: 1.75.

more dominant, in contrast to the stratiform rainfall arising from the meeting of air masses, which will tend to dominate at lower temperatures. Higher temperatures affect convective rainfall, not just because of increased moisture in the air, but also because the strength of updraughts is increased as the land is subject to greater heating.

- As temperatures increase, we also see that storms are generally shorter (γ tends to be higher), although there is a peak at around 15 degrees. There are also fewer cells per rain event (represented by $1+\beta/\gamma$ in the BLRP model), and the cells are shorter. These effects occur at all levels of sea-level pressure, and tie in with the fact that convective storms tend to have fewer cells than stratiform.

- The storm arrival rate, λ, decreases almost linearly with increasing sea-level pressure, at all temperatures i.e. we have fewer storms with increasing pressure. Again, this is in line with intuition, since rainfall is related to low pressure systems. The relationship between temperature and λ is less clear — the two peaks persist at all levels of sea-level pressure, but may be related to some other factor that we have not considered. Note also that the shape of this curve mimics the peaks in the density of temperature. This is something to investigate further.

- The key impact of the zonal wind component is an increase in the storm arrival rate, which reflects the fact that strong westerly winds across central Europe are usually associated with cyclonic conditions (Beckmann & Buishand 2002).

6.5.2 Uncertainty and the curse of dimensionality

Variability bands of $\pm$ 2 standard errors have been shown in all the plots with a single covariate (Figures 6.16 and D.1). As discussed in Section 4.2.5, these cannot be taken as 95% confidence intervals for the parameter estimates. However, from a practical perspective, they are probably sufficient to give an idea of the level of uncertainty involved. As expected, uncertainty is much higher in areas where the data are sparse, which tends to be at the ends of the ranges, and also in areas where there is a high curvature. Overall though, the levels of uncertainty over the central range of data are not unreasonable.

It is interesting to consider how this uncertainty compares with the uncertainty involved in fitting separate models for each calendar month, and also to assess how the variance increases as the covariate increases in dimension: the so-called 'curse of dimensionality'. As mentioned earlier, comparison of the standard errors when the covariate is multi-dimensional is not straightforward. It is difficult to find a visually appealing way of showing it, and any approach that compares standard errors for the same observation point suffers from the fact that the fits themselves may be rather different as more covariates are added.

One approach, that is quite useful, is to consider the distribution of the 'effective sample

size' in respect of each fit. Consider again the equation for the variance, given by:

$$\begin{aligned}\mathrm{Var}[\hat{\theta}(x_0)] \approx & \frac{\sum_{t=1}^{n} K_h^2(X_t - x_0)}{\{\sum_{t=1}^{n} K_h(X_t - x_0)\}^2} \\ & \times \left[\left[\frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_n(x_0) \frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}\right]^{-1} \left[\frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_n(x_0)\, \mathrm{Var}[T(Y)|\, x_0] \\ & \times W_n(x_0)\, \frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta} \left[\left[\frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}\right]^{\mathrm{T}} W_n(x_0) \frac{\partial \tau(\hat{\theta}(x_0))}{\partial \theta}\right]^{-1}.\end{aligned}$$

In the case of the discrete covariate, calendar month, where there are n_m observations available for month m, then the sample size in respect of month m is clearly just n_m. The equation for the variance in this case is given by Equation (2.4) as:

$$\begin{aligned}\mathrm{Var}[\hat{\theta}(x_0)] \approx & \left[\left[\frac{\partial \tau(\hat{\theta}_m)}{\partial \theta}\right]^{\mathrm{T}} W_m \frac{\partial \tau(\hat{\theta}_m)}{\partial \theta}\right]^{-1} \left[\frac{\partial \tau(\hat{\theta}_m)}{\partial \theta}\right]^{\mathrm{T}} W_m\, \frac{\mathrm{Var}[T(Y)|\, m]}{n_m} \\ & \times W_m\, \frac{\partial \tau(\hat{\theta}_m)}{\partial \theta} \left[\left[\frac{\partial \tau(\hat{\theta}_m)}{\partial \theta}\right]^{\mathrm{T}} W_m \frac{\partial \tau(\hat{\theta}_m)}{\partial \theta}\right]^{-1}.\end{aligned}$$

Here the variance of T is divided by n_m, because we are estimating the covariance of the mean summary statistics. By analogy, the inverse of the factor $\frac{\sum_{t=1}^{n} K_h^2(X_t - x_0)}{\{\sum_{t=1}^{n} K_h(X_t - x_0)\}^2}$ can be treated as the effective sample size for a continuous covariate at the observation with covariate value x_0.

Figure 6.18 shows a boxplot of the results using this approach, across all observation points for each of our models of interest. This shows that, at the (approximately) mean-square optimal bandwidths, the effective sample sizes are generally reasonable for the majority of points, and indeed higher than for the current approach. Clearly, some care needs to be taken with the fits at some observation points, particularly in respect of those with 2 or 3 dimensional covariates, but it should be clear which these are. Possibly, fitted models from other locations with slightly different ranges of the covariates could be used to supplement the sparse information here, or alternatively the bandwidth close to the boundaries could be increased (although of course this would lead to a corresponding increase in bias). The graphs would also tend to suggest that three is the maximum number of covariates that could reasonably be used, noting that we have also had to increase bandwidths (and hence bias) as the dimension has increased.

6.5.3 Assessment of the fit of the models

In this section, we assess the fit of our selected models in more detail, using various graphical representations. As we have seen, the fit of a point process-based rainfall model depends on a number of different factors, including the basic model structure (reviewed in

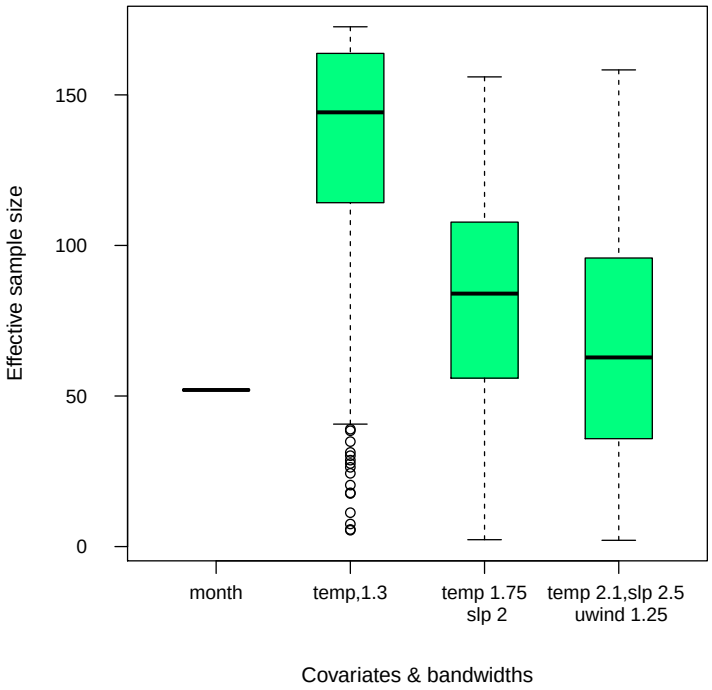


Figure 6.18: Effective sample sizes in respect of selected models, and all modelled observation points.

Chapter 3), the fitting properties selected (discussed in Sections 2.4.2 and 3.10), the choice of covariates, and the smoothing method and smoothing parameters. Numerous different plots are possible, depending in particular on what aspect of the performance we wish to focus. Here we are interested primarily in the impact of relating model parameters to atmospheric covariates. Note that the purpose is not to estimate accurately the size of the errors, but to get an idea of the relative performance of models with different covariates across the properties of interest, and so (except where stated otherwise) we now consider the whole data set.

In Figures 6.2 and 6.3, we showed scatterplots of the monthly fitting statistics against temperature, and used the plots for a visual assessment of the bandwidth used in the fit. We show a similar plot in Figure 6.19 for the model with the optimal pair of covariates: sea-level pressure and temperature. The results are broadly in line with expectations: the relationships between the fitted statistics and the covariates are similar to those of the observed data, but a great deal of the noise has been smoothed out. There is also some evidence of slight understatement of the one hour coefficient of variation and the one hour skewness at higher temperatures.

Rather than considering the raw statistics over individual observation months, the performance of the point process-based rainfall models is usually assessed by plotting mean

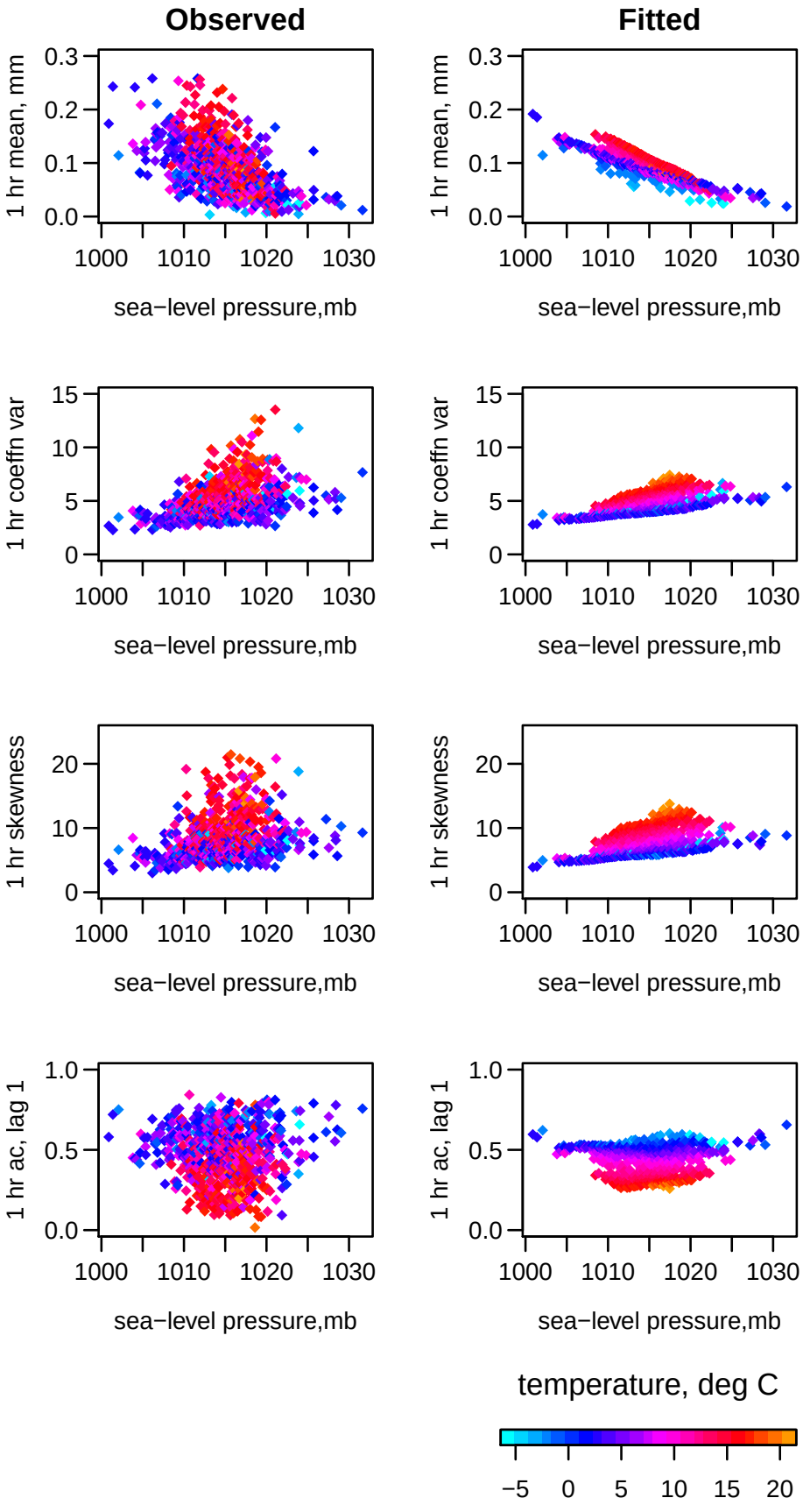


Figure 6.19: Observed (left column) and fitted (right column) statistics at the one hour timescale, with sea-level pressure and temperature as covariates (bandwidths of 1.75 and 1.5 respectively).

statistics over each calendar month, as discussed in Chapters 2 and 3. Here, as well as considering the performance by month, we are also interested in how the observed and fitted properties vary over the covariates. In order to produce comparable plots, we have created factors for each of our selected covariates in turn, by binning the observations into twelve bins (with equal numbers in each bin), based on the covariate value. We then average both observed and fitted properties over the bins. On each of the graphs, we compare four models. The first is a separate fit for each calendar month, the others are the three nested models of one, two and three continuous covariates i.e. the four sets of covariates are:

1. calendar month,
2. temperature,
3. temperature and sea-level pressure,
4. temperature, sea-level pressure, and the zonal wind component.

Figure 6.20 shows the plots in respect of the mean hourly rainfall. The current practice of

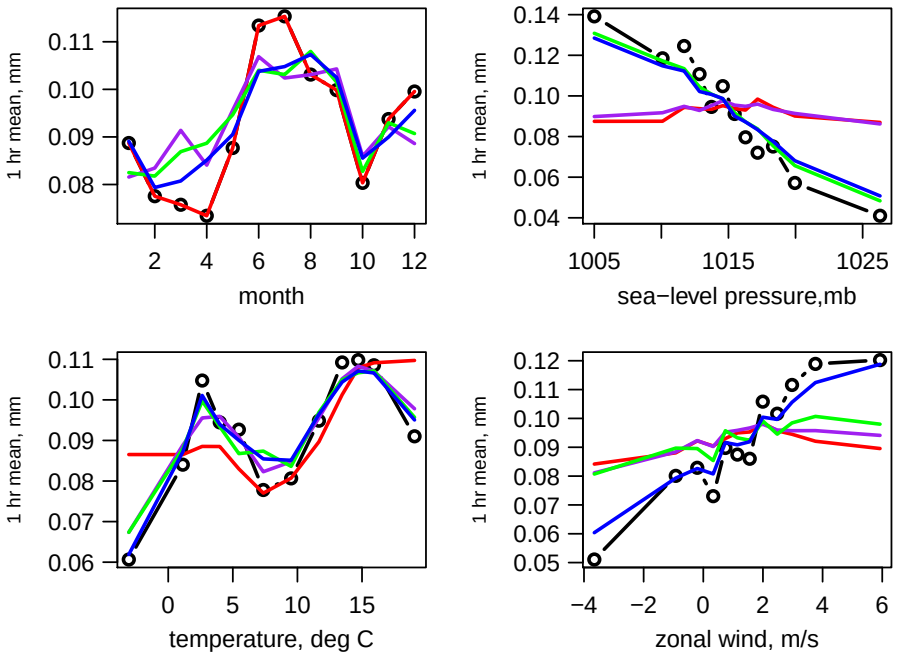


Figure 6.20: Mean hourly rainfall v binned values of selected covariates. Observed values are shown in black; The fitted values correspond to the BLRP model with covariates: month (red); temperature (purple); sea-level pressure and temperature (green); sea-level pressure, temperaure and zonal wind (blue).

a separate model for each calendar month reproduces the monthly means exactly, as we

saw earlier in Figure 3.3 of Chapter 3. It also gives a reasonable fit across the temperature bands, except at the lowest and highest values. If, however, the seasonal pattern of temperatures in the future is different to that in the data, then simulations from this model will not correctly reflect this, as already discussed. The mean rainfall can be seen to show much greater variation over the sea-level pressure bins than over calendar month or the temperature bins, and this variation is not reflected at all in the current approach. The model with the optimal pair of covariates: sea-level pressure and temperature reflects the variation in mean rainfall well across month, pressure bands and temperature bands. Adding the zonal wind component also reflects the variation in the wind velocity bands, as expected. Given the curse of dimensionality, and the consequent increase in bandwidths required with the addition of a further covariate, it is expected that this will adversely affect the fit across the other factors, due to an increase in bias. From the graphs, it can be seen that this effect is minimal (at this level of aggregation), and in fact, adding the wind component has also improved the fit by month, for the winter months, when compared with the model with just the two continuous covariates. Similar graphs could be plotted in respect of other properties of interest.

Given the various covariate options to be compared, and the numerous statistics of interest, a fairly large number of plots of this type would be required in order to examine them all, and plotting against single-dimensional factors only can be misleading. Alternatively, we can get a broad idea by breaking down our error statistic (the mean weighted sum of squared residuals of the fitted statistics) into its component parts. This will indicate how the improvements in fit shown in Figure 6.15 are distributed across the thirteen fitting properties. Figure 6.21 shows the mean error statistic over the 25 hold-out samples, broken down into the thirteen fitting properties. Again, we have scaled the errors, so that here we have an error of 100 for the global model. Note that it is the change in the component values as covariates are added, rather than the absolute values of the components, that are of primary interest. The figure shows that the optimal model with covariates: sea-level pressure, temperature and zonal wind component, has substantially improved the fit to the mean rainfall, and to the coefficient of variation at all timescales, with a reduction in error of the order of 40% to 50% for these properties, The skewness and auto-correlations at the shorter timescales also show a reasonable improvement (of the order of 25% to 35%). The properties showing the least improvement are the daily skewness and autocorrelation.

Finally in this section, we consider interannual variability, the underestimation of which is one of the criticisms of many of the rainfall models. Figure 6.22 shows a range of percentiles between the 5th and the 95th in respect of the mean hourly rainfall, based on 200 simulations, over 52 years. In the first graph, the simulations have been based on the twelve monthly sets of parameters, sampling from the appropriate calendar month's parameter distribution for each observation month in turn. In the second graph, a different parameter distribution has been used for each observation month, reflecting that month's

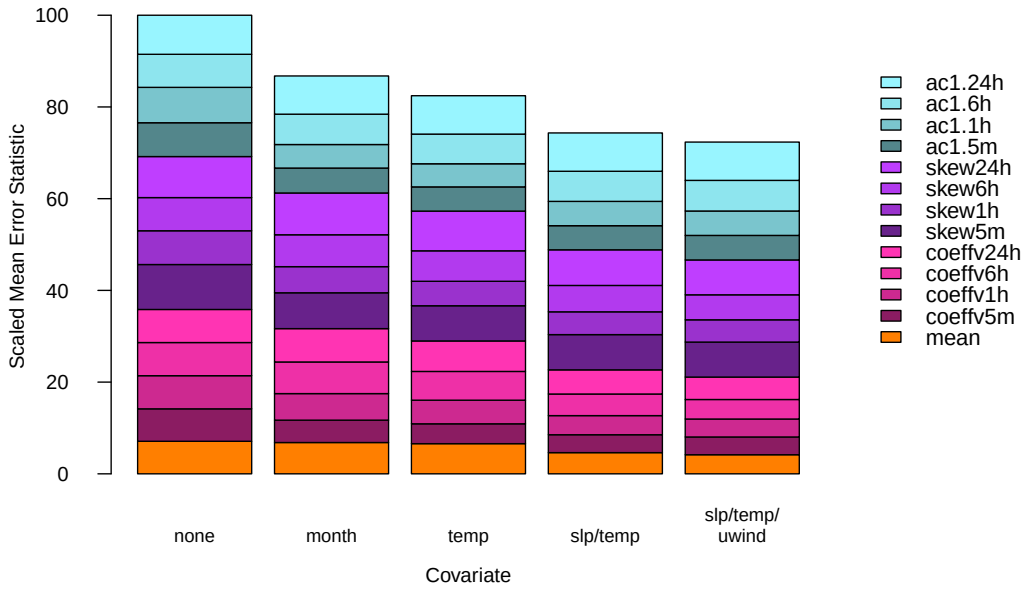


Figure 6.21: Breakdown of the mean weighted sum of squared errors over the individual components; results based on the mean over 25 hold-out samples

covariate values. It can be seen that allowing the parameters to depend on these covariates gives a much improved representation of the interannual variability. Similar simulations were also used to compare the fit to extreme values. However, no improvement was seen, compared with a separate model for each calendar month, so these graphs are not included here. While this is a little disappointing, it primarily reflects the issues already identified with respect to the BLRP model's fit to extremes, and the set of fitting properties used here, as discussed in Sections 3.8.3 and 3.10 respectively. Note that the new approach should nevertheless outperform the current one in a warming climate, since simulations in respect of future time periods can reflect the higher mean temperatures, and this will lead to a greater incidence of extreme values.

Overall, the assessment of the performance has shown that, with just two or three covariates, we can improve considerably the explanatory power of the BLRP model. If the same approach can be extended to more complex models, then a combination of the optimal model structure, and the new covariate-dependent parameters should lead to an important improvement in hydrologists' ability to simulate realistic rainfall, allowing for the impact of climate change. We consider the application of the new methodology to the BLRPR$_X$ model in Section 6.7. First though, we consider briefly two alternative datasets in order to check that the methodology can be extended to other types of location and climate.

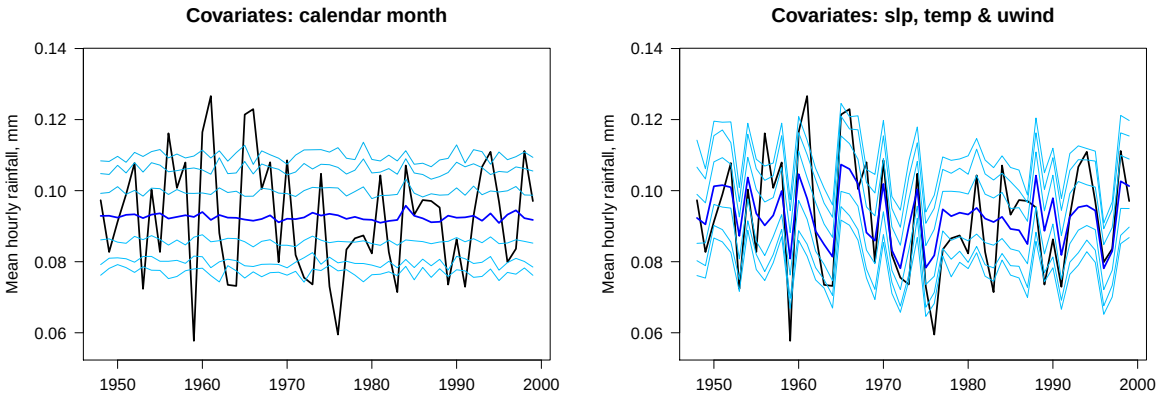


Figure 6.22: Simulated distributions of mean annual rainfall (expressed in mm per hour) for Bochum. The blue bands show the 5th, 10th, 25th, 50th (thicker blue line), 75th, 90th, and 95th percentiles from 200 simulations, and the thick black line shows the observed values.

6.6 Other datasets

In this section we start by considering the five-minute rainfall series from Kelburn, New Zealand, introduced in Section 3.11. Combined with the appropriate monthly mean predictors from the NCEP reanalysis data, we have 57 complete years of data, from 1948 to 2004. Initial analysis indicates that the climate of Kelburn shows far less seasonality than Bochum, with very mild winters. Mean monthly temperature values range from 6.3°C to 18°C, compared with a range of -6.4°C to 21.5°C at Bochum. Rainfall levels are greater than in Bochum, with mean hourly rainfall varying from around 0.1mm in January and February, to 0.19mm in the wettest month, June. Although the range of the mean hourly rainfall is wider than at Bochum, the other fitting statistics show far less variability across months, particularly at the shorter timescales. Lag-1 autocorrelation at the 5 minute level is almost constant over the year, and skewness levels are much lower. Relationships between the fitting statistics and the monthly mean values of the atmospheric variables are generally weaker (with the exception of relative humidity, which here shows a stronger relationship with the mean hourly rainfall). We surmise that rainfall here may be primarily affected by individual weather systems; these are not well captured within the monthly mean data.

Note that our primary purpose here is to validate the methodology, rather than to find the optimal model for Kelburn. Therefore, we restrict models to the best choice of one or two covariates from the three covariates: temperature, sea-level pressure and relative humidity, and consider mean parameter estimates only. It is also of interest to consider to what extent the relationships found at Bochum apply in a different climate, and to identify any potential new issues. Using the same methodology as before to find optimal bandwidths, we find a similar level for temperature, but a higher value in respect of sea-

level pressure (2.5 compared with 1.5 as a single covariate), and a lower value in respect of relative humidity (1.5 compared with 2.5 for Bochum). The selected bandwidths for the Kelburn data are shown in Table 6.3.

Covariate/s	Bandwidth/s
temp	1.3
slp	2.5
rhum	1.5
temp/slp	1.5/2.9
temp/rhum	1.75/2
rhum/slp	1.75/2.9

Table 6.3: Estimated mean-square optimal bandwidths for Kelburn, based on twenty-five hold-out samples.

A comparison of the mean weighted sum of squared errors in respect of the six possible models, using these bandwidths, is shown in Figure 6.23, again including also a fit with no covariates, and one with calendar month. This shows that, as expected, the introduction

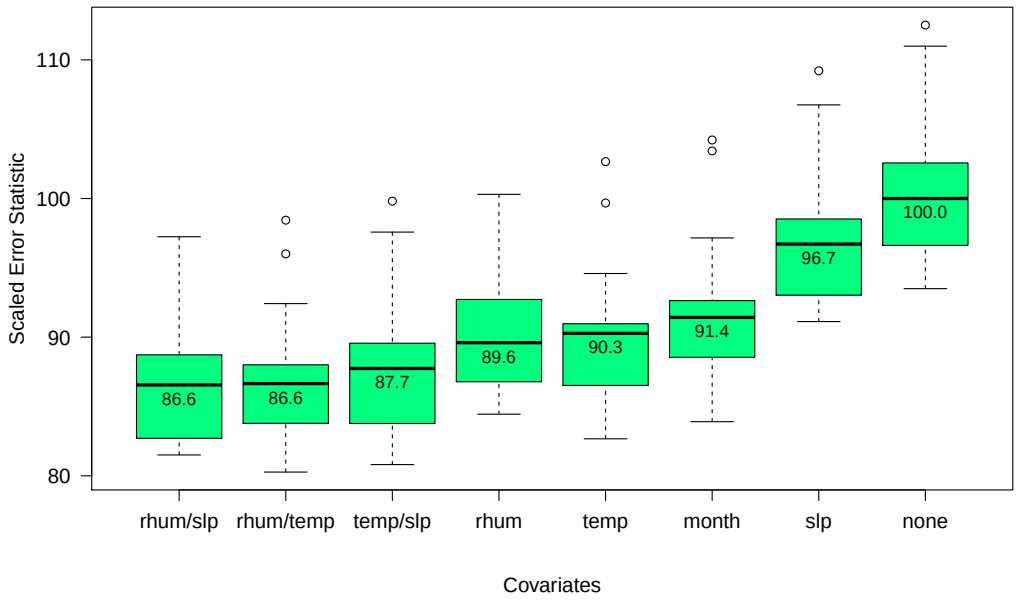


Figure 6.23: Comparison of the mean weighted sum of squared errors for Kelburn for selected covariates over 25 hold-out samples of 250 observations; figures shown are median errors. The bandwidths are as in Table 6.3.

of covariates into the model has less impact in respect of Kelburn data, than for Bochum. Here there is only a 13% reduction in the mean weighted sum of squared errors for the optimal two-covariate model, when compared with the global model. The corresponding reduction for Bochum is 25%. Nevertheless, a combination of continuous atmospheric

covariates does show a better performance than using calendar month, and allows for parameters to reflect future climate change, which was our original motivation.

There is little to choose between the three pairs of covariates, and we consider the optimal model (sea-level pressure and relative humidity), and the model with covariates sea-level pressure and temperature in a little more detail. Figure 6.24 gives a breakdown of the mean

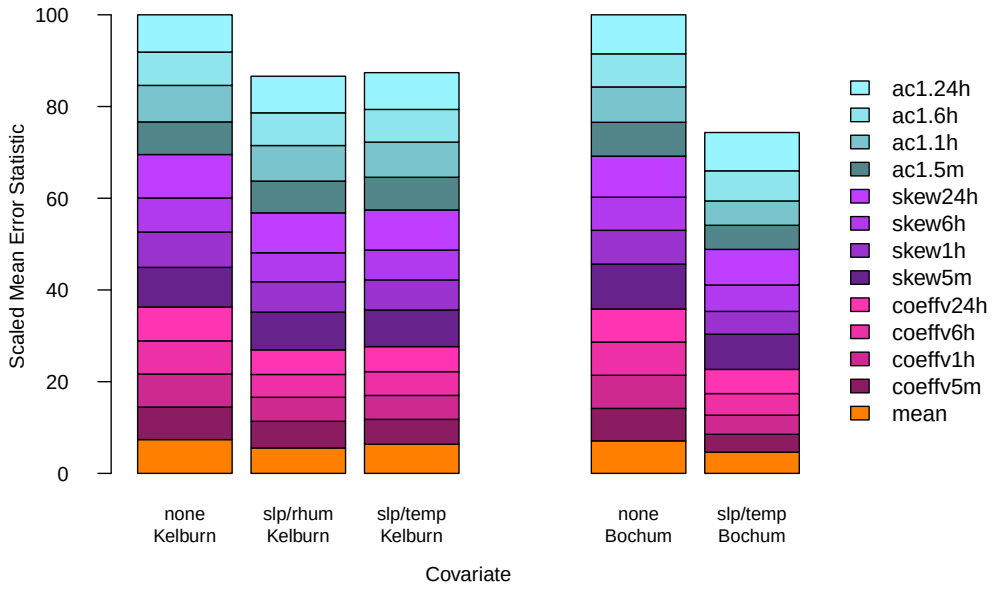
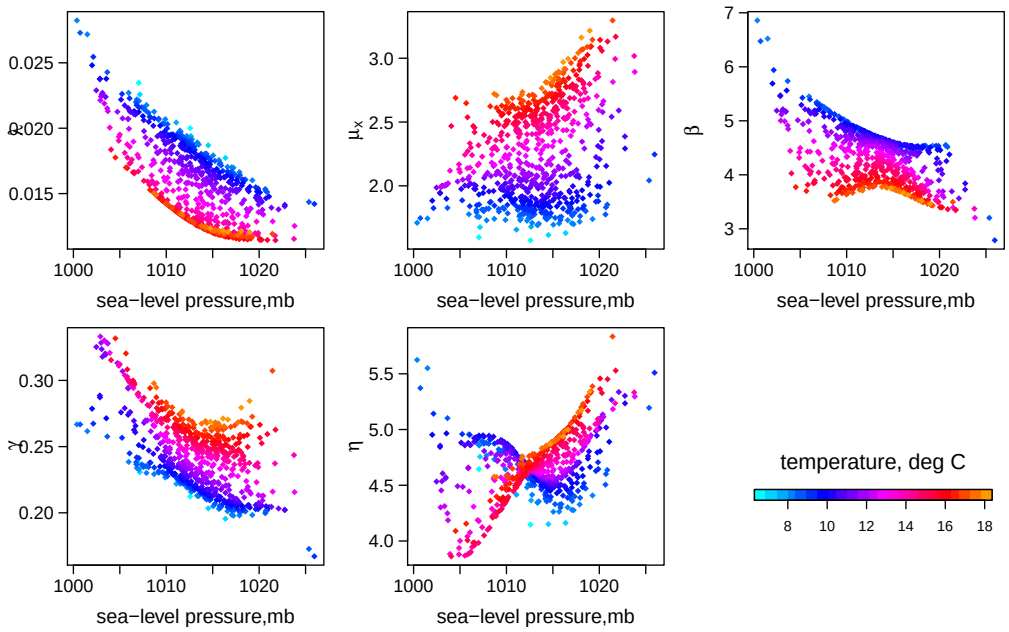


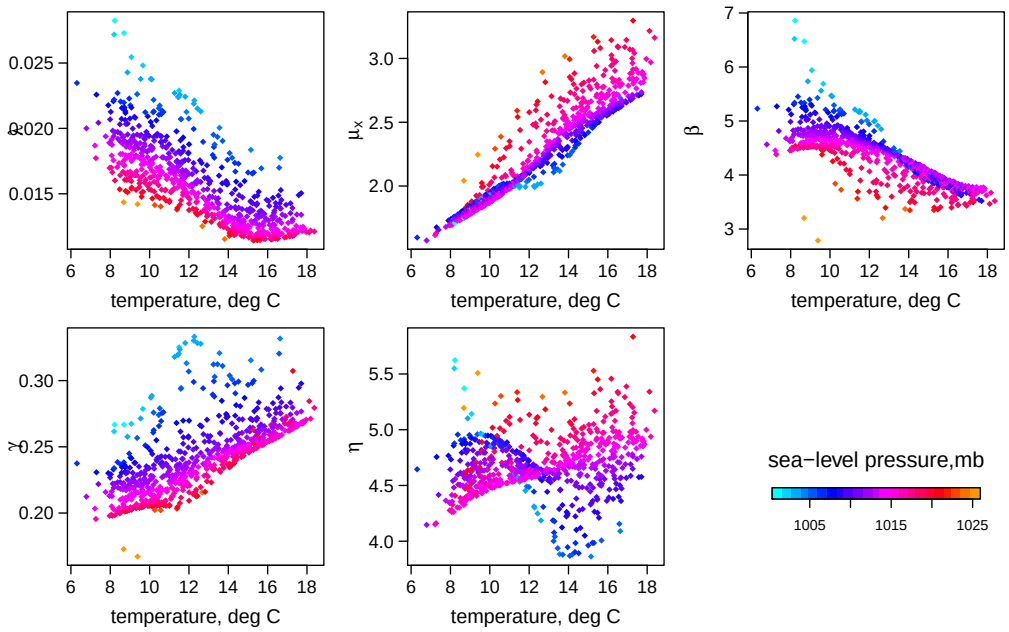
Figure 6.24: Breakdown of the mean weighted sum of squared errors over the individual components; results based on the mean over 25 hold-out samples. For each of the locations, the errors have been scaled so that the model with no covariates has mean 100.

weighted sum of squared errors, over the components for these two models, and for the model with no covariates. Plots for Bochum are also shown for comparison. At Kelburn, the model with sea-level pressure and relative humidity gives a greater improvement to the fit of mean rainfall, but otherwise the two pairs of covariates perform similarly, reducing the error in the coefficient of variation at all timescales by of the order of 20%–30%. There is a small improvement in respect of skewness, but no change at all in respect of lag-1 autocorrelation.

In order to assess whether the nature of the relationships between the parameters and the covariates are similar to what we have seen at Bochum, we plot the parameter sets for the model with sea-level pressure and temperature as covariates, and compare these with the Bochum parameters of Figure 6.17. In Figure 6.25 we plot the Kelburn parameters, using 'natural' scales which reflect the actual ranges of the data. In Figure 6.26, we show the same results, but here the scales have been adjusted to allow direct comparison

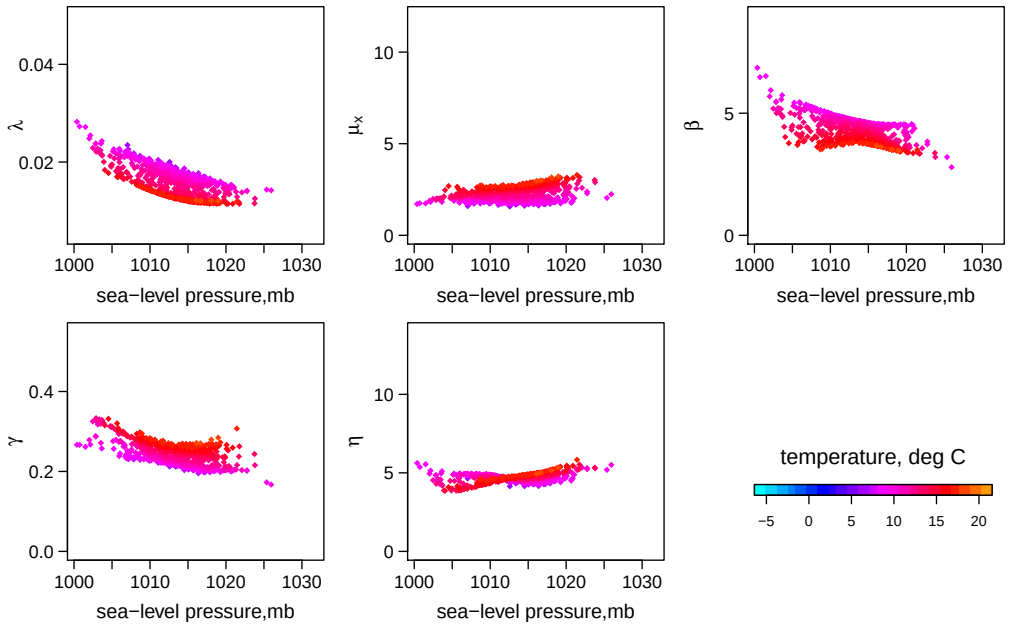


(a) sea-level pressure on x-axis

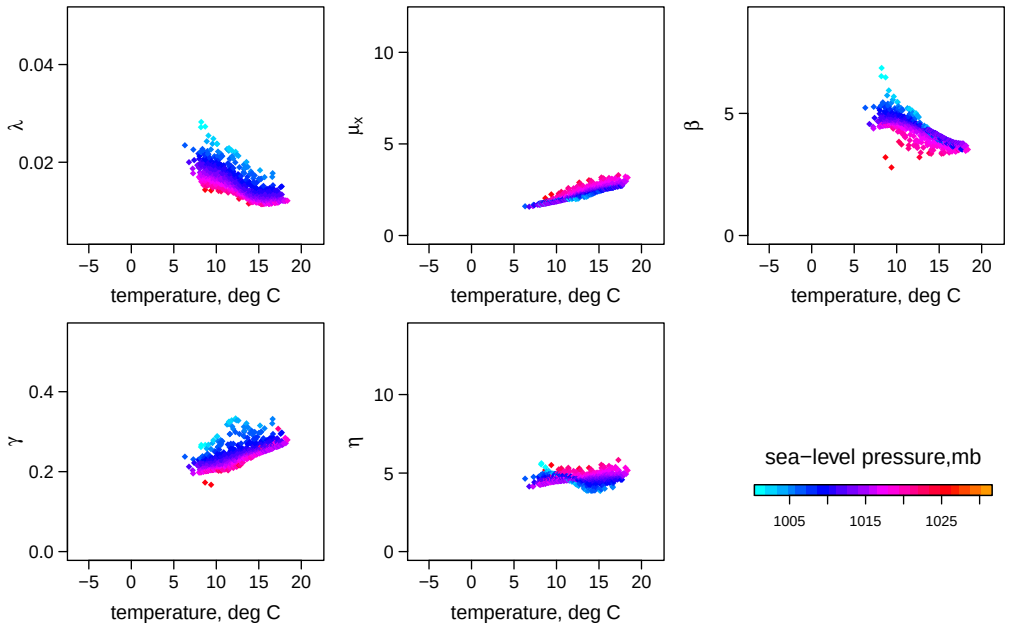


(b) temperature on x-axis

Figure 6.25: Scatterplots of fitted parameters vs sea-level pressure and temperature for Kelburn; bandwidths: sea-level pressure: 2.9; temperature: 1.5.



(a) sea-level pressure on x-axis



(b) temperature on x-axis

Figure 6.26: Scatterplots of fitted parameters vs sea-level pressure and temperature for Kelburn; bandwidths: sea-level pressure: 2.9; temperature: 1.5; scales adjusted to match those for Bochum in Figure 6.17.

with Bochum. The two most important relationships from Bochum are seen to hold at Kelburn too. The storm arrival rate, λ can be seen to decrease with increasing sea-level pressure, and the mean rainfall intensity, μ_X increases with increasing temperature, although the latter, in particular, shows far less variation (even when considered over the same temperature range). In terms of the other parameters, most of the relationships are similar. It is interesting to note, though, that the peaks in the Bochum plot of λ against temperature are not repeated here, with λ decreasing with increasing temperature up to around 15°C, and then remaining level.

The other dataset that we consider is a rainfall series from Heathrow in England. This series is at an hourly, rather than a five-minute level, and runs for 52 years from 1949 to 2000. The climate at Heathrow shows more similarity to Bochum, although rainfall levels are a little lower, and the range of mean hourly rainfall across calendar months is slightly narrower (between 0.054mm (February) and 0.084mm (October)). An initial analysis of relationships between the potential predictors and the nine fitting statistics (now excluding those at the five minute timescale) indicates that these are similar to those seen at Bochum.

We have not considered the Heathrow dataset in the same level of detail as Kelburn, but found that some initial fits, using the same bandwidths as for Bochum, raised a new issue, which we discuss here. The fitted parameters with temperature and sea-level pressure as covariates, using the same bandwidths as for Bochum, are shown in Figure 6.27. As can be seen, although the plots for λ, β and γ are comparable to those from Bochum, the mean rainfall intensity, μ_X, and the cell duration parameter, η, if unconstrained, become extremely large, so that storms effectively comprise a few very intense instantaneous bursts. This is clearly unrealistic. Since the most significant difference between the Heathrow fit and that of Bochum, is that the latter includes statistics also at the five-minute timescale, we investigate this, by repeating the fit for Bochum, but now including only the statistics at hourly and longer timescales. The results, in Figure 6.28, indicate a similar effect to that at Heathrow. Further investigation, including fitting models with sea-level pressure and temperature as single covariates, revealed that the problem arises particularly at the higher values of temperature (although at Heathrow, with both covariates, it occurs at all except the most central observations). An intuitive explanation for the problem, is that, over a longer timescale, it is not possible to differentiate between an instantaneous high-intensity pulse, and a longer low-intensity rain cell, particularly if the overall duration of the storm is short (i.e. if it does not last beyond a single hourly interval). There is therefore no longer a unique solution for θ. The problem is exacerbated by the fact that, after the initial fit, we use the parameter set from the previous fit as a starting point for the numerical optimisation of the next fit. Thus, once this region of the parameter space is entered, later fits tend to be similar. When each hour is split into twelve intervals, however, the smoother nature of the rainfall is apparent, and the more realistic parameter

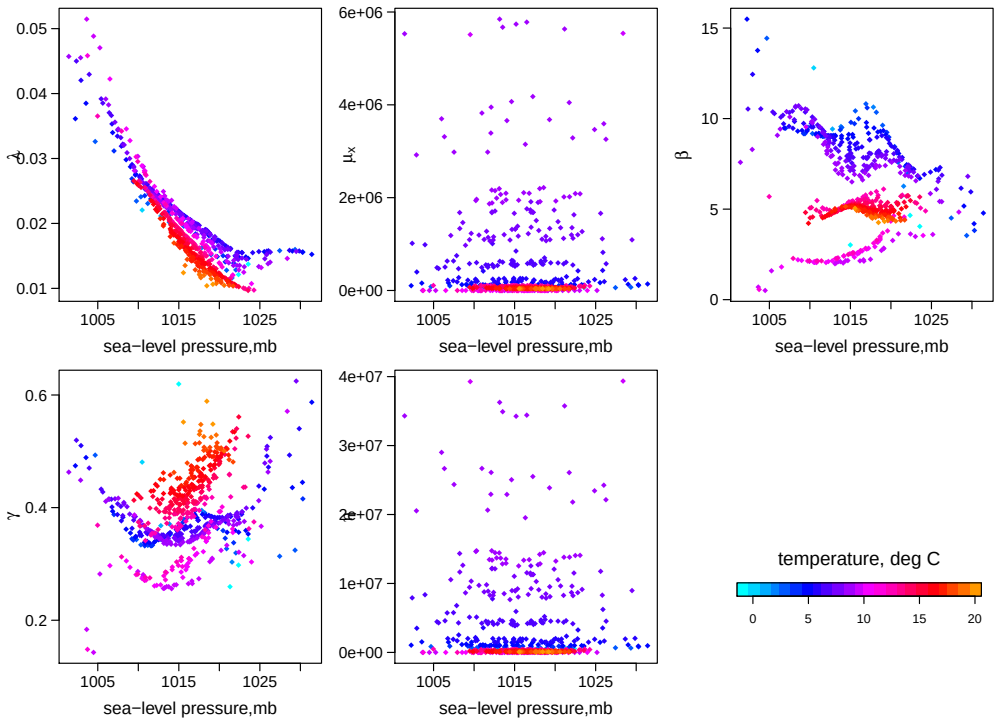


Figure 6.27: Scatterplots of fitted parameters vs sea-level pressure and temperature for Heathrow; bandwidths: sea-level pressure: 2; temperature: 1.75

sets that we had before are found. The solution to this problem could involve one or more of the following:

- Restricting the parameter space by introducing constraints in the fit, to prevent the very large values (although in our experience it is difficult to pick these individually for parameters without finding solutions that lie on the boundaries).
- Widening the bandwidths, or allowing these to vary locally, so that they can be wider at the observations more prone to this type of solution.
- Trying more than just the single initial value at each fit, to prevent the parameters getting 'stuck' in an undesirable part of the parameter space.
- Creating statistics at the five-minute timescale using 'scaling relationships' between statistics at different timescales. This type of methodology was discussed in Section 1.3, although it was noted that such relationships had not yet been developed in respect of sub-hourly rainfall.

The last of these options would be the preferred approach, but it is not known if such relationships can be found down to the fine scale required. This would require further development. For Bochum, we found that simply increasing both of the bandwidths to

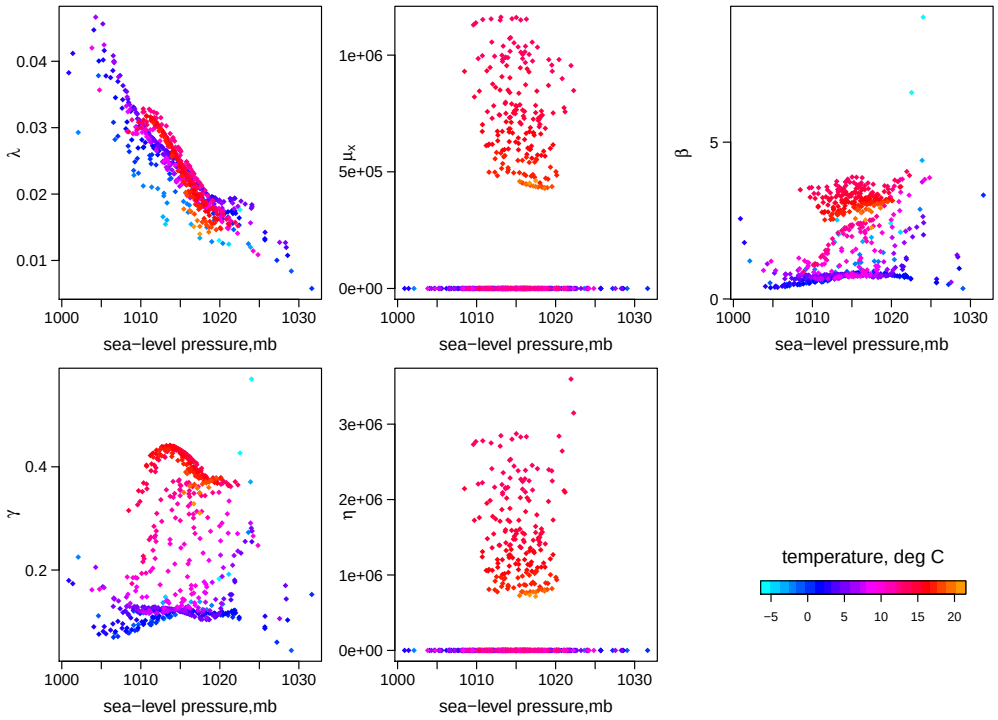


Figure 6.28: Scatterplots of fitted parameters vs sea-level pressure and temperature for Bochum, excluding all statistics at a five-minute timescale; bandwidths: sea-level pressure: 2; temperature: 1.75

around 3 gave a realistic parameter set. For Heathrow, even with much wider bandwidths, we obtained extreme values for μ_X and η for some observations. No suitable constraints could be found that did not lead to some observations lying on a constraint boundary. Both of the bandwidths had to be increased to 4.5 before parameter sets with no extreme values were found, and the resulting plot is shown in Figure 6.29. It is not suggested that this is an effective solution, and a better solution would retain the narrower bandwidth in the central range of temperatures, widening it only close to the boundaries. Alternatively, rather than setting a global bandwidth, a global span would probably be more appropriate. However, the plot does give a rough indication of the nature of the relationships that hold at Heathrow, which are broadly similar to those at Bochum. The storm arrival rate decreases with increasing sea-level pressure, as at Bochum. Mean rainfall intensity increases and storms and cells become shorter with increasing temperatures, again as at Bochum. Differences here are that the storm arrival rate decreases with increasing temperatures across the whole range of temperatures (in line with Kelburn), and the cell arrival rate increases. The latter effect is seen also in the Bochum fit, when 5-minute data are excluded, so appears to be related to the minimum timescale of the statistics, rather than any differences in the locations.

Overall, the analyses of Kelburn and Heathrow show that the local GMM methodology

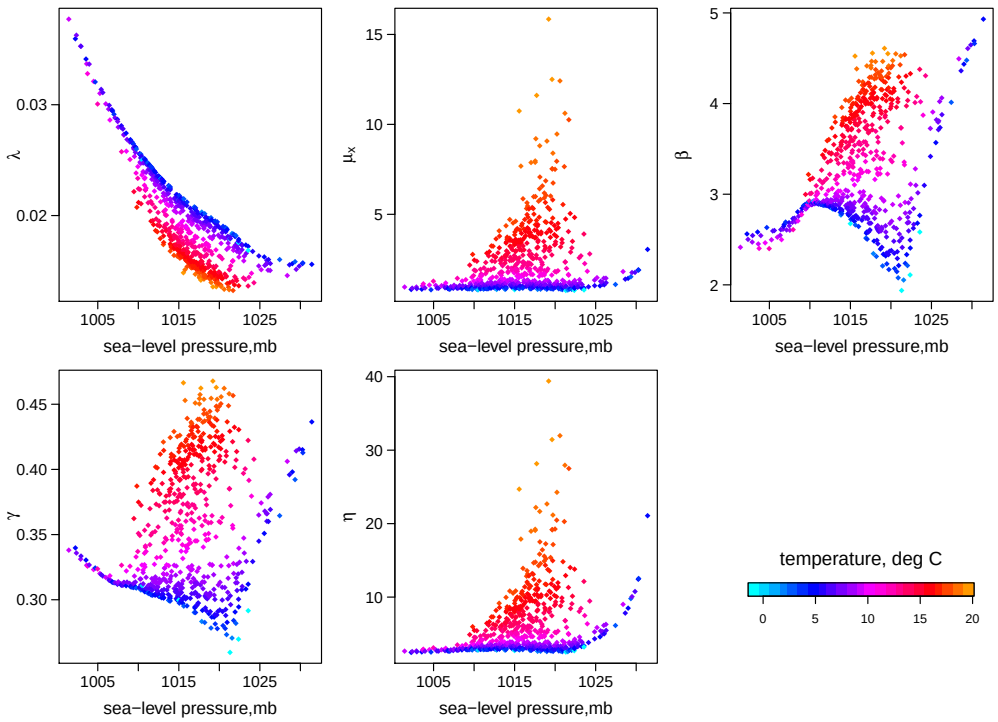


Figure 6.29: Scatterplots of fitted parameters vs sea-level pressure and temperature for Heathrow; bandwidths: sea-level pressure: 4.5; temperature: 4.5

may usefully be applied at other locations, and that parameter relationships are expected to be similar. While there will undoubtedly be issues to address in respect of any new locations, these do not appear to be insurmountable.

6.7 More complex model variants

We selected the BLRP model (with an exponential intensity distribution) for the development of the local GMM methodology, as the simplest of the Bartlett-Lewis clustered models. It also has the advantage, compared to the models with randomised cell duration parameter, that the parameters are easy to interpret. However, the results in Chapter 3 indicate that the BLRPR$_X$ model gives a better fit to the data. In this section therefore, we apply the local GMM methodology to this model, and compare results with those for the BLRP model.

The same bandwidths as for the BLRP model are used for all the fits. Figure 6.30 compares the mean weighted sum of squared errors of the BLRPR$_X$ model, against the BLRP model. The former can be seen to outperform the latter consistently by around 2%–3% across all the covariate options. The BLRPR$_X$ model is preferred not only on the basis of its fit to the properties used in the fitting process, but also because other important properties,

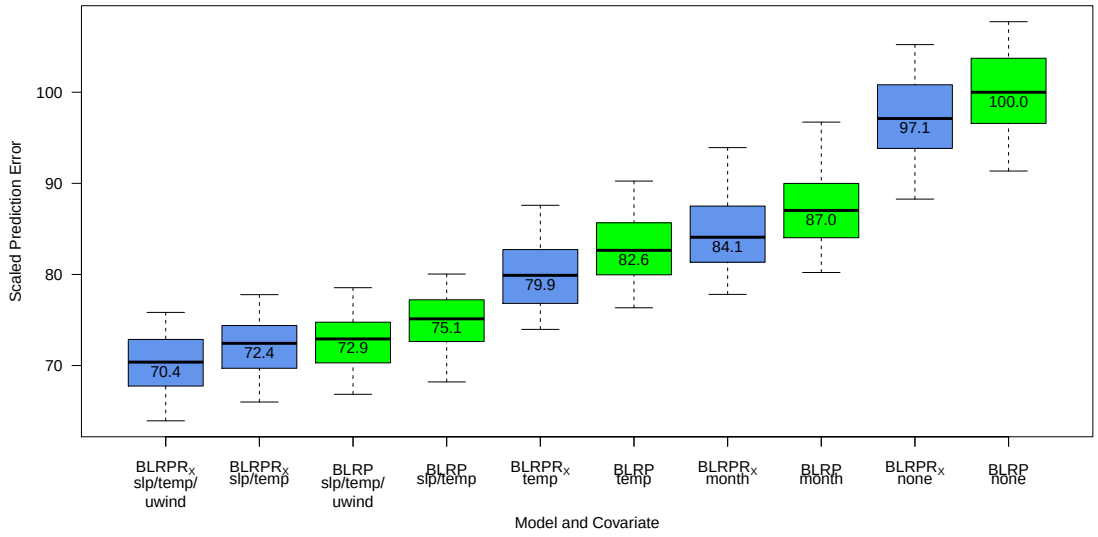


Figure 6.30: Comparison of the mean weighted sum of squared errors for selected covariates over 25 hold-out samples of 225 observations, for Bochum; figures shown are median errors. The bandwidths are as in Table 6.2. The green boxes are in respect of the BLRP model, and the blue the BLRPR$_X$ model.

not included in the fit (such as the proportion of intervals that are dry), are also better represented. The fits of the two models in respect of further properties of interest may be compared using plots similar to Figure 6.20. As an example, in Figure 6.31, we compare the fit to 1-hour proportion dry of the two models. This indicates that the BLRPR$_X$ model outperforms the BLRP model, as expected, and that the impact of the covariates is essentially independent of the impact of the chosen model structure. Here we have used the optimal set of covariates i.e. sea-level pressure, temperature and zonal wind.

Finally, in Figure 6.32, we plot the fitted parameters with the optimal pair of covariates, sea-level pressure and temperature. Recall that the parameter α/ν here represents the mean of the cell duration parameter, η, which varies randomly over storms, and which is distributed as Gamma(α, ν). Also, ι, κ and ϕ are the ratios of μ_X (the mean rainfall intensity), β (the cell arrival rate) and γ (the storm termination rate) to η respectively. The relationships of the parameters to the covariates are as expected given the earlier BLRP fits. The storm arrival rate, λ, decreases with increasing sea-level pressure, exactly as before, and the parameter representing the mean cell duration, α/ν follows the behaviour of η in the BLRP model. The parameter ι increases with increasing temperature, so that the mean rainfall intensity, μ_X, increases with temperature more steeply than η, again as before. Similarly, the behaviour of κ and ϕ can also be seen to be consistent with earlier results. Note that, as discussed in Section 3.3.1, α has been constrained to be greater than 2, and lies at this constraint level for the majority of the points. Interestingly, when

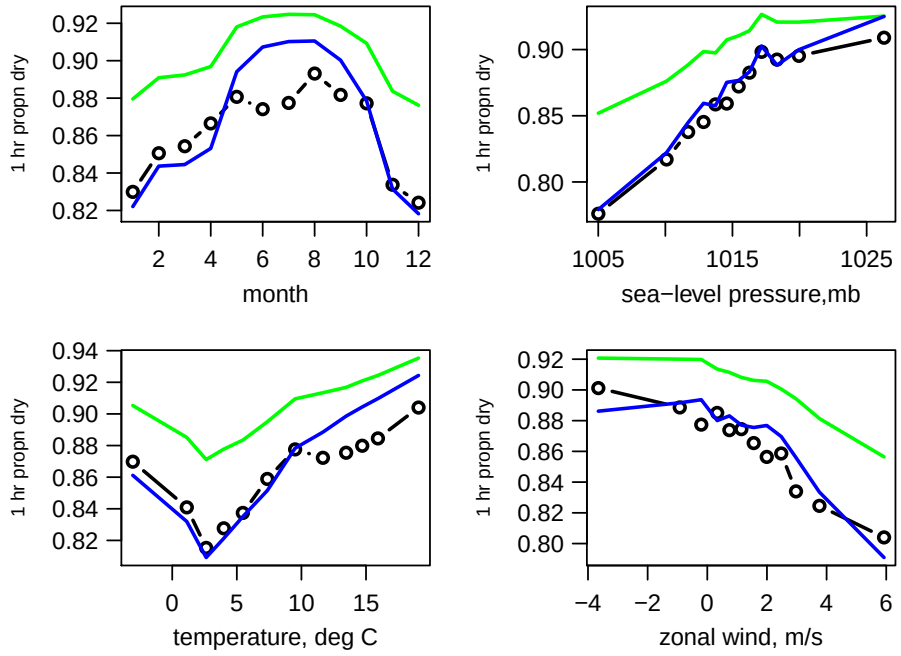
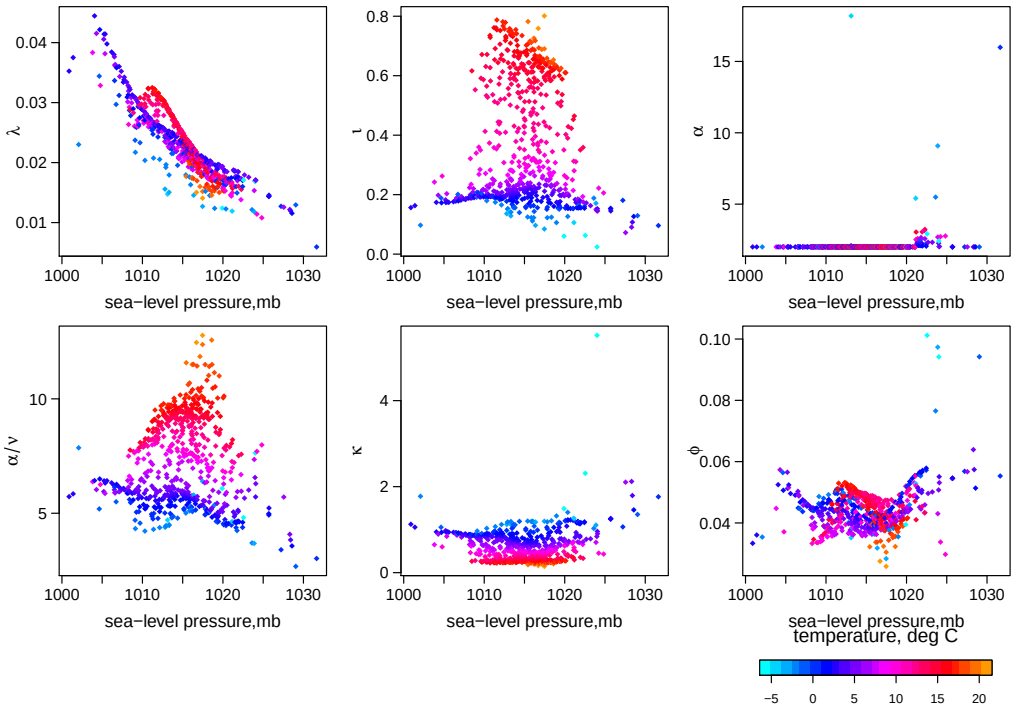


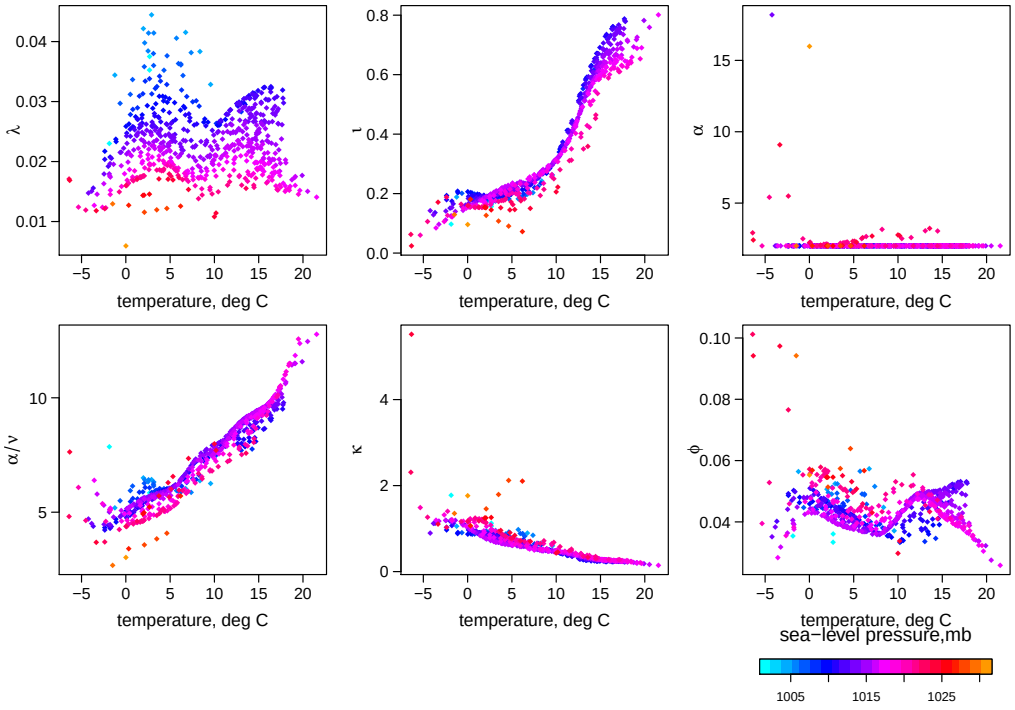
Figure 6.31: Mean observed and fitted 1-hour proportion dry v binned values of selected covariates at Bochum. Observed values are shown in black; The fitted values correspond to the BLRP model (green), and the BLRPR$_X$ model (blue), both with covariates: sea-level pressure and temperature.

we added zonal wind as a third covariate, the fitted solution at a few points close to one of the boundaries reverted to the 'fixed η' fit. This effect could be eliminated, if desired, by allowing some local variation to the bandwidth, for example allowing the bandwidth to increase close to the boundaries, as discussed earlier. The effect was found to be more pronounced when we replaced the exponential intensity distribution with the Weibull, with a shape parameter of 0.6. Recall from Section 3.10 that we had a slight preference for the latter. Here we find essentially no difference between the two options in terms of the out-of-sample errors, and all the figures in respect of the BLRPR$_X$ model assume the exponential distribution.

In conclusion, the local GMM fitting methodology, with large scale atmospheric variables as covariates, is capable of being applied to different climates and model structures. Issues with non-unique parameter sets and numerical instabilities may arise, as indeed they do for the monthly models, particularly where models have a high number of parameters and where observed rainfall series are relatively short. These issues may be addressed pragmatically as and when they arise, typically by the local adjustment of bandwidths, or by restrictions on the parameter space (i.e. fixing or constraining parameters). Some further thoughts on these issues, and suggestions for potential further developments are given in Chapter 7.



(a) sea-level pressure on x-axis



(b) temperature on x-axis

Figure 6.32: Scatterplots of fitted parameters vs sea-level pressure and temperature for the BLRPR_X model ; bandwidths: sea-level pressure: 2.0; temperature: 1.75.

Chapter 7

Conclusions and future research

7.1 What have we achieved?

In this thesis we have considered both the structure and the fitting methodology of the point process-based rainfall models. We have introduced a new model version, and developed a local fitting approach that allows for a nonstationary climate. We believe that both of these developments will be useful to the hydrological community and to other practitioners who require artificial rainfall simulations for their applications. Importantly, they are also readily implementable.

Using a structured comparison of different versions of the Bartlett Lewis clustered point process-based models, including two new model extensions, we have clarified some key aspects of performance. Our focus here has been on fine scale data, for which models with rectangular pulses were previously assumed inappropriate. We showed that the main driver behind improved performance, particularly in respect of skewness and extremes at short timescales, is the introduction of an inverse dependence between rainfall intensity and cell duration. Our proposed new model, which is an extension of the Random Parameter Bartlett-Lewis Rectangular Pulse model, gives a simple but effective way of introducing such dependence, with no increase in the number of parameters. It allows the rainfall intensity parameter, μ_X, previously assumed to be constant, to vary in proportion to the cell duration parameter, η, which itself varies randomly between storms. Although instantaneous pulses were useful in leading us to this conclusion, ultimately we discovered that they are not required, and the computationally simpler rectangular pulse version is preferred.

We believe that acknowledging the limitations of all such models is important. Modelling a complex process by a relatively simple statistical model invariably requires some compromises in the quality of the fit to different features of the process. Adding ever-increasing complexity to such models in order to gain marginal improvements in fit is unlikely to be

desirable, since it leads to parameters that are not well-identified. Appropriately allowing for the uncertainty in the parameter estimates is important, and this is generally not addressed in the hydrological literature. We do this here, when generating simulations, by sampling parameters from the multivariate lognormal distribution. In this way, rare, but potentially damaging scenarios should be better represented in simulations, particularly if, as here, extreme values tend to be underestimated by the model.

Although finding an optimal model structure is important to practitioners, of greater concern currently is the ability to allow for a changing climate. We have shown that a local generalised method of moments methodology offers a useful new approach to fitting point process-based rainfall models. With just two or three covariates, we can produce a model with better explanatory power than the current approach, with more realistic interannual variation, and the ability to generate simulations that reflect future climate change scenarios. Note that, while replacing the BLRP model with the optimal $BLRPR_X$ model reduces the mean weighted sum of squared errors over a validation set by around 2%–3%, replacing the default covariate, calendar month, with the optimal three continuous covariates (sea-level pressure, temperature and zonal wind velocity) can reduce errors by of the order of 25% (depending on the location and climate).

A key advantage over other types of rainfall model, is that the fitting process here allows various properties of rainfall to be related to large-scale atmospheric covariates, not just the mean rainfall or probability of occurrence, thus imposing fewer restrictions on future rainfall patterns. The methodology is flexible, so that the amount and variability of the data can be adjusted for, using a tuning parameter, ensuring an appropriate degree of smoothness in the relationships between the parameter estimates and the covariates. The framework also allows estimation of uncertainty.

7.2 Future Research

Throughout the thesis, we have highlighted possible areas for further investigation. These are summarised here, together with additional thoughts on potential future work. First we consider aspects relating to point process-based models in general, and then focus on research ideas relating specifically to the local fitting methodology.

7.2.1 Point process-based models

Point process-based rainfall models have evolved since their introduction by Rodriguez-Iturbe et al. (1987) via a mixture of theoretical and empirical developments. The recent research of Jesus & Chandler (2011) considered the theory behind the generalised method of moments fitting methodology, and focused in particular on the choice of the weighting

matrix. The optimal number and choice of moments together with model and moment selection criteria is another area that, in our view, merits further research.

As discussed in Chapter 2, the test for model misspecification, known as the 'Hansen J-test' or the 'test for overidentifying restrictions', is frequently used in the econometrics literature. If a model fails the test, then in theory at least, the moment conditions must be amended, or (typically) reduced in number, until the model is deemed correctly specified. In the context of our simple rainfall models, this is likely to lead to a model where the number of moment conditions exactly matches the number of unknown parameters (which is always correctly specified within the GMM framework), particularly where we have a reasonably large amount of data. This is unlikely to be considered desirable by practitioners, however. We know that our model is just a useful approximation for a specific purpose (typically the creation of artificial rainfall time series for testing hydrological designs), and the hydrological literature tends to suggest a preference for a model which fits a greater number of properties less well, over one that fits a small number exactly. In this case, as discussed in Section 7.1, we should acknowledge that the model is misspecified, and consider how this impacts on the asymptotic results and related inference, and consider potential moment and model selection criteria in this context. The paper by Hall & Inoue (2003), which discusses the impact of misspecification with four different options for the weighting matrix, may be a good starting point.

The greater difficulties with the local fitting encountered with the Heathrow dataset, compared with Bochum, were found to relate to the absence of five-minute statistics. Simulations at this timescale are also increasingly required, not just for hydrological applications such as urban drainage design, but also for radio telecommunications, which are adversely affected by heavy rainfall. Unfortunately, in practice, data at this timescale, and even at an hourly timescale, are not often available. As discussed in Section 1.3, various methods exist to disaggregate or downscale data to a shorter timescale. The use of scaling relationships (Kilsby et al. 2007, Chandler et al. 2007b) is particularly appropriate if, as for the local modelling, we require statistics at this level prior to local fitting. Investigating whether suitable relationships can be found down the the sub-hourly level is therefore another useful area for further research.

7.2.2 Potential improvements to our local fitting methodology

We decided to use a global bandwidth in our empirical work, on the basis that our data were unlikely to be sufficient to justify a local approach, and we did not expect the relationships to be unduly complex. While this appeared adequate for the Bochum data, on which most of our analysis was carried out, the other datasets led us to question whether this was indeed the most appropriate option. The problem is that, in order to avoid high variability and parameter identification problems near to the boundaries, a global bandwidth may

force us to have a high bandwidth across the whole range. Close to the boundaries, not only is there an issue because neighbouring points lie only to one side, but the density of observation points also tends to be very sparse, making estimates highly variable, and leading to identification problems. As discussed in Section 6.6, this could be addressed by allowing the bandwidth to vary locally in a limited way in order to overcome this problem. However, a better and simpler solution might be to replace the global bandwidth with a global span. Here, a fixed proportion of the data is given a positive weight at each evaluation point, which therefore gives a broadly constant variance (albeit at the expense of more variable bias). A compactly supported kernel function is required, such as the tri-cube or Epanechnikov kernels. This alternative method of adjusting the smoothness should be investigated.

Another practical issue that requires further investigation, relates to the numerical optimisations. In order to ensure that a global, rather than a local, minimum is found, it is usual practice to try different starting values. Here, in order to reduce computation time and ensure a smooth curve, we did this only in respect of the fit at the first evaluation point, with subsequent points fitted using just a single optimisation, with starting values based on the estimates from a neighbouring point. The risk with this approach is that, once an undesirable region of the parameter space is entered, later fits will tend to be similar. In practice, this is likely to occur close to the boundaries, where a sparsity of design points may lead to poorly identified parameters. Therefore, if the first evaluation point is chosen appropriately, it should not cause significant problems, particularly if, as discussed, we use a global span rather than a bandwidth. However, some minor adjustments to the fitting approach might be desirable, for example trying parameter estimates from a small number of previous fits as starting values, rather than just using a single set. Another related issue here is the choice of the first evaluation point, and the order in which subsequent fits are undertaken. While this is straightforward in the case of a univariate predictor, it is less so in the multivariate case. Our approach has been to base these on an arbitrarily chosen single covariate, but other approaches may be preferable. One idea might be to use the first principal component (with suitably standardised covariates) for this purpose, so that we move between evaluation points in the direction of maximum variability.

We also made an arbitrary choice of a single covariate when calculating smoothed sample variances, for the purpose of deriving a weighting matrix. Here too, it should be relatively straightforward to improve the methodology in the case of multivariate predictors. For example, this could be done by applying K-means clustering to the (standardised) covariates, and calculating a sample variance in respect of each cluster. The conditional variance in respect of each individual evaluation point could then be derived using a local mean estimator, assuming that the cluster sample variances apply to the mean point of each cluster.

In addition to these suggested potential improvements, another area that might warrant

some further thought is the visual representation of results, which is particularly problematic with multiple dimensions, particularly where we wish to demonstrate levels of uncertainty. Given the models are intended to be applied in practice by non-statisticians, it is particularly important that model output is readily interpretable.

7.2.3 Practical application

So far we have only considered reanalysis data, as we have focused on developing the methodology. The ultimate aim, however, is to apply this methodology to data projected by climate models, whereby the local point process-based model is trained on the reanalysis data, but fitted at evaluation points given by the climate model output. Most of the issues involved, such as climate model biases and uncertainties, different grid scales between the reanalysis data and the climate model etc., are not particular to our model, and research already exists in the literature which should inform these aspects. However, it is likely that other issues relating to climate model output data will also arise and require consideration. One aspect that is clearly a problem is that we have very sparse data at the boundaries, but it is likely that these areas of the design density will be encountered more often in a future climate, for example the upper end of the temperature range. A possible way of addressing this is to use relationships found across a variety of locations and climates, to inform the extrapolation at a given location. The methodology may also be useful for providing simulations at a location for which no suitable rainfall series exists, by using large scale atmospheric data from that location, coupled with a model fitted at an alternative location, with a similar climate.

In this thesis we have focused purely on temporal models i.e those fitted to a single site. Such models are appropriate given the sort of urban areas for which fine-scale models are generally required. However, much hydrological design relates to much more extensive regions, for which models with a spatial dimension are required. The point process-based rainfall models may readily be extended to the spatio-temporal domain and fitted to multiple rain-gauges, following the approach of Cowpertwait (1995) or Chapter 5 of Wheater et al. (2000). In these models, the temporal structure replicates that of the corresponding single-site model, and the same fitting methodology is used, so it should be relatively straightforward to extend our methodology to these models and to incorporate the relationships between the temporal parameters and the large-scale atmospheric variables. In the simplest case (circular cells moving with zero velocity), there is just one additional parameter: the mean cell radius, which could be assumed constant or also related to covariates.

7.2.4 Curse of dimensionality

Finally, we consider the curse of dimensionality, which is a significant constraint for this type of model, limiting the number of covariates that we can reasonably incorporate to a maximum of three. Two possible approaches could be used to address this issue. The first, simpler alternative is to use some form of dimension reduction, such as principal components analysis or factor analysis, on the desired set of covariates, so that exactly the same fitting approach is used as before, and we are still limited to the same number of covariates. It is not clear how effective this would be. Combining predictors with different units of measurement is not appealing, nor are the created predictors likely to be easy to interpret. However, this approach could be used in a limited way, such that these disadvantages are minimised: for example by combining the same predictor at a range of pressure levels, or in respect of several neighbouring grid-squares, in order to capture more information.

The alternative approach is to impose some additional structure on the model itself. Our original intention, before starting this research, was to use the local fitting as an exploratory tool, ultimately assuming parametric forms for the relationships. However, we have reached the conclusion that this is not appropriate as it requires many global parameters, even assuming very simple relationships. We prefer the more flexible nonparametric approach, with a small number of local parameters. In the case of local regression, an approach that adds some structure, whilst still retaining flexibility, is the additive model (Hastie & Tibshirani 1990). In its simplest form, the additive model assumes that there are no interactions between the covariates, and the equation has the form:

$$Y = \alpha + \theta_1(x_1) + \ldots \theta_d(x_d) + \epsilon,$$

for some response vector Y, and d-dimensional covariate, $X = (X_1, \ldots X_d)^{\mathrm{T}}$, with ϵ representing the error. The θ_j are flexible functions, which may be nonparametric. In order to ensure identifiability, the functions are standardised so that $\mathrm{E}[\theta_j(X_j)] = 0$. It can be seen that this is essentially an extension of multiple regression, since the contribution of each covariate is assumed to be additive, but here the relationships between the response and the covariates are no longer constrained to be linear. Limited interaction terms can be added, if desired, by making some of the θ_j functions of more than one covariate. This type of model is solved via a 'back-fitting' algorithm. At each iteration we fit a univariate regression in respect of one of the covariates, cycling through the covariates in turn, and continuing until convergence. Each univariate regression takes the residual after fitting all the other covariates as the response. So, for the lth covariate, for example, we fit $Y - \sum_{j \neq l} \theta_j^*(x_j) = \theta_l(x_l) + \epsilon$, where θ^* denotes estimates from the previous iteration.

In the same way that the generalised linear model (GLM) extends linear regression to cases where the response is not normally distributed, so the generalised additive model

(GAM) is an extension of the additive model. The GLM has greater similarity with our scenario than normal regression, because its equations also cannot be solved explicitly, and a Newton-type iterative method must be used. For the GAM, fitting requires two calculation loops: the inner carrying out the backfitting algorithm, and the outer the Newton-type iteration. A similar approach should be possible for our model, although the simplifications that arise from the exponential family assumption in GLMs will not apply, and so the process is expected to be more complex still.

An alternative idea that retains the assumption of additive covariate effects, but avoids the inner backfitting loop, is to use regression splines, as discussed briefly in Section 4.1.2. Although splines are more difficult to extend to multiple covariates than kernels in the general case, they are more straightforward if the constraint of additive effects is applied. A choice of knots and the corresponding set of B-splines would be required for each dimension, with the coefficients found by minimising the following objective function:

$$S = \frac{1}{n}\sum_{t=1}^{n}[T(y_t) - \tau(\theta(x_t))]^{\mathrm{T}}\, W_n(x_t)\, [T(y_t) - \tau(\theta(x_t))], \tag{7.1}$$

with $\theta(x_t) = a_0 + \sum_{l=1}^{d}\sum_{m=1}^{k_d} a_{lm}B_{lm}(x_{lt})$ where $B_{lm}(x_{lt})$ is the m^{th} of k_l B-splines for the l^{th} covariate, evaluated at the t^{th} data point, and a_{lm} denotes the respective coefficient (with a_0 the intercept). However, regression splines have some undesirable properties. In order to keep the dimensions reasonably low, a small number of knots is required, but poor placement of these can lead to misleading results, and strange local behaviour (Hastie & Tibshirani 1990). Splines with a large number of knots with penalties to reduce the dimensionality show better local behaviour. Eilers & Marx (1996) suggest that P-splines could also be used in this simple additive way, with a separate penalty introduced for each dimension, and this is an approach that is worth investigating further.

It should be noted, however, that our principal aim is for a modelling methodology that is useful to practitioners. Therefore, any additional complexity, particularly where it leads to significantly increased computation times, should be carefully considered against any additional benefit that it generates. Identifying ways to improve the computational speed in respect of any of the methods discussed (local mean or linear, fully flexible or additive versions) is of course, also a potentially useful area of future work.

Appendix A

Formulae for fitting properties

A.1 Introduction

The generalised method of moments fitting methodology requires a set of moments or 'fitting properties' in respect of each model. The set of properties used for all the models fitted here consists of the mean, and the dimensionless properties: the coefficient of variation, skewness coefficient and autocorrelation at lag 1. The dimensionless properties have been included at four levels of aggregation: 5 minutes, and 1, 6 and 24 hours.

In this Appendix, we give the formulae for the mean, variance, lag-1 autocovariance and 3rd central moment of the discrete-time aggregated process in respect of each of the Bartlett-Lewis model variants fitted in this thesis. Throughout, the timescale to which the continuous process is aggregated is denoted as h. The required fitting properties can be derived from those given here, as follows:

$$\text{Coefficient of variation} = \frac{\sqrt{\text{Var}[Y_i^h]}}{\text{E}[Y_i^h]}, \tag{A.1}$$

$$\text{Skewness coefficient} = \frac{\text{E}[(Y_i^h - \text{E}(Y_i^h))^3]}{\text{Var}[Y_i^h]^{3/2}}, \tag{A.2}$$

$$\text{Lag 1 autocorrelation} = \frac{\text{Cov}(Y_i^h, Y_{i+1}^h)}{\text{Var}[Y_i^h]}. \tag{A.3}$$

A number of other properties, not involved in the fitting, are also examined within the practical study in order to assess the goodness of fit of the models. These include the proportion of intervals with no rain, as well as wet and dry transition probabilities and extreme values. These properties are generally derived by simulation.

The random cell intensity (or random pulse depth in the case of the instantaneous pulse models), denoted X, has been assumed to have either a one or a two parameter distribution. In terms of the mathematical expressions for the fitting properties, an actual distribution does not need to be specified. However, since skewness, which involves terms in $\mathrm{E}(X^3)$, represents one of the fitting properties, we do at least need to specify the relationship between the 3rd moment and the two chosen parameters. Of course, the intensity distributions need to be fully specified when carrying out simulations. The intensity distributions considered, and their parameterisations have been as follows:

Gamma:

Parameters: μ_X, $r = \sigma_X/\mu_X$

Moments: $\mathrm{E}(X) = \mu_X$.

$\mathrm{E}(X^2) = (r^2 + 1)\mu_X^2$.

$\mathrm{E}(X^3) = (1 + 3r^2 + 2r^4)\mu_X^3$.

Note: the **Exponential** distribution is a special case of the Gamma, with r constrained to be 1. In terms of the usual parameterisation of a Gamma distribution with shape parameter α and rate parameter β, our parameters are given by $\mu_X = \alpha/\beta$ and $r = 1/\sqrt{\alpha}$. Therefore, the usual parameterisation would give: $\alpha = 1/r^2$ and $\beta = 1/(r^2\mu_X)$.

Weibull:

Parameters: μ_X, ϱ

Moments: $\mathrm{E}(X) = \mu_X$.

$\mathrm{E}(X^2) = \mu_X^2\ \Gamma(1 + 2/\varrho)\ /\ [\Gamma(1 + 1/\varrho)]^2$.

$\mathrm{E}(X^3) = \mu_X^3\ \Gamma(1 + 3/\varrho)\ /\ [\Gamma(1 + 1/\varrho)]^3$.

A.2 Bartlett-Lewis Rectangular Pulse (BLRP) model: Derivations

In this section we show the derivation of the first and second order fitting properties in respect of the BLRP model. This is included here as an example of the general approach used for the models, and does not constitute new theory. The properties were originally derived by Rodriguez-Iturbe et al. (1987), and the same approach is followed here, but with the addition of the detailed steps, omitted in the original.

Rainfall data are available only in aggregated form, so the properties are required in respect of rainfall totals over disjoint intervals of some fixed length, h say. In order to calculate these, however, we first need to derive the properties of the continuous rainfall process.

Notation

- λ - storm arrival rate
- β - cell arrival rate
- γ - exponential parameter of the storm duration
- η - exponential parameter of the cell duration
- μ_X - mean cell intensity
- $\mathrm{E}(X^2)$ - mean of squares of cell intensities
- $\mathrm{E}(X^3)$ - mean of cubes of cell intensities
- $\mu_C = 1 + \beta/\gamma$ - mean number of cells per storm

First and second order properties of the continuous process

Let $Y(t)$ be a random variable denoting the rainfall intensity at time t. Let X_{t-u} denote the intensity of a rain cell, beginning at time $t-u$, measured a time u later, and N denote the counting process of cell origins. The shorthand notation $\mathrm{d}N(t)$ denotes the count in the small interval of time $(t - \mathrm{d}t, t]$. We assume that the sequence of X's is independent of the counting process, and independent of each other, i.e. the X's are i.i.d.

Then, $Y(t)$ is given by the summation of the intensities of all cells active at time t, and can be expressed as:

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}\, \mathrm{d}N(t-u), \tag{A.4}$$

where:

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

The mean of the continuous process $Y(t)$ is given by:

$$\begin{aligned} \mathrm{E}[Y(t)] &= \int_{u=0}^{\infty} \mathrm{E}[X_{t-u}]\, \mathrm{E}[\mathrm{d}N(t-u)] \quad \text{(by independence of } \{X\} \text{ and } \{N\}) \\ &= \int_{u=0}^{\infty} \mu_X e^{-\eta u}\, \lambda \mu_C \mathrm{d}u \\ &= \frac{\lambda\, \mu_C \mu_X}{\eta}. \end{aligned} \tag{A.5}$$

The covariance function of the continuous process $Y(t)$ is given by:

$$\begin{aligned} \mathrm{Cov}(Y(t), Y(t+\tau)) &= \mathrm{Cov}\bigg(\int_{u=0}^{\infty} X_{t-u}(u)\mathrm{d}N(t-u), \int_{v=0}^{\infty} X_{t+\tau-v}(v)\mathrm{d}N(t+\tau-v) \bigg) \\ &= \int_{v=0}^{\infty} \int_{u=0}^{\infty} \mathrm{Cov}\big(X_{t-u}(u)\mathrm{d}N(t-u),\, X_{t+\tau-v}(v)\mathrm{d}N(t+\tau-v) \big) \\ &= \int_{v=0}^{\infty} \int_{u=0}^{\infty} \Big\{ \mathrm{E}\big(X_{t-u}(u) X_{t+\tau-v}(v)\big)\, \mathrm{E}\big(\mathrm{d}N(t-u)\mathrm{d}N(t+\tau-v)\big) \\ &\qquad - \mathrm{E}\big(X_{t-u}(u)\big)\, \mathrm{E}\big(\mathrm{d}N(t-u)\big)\, \mathrm{E}\big(X_{t+\tau-v}(v)\big)\, \mathrm{E}\big(\mathrm{d}N(t+\tau-v)\big) \Big\}, \end{aligned}$$

again by independence of $\{X\}$ and $\{N\}$.

Now $\mathrm{E}\big(X_{t-u}(u)\, X_{t+\tau-v}(v)\big) = \mathrm{E}\big(X_{t-u}(u)\big)\, \mathrm{E}\big(X_{t+\tau-v}(v)\big)$ except for the set u, v such that $u = v - \tau$ (i.e. the cell intensities are independent unless they relate to the same cell). Therefore we can change $\mathrm{E}\big(X_{t-u}(u)\big)\, \mathrm{E}\big(X_{t+\tau-v}(v)\big)$ to $\mathrm{E}\big(X_{t-u}(u)\, X_{t+\tau-v}(v)\big)$ in the integral, since they are only distinct on a set of measure zero.

So:

$$\mathrm{Cov}(Y(t), Y(t+\tau)) = \int_{v=0}^{\infty} \int_{u=0}^{\infty} \mathrm{E}[X_{t-u}(u)\, X_{t+\tau-v}(v)]\, \mathrm{Cov}(\mathrm{d}N(t-u), \mathrm{d}N(t+\tau-v)). \tag{A.6}$$

Now,

$$
\begin{aligned}
&\mathrm{Cov}\{\mathrm{d}N(t-u), \mathrm{d}N(t+\tau-v)\} \\
&= \mathrm{E}[\mathrm{d}N(t-u)\,\mathrm{d}N(t+\tau-v)] - E[\mathrm{d}N(t-u)]\,E[\mathrm{d}N(t+\tau-v)] \\
&= P[\mathrm{d}N(t-u) = \mathrm{d}N(t+\tau-v) = 1] \\
&\quad - P[\mathrm{d}N(t-u) = 1]\,P[\mathrm{d}N(t+\tau-v) = 1] + o(\mathrm{d}u\,\mathrm{d}v) \\
&= \{\lambda\mu_C\,\delta(u+\tau-v) + (\lambda\mu_C)^2 + \lambda\mu_C\,\beta\mathrm{e}^{-\gamma|u+\tau-v|} - (\lambda\mu_C)^2\}\,\mathrm{d}u\mathrm{d}v + o(\mathrm{d}u\,\mathrm{d}v) \\
&= \{\lambda\mu_C\delta(u+\tau-v) + \lambda\mu_C\,\beta\mathrm{e}^{-\gamma|u+\tau-v|}\}\,\mathrm{d}u\mathrm{d}v + o(\mathrm{d}u\,\mathrm{d}v), \qquad \text{(A.7)}
\end{aligned}
$$

where we use the stationarity of the process, so that the covariance does not depend on t. In the third step, the first term relates to the case where $u = v - \tau$, and $\delta(z)$ is the Dirac delta function, such that $\delta(z)$ is zero everywhere except at $z = 0$ and integrates to 1. The second term relates to two cell origins from separate storms which are therefore placed independently, whereas the third term relates to two cell origins from the same storm, in which case the rate at a time $|u+\tau-v|$ after the first origin is β as long as the storm has not yet terminated, which has probability $\mathrm{e}^{-\gamma|u+\tau-v|}$.

Substituting the results of Equation (A.7) into Equation (A.6):

$$
\begin{aligned}
c(\tau) &\equiv \mathrm{Cov}(Y(t), Y(t+\tau)) \\
&= \int_0^\infty E(X^2)e^{-\eta(u+\tau)}\lambda\mu_C\mathrm{d}u + \lambda\mu_C\beta\int_0^\infty\int_{v=0}^{\tau+u}\mu_X^2 e^{-\eta u}e^{-\eta v}e^{-\gamma(\tau+u-v)}\mathrm{d}v\mathrm{d}u \\
&\quad + \lambda\mu_C\beta\int_0^\infty\int_{v=\tau+u}^\infty\mu_X^2 e^{-\eta u}e^{-\eta v}e^{\gamma(\tau+u-v)}\mathrm{d}v\mathrm{d}u \\
&= \frac{\lambda}{\eta}\mu_C E(X^2)e^{-\eta\tau} + \frac{\lambda\mu_C\beta\mu_X^2}{\gamma-\eta}\int_0^\infty(e^{-2\eta u-\eta\tau} - e^{-(\eta+\gamma)u-\gamma\tau})\mathrm{d}u \\
&\quad + \frac{\lambda\mu_C\beta\mu_X^2}{\gamma+\eta}\int_0^\infty e^{-2\eta u-\eta\tau}\mathrm{d}u \\
&= \frac{\lambda}{\eta}\mu_C E(X^2)e^{-\eta\tau} + \frac{\lambda\mu_C\beta\mu_X^2}{\gamma-\eta}\left[\frac{e^{-\eta\tau}}{2\eta} - \frac{e^{-\gamma\tau}}{\gamma+\eta}\right] + \frac{\lambda\mu_C\beta\mu_X^2}{\gamma+\eta}\left[\frac{e^{-\eta\tau}}{2\eta}\right] \\
&= \frac{\lambda}{\eta}\mu_C E(X^2)e^{-\eta\tau} + \frac{\lambda\mu_C\beta\mu_X^2 e^{-\eta\tau}\gamma}{\eta(\gamma^2-\eta^2)} - \frac{\lambda\mu_C\beta\mu_X^2 e^{-\gamma\tau}}{\gamma^2-\eta^2} \\
&= \frac{\lambda\mu_C}{\eta}\left\{E(X^2)e^{-\eta\tau} + \frac{\beta\mu_X^2\gamma e^{-\eta\tau}}{\gamma^2-\eta^2} - \frac{\beta\mu_X^2\eta e^{-\gamma\tau}}{\gamma^2-\eta^2}\right\}. \qquad \text{(A.8)}
\end{aligned}
$$

First and second order properties of the aggregated process: generic equations

Now let Y_i^h be the total aggregated rainfall over the ith interval of length h. First we derive the generic formulae:

$$\begin{aligned} \mathrm{E}(Y_i^{(h)}) &= \int_{(i-1)h}^{ih} \mathrm{E}(Y(w))\mathrm{d}w \\ &= h\,\mathrm{E}(Y(t)). \end{aligned} \tag{A.9}$$

$$\begin{aligned} \mathrm{Var}(Y_i^{(h)}) &= \mathrm{Cov}(Y_1^{(h)}, Y_1^{(h)}) \\ &= \int_0^h \int_0^h \mathrm{Cov}(Y(s), Y(t))\,\mathrm{d}s\,\mathrm{d}t \\ &= 2\int_0^h \int_0^t \mathrm{Cov}(Y(s), Y(t))\,\mathrm{d}s\,\mathrm{d}t \quad \text{(by symmetry)} \\ &= 2\int_0^h \int_0^t \mathrm{Cov}(Y(t-w), Y(t))\,\mathrm{d}w\,\mathrm{d}t \quad \text{(substituting w = t - s)} \\ &= 2\int_0^h \int_w^h \mathrm{Cov}(Y(t-w), Y(t))\,\mathrm{d}t\,\mathrm{d}w \\ &= 2\int_0^h (h-w)\,c(w)\mathrm{d}w \quad \text{(by stationarity).} \end{aligned} \tag{A.10}$$

$$\begin{aligned} \mathrm{Cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) &= \mathrm{Cov}(Y_1^{(h)}, Y_{1+k}^{(h)}) \qquad \text{for } k \geq 1 \\ &= \int_{kh}^{(k+1)h} \int_0^h \mathrm{Cov}(Y(s), Y(t))\,\mathrm{d}s\,\mathrm{d}t \\ &= \int_{kh}^{(k+1)h} \int_{t-h}^{t} c(w)\mathrm{d}w\,\mathrm{d}t \quad \text{(substituting } w = t - s) \\ &= \int_{(k-1)h}^{kh} \int_{kh}^{w+h} c(w)\mathrm{d}t\,\mathrm{d}w + \int_{kh}^{(k+1)h} \int_{w}^{(k+1)h} c(w)\mathrm{d}t\,\mathrm{d}w \\ &= \int_{(k-1)h}^{kh} (w+h-kh)\,c(w)\mathrm{d}w + \int_{kh}^{(k+1)h} ((k+1)h-w)\,c(w)\mathrm{d}w \\ &= \int_{-h}^{0} (h+z)\,c(z+kh)\,\mathrm{d}z + \int_0^h (h-z)\,c(z+kh)\,\mathrm{d}z \\ &\qquad \text{(substituting } z = w - kh). \end{aligned} \tag{A.11}$$

First and second order properties of the aggregated process: BLRP equations

Now, substituting the results of Equation (A.5) into Equation (A.9), and Equation (A.8) into Equations (A.10) and (A.11) gives the first and second order moments of the aggre-

gated process. So, we have:

$$\begin{aligned} \mathrm{E}(Y_i^{(h)}) &= h\,\mathrm{E}(Y(t)) \\ &= \frac{\lambda\,\mu_X\,\mu_C\,h}{\eta}. \end{aligned} \tag{A.12}$$

$$\begin{aligned} \mathrm{Var}(Y_i^{(h)}) &= 2\int_0^h (h-w)\,c(w)\mathrm{d}w \\ &= \frac{2\lambda\mu_C}{\eta}\int_0^h (h-w)\left\{E(X^2)e^{-\eta w} + \frac{\beta\mu_X^2\gamma e^{-\eta w}}{\gamma^2-\eta^2} - \frac{\beta\mu_X^2\eta e^{-\gamma w}}{\gamma^2-\eta^2}\right\}\mathrm{d}w. \end{aligned} \tag{A.13}$$

Integrating by parts:

$$\begin{aligned} \int_0^h (h-w)e^{-\eta w}\mathrm{d}w &= \frac{1}{\eta}\Big[-(h-w)e^{-\eta w} + \frac{1}{\eta}e^{-\eta w}\Big]_0^h \\ &= \frac{h}{\eta} + \frac{1}{\eta^2}(e^{-\eta h}-1). \end{aligned} \tag{A.14}$$

Similarly,

$$\int_0^h (h-w)e^{-\gamma w}\mathrm{d}w = \frac{h}{\gamma} + \frac{1}{\gamma^2}(e^{-\gamma h}-1). \tag{A.15}$$

So, substituting the results of Equations (A.14) and (A.15) into Equation (A.13):

$$\begin{aligned} \mathrm{Var}(Y_i^{(h)}) &= \frac{2\lambda\mu_C}{\eta}\left\{\left(E(X^2) + \frac{\beta\mu_X^2\gamma}{\gamma^2-\eta^2}\right)\left(\frac{h}{\eta} + \frac{(e^{-\eta h}-1)}{\eta^2}\right) - \frac{\beta\mu_X^2\eta}{\gamma^2-\eta^2}\left(\frac{h}{\gamma} + \frac{(e^{-\gamma h}-1)}{\gamma^2}\right)\right\} \\ &= \frac{2\lambda\mu_C}{\eta}\left[\frac{(\mathrm{E}(X^2) + \beta\mu_X^2/\gamma)h}{\eta} + \frac{\beta\mu_X^2\eta(1-e^{-\gamma h})}{\gamma^2(\gamma^2-\eta^2)}\right. \\ &\qquad \left. - \left(\mathrm{E}(X^2) + \frac{\beta\gamma\mu_X^2}{\gamma^2-\eta^2}\right)\left(\frac{1-e^{-\eta h}}{\eta^2}\right)\right]. \end{aligned} \tag{A.16}$$

Similarly,

$$\begin{aligned} \mathrm{Cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) &= \int_{-h}^0 (h+z)\,c_Y(z+kh)\,\mathrm{d}z + \int_0^h (h-z)\,c_Y(z+kh)\,\mathrm{d}z \\ &= \frac{\lambda\mu_C}{\eta}\left[E(X^2) + \frac{\beta\mu_X^2\gamma}{\gamma^2-\eta^2}\right]\left\{\int_{-h}^0 (h+z)\,e^{-\eta(z+kh)} + \int_0^h (h-z)\,e^{-\eta(z+kh)}\right\}\mathrm{d}z \\ &\quad - \frac{\lambda\mu_C\beta\mu_X^2}{\gamma^2-\eta^2}\left\{\int_{-h}^0 (h+z)\,e^{-\gamma(z+kh)} + \int_0^h (h-z)\,e^{-\gamma(z+kh)}\right\}\mathrm{d}z. \end{aligned} \tag{A.17}$$

Integrating by parts, we have:

$$\begin{aligned}\int_{-h}^{0}(h+z)\,e^{-\eta(z+kh)}\mathrm{d}z &= \left[-\frac{h+z}{\eta}e^{-\eta(z+kh)}-\frac{1}{\eta^2}e^{-\eta(z+kh)}\right]_{-h}^{0}\\ &= -\frac{h}{\eta}e^{-\eta kh}-\frac{1}{\eta^2}e^{-\eta kh}+\frac{1}{\eta^2}e^{-\eta(k-1)h},\end{aligned}$$

and:

$$\begin{aligned}\int_{0}^{h}(h-z)\,e^{-\eta(z+kh)}\mathrm{d}z &= \left[-\frac{h-z}{\eta}e^{-\eta(z+kh)}+\frac{1}{\eta^2}e^{-\eta(z+kh)}\right]_{0}^{h}\\ &= \frac{1}{\eta^2}e^{-\eta(k+1)h}+\frac{h}{\eta}e^{-\eta kh}-\frac{1}{\eta^2}e^{-\eta kh}.\end{aligned}$$

And so:

$$\begin{aligned}\int_{-h}^{0}(h+z)\,e^{-\eta(z+kh)}\mathrm{d}z+\int_{0}^{h}(h-z)\,e^{-\eta(z+kh)}\mathrm{d}z &= -\frac{2}{\eta^2}e^{-\eta kh}+\frac{1}{\eta^2}e^{-\eta(k-1)h}+\frac{1}{\eta^2}e^{-\eta(k+1)h}\\ &= \frac{e^{-\eta(k-1)h}}{\eta^2}(1-e^{-\eta h})^2. \qquad \text{(A.18)}\end{aligned}$$

Similarly, with γ replacing η:

$$\int_{-h}^{0}(h+z)\,e^{-\gamma(z+kh)}\mathrm{d}z+\int_{0}^{h}(h-z)\,e^{-\gamma(z+kh)}\mathrm{d}z=\frac{e^{-\gamma(k-1)h}}{\gamma^2}(1-e^{-\gamma h})^2. \qquad \text{(A.19)}$$

Finally, substituting the results of Equations (A.18) and (A.19) into Equation (A.17):

$$\begin{aligned}\mathrm{Cov}(Y_i^{(h)},Y_{i+k}^{(h)}) = \frac{\lambda\mu_C}{\eta}\Bigg[&\left(\mathrm{E}(X^2)+\frac{\beta\gamma\mu_X^2}{\gamma^2-\eta^2}\right)\left(\frac{(1-e^{-\eta h})^2e^{-\eta(k-1)h}}{\eta^2}\right)\\ &-\frac{\mu_X^2\beta\eta(1-e^{-\gamma h})^2e^{-\gamma(k-1)h}}{\gamma^2(\gamma^2-\eta^2)}\Bigg] \qquad \text{(A.20)}\end{aligned}$$

A similar, but much lengthier calculation is required for the 3rd central moment. The final properties are restated in Appendix A.3, which also includes the 3rd central moment.

A.3 Bartlett-Lewis Rectangular Pulse (BLRP) model: Fitting Properties

Sources: The mean, variance, and covariance were derived by Rodriguez-Iturbe et al. (1987). The 3rd central moment is derived from Wheater et al. (2006)'s 3rd moment about zero.

Notation

- λ - storm arrival rate
- β - cell arrival rate
- γ - exponential parameter of the storm duration
- η - exponential parameter of the cell duration
- μ_X - mean cell intensity
- $\mathrm{E}(X^2)$ - mean of squares of cell intensities
- $\mathrm{E}(X^3)$ - mean of cubes of cell intensities
- $\mu_C = 1 + \beta/\gamma$ - mean number of cells per storm

Mean

$$\mathrm{E}[Y_i^h] = \frac{\lambda\,\mu_X\,\mu_C\,h}{\eta}. \tag{A.21}$$

Variance

$$\begin{aligned}\mathrm{Var}[Y_i^h] = \frac{2\lambda\mu_C}{\eta}\Bigg[&\frac{(\mathrm{E}(X^2)+\beta\mu_X^2/\gamma)h}{\eta} + \frac{\mu_X^2\beta\eta(1-e^{-\gamma h})}{\gamma^2(\gamma^2-\eta^2)} \\ &- \left(\mathrm{E}(X^2)+\frac{\beta\gamma\mu_X^2}{\gamma^2-\eta^2}\right)\left(\frac{1-e^{-\eta h}}{\eta^2}\right)\Bigg].\end{aligned} \tag{A.22}$$

Covariance at lag $k \geq 1$

$$\begin{aligned}&\mathrm{Cov}(Y_i^h, Y_{i+k}^h) \\ &= \frac{\lambda\mu_C}{\eta}\left[\left(\mathrm{E}(X^2)+\frac{\beta\gamma\mu_X^2}{\gamma^2-\eta^2}\right)\left(\frac{(1-e^{-\eta h})^2 e^{-\eta(k-1)h}}{\eta^2}\right) - \frac{\mu_X^2\beta\eta(1-e^{-\gamma h})^2 e^{-\gamma(k-1)h}}{\gamma^2(\gamma^2-\eta^2)}\right].\end{aligned} \tag{A.23}$$

3rd central moment

$$
\begin{aligned}
&\mathrm{E}[(Y_i^h - \mathrm{E}(Y_i^h))^3] \\
&= \frac{\lambda\mu_c}{(\eta^2+2\gamma\eta+\gamma^2)(\gamma^4-2\eta\gamma^3-3\eta^2\gamma^2+8\eta^3\gamma-4\eta^4)\eta^4\gamma^3} \\
&\times \Big\{ 12\gamma^7\mu_X^3\beta^2 e^{-\eta h} - 48\mu_X^3 e^{-\gamma h}\beta^2\eta^7 + 72\gamma^7\mathrm{E}(X^3)\eta^2 + 48\gamma\mu_X\mathrm{E}(X^2)\beta\eta^7 \\
&- 24\gamma h\mu_X^3\eta^7\beta^2 + 24\mu_X\mathrm{E}(X^2)\gamma^4 h e^{-\eta h}\beta\eta^5 + 24\mu_X\mathrm{E}(X^2)\gamma^2\beta\eta^6 - 36\mu_X\mathrm{E}(X^2)\gamma^3\beta\eta^5 \\
&- 72\mu_X\mathrm{E}(X^2)\gamma^6 h\beta\eta^3 - 84\gamma^2\mu_X^3\beta^2\eta^5 - 36\gamma^5 h\mu_X^3\beta^2\eta^3 + 54\gamma^3 h\mu_X^3\eta^5\beta^2 \\
&+ 6\mu_X\mathrm{E}(X^2)\gamma^5\beta\eta^3 + 6\gamma^7 h\mu_X^3\eta\beta^2 - 24\mu_X\mathrm{E}(X^2)\gamma^2 e^{-\eta h}\beta\eta^6 + 117\mu_X\mathrm{E}(X^2)\gamma^6\beta\eta^2 \\
&- 18\gamma^4\mu_X^3 e^{-\eta h}\beta^2\eta^3 + 54\gamma^5 h\mathrm{E}(X^3)\eta^5 + 39\gamma^5\mu_X^3\beta^2\eta^2 - 36\gamma^7 h\mathrm{E}(X^3)\eta^3 \\
&- 24\gamma^3 h\mathrm{E}(X^3)\eta^7 - 12\gamma^9\mathrm{E}(X^3) + 6\eta\gamma^9 h\mathrm{E}(X^3)e^{-\eta h} \\
&+ 6\mu_X\mathrm{E}(X^2)\gamma^4 e^{-\gamma h}\beta\eta^4 - 30\mu_X\mathrm{E}(X^2)\gamma^6 h e^{-\eta h}\beta\eta^3 - 48\mu_X\mathrm{E}(X^2)\gamma^2 h\beta\eta^7 \\
&- 48\gamma\mu_X\mathrm{E}(X^2)e^{-\gamma h}\beta\eta^7 - 24\gamma h\mu_X^3 e^{-\gamma h}\eta^7\beta^2 + 6\gamma^5\mu_X^3 e^{-\gamma h}\beta^2\eta^2 \\
&+ 6\mu_X\mathrm{E}(X^2)\gamma^8 h e^{-\eta h}\beta\eta - 138\mu_X\mathrm{E}(X^2)\gamma^4\beta\eta^4 + 48\mu_X^3\beta^2\eta^7 + 30\gamma^3 h\mu_X^3 e^{-\gamma h}\eta^5\beta^2 \\
&- 24\mu_X\mathrm{E}(X^2)\gamma^2 e^{-\gamma h}\beta\eta^6 + 36\mu_X\mathrm{E}(X^2)\gamma^3 e^{-\gamma h}\beta\eta^5 + 24\mu_X\mathrm{E}(X^2)\gamma^3 e^{-\eta h}\beta\eta^5 \\
&- 132\mu_X\mathrm{E}(X^2)\gamma^6 e^{-\eta h}\beta\eta^2 - 6\mu_X\mathrm{E}(X^2)\gamma^5 e^{-\gamma h}\beta\eta^3 + 150\mu_X\mathrm{E}(X^2)\gamma^4 e^{-\eta h}\beta\eta^4 \\
&- 42\gamma^5\mu_X^3 e^{-\eta h}\beta^2\eta^2 - 6\gamma^5 h\mu_X^3 e^{-\gamma h}\beta^2\eta^3 + 12\mu_X\mathrm{E}(X^2)\gamma^8 h\beta\eta \\
&- 6\mu_X\mathrm{E}(X^2)\gamma^5 e^{-\eta h}\beta\eta^3 - 24\mathrm{E}(X^2)\mu_X\beta\gamma^3 e^{-h(\eta+\gamma)}\eta^5 - 12\mathrm{E}(X^2)\mu_X\gamma^4\beta e^{-2\eta h}\eta^4 \\
&- 6\mathrm{E}(X^2)\mu_X\gamma^4\beta e^{-h(\eta+\gamma)}\eta^4 + 6\gamma^5\mathrm{E}(X^2)\mu_X\beta e^{-h(\eta+\gamma)}\eta^3 - 3\mathrm{E}(X^2)\mu_X\gamma^8\beta e^{-2\eta h} \\
&+ 24\mathrm{E}(X^2)\mu_X\beta\gamma^2 e^{-h(\eta+\gamma)}\eta^6 + 15\mathrm{E}(X^2)\mu_X\gamma^6\beta e^{-2\eta h}\eta^2 - 3\gamma^7\mu_X^3\beta^2 e^{-2\eta h} \\
&+ 18\gamma^4\mu_X^3\beta^2 e^{-h(\eta+\gamma)}\eta^3 - 12\gamma^3\mu_X^3\beta^2 e^{-h(\eta+\gamma)}\eta^4 - 6\gamma^5\mu_X^3\beta^2 e^{-h(\eta+\gamma)}\eta^2 \\
&+ 3\gamma^5\mu_X^3\beta^2 e^{-2\eta h}\eta^2 - 9\gamma^7\mu_X^3\beta^2 + 108\gamma^5\mathrm{E}(X^3)e^{-\eta h}\eta^4 + 48\gamma^3\mathrm{E}(X^3)\eta^6 \\
&- 72\gamma^7\mathrm{E}(X^3)e^{-\eta h}\eta^2 - 48\mathrm{E}(X^3)\gamma^3 e^{-\eta h}\eta^6 + 84\gamma^2\mu_X^3 e^{-\gamma h}\beta^2\eta^5 + 18\gamma^4\mu_X^3\beta^2\eta^3 \\
&+ 24\mathrm{E}(X^2)\mu_X\gamma^8\beta e^{-\eta h} + 54\gamma^5 h\mathrm{E}(X^3)e^{-\eta h}\eta^5 - 24\eta^7 h\mathrm{E}(X^3)\gamma^3 e^{-\eta h} \\
&- 36\eta^3\gamma^7 h\mathrm{E}(X^3)e^{-\eta h} - 21\gamma^8\mathrm{E}(X^2)\mu_X\beta + 6\gamma^9 h\mathrm{E}(X^3)\eta + 12\gamma^3\mu_X^3 e^{-\eta h}\beta^2\eta^4 \\
&+ 12\gamma^3\mu_X^3 e^{-\gamma h}\beta^2\eta^4 - 18\gamma^4\mu_X^3 e^{-\gamma h}\beta^2\eta^3 - 12\gamma^3\mu_X^3\beta^2\eta^4 - 108\gamma^5\mathrm{E}(X^3)\eta^4 \\
&+ 108\mu_X\mathrm{E}(X^2)\gamma^4 h\beta\eta^5 + 12\gamma^9\mathrm{E}(X^3)e^{-\eta h} \Big\}. \qquad \text{(A.24)}
\end{aligned}
$$

A.4 Random Parameter Bartlett-Lewis Rectangular Pulse (BLRPR) model

Sources: The mean, variance, and covariance were originally derived by Rodriguez-Iturbe et al. (1988). Here all moments are expressed by first re-parameterising the expressions for the BLRP model in Appendix A.3, then grouping together terms in $\eta^{-k}e^{-\eta s}$ for possible values of k and s, and taking expectations over η. The expectations are left in the form $E_\eta\left[\eta^{-k}\, e^{-\eta s}\right]$, which is convenient for coding. They may be evaluated as:

$$E_\eta\left[\eta^{-k}e^{-\eta s}\right] = \frac{\nu^\alpha}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha-k)}{(\nu+s)^{\alpha-k}}, \qquad \text{for } \alpha > k.$$

Notation

- λ - storm arrival rate
- α - shape parameter for the Gamma distribution of the cell duration parameter, η
- ν - scale parameter for the Gamma distribution of η
- κ - ratio of the cell arrival rate to η (i.e. β/η)
- ϕ - ratio of the storm (cell process) termination rate to η (i.e. γ/η)
- μ_X - mean cell intensity
- $\mathrm{E}(X^2)$ - mean of squares of cell intensities
- $\mathrm{E}(X^3)$ - mean of cubes of cell intensities
- $\mu_C = 1 + \kappa/\phi$ - mean number of cells per storm

Mean

$$\mathrm{E}[Y_i^h] = \lambda h \mu_x \mu_c E_\eta\left(\eta^{-1}\right). \tag{A.25}$$

Variance

$$\begin{aligned}\mathrm{Var}[Y_i^h] = 2\lambda\mu_c\Bigg\{&\mathrm{E}_\eta(\eta^{-2})\left\{\mathrm{E}(X^2) + \frac{\kappa\mu_x^2}{\phi}\right\}h + \mathrm{E}_\eta(\eta^{-3})\left\{\frac{\mu_x^2\kappa(1-\phi^3)}{\phi^2(\phi^2-1)} - \mathrm{E}(X^2)\right\} \\ &- \mathrm{E}_\eta(\eta^{-3}e^{-\phi\eta h})\frac{\mu_x^2\kappa}{\phi^2(\phi^2-1)} + \mathrm{E}_\eta(\eta^{-3}e^{-\eta h})\left\{\mathrm{E}(X^2) + \frac{\kappa\phi\mu_x^2}{\phi^2-1}\right\}\Bigg\}.\end{aligned} \tag{A.26}$$

Covariance at lag $k \geq 1$

$$\begin{aligned}\mathrm{Cov}(Y_i^h, Y_{i+k}^h) = \lambda\mu_c \Bigg\{ \left(\mathrm{E}(X^2) + \frac{\kappa\phi\mu_x^2}{\phi^2 - 1} \right) \Big[\mathrm{E}_\eta(\eta^{-3} e^{-\eta(k-1)h}) - 2\,\mathrm{E}_\eta(\eta^{-3} e^{-\eta k h}) \\ + \mathrm{E}_\eta(\eta^{-3} e^{-\eta(k+1)h}) \Big] - \frac{\mu_x^2 \kappa}{\phi^2(\phi^2 - 1)} \Big[\mathrm{E}_\eta(\eta^{-3} e^{-\phi\eta(k-1)h}) - 2\,\mathrm{E}_\eta(\eta^{-3} e^{-\phi\eta k h}) \\ + \mathrm{E}_\eta(\eta^{-3} e^{-\phi\eta(k+1)h}) \Big] \Bigg\}. \end{aligned} \tag{A.27}$$

3rd central moment

$$
\begin{aligned}
&\mathrm{E}_\eta[(Y_i^h - \mathrm{E}(Y_i^h))^3] \\
&= \frac{\lambda\mu_c}{(1+2\phi+\phi^2)(\phi^4-2\phi^3-3\phi^2+8\phi-4)\phi^3} \\
&\times \Bigg\{ \mathrm{E}_\eta\left[\eta^{-4}e^{-\eta h}\right] \Bigg(12\phi^7\mu_x^3\kappa^2 - 24\mu_x\mathrm{E}(X^2)\phi^2\kappa - 18\phi^4\mu_x^3\kappa^2 + 24\mu_x\mathrm{E}(X^2)\phi^3\kappa \\
&- 132\mu_x\mathrm{E}(X^2)\phi^6\kappa + 150\mu_x\mathrm{E}(X^2)\phi^4\kappa - 42\phi^5\mu_x^3\kappa^2 - 6\mu_x\mathrm{E}(X^2)\phi^5\kappa + 108\phi^5\mathrm{E}(X^3) \\
&- 72\phi^7\mathrm{E}(X^3) - 48\mathrm{E}(X^3)\phi^3 + 24\mathrm{E}(X^2)\mu_x\phi^8\kappa + 12\phi^3\mu_x^3\kappa^2 + 12\phi^9\mathrm{E}(X^3) \Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-3}e^{-\eta h}\right] \Bigg(24\mu_x\mathrm{E}(X^2)\phi^4 h\kappa + 6\phi^9 h\mathrm{E}(X^3) - 30\mu_x\mathrm{E}(X^2)\phi^6 h\kappa + 6\mu_x\mathrm{E}(X^2)\phi^8 h\kappa \\
&+ 54\phi^5 h\mathrm{E}(X^3) - 24h\mathrm{E}(X^3)\phi^3 - 36\phi^7 h\mathrm{E}(X^3) \Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-4}e^{-\eta\phi h}\right] \Bigg(-48\mu_x^3\kappa^2 + 6\mu_x\mathrm{E}(X^2)\phi^4\kappa - 48\phi\mu_x\mathrm{E}(X^2)\kappa + 6\phi^5\mu_x^3\kappa^2 \\
&- 24\mu_x\mathrm{E}(X^2)\phi^2\kappa + 36\mu_x\mathrm{E}(X^2)\phi^3\kappa - 6\mu_x\mathrm{E}(X^2)\phi^5\kappa + 84\phi^2\mu_x^3\kappa^2 + 12\phi^3\mu_x^3\kappa^2 - 18\phi^4\mu_x^3\kappa^2 \Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-3}e^{-\eta\phi h}\right] \Bigg(-24\phi h\mu_x^3\kappa^2 + 30\phi^3 h\mu_x^3\kappa^2 - 6\phi^5 h\mu_x^3\kappa^2 \Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-4}\right] \Bigg(72\phi^7\mathrm{E}(X^3) + 48\phi\mu_x\mathrm{E}(X^2)\kappa + 24\mu_x\mathrm{E}(X^2)\phi^2\kappa - 36\mu_x\mathrm{E}(X^2)\phi^3\kappa - 84\phi^2\mu_x^3\kappa^2 \\
&+ 6\mu_x\mathrm{E}(X^2)\phi^5\kappa + 117\mu_x\mathrm{E}(X^2)\phi^6\kappa + 39\phi^5\mu_x^3\kappa^2 - 12\phi^9\mathrm{E}(X^3) - 138\mu_x\mathrm{E}(X^2)\phi^4\kappa + 48\mu_x^3\kappa^2 \\
&- 9\phi^7\mu_x^3\kappa^2 + 48\phi^3\mathrm{E}(X^3) + 18\phi^4\mu_x^3\kappa^2 - 21\phi^8\mathrm{E}(X^2)\mu_x\kappa - 12\phi^3\mu_x^3\kappa^2 - 108\phi^5\mathrm{E}(X^3) \Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-3}\right] \Bigg(-24\phi h\mu_x^3\kappa^2 - 72\mu_x\mathrm{E}(X^2)\phi^6 h\kappa - 36\phi^5 h\mu_x^3\kappa^2 + 54\phi^3 h\mu_x^3\kappa^2 + 6\phi^7 h\mu_x^3\kappa^2 \\
&+ 54\phi^5 h\mathrm{E}(X^3) - 36\phi^7 h\mathrm{E}(X^3) - 24\phi^3 h\mathrm{E}(X^3) - 48\mu_x\mathrm{E}(X^2)\phi^2 h\kappa + 12\mu_x\mathrm{E}(X^2)\phi^8 h\kappa \\
&+ 6\phi^9 h\mathrm{E}(X^3) + 108\mu_x\mathrm{E}(X^2)\phi^4 h\kappa \Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-4}e^{-2\eta h}\right] \Bigg(-12\mathrm{E}(X^2)\mu_x\phi^4\kappa - 3\mathrm{E}(X^2)\mu_x\phi^8\kappa + 15\mathrm{E}(X^2)\mu_x\phi^6\kappa - 3\phi^7\mu_x^3\kappa^2 \\
&+ 3\phi^5\mu_x^3\kappa^2 \Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-4}e^{-\eta h(1+\phi)}\right] \Bigg(-24\mathrm{E}(X^2)\mu_x\kappa\phi^3 - 6\mathrm{E}(X^2)\mu_x\phi^4\kappa + 6\phi^5\mathrm{E}(X^2)\mu_x\kappa \\
&+ 24\mathrm{E}(X^2)\mu_x\kappa\phi^2 + 18\phi^4\mu_x^3\kappa^2 - 12\phi^3\mu_x^3\kappa^2 - 6\phi^5\mu_x^3\kappa^2 \Bigg) \Bigg\}. \qquad \text{(A.28)}
\end{aligned}
$$

A.5 Random Parameter Bartlett-Lewis Rectangular Pulse model with dependent intensity-duration (BLRPR$_X$)

Source: derived from the BLRPR model as described in Section 3.7. All expectations are left in the form $E_\eta\left[\eta^{-k}\, e^{-\eta s}\right]$ for various values of k and s, and may be evaluated as:

$$E_\eta\left[\eta^{-k}e^{-\eta s}\right] = \frac{\nu^\alpha}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha-k)}{(\nu+s)^{\alpha-k}}, \qquad \text{for } \alpha > k.$$

Notation

- λ - storm arrival rate
- α - shape parameter for the Gamma distribution of the cell duration parameter, η
- ν - scale parameter for the Gamma distribution of η
- κ - ratio of the cell arrival rate to η (i.e. β/η)
- ϕ - ratio of the storm (cell process) termination rate to η (i.e. γ/η)
- ι - ratio of mean cell intensity to η (i.e. μ_X/η)
- f_1 - $\mathrm{E}(X^2)/\mu_x^2$ (depends on distribution of X, see Appendix A.1)
- f_2 - $\mathrm{E}(X^3)/\mu_x^3$ (depends on distribution of X, see Appendix A.1)
- $\mu_C = 1 + \kappa/\phi$ - mean number of cells per storm

Mean

$$\mathrm{E}[Y_i^h] = \lambda h \iota \mu_c. \tag{A.29}$$

Variance

$$\begin{aligned}\mathrm{Var}[Y_i^h] = 2\lambda\mu_c\iota^2\Bigg\{\left\{f_1 + \frac{\kappa}{\phi}\right\}h + \mathrm{E}_\eta(\eta^{-1})\left\{\frac{\kappa(1-\phi^3)}{\phi^2(\phi^2-1)} - f_1\right\} \\ - \mathrm{E}_\eta(\eta^{-1}e^{-\phi\eta h})\frac{\kappa}{\phi^2(\phi^2-1)} + \mathrm{E}_\eta(\eta^{-1}e^{-\eta h})\left\{f_1 + \frac{\kappa\phi}{\phi^2-1}\right\}\Bigg\}.\end{aligned} \tag{A.30}$$

Covariance at lag $k \geq 1$

$$\begin{aligned}\mathrm{Cov}(Y_i^h, Y_{i+k}^h) = \lambda\mu_c\iota^2\Bigg\{\left(f_1 + \frac{\kappa\phi}{\phi^2-1}\right)\Big[\mathrm{E}_\eta(\eta^{-1}e^{-\eta(k-1)h}) - 2\,\mathrm{E}_\eta(\eta^{-1}e^{-\eta k h}) \\ + \mathrm{E}_\eta(\eta^{-1}e^{-\eta(k+1)h})\Big] - \frac{\kappa}{\phi^2(\phi^2-1)}\Big[\mathrm{E}_\eta(\eta^{-1}e^{-\phi\eta(k-1)h}) - 2\,\mathrm{E}_\eta(\eta^{-1}e^{-\phi\eta k h}) \\ + \mathrm{E}_\eta(\eta^{-1}e^{-\phi\eta(k+1)h})\Big]\Bigg\}.\end{aligned} \tag{A.31}$$

3rd central moment

$$
\begin{aligned}
&\mathrm{E}_\eta[(Y_i^h - \mathrm{E}(Y_i^h))^3] \\
&= \frac{\lambda\mu_c\iota^3}{(1+2\phi+\phi^2)(\phi^4-2\phi^3-3\phi^2+8\phi-4)\phi^3} \\
&\times \Bigg\{ \mathrm{E}_\eta\left[\eta^{-1}e^{-\eta h}\right]\Bigg(12\phi^7\kappa^2 - 24f_1\phi^2\kappa - 18\phi^4\kappa^2 + 24f_1\phi^3\kappa - 132f_1\phi^6\kappa + 150f_1\phi^4\kappa \\
&- 42\phi^5\kappa^2 - 6f_1\phi^5\kappa + 108\phi^5 f_2 - 72\phi^7 f_2 - 48\phi^3 f_2 + 24f_1\mu_x\phi^8\kappa + 12\phi^3\kappa^2 + 12\phi^9 f_2\Bigg) \\
&+ \mathrm{E}_\eta\left[e^{-\eta h}\right]\Bigg(24f_1\phi^4 h\kappa + 6\phi^9 h f_2 - 30f_1\phi^6 h\kappa + 6f_1\phi^8 h\kappa + 54\phi^5 h f_2 - 24h f_2\phi^3 - 36\phi^7 h f_2\Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-1}e^{-\eta\phi h}\right]\Bigg(- 48\kappa^2 + 6f_1\phi^4\kappa - 48\phi f_1\kappa + 6\phi^5\kappa^2 - 24f_1\phi^2\kappa + 36f_1\phi^3\kappa \\
&- 6f_1\phi^5\kappa + 84\phi^2\kappa^2 + 12\phi^3\kappa^2 - 18\phi^4\kappa^2\Bigg) \\
&+ \mathrm{E}_\eta\left[e^{-\eta\phi h}\right]\Bigg(- 24\phi h\kappa^2 + 30\phi^3 h\kappa^2 - 6\phi^5 h\kappa^2\Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-1}\right]\Bigg(72\phi^7 f_2 + 48\phi f_1\kappa + 24f_1\phi^2\kappa - 36f_1\phi^3\kappa - 84\phi^2\kappa^2 + 6f_1\phi^5\kappa + 117f_1\phi^6\kappa \\
&+ 39\phi^5\kappa^2 - 12\phi^9 f_2 - 138f_1\phi^4\kappa + 48\kappa^2 - 9\phi^7\kappa^2 + 48\phi^3 f_2 + 18\phi^4\kappa^2 - 21\phi^8 f_1\kappa \\
&- 12\phi^3\kappa^2 - 108\phi^5 f_2\Bigg) \\
&+ \Bigg(- 24\phi h\kappa^2 - 72f_1\phi^6 h\kappa - 36\phi^5 h\kappa^2 + 54\phi^3 h\kappa^2 + 6\phi^7 h\kappa^2 + 54\phi^5 h f_2 - 36\phi^7 h f_2 \\
&- 24\phi^3 h f_2 - 48f_1\phi^2 h\kappa + 12f_1\phi^8 h\kappa + 6\phi^9 h f_2 + 108f_1\phi^4 h\kappa\Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-1}e^{-2\eta h}\right]\Bigg(- 12f_1\phi^4\kappa - 3f_1\phi^8\kappa + 15f_1\phi^6\kappa - 3\phi^7\kappa^2 + 3\phi^5\kappa^2\Bigg) \\
&+ \mathrm{E}_\eta\left[\eta^{-1}e^{-\eta h(1+\phi)}\right]\Bigg(- 24f_1\kappa\phi^3 - 6f_1\phi^4\kappa + 6\phi^5 f_1\kappa + 24f_1\kappa\phi^2 + 18\phi^4\kappa^2 \\
&- 12\phi^3\kappa^2 - 6\phi^5\kappa^2\Bigg)\Bigg\}. \qquad \text{(A.32)}
\end{aligned}
$$

A.6 Bartlett-Lewis Instantaneous Pulse (BLIP) model

Source: Cowpertwait et al. (2007)

Notation

- λ - storm arrival rate
- β - cell arrival rate
- γ - exponential parameter of the storm duration
- η - exponential parameter of the cell duration
- ξ - pulse arrival rate
- μ_X - mean pulse depth
- $\mathrm{E}(X^2)$ - mean of squares of pulse depths
- $\mathrm{E}(X^3)$ - mean of cubes of pulse depths
- $\mathrm{E}(X_{ijk}X_{ijl})$ - product moment of the depths of 2 pulses within the same cell
- $\mathrm{E}(X_{ijk}X_{ijl}X_{ijm})$ - product moment of the depths of 3 pulses within the same cell
- $\mu_p = \frac{\beta\xi}{\gamma(\gamma+\eta)}$ - mean number of pulses per storm

Mean

$$\mathrm{E}[Y_i^h] = \lambda\mu_p\mu_X h. \tag{A.33}$$

Variance

$$\begin{aligned}\mathrm{Var}[Y_i^h] = \lambda\mu_p\Bigg\{&\mathrm{E}(X^2)h + 2\mu_X^2\beta\xi\left(\frac{e^{-\gamma h}-1+\gamma h}{\eta\gamma^2}\right)\\ &+ 2\xi\left[\mathrm{E}(X_{ijk}X_{ijl}) - \mu_X^2\beta\frac{\gamma}{\eta(\gamma+2\eta)}\right]\left(\frac{e^{-(\gamma+\eta)h}-1+(\gamma+\eta)h}{(\gamma+\eta)^2}\right)\Bigg\}.\end{aligned} \tag{A.34}$$

Covariance at lag $k \geq 1$

$$\begin{aligned}&\mathrm{Cov}(Y_i^h, Y_{i+k}^h)\\ &= \lambda\mu_p\xi\Bigg[\mu_X^2\beta\left(\frac{e^{-\gamma(k-1)h}(1-e^{-\gamma h})^2}{\eta\gamma^2}\right)\\ &+ \left(\mathrm{E}(X_{ijk}X_{ijl}) - \mu_X^2\beta\frac{\gamma}{\eta(\gamma+2\eta)}\right)\left(\frac{e^{-(\gamma+\eta)(k-1)h}(1-e^{-(\gamma+\eta)h})^2}{(\gamma+\eta)^2}\right)\Bigg].\end{aligned} \tag{A.35}$$

3rd central moment

$$
\begin{aligned}
&\mathrm{E}[(Y_i^h - \mathrm{E}(Y_i^h))^3] \\
&= \lambda\beta\xi^3\Bigg\{ \frac{6}{(\eta+\gamma)^2}\left[\frac{\mathrm{E}(X_{ijk}X_{ijl}X_{ijm})}{\gamma(\eta+\gamma)} + \frac{\mathrm{E}(X_{ijk}X_{ijl})\mu_X\beta}{\gamma(\eta+\gamma)}\left(\frac{1}{\eta} - \frac{\gamma}{\eta(2\eta+\gamma)}\right)\right. \\
&\qquad \left. - \frac{\mu_X^3\beta^2}{\eta^2(\eta+\gamma)(2\eta+\gamma)}\right]\left[h - \frac{2}{(\eta+\gamma)} + \frac{2e^{-(\eta+\gamma)h}}{(\eta+\gamma)} + he^{-(\eta+\gamma)h}\right] \\
&+ \frac{6}{(\eta+\gamma)(2\eta+\gamma)}\left[-\frac{2\mathrm{E}(X_{ijk}X_{ijl})\mu_X\beta}{\eta(\eta+\gamma)(2\eta+\gamma)} + \frac{\mu_X^3\beta^2}{\eta^2(2\eta+\gamma)(3\eta+\gamma)}\right] \\
&\qquad \times \left[h - \frac{3\eta+2\gamma}{(\eta+\gamma)(2\eta+\gamma)} + \frac{(2\eta+\gamma)e^{-(\eta+\gamma)h}}{\eta(\eta+\gamma)} - \frac{(\eta+\gamma)e^{-(2\eta+\gamma)h}}{\eta(2\eta+\gamma)}\right] \\
&+ \frac{6\mu_X^3\beta^2}{\eta^2\gamma^3(\eta+\gamma)}\left[h - \frac{2}{\gamma} + \frac{2e^{-\gamma h}}{\gamma} + he^{-\gamma h}\right] \\
&+ \frac{6}{\gamma(\eta+\gamma)}\left[\frac{2\mathrm{E}(X_{ijk}X_{ijl})\mu_X\beta}{\eta\gamma(\eta+\gamma)} - \frac{\mu_X^3\beta^2}{\eta^2(\eta+\gamma)(2\eta+\gamma)}\right]\left[h - \frac{\eta+2\gamma}{\gamma(\eta+\gamma)} + \frac{(\eta+\gamma)e^{-\gamma h}}{\eta\gamma}\right. \\
&\qquad \left. - \frac{\gamma e^{-(\eta+\gamma)h}}{\eta(\eta+\gamma)}\right] + \frac{6\mathrm{E}(X_{ijk}^2X_{ijl})}{\xi\gamma(\eta+\gamma)^2}\left[h - \frac{1}{\eta+\gamma} + \frac{e^{-(\eta+\gamma)h}}{\eta+\gamma}\right] \\
&+ \frac{6\mathrm{E}(X^2)\mu_X\beta}{\xi\eta\gamma^2(\eta+\gamma)}\left[h - \frac{1}{\gamma} + \frac{e^{-\gamma h}}{\gamma} - \frac{\gamma^2}{(\eta+\gamma)(2\eta+\gamma)}\left(h - \frac{1}{\eta+\gamma} + \frac{e^{-(\eta+\gamma)h}}{\eta+\gamma}\right)\right] \\
&+ \frac{\mathrm{E}(X^3)h}{\xi^2\gamma(\eta+\gamma)}\Bigg\}. \qquad \text{(A.36)}
\end{aligned}
$$

A.7 Random Parameter Bartlett-Lewis Instantaneous Pulse (BLIPR) model

Source: derived from the BLIP model as described in Section 3.4.

Notation

- λ - storm arrival rate
- α - shape parameter for the Gamma distribution of the cell duration parameter, η
- ν - scale parameter for the Gamma distribution of η
- κ - ratio of the cell arrival rate to η (i.e. β/η)
- ϕ - ratio of the storm (cell process) termination rate to η (i.e. γ/η)
- ω - ratio of the pulse arrival rate to η (i.e. ξ/η)
- μ_X - mean pulse depth
- $\mathrm{E}(X^2)$ - mean of squares of pulse depths
- $\mathrm{E}(X^3)$ - mean of cubes of pulse depths
- $\mathrm{E}(X_{ijk}X_{ijl})$ - product moment of the depths of 2 pulses within the same cell
- $\mathrm{E}(X_{ijk}X_{ijl}X_{ijm})$ - product moment of the depths of 3 pulses within the same cell
- $\mu_p = \frac{\kappa\omega}{\phi(\phi+1)}$ - mean number of pulses per storm

Mean

$$\mathrm{E}[Y_i^h] = \lambda\mu_p\mu_X h. \tag{A.37}$$

Variance

$$\begin{aligned}\mathrm{Var}[Y_i^h] = \lambda\mu_p\Bigg\{&\mathrm{E}(X^2)h + \frac{2\mu_X^2\kappa\omega}{\phi^2}\mathrm{E}_\eta\left(\frac{1}{\eta}e^{-\phi\eta h} - \frac{1}{\eta} + \phi h\right)\\ &+ \frac{2\omega}{(\phi+1)^2}\left[\mathrm{E}(X_{ijk}X_{ijl}) - \mu_X^2\kappa\frac{\phi}{\phi+2}\right]\mathrm{E}_\eta\left(\frac{1}{\eta}e^{-(\phi+1)\eta h} - \frac{1}{\eta} + (\phi+1)h\right)\Bigg\}\\ = \lambda\mu_p\Bigg\{&\mathrm{E}(X^2)h + \frac{2\mu_X^2\kappa\omega}{\phi^2}\left(\frac{\nu^\alpha}{(\alpha-1)(\nu+\phi h)^{\alpha-1}} - \frac{\nu}{\alpha-1} + \phi h\right)\\ &+ \frac{2\omega}{(\phi+1)^2}\left[\mathrm{E}(X_{ijk}X_{ijl}) - \mu_X^2\kappa\frac{\phi}{\phi+2}\right]\left(\frac{\nu^\alpha}{(\alpha-1)(\nu+(\phi+1)h)^{\alpha-1}} - \frac{\nu}{\alpha-1}\right.\\ &\left.+ (\phi+1)h\right)\Bigg\}.\end{aligned} \tag{A.38}$$

Covariance at lag $k \geq 1$

$$
\begin{aligned}
&\mathrm{Cov}(Y_i^h, Y_{i+k}^h) \\
&= \lambda\mu_p\omega\left[\frac{\mu_X^2\kappa}{\phi^2}\mathrm{E}_\eta\left(\frac{e^{-\phi\eta(k-1)h} - 2e^{-\phi\eta kh} + e^{-\phi\eta(k+1)h}}{\eta}\right)\right. \\
&\quad \left.+ \left(\mathrm{E}(X_{ijk}X_{ijl}) - \mu_X^2\kappa\frac{\phi}{(\phi+2)}\right)\mathrm{E}_\eta\left(\frac{e^{-(\phi+1)\eta(k-1)h} - 2e^{-(\phi+1)\eta kh} + e^{-(\phi+1)\eta(k+1)h}}{(1+\phi)^2\eta}\right)\right] \\
&= \lambda\mu_p\omega\left(\frac{\nu}{\alpha-1}\right)\left[\frac{\mu_X^2\kappa}{\phi^2}\left\{\left(\frac{\nu}{\nu+\phi(k-1)h}\right)^{\alpha-1}\right.\right. \\
&\quad \left.- 2\left(\frac{\nu}{\nu+\phi kh}\right)^{\alpha-1} + \left(\frac{\nu}{\nu+\phi(k+1)h}\right)^{\alpha-1}\right\} + \left(\mathrm{E}(X_{ijk}X_{ijl}) - \mu_X^2\kappa\frac{\phi}{(\phi+2)}\right) \\
&\quad \times\left\{\left(\frac{\nu}{\nu+(\phi+1)(k-1)h}\right)^{\alpha-1} - 2\left(\frac{\nu}{\nu+(\phi+1)kh}\right)^{\alpha-1}\right. \\
&\quad \left.\left.+ \left(\frac{\nu}{\nu+(\phi+1)(k+1)h}\right)^{\alpha-1}\right\}\right]. \qquad \text{(A.39)}
\end{aligned}
$$

3rd central moment

$$
\begin{aligned}
&\mathrm{E}[(Y_i^h - \mathrm{E}(Y_i^h))^3] \\
&= \lambda\kappa\omega^3 \Bigg\{ \frac{6}{(1+\phi)^3} \left[\frac{\mathrm{E}(X_{ijk}X_{ijl}X_{ijm})}{\phi} + \frac{2\mathrm{E}(X_{ijk}X_{ijl})\mu_X\kappa}{\phi(2+\phi)} - \frac{\mu_X^3\kappa^2}{(2+\phi)} \right] \\
&\times \left[h - \frac{2\nu}{(\alpha-1)(1+\phi)} + \frac{2\nu}{(1+\phi)(\alpha-1)} \left(\frac{\nu}{\nu+(1+\phi)h} \right)^{\alpha-1} + h \left(\frac{\nu}{\nu+(1+\phi)h} \right)^{\alpha} \right] \\
&+ \frac{6}{(1+\phi)(2+\phi)^2} \left[- \frac{2\mathrm{E}(X_{ijk}X_{ijl})\mu_X\kappa}{(1+\phi)} + \frac{\mu_X^3\kappa^2}{(3+\phi)} \right] \\
&\times \left[h - \frac{\nu}{(\alpha-1)} \left\{ \frac{3+2\phi}{(1+\phi)(2+\phi)} - \left(\frac{2+\phi}{1+\phi} \right) \left(\frac{\nu}{\nu+(1+\phi)h} \right)^{\alpha-1} \right. \right. \\
&\left. \left. + \left(\frac{1+\phi}{2+\phi} \right) \left(\frac{\nu}{\nu+(2+\phi)h} \right)^{\alpha-1} \right\} \right] \\
&+ \frac{6\mu_X^3\kappa^2}{\phi^3(1+\phi)} \left[h - \frac{2\nu}{\phi(\alpha-1)} + \frac{2\nu}{\phi(\alpha-1)} \left(\frac{\nu}{\nu+\phi h} \right)^{\alpha-1} + h \left(\frac{\nu}{\nu+\phi h} \right)^{\alpha} \right] \\
&+ \frac{6}{\phi(1+\phi)^2} \left[\frac{2\mathrm{E}(X_{ijk}X_{ijl})\mu_X\kappa}{\phi} - \frac{\mu_X^3\kappa^2}{(2+\phi)} \right] \\
&\times \left[h - \frac{\nu}{(\alpha-1)} \left\{ \frac{1+2\phi}{\phi(1+\phi)} - \frac{(1+\phi)}{\phi} \left(\frac{\nu}{\nu+\phi h} \right)^{\alpha-1} + \frac{\phi}{(1+\phi)} \left(\frac{\nu}{\nu+(1+\phi)h} \right)^{\alpha-1} \right\} \right] \\
&+ \frac{6\mathrm{E}(X_{ijk}^2X_{ijl})}{\omega\phi(1+\phi)^2} \left[h - \frac{\nu}{(1+\phi)(\alpha-1)} \left\{ 1 - \left(\frac{\nu}{\nu+(1+\phi)h} \right)^{\alpha-1} \right\} \right] \\
&+ \frac{6\mathrm{E}(X^2)\mu_X\kappa}{\omega\phi^2(1+\phi)} \left[h - \frac{\nu}{\phi(\alpha-1)} + \frac{\nu}{\phi(\alpha-1)} \left(\frac{\nu}{\nu+\phi h} \right)^{\alpha-1} - \frac{\phi^2}{(1+\phi)(2+\phi)} \right. \\
&\left. \times \left(h - \frac{\nu}{(1+\phi)(\alpha-1)} + \frac{\nu}{(1+\phi)(\alpha-1)} \left(\frac{\nu}{\nu+(1+\phi)h} \right)^{\alpha-1} \right) \right] + \frac{\mathrm{E}(X^3)h}{\omega^2\phi(1+\phi)} \Bigg\}.
\end{aligned}
\tag{A.40}
$$

Appendix B

Fitted parameters in respect of selected monthly models

In the tables below, we show the fitted parameter sets for selected models, fitted to a five minute time series of rainfall data from 1931 to 1999 from Bochum in Germany. A separate model is fitted in respect of each calendar month. The fitting properties are the hourly mean, plus the coefficient of variation, lag-1 autocorrelation and skewness at timescales of 5 minutes, 1 hour, 6 hours and 24 hours.

For each model, as well as the fitted parameters, we show a number of key properties, in order to allow a better comparison of models with different parameterisations. The acronyms used for these properties are given below:

MSIT	mean storm inter-arrival time, hours
MSD	mean duration of storm activity, hours
MCIT	mean cell inter-arrival time, minutes
MCD	mean cell duration, minutes
MCS	mean number of cells per storm ($= \mu_C$)
MPC	mean number of pulses per cell

	λ	μ_X	β	γ	η	MSIT	MSD	MCIT	MCD	MCS
Jan	0.022	0.960	5.422	0.231	5.975	45.0	4.3	11.1	10.0	24.5
Feb	0.021	0.942	5.142	0.260	5.310	47.1	3.8	11.7	11.3	20.7
Mar	0.021	1.334	4.478	0.262	7.061	47.2	3.8	13.4	8.5	18.1
Apr	0.022	1.944	3.829	0.271	8.387	45.7	3.7	15.7	7.2	15.1
May	0.023	3.662	3.157	0.370	9.239	44.3	2.7	19.0	6.5	9.5
Jun	0.025	6.431	2.694	0.413	11.154	39.2	2.4	22.3	5.4	7.5
Jul	0.023	10.136	1.672	0.356	12.011	43.5	2.8	35.9	5.0	5.7
Aug	0.023	7.072	2.411	0.408	11.066	43.4	2.5	24.9	5.4	6.9
Sep	0.021	5.306	2.945	0.379	10.470	47.1	2.6	20.4	5.7	8.8
Oct	0.019	2.209	4.071	0.275	8.104	53.3	3.6	14.7	7.4	15.8
Nov	0.023	1.207	5.884	0.276	6.741	42.8	3.6	10.2	8.9	22.3
Dec	0.024	1.059	5.475	0.265	5.906	41.1	3.8	11.0	10.2	21.7

Table B.1: Parameters for Bartlett-Lewis Rectangular Pulse model, with exponential intensity distribution.

	λ	μ_X	β	γ	η	ξ	MSIT	MSD	MCIT	MCD	MCS	MPC
Jan	0.023	0.013	0.220	0.078	1.166	124.9	43.0	12.9	272.2	51.5	2.8	100
Feb	0.025	0.008	1.387	0.239	2.547	182.7	39.7	4.2	43.3	23.6	5.8	66
Mar	0.022	0.020	0.188	0.079	1.393	97.8	44.7	12.7	319.8	43.1	2.4	66
Apr	0.024	0.033	0.209	0.094	1.684	77.3	41.0	10.7	287.6	35.6	2.2	43
May	0.028	0.038	1.452	0.420	5.696	144.1	35.8	2.4	41.3	10.5	3.5	24
Jun	0.033	0.086	1.237	0.488	6.101	100.8	30.0	2.1	48.5	9.8	2.5	15
Jul	0.032	0.141	0.707	0.423	6.558	100.9	30.8	2.4	84.9	9.1	1.7	14
Aug	0.031	0.095	1.042	0.477	6.023	103.2	32.2	2.1	57.6	10.0	2.2	16
Sep	0.027	0.068	1.355	0.442	5.826	105.3	37.1	2.3	44.3	10.3	3.1	17
Oct	0.021	0.022	1.652	0.282	4.758	145.8	46.5	3.6	36.3	12.6	5.9	29
Nov	0.029	0.018	0.237	0.107	1.208	107.4	34.8	9.4	253.0	49.7	2.2	82
Dec	0.028	0.014	0.213	0.093	1.183	129.4	35.2	10.8	281.4	50.7	2.3	101

Table B.2: Parameters for Bartlett-Lewis Instantaneous Pulse model, with independent within-cell pulse depths and an exponential intensity distribution.

	λ	μ_X	α	α/ν	κ	ϕ	ω	MSIT	MSD	MCIT	MCD	MCS	MPC
Jan	0.024	0.001	2.147	4.591	1.027	0.046	173	42.3	8.9	23.8	24.5	22.4	165
Feb	0.023	0.001	3.680	4.394	1.096	0.058	187	42.6	5.4	17.1	18.8	18.8	177
Mar	0.023	0.001	2.000	5.525	0.712	0.043	204	44.1	8.3	30.5	21.7	16.4	195
Apr	0.024	0.001	2.000	6.740	0.517	0.039	248	41.7	7.7	34.4	17.8	13.4	239
May	0.027	0.001	2.000	7.760	0.437	0.054	413	37.3	4.8	35.4	15.5	8.1	392
Jun	0.031	0.001	2.000	9.607	0.310	0.050	606	32.1	4.1	40.3	12.5	6.2	576
Jul	0.030	0.001	2.000	10.413	0.167	0.039	908	33.4	4.9	69.2	11.5	4.2	874
Aug	0.029	0.001	2.000	9.683	0.293	0.053	663	34.4	3.9	42.2	12.4	5.6	630
Sep	0.025	0.001	2.000	8.901	0.345	0.047	534	40.1	4.8	39.1	13.5	7.4	510
Oct	0.021	0.001	2.126	6.698	0.580	0.041	286	48.4	6.9	29.2	16.9	14.3	274
Nov	0.025	0.001	2.000	5.389	1.055	0.049	182	39.9	7.6	21.1	22.3	21.5	173
Dec	0.026	0.001	2.035	4.584	1.093	0.054	188	37.9	7.9	23.6	25.7	20.1	179

Table B.3: Parameters for Random Parameter Bartlett-Lewis Instantaneous Pulse model, with common within-cell pulse depths and an exponential intensity distribution (conditional on η); constraints: $\alpha > 2$, $\mu_X = 001$.

	λ	ι	α	α/ν	κ	ϕ	MSIT	MSD	MCIT	MCD	MCS
Jan	0.022	0.164	2.075	5.014	0.996	0.042	46.2	9.1	23.2	23.1	24.6
Feb	0.021	0.177	3.451	4.818	1.063	0.053	47.5	5.5	16.5	17.5	20.9
Mar	0.020	0.196	2.000	5.910	0.695	0.041	48.8	8.3	29.2	20.3	18.0
Apr	0.022	0.241	2.000	7.083	0.509	0.037	46.5	7.6	33.3	16.9	14.8
May	0.023	0.400	2.000	8.127	0.434	0.052	43.9	4.7	34.0	14.8	9.4
Jun	0.026	0.586	2.000	10.015	0.311	0.049	38.9	4.1	38.5	12.0	7.3
Jul	0.024	0.879	2.000	10.777	0.173	0.040	42.3	4.6	64.3	11.1	5.3
Aug	0.024	0.639	2.000	10.109	0.299	0.052	42.3	3.8	39.7	11.9	6.8
Sep	0.021	0.518	2.000	9.257	0.343	0.045	47.4	4.8	37.7	13.0	8.6
Oct	0.019	0.277	2.051	7.006	0.575	0.039	53.8	7.1	29.1	16.7	15.7
Nov	0.023	0.175	2.000	5.832	1.018	0.045	43.9	7.6	20.2	20.6	23.5
Dec	0.024	0.179	2.000	5.018	1.056	0.050	42.0	8.0	22.6	23.9	22.2

Table B.4: Parameters for Random Parameter Bartlett-Lewis Rectangular Pulse model, with dependent intensity/duration ($\mu_X \propto \eta$) and an exponential intensity distribution (conditional on η); constraint: $\alpha > 2$.

	λ	ι	ϱ	α	α/ν	κ	ϕ	MSIT	MSD	MCIT	MCD	MCS
Jan	0.022	0.096	0.690	2.433	5.284	1.627	0.041	45.0	7.9	11.9	19.3	41.0
Feb	0.022	0.078	0.598	5.877	5.240	2.209	0.050	45.6	4.6	6.2	13.8	45.4
Mar	0.021	0.102	0.661	2.400	6.389	1.253	0.039	46.7	6.9	12.8	16.1	33.1
Apr	0.023	0.086	0.560	3.774	8.185	1.268	0.034	43.4	5.0	7.9	10.0	38.8
May	0.025	0.119	0.531	4.307	9.768	1.272	0.046	40.0	2.9	6.3	8.0	28.8
Jun	0.029	0.136	0.528	2.000	13.819	1.679	0.063	34.3	2.3	5.2	8.7	27.8
Jul	0.025	0.720	0.887	2.000	11.271	0.228	0.044	40.6	4.1	46.7	10.6	6.2
Aug	0.026	0.297	0.688	2.000	12.270	0.802	0.066	38.5	2.5	12.2	9.8	13.2
Sep	0.023	0.205	0.609	2.336	10.782	0.883	0.047	43.0	3.4	11.0	9.7	19.8
Oct	0.020	0.082	0.518	4.964	8.264	1.676	0.035	50.2	4.4	5.4	9.1	49.3
Nov	0.024	0.077	0.595	2.450	6.247	2.169	0.043	42.1	6.3	7.5	16.2	51.2
Dec	0.025	0.102	0.679	2.353	5.297	1.767	0.048	40.7	6.8	11.1	19.7	37.9

Table B.5: Parameters for Random Parameter Bartlett-Lewis Rectangular Pulse model, with dependent intensity/duration ($\mu_X \propto \eta$) and a Weibull intensity distribution (conditional on η) with no constraints on the shape parameter; constraint: $\alpha > 2$.

	λ	ι	α	α/ν	κ	ϕ	c	MSIT	MSD	MCIT	MCD	MCS
Jan	0.027	0.116	2.001	4.175	0.973	0.043	1.13	36.9	11.2	29.5	28.7	23.7
Feb	0.031	0.103	2.001	3.401	1.077	0.058	1.23	32.7	10.2	32.7	35.3	19.6
Mar	0.024	0.154	2.001	5.144	0.673	0.041	1.09	41.8	9.5	34.7	23.3	17.5
Apr	0.026	0.179	2.001	6.048	0.490	0.037	1.09	38.7	9.0	40.4	19.8	14.3
May	0.026	0.331	2.002	6.897	0.387	0.049	1.07	38.5	5.9	44.9	17.4	8.9
Jun	0.028	0.513	2.001	8.849	0.275	0.046	1.04	35.6	5.0	49.3	13.6	7.0
Jul	0.027	0.728	2.001	9.055	0.138	0.034	1.06	37.5	6.5	95.9	13.2	5.1
Aug	0.027	0.538	2.001	8.365	0.236	0.045	1.06	37.5	5.4	60.8	14.3	6.3
Sep	0.023	0.462	2.001	8.352	0.318	0.043	1.04	43.9	5.5	45.2	14.4	8.3
Oct	0.022	0.216	2.001	6.024	0.554	0.039	1.08	46.1	8.5	35.9	19.9	15.2
Nov	0.026	0.145	2.001	5.225	0.990	0.046	1.07	38.9	8.4	23.2	23.0	22.7
Dec	0.028	0.140	2.001	4.305	1.017	0.050	1.10	35.5	9.2	27.4	27.9	21.2

Table B.6: Parameters for Random Parameter Bartlett-Lewis Rectangular Pulse model, with dependent intensity-duration relationship $\iota = \mu_X/\eta^c$ and an exponential intensity distribution (conditional on η); constraint: $\alpha > 2$.

Appendix C

Plots of observed v fitted properties for variants of the BLRPR$_X$ model

The models plotted in this Appendix are all versions of the Random Parameter Bartlett-Lewis Rectangular Pulse model with dependent intensity/duration. In all cases, we have constrained the shape parameter, α, of the distribution of the cell duration parameter, η, to be greater than 2. Further details of the individual model variants are given below:

BLRPR$_X$ (E)	Exponential intensity distribution conditional on duration parameter, mean cell intensity, μ_X, proportional to η.
BLRPR$_X$ (W)	Weibull intensity distribution conditional on duration parameter, with fixed shape parameter of 0.6; mean cell intensity, μ_X, proportional to η
BLRPR$_{X_c}$ (E)	Exponential intensity distribution conditional on duration parameter, mean cell intensity, μ_X, proportional to η^c for some additional parameter, c.

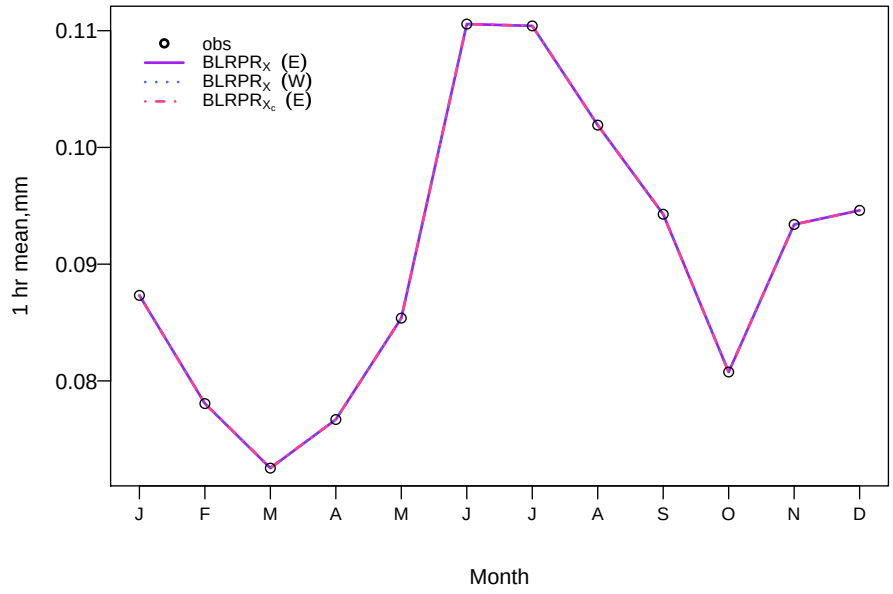


Figure C.1: Mean 1-hour rainfall by month, fitted v observed (new model variants).

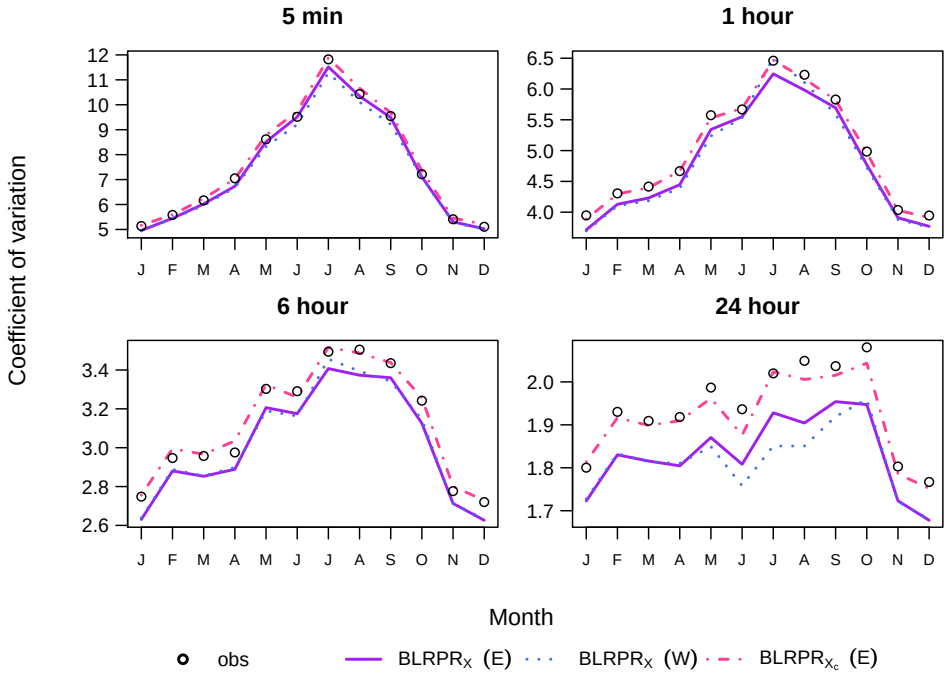


Figure C.2: Coefficient of variation by month, fitted v observed (new model variants).

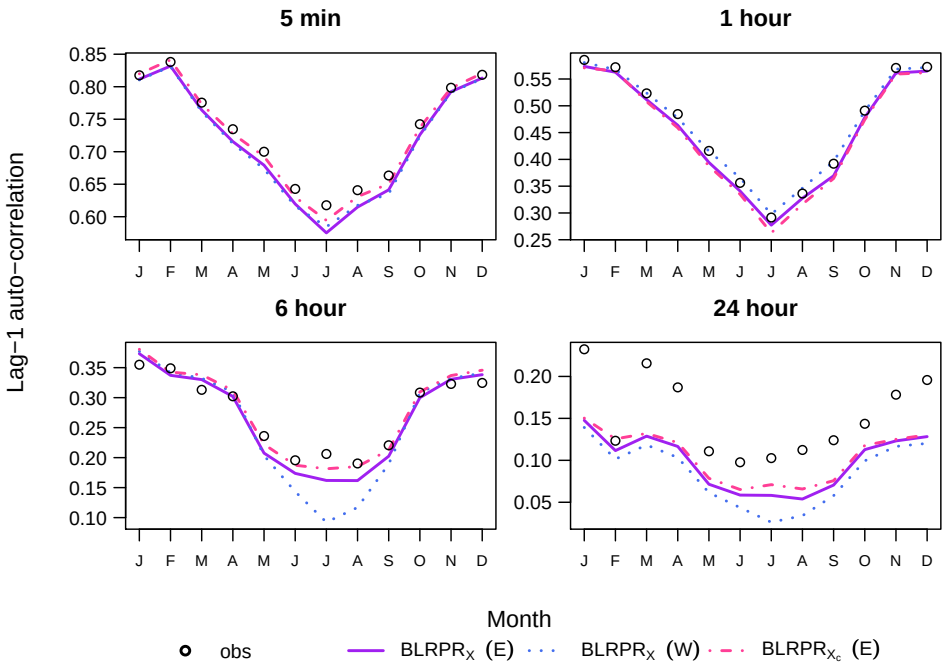


Figure C.3: Lag-1 autocorrelation by month, fitted v observed (new model variants).

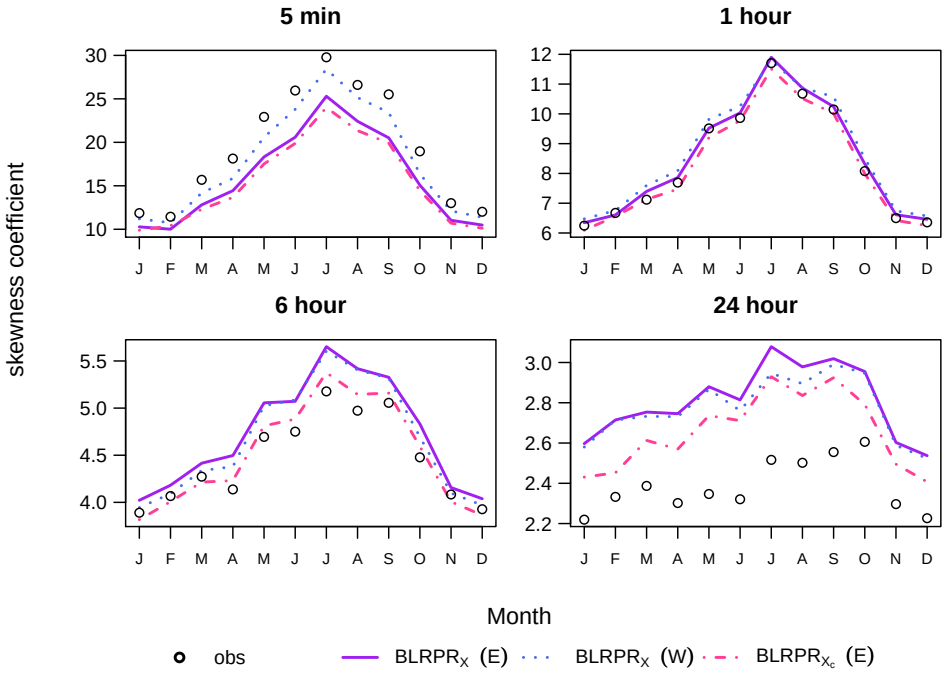


Figure C.4: Coefficient of skewness by month, fitted v observed (new model variants).

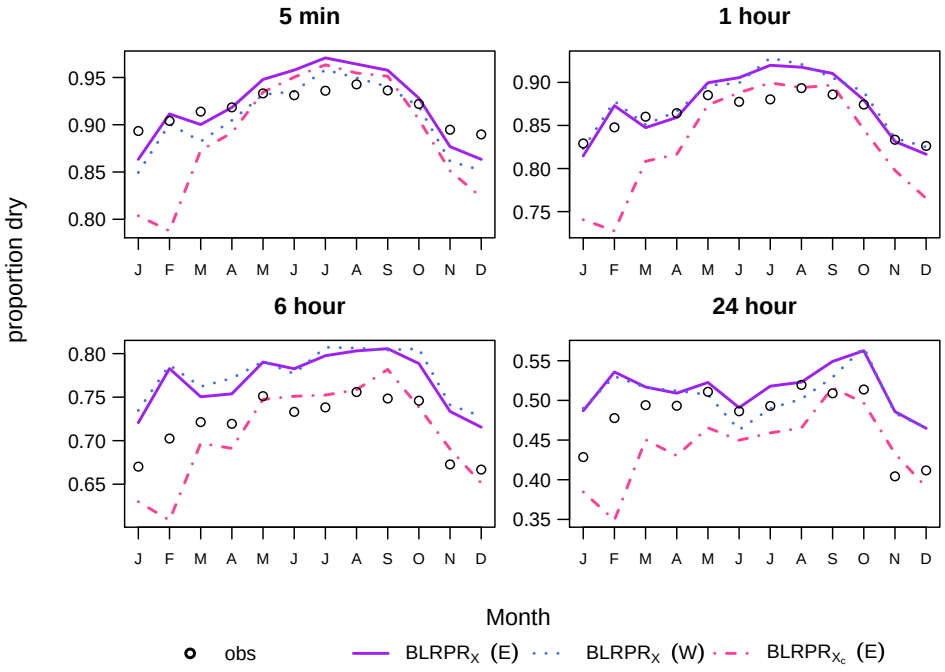
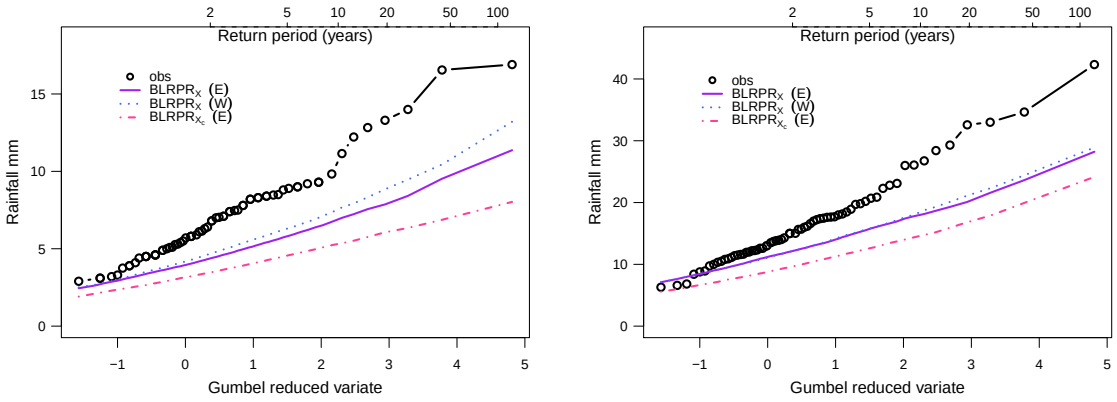


Figure C.5: Proportion dry by month, fitted v observed (new model variants).

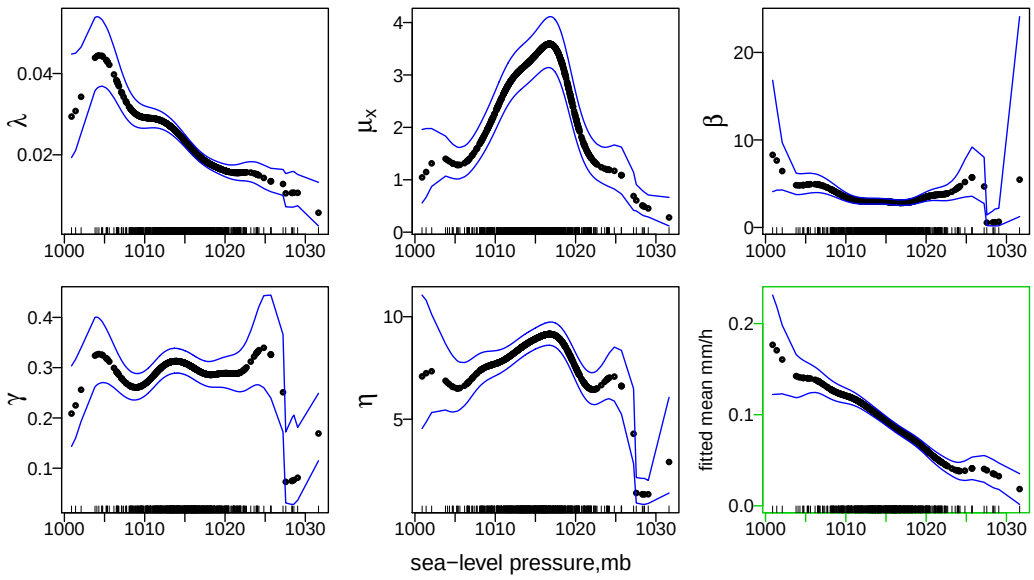


(a) 5 minute **(b)** 1 hour

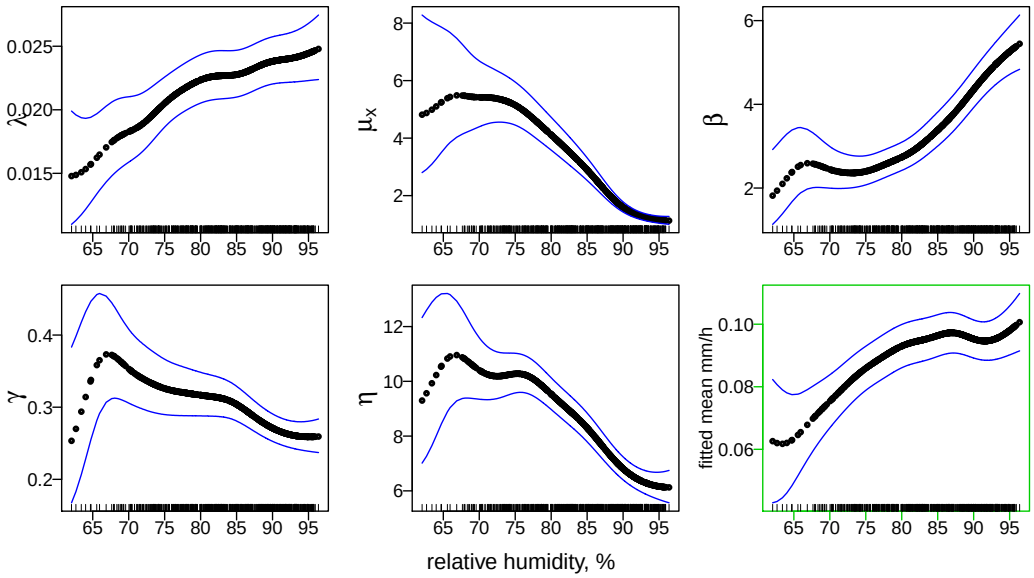
Figure C.6: Annual Gumbel plots of observed v simulated extremes for variants of the BLRPR$_X$ model.

Appendix D

Plots of fitted parameters of the BLRP model against selected continuous predictors

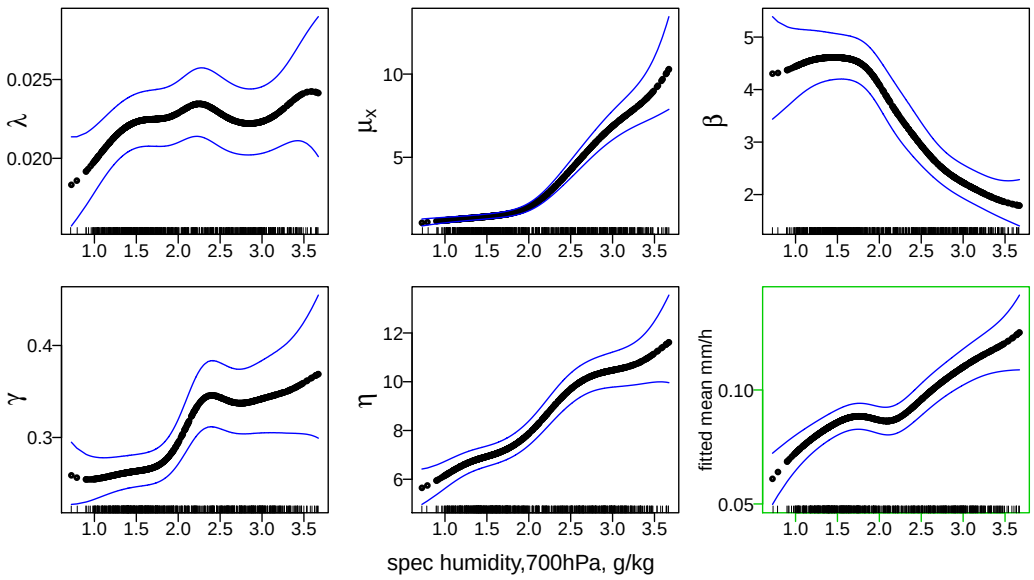


(a) sea-level pressure, h = 1.5

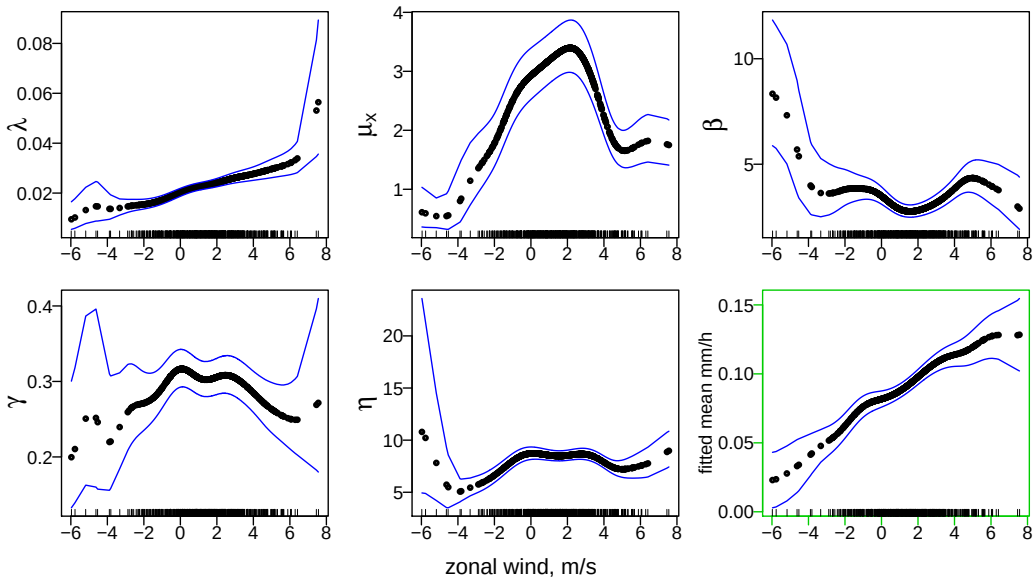


(b) relative humidity, h = 2.75

Figure D.1: Fitted parameters of the BLRP model (with exponential intensity distribution) plus mean rainfall v selected single covariates. Variability bands plotted at ± 2 standard errors.



(c) specific humidity, h = 0.25



(d) zonal wind component, h = 0.75

Figure D.1: Fitted parameters of the BLRP model (with exponential intensity distribution) plus mean rainfall v selected single covariates. Variability bands plotted at ± 2 standard errors.

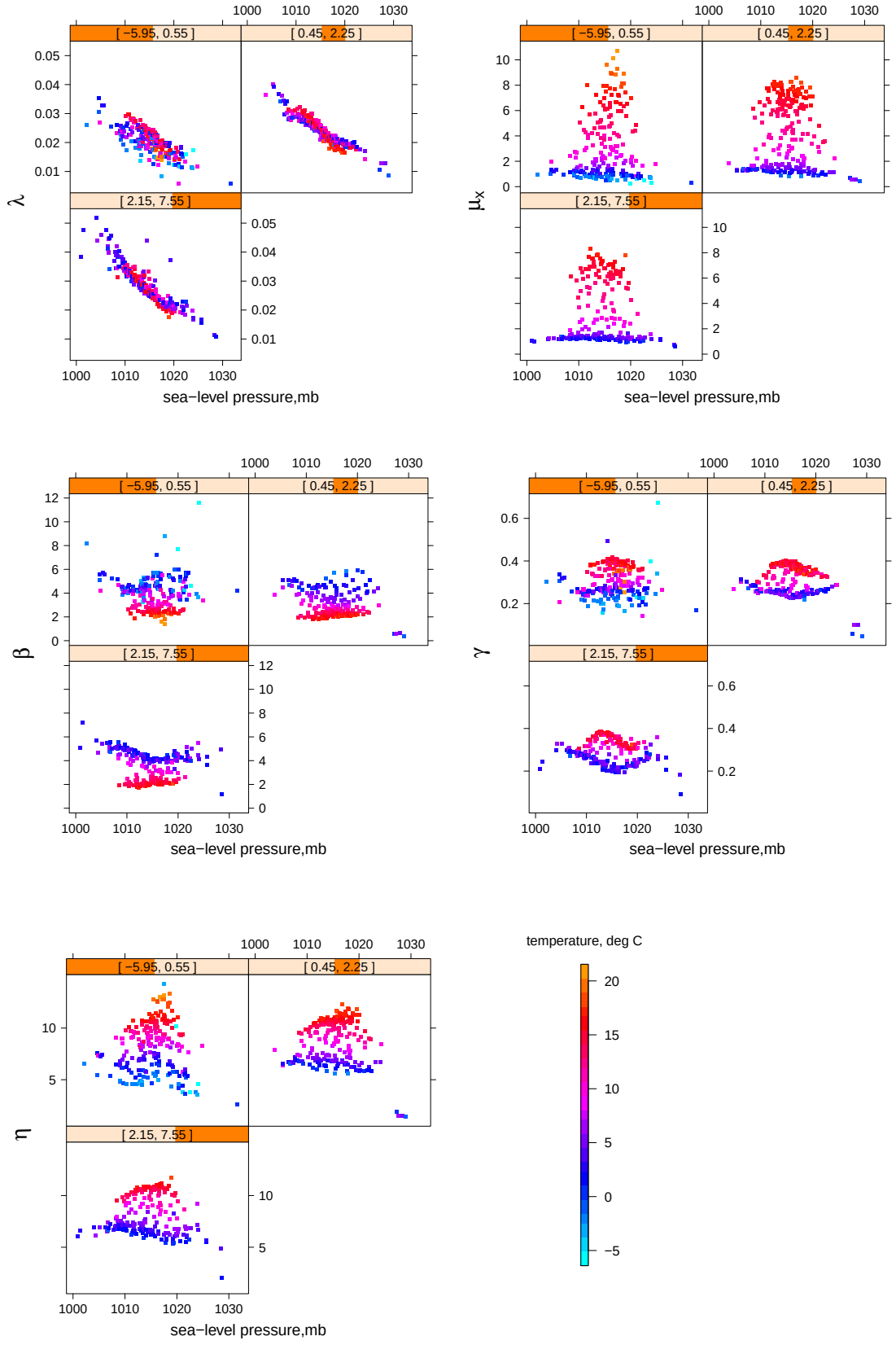


Figure D.2: Fitted parameters of the BLRP model (with exponential intensity distribution) v sea-level pressure (shown on the x-axis), temperature (indicated by different colours), and the zonal wind component m/s (split into 3 bands in the panels).

Bibliography

Andrews, D. W. K. & Lu, B. (2001), 'Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models', *Journal of Econometrics* **101**(1), 123–164.

Apelian, C. & Surace, S. (2009), *Real and Complex Analysis*, Pure and Applied Mathematics, Taylor & Francis.

Arlot, S. & Celisse, A. (2011), 'A survey of cross-validation procedures for model selection', *Statistics Surveys* **4**, 40–79.

Austin, P. M. & Houze, R. A. (1972), 'Analysis of the structure of precipitation patterns in New England', *Journal of Applied Meteorology* **11**, 926–935.

Bardossy, A. & Plate, E. J. (1992), 'Space-time model for daily rainfall using atmospheric circulation patterns', *Water Resources Research* **28**(5), 1247–1259.

Barry, R. G. & Chorley, R. J. (2003), *Atmosphere, weather and climate*, Routledge, London.

Bates, B. C., Kundzewicz, Z. W., Wu, S. & Palutikof, J. P., eds (2008), *Climate Change and Water. Technical Paper of the Intergovernmental Panel on Climate Change*, IPCC Secretariat, Geneva.

Beckmann, B. R. & Buishand, T. A. (2002), 'Statistical downscaling relationships for precipitation in the Netherlands and North Germany', *International Journal of Climatology* **22**(1), 15–32.

Bellman, R. E. (1961), *Adaptive Control Processes*, Princeton University Press, Princeton.

Bowman, A. W. & Azzalini, A. (1997), *Applied Smoothing Techniques for Data Analysis*, Oxford Statistical Science Series, Clarendon Press, Oxford.

Buishand, T. A. & Brandsma, T. (1997), 'Comparison of circulation classification schemes for predicting temperature and precipitation in the Netherlands', *International Journal of Climatology* **17**(8), 875–889.

Burton, A., Fowler, H. J., Blenkinsop, S. & Kilsby, C. G. (2010), 'Downscaling transient climate change using a Neyman-Scott Rectangular Pulses stochastic rainfall model', *Journal of Hydrology* **381**(1-2), 18–32.

Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P. & O'Connell, P. E. (2008), 'Rainsim: A spatial-temporal stochastic rainfall modelling system', *Environmental Modelling & Software* **23**(12), 1356–1369.

Cai, Z. & Li, Q. (2008), 'Nonparametric estimation of varying coefficient dynamic panel data models', *Econometric Theory* **24**(5), 1321–1342.

Carroll, R. J., Ruppert, D. & Welsh, A. H. (1998), 'Local estimating equations', *Journal of the American Statistical Association* **93**(441), 214–227.

Cavazos, T. & Hewitson, B. C. (2005), 'Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation', *Climate Research* **28**, 95–107.

Chandler, R. E. (1997), 'A spectral method for estimating parameters in rainfall models', *Bernouilli* **3**(3), 301–322.

Chandler, R. E. (2003), Moment-based inference for stochastic-mechanistic models, Technical Report 7, DEFRA project FD2105. Improved Methods for national spatial-temporal rainfall and evaporation modelling for BSM.

Chandler, R. E. (2011), 'Exploiting strength, discounting weakness: combining information from multiple climate simulators'. Research Report No. 311, Department of Statistical Science, University College London.

Chandler, R. E., Isham, V., Bellone, E., Yang, C. & Northrop, P. (2007a), Space-time modelling of rainfall for continuous simulation, *in* 'Statistical Methods for Spatio-Temporal Systems', Chapman and Hall/CRC, pp. 177–215.

Chandler, R. E., Isham, V. S., Wheater, H. S., Onof, C. J., Leith, N., Frost, A. J. & Segond, M.-L. (2007b), Spatial-temporal rainfall modelling with climate change scenarios, Technical Report FD2113, DEFRA/EA.

Chandler, R. E., Lourmas, G. & Jesus, J. (2010), 'Momfit - software for moment-based fitting of single-site stochastic rainfall models'.
URL: *http://www.ucl.ac.uk/ ucakarc/work/momfit.html*

Chandler, R. E. & Wheater, H. S. (2002), 'Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland', *Water Resources Research* **38**(10), 1192.

Charles, S. P., Bates, B. C., Smith, I. N. & Hughes, J. P. (2004), 'Statistical downscaling of daily precipitation from observed and modelled atmospheric fields', *Hydrological Processes* **18**(8), 1373–1394.

Christensen, J. H., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., Kolli, R. K., Kwon, W.-T., Laprise, R., Magaña Rueda, V., Mearns, L., Menndez, C. G., Risnen, J., Rinke, A., Sarr, A. & Whetton, P. (2007), Regional climate projections, *in* 'Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press, U.K.

Cleveland, W. S. (1979), 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association* **74**(368), 829–836.

Cleveland, W. S. & Devlin, S. J. (1988), 'Locally weighted regression: An approach to regression analysis by local fitting', *Journal of the American Statistical Association* **83**(403), 596–610.

Coles, S. (2001), *An Introduction to Statistical Modelling of Extreme Values*, Springer Series in Statistics, Springer-Verlag, London.

Cowpertwait, P., Isham, V. & Onof, C. (2007), 'Point process models of rainfall: Developments for fine-scale structure', *Proceedings of The Royal Society of London, Series A* **463**(2086), 2569–2587.

Cowpertwait, P. S. P. (1994), 'A generalized point process model for rainfall', *Proceedings of The Royal Society of London, Series A* **447**(1929), 23–37.

Cowpertwait, P. S. P. (1995), 'A generalized spatial-temporal model of rainfall based on a clustered point process', *Proceedings of The Royal Society of London, Series A* **450**(1938), 163–175.

Cowpertwait, P. S. P. (1998), 'A Poisson-cluster model of rainfall: high-order moments and extreme values', *Proceedings of The Royal Society of London, Series A* **454**(1971), 885–898.

Cowpertwait, P. S. P. (2004), 'Mixed rectangular pulses models of rainfall', *Hydrology and Earth System Sciences* **8**(5), 993–1000.

Cowpertwait, P. S. P. (2006), 'A spatial-temporal point process model of rainfall for the Thames catchment, UK', *Journal of Hydrology* **330**(3-4), 586–595.

Cowpertwait, P. S. P., Kilsby, C. G. & O'Connell, P. E. (2002), 'A space-time Neyman-Scott model of rainfall: empirical analysis of extremes', *Water Resources Research* **38**(8), 1131.

Cowpertwait, P. S. P., O'Connell, P. E., Metcalfe, A. V. & Mawdsley, J. A. (1996), 'Stochastic point process modelling of rainfall. I. Single-site fitting and validation', *Journal of Hydrology* **175**(1-4), 17–46.

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

Dai, J. & Sperlich, S. (2010), 'Simple and effective boundary correction for kernel densities and regression with an application to the world income and Engel curve estimation', *Computational Statistics & Data Analysis* **54**(11), 2487–2497.

Davison, A. C. (2003), *Statistical Models*, Cambridge University Press, New York.

de Boor, C. (2001), *A practical guide to splines*, Springer-Verlag, New-York.

Eilers, P. H. C. & Marx, B. D. (1996), 'Flexible smoothing with B-splines and penalties', *Statistical Science* **11**(2), 89–121.

Entekhabi, D., Rodriguez-Iturbe, I. & Eagleson, P. S. (1989), 'Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: parameter estimation and validation', *Water Resources Research* **25**(2), 295–302.

Fan, J. & Gijbels, I. (1995), 'Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation', *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(2), 371–394.

Fan, J. & Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, Chapman and Hall, London.

Fowler, H. J., Blenkinsop, S. & Tebaldi, C. (2007), 'Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling', *International Journal of Climatology* **27**(12), 1547–1578.

Fowler, H. J., Kilsby, C. G. & O'Connell, P. E. (2000), 'A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions', *Hydrology and Earth System Sciences* **4**(2), 263–281.

Furrer, E. M. & Katz, R. W. (2007), 'Generalized linear modeling approach to stochastic weather generators', *Climate Research* **34**, 129–144.

Furrer, R., Nychka, D. & Sain, S. (2011), *Fields: Tools for spatial data*. R package version 6.6.2.
URL: *http://cran.r-project.org/web/packages/fields/index.html*

Gasser, T., Kneip, A. & Köhler, W. (1991), 'A flexible and fast method for automatic smoothing', *Journal of the American Statistical Association* **86**(415), 643–652.

Genz, A. & Bretz, F. (2009), *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statistics, Springer-Verlag, Heidelberg.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F. & Hothorn, T. (2012), *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-9993.
URL: *http://CRAN.R-project.org/package=mvtnorm*

Glasbey, C. A., Cooper, G. & McGechan, M. B. (1995), 'Disaggregation of daily rainfall by conditional simulation from a point process model', *Journal of Hydrology* **165**(1-4), 1–9.

Gozalo, P. & Linton, O. (2000), 'Local nonlinear least squares: Using parametric information in nonparametric regression', *Journal of Econometrics* **99**(1), 63–106.

Gupta, V. K. & Waymire, E. C. (1993), 'A statistical analysis of mesoscale rainfall as a random cascade', *Journal of Applied Meteorology* **32**, 251–267.

Hall, A. R. (2005), *Generalized Method of Moments*, Oxford University Press, Oxford.

Hall, A. R. (2010), Generalized method of moments, *in* 'Encyclopaedia of Quantitative Finance', John Wiley & Sons Ltd, Chichester, UK, pp. 836–840.

Hall, A. R. & Inoue, A. (2003), 'The large sample behaviour of the generalized method of moments estimator in misspecified models', *Journal of Econometrics* **114**(2), 361–394.

Hansen, B. E. (2005), 'Challenges for econometric model selection', *Econometric Theory* **21**(1), 60–68.

Hansen, B. E. (2012), Econometrics. Chapter 11 of: Unpublished book manuscript, University of Wisconsin-Madison.
URL: *http://www.ssc.wisc.edu/ bhansen/econometrics/*

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**(4), 1029–1054.

Hastie, T. & Tibshirani, R. (1990), *Generalized additive models*, Chapman and Hall, London.

Hastie, T., Tibshirani, R. & Friedman, J. (2001), *The Elements of Statistical Learning*, Springer Series in Statistics, Springer-Verlag, New York.

Haylock, M. R., Cawley, G. C., Harpham, C., Wilby, R. L. & Goodess, C. M. (2006), 'Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios', *International Journal of Climatology* **26**(10), 1397–1415.

Hengartner, N. W. & Wegkamp, M. H. (2002), 'Bandwidth selection for local linear regression', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **64**(4), 791–804.

Houze, R. A. & Hobbs, P. V. (1982), 'Organization and structure of precipitating cloud systems', *Advances in Geophysics* **24**, 225–315.

Hulme, M., Jenkins, G. J., Lu, X., Turnpenny, J. R., Mitchell, T. D., Jones, R. G., Lowe, J., Murphy, J. M., Hassell, D., Boorman, P., McDonald, R. & Hill, S. (2002), 'Climate

Change Scenarios for the United Kingdom: The UKCIP02 Scientific Report'. Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich, UK, 120 pp.

IPCC (2007), *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, IPCC, Geneva. (Core writing team: Pachauri, R. K. & Reisinger, A. (eds.)).

Jesus, J. (2012), Contributions to inference without likelihoods. PhD Thesis.

Jesus, J. & Chandler, R. E. (2011), 'Estimating functions and the generalized method of moments', *Interface Focus* **1**(6), 871–885.

Jones, M. C. (1993), 'Simple boundary correction in kernel density estimation', *Statistics and Computing* **3**, 135–146.

Jones, P. D., Hulme, M. & Briffa, K. R. (1993), 'A comparison of Lamb circulation types with an objective classification scheme', *International Journal of Climatology* **13**(6), 655–663.

Jones, P. D., Kilsby, C. G., Harpham, C., Glenis, V. & Burton, A. (2009), UK Climate Projections science report: Projections of future daily climate for the UK from the Weather Generator, Technical report, University of Newcastle, UK.

Jones, P. G. & Thornton, P. K. (1997), 'Spatial and temporal variability of rainfall related to a third-order Markov model', *Agricultural and Forest Meteorology* **86**(1–2), 127–138.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. & Joseph, D. (1996), 'The NCEP/NCAR 40-year reanalysis project', *Bulletin of the American Meteorological Society* **77**, 437.

Katz, R. W. (1977), 'Precipitation as a chain-dependent process', *Journal of Applied Meteorology* **16**(7), 671–676.

Kauermann, G., Müller, M. & Carroll, R. J. (1998), 'The efficiency of bias-corrected estimators for nonparametric kernel estimation based on local estimating equations', *Statistics & Probability Letters* **37**(1), 41–47.

Kilsby, C. G., Cowpertwait, P. S. P., O'Connell, P. E. & Jones, P. D. (1998), 'Predicting rainfall statistics in England and Wales using atmospheric circulation variables', *International Journal of Climatology* **18**(5), 523–539.

Kilsby, C. G., Jones, P. D., Burton, A., Ford, A. C., Fowler, H. J., Harpham, C., James, P., Smith, A. & Wilby, R. L. (2007), 'A daily weather generator for use in climate change studies', *Environmental Modelling & Software* **22**(12), 1705–1719.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R. & Fiorino, M. (2001), 'The NCEP/NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation', *Bulletin of the American Meteorological Society* **82**, 247.

Klein, W. H. & Glahn, H. R. (1974), 'Forecasting local weather by means of model output statistics', *Bulletin of the American Meteorological Society* **55**, 1217–1227.

Knutti, R., Reinhard, F., Tebaldi, C., Cermak, J. & Meehl, G. A. (2010), 'Challenges in combining projections from multiple climate models', *Journal of Climate* **23**(10), 2739–2758.

Koutsoyiannis, D. & Onof, C. (2000), 'HYETOS - a computer program for stochastic disaggregation of fine-scale rainfall'. http://www.itia.ntua.gr/e/softinfo/3/.

Koutsoyiannis, D. & Onof, C. (2001), 'Rainfall disaggregation using adjusting procedures on a Poisson cluster model', *Journal of Hydrology* **246**(1-4), 109–122.

Kuonen, D. (1999), 'Saddlepoint approximations for distributions of quadratic forms in normal variables', *Biometrika* **86**(4), 929–935.

Lewbel, A. (2007), 'A local generalized method of moments estimator', *Economics Letters* **94**(1), 124–128.

Loader, C. R. (1999), 'Bandwidth selection: classical or plug-in?', *The Annals of Statistics* **27**(2), 415–438.

Lovejoy, S. & Schertzer, D. (1995), Multifractals and rain, in uncertainty concepts, *in* A. W. Kundzewicz, ed., 'Hydrology and hydrological modelling', Cambridge University Press, pp. 62–103.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M. & Thiele-Eich, I. (2010), 'Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user', *Reviews of Geophysics* **48**(3), RG3003.

Michna, P. (2012), *RNetCDF: R Interface to NetCDF Datasets*. R package version 1.6.1-2. **URL:** *http://cran.r-project.org/web/packages/RNetCDF/index.html*

Müller, H.-G. (1991), 'Optimum kernel estimators near endpoints', *Biometrika* **78**(3), 521–530.

Nadaraya, E. A. (1964), 'On estimating regression', *Theory of Probability & Its Applications* **9**(1), 141–142.

Nakićenović, N. & Swart, R., E. (2000), *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, U.K.

Newey, W. K. (1991), 'Uniform convergence in probability and stochastic equicontinuity', *Econometrica* **59**(4), 1161–1167.

Newey, W. K. & McFadden, D. L. (1986), Large sample estimation and hypothesis testing, *in* R. F. Engle & D. McFadden, eds, 'Handbook of Econometrics', 1 edn, Vol. 4, Elsevier, chapter 36, pp. 2111–2245.
URL: *http://EconPapers.repec.org/RePEc:eee:ecochp:4-36*

Northrop, P. (2006), 'Estimating the parameters of rainfall models using maximum marginal likelihood', *Student* **5**(3/4), 173–183.

Onof, C. & Arnbjerg-Nielsen, K. (2009), 'Quantification of anticipated future changes in high resolution design rainfall for urban areas', *Atmospheric Research* **92**(3), 350–363.

Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheater, H. S. & Isham, V. (2000), 'Rainfall modelling using Poisson-cluster processes: a review of developments', *Stochastic Environmental Research and Risk Assessment* **14**, 384–411.

Onof, C., Townsend, J. & Kee, R. (2005), 'Comparison of two hourly to 5-min rainfall disaggregators', *Atmospheric Research* **77**, 176–187.

Onof, C. & Wheater, H. S. (1993), 'Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model', *Journal of Hydrology* **149**(1-4), 67–95.

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org/*

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R. J., Sumi, A. & Taylor, K. E. (2007), Climate models and their evaluation., *in* 'Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press, U.K.

Richardson, C. W. (1981), 'Stochastic simulation of daily precipitation, temperature, and solar radiation', *Water Resources Research* **17**(1), 182–190.

Rodriguez-Iturbe, I., Cox, D. R. & Isham, V. (1987), 'Some models for rainfall based on stochastic point processes', *Proceedings of The Royal Society of London, Series A* **410**, 269–288.

Rodriguez-Iturbe, I., Cox, D. R. & Isham, V. (1988), 'A point process model for rainfall: further developments', *Proceedings of The Royal Society of London, Series A* **417**, 283–298.

Ruppert, D. (1997), 'Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation', *Journal of the American Statistical Association* **92**(439), 1049–1062.

Ruppert, D., Sheather, S. J. & Wand, M. P. (1995), 'An effective bandwidth selector for local least squares regression', *Journal of the American Statistical Association* **90**(432), 1257–1269.

Ruppert, D. & Wand, M. P. (1994), 'Multivariate locally weighted least squares regression', *The Annals of Statistics* **22**(3), 1346–1370.

Salim, A. & Pawitan, Y. (2003), 'Extensions of the Bartlett-Lewis model for rainfall processes', *Statistical Modelling* **3**(2), 79–98.

Sarkar, D. (2010), *Lattice: Lattice Graphics*. R package version 0.20-10.
URL: *http://cran.r-project.org/web/packages/lattice/index.html*

Schuster, E. F. (1972), 'Joint asymptotic distribution of the estimated regression function at a finite number of distinct points', *The Annals of Mathematical Statistics* **43**(1), 84–88.

Shaw, E. M., Beven, K. J., Chappell, N. A. & Lamb, R. (2011), *Hydrology in Practice*, Spon Press, Abingdon, Oxon.

Stern, R. D. & Coe, R. (1984), 'A model fitting analysis of daily rainfall data', *Journal of the Royal Statistical Society. Series A (General)* **147**(1), 1–34.

Stuetzle, W. & Mittal, Y. (1979), Some comments on the asymptotic behaviour of robust smoothers., *in* T. Gasser & M. Rosenblatt, eds, 'Smoothing Techniques for Curve Estimation', Vol. 757, Springer Lecture Notes, pp. 191–195.

Tibshirani, R. & Hastie, T. (1987), 'Local likelihood estimation', *Journal of the American Statistical Association* **82**(398), 559–567.

Tran, K. C. & Tsionas, E. G. (2010), 'Local GMM estimation of semiparametric panel data with smooth coefficient models', *Econometric Reviews* **29**(1), 39 – 61.

Trenberth, K. E., Jones, P. D., Ambenje, P., Bojariu, R., Easterling, D., Klein Tank, A., Parker, D., Rahimzadeh, F., Renwick, J. A., Rusticucci, M., Soden, B. & Zhai, P. (2007), Observations: Surface and Atmospheric Climate Change, *in* 'Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press, U.K.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, U.K.

Verhoest, N. E. C., Vandenberghe, S., Cabus, P., Onof, C., Meca-Figueras, T. & Jameleddine, S. (2010), 'Are stochastic point rainfall models able to preserve extreme flood statistics?', *Hydrological Processes* **24**(23), 3439–3445.

von Storch, H. & Zwiers, F. W. (1999), *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, U.K.

Wand, M. P. & Jones, M. C. (1995), *Kernel Smoothing*, Chapman & Hall, London.

Wasserman, L. (2006), *All of Nonparametric Statistics*, Springer, New York.

Watson, G. S. (1964), 'Smooth regression analysis', *Sankhyā: The Indian Journal of Statistics, Series A* **26**(4), 359–372.

Wheater, H. S., Chandler, R. E., Onof, C. J., Isham, V. S., Bellone, E., Yang, C., Lekkas, D., Lourmas, G. & Segond, M.-L. (2005), 'Spatial-temporal rainfall modelling for flood risk estimation', *Stochastic Environmental Research and Risk Assessment* **19**(6), 403–416.

Wheater, H. S., Isham, V. S., Chandler, R. E., Onof, C. J. & Stewart, E. J. (2006), Improved methods for national spatial-temporal rainfall and evaporation modelling for BSM, Technical report, DEFRA project F2105.

Wheater, H. S., Isham, V. S., Onof, C., Chandler, R. E., Northrop, P. J., Guiblin, P., Bate, S. M., Cox, D. R. & Koutsoyiannis, D. (2000), 'Generation of spatially consistent rainfall data. Report to the Ministry of Agriculture, Fisheries and Food (2 volumes)'. Also available as Research Report 204, Department of Statistical Science, University College London http://www.ucl.ac.uk/Stats/research/reports.

Wilby, R. L. (1997), 'Non-stationarity in daily precipitation series: implications for GCM down-scaling using atmospheric circulation indices', *International Journal of Climatology* **17**(4), 439–454.

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P. & Mearns, L. O. (2004), Guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change (IPCC), prepared on behalf of Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA), Technical report, IPCC.

Wilby, R. L. & Wigley, T. M. L. (2000), 'Precipitation predictors for downscaling: observed and General Circulation Model relationships', *International Journal of Climatology* **20**(6), 641–661.