Single site rainfall generators: A look at some developments

Berlin Workshop on Weather Generators
September 2017

Chris Onof[1,2], David Cross[1], Dongkyun Kim[3], Lipen Wang[2,1]

1: Imperial College London
2: RainPlusPlus, Derby
3: Hongik University, Seoul

I. Outline

- Approaches to rainfall modelling
- The Poisson-cluster process approach
- Main challenges
- Some recent and contemporary work
- Conclusion

II - Approaches to rainfall modelling

1. Statistically governed deterministic modelling approaches

Resampling methods (Method of Analogues, e.g. Wetterhall) together with a model for weather type generation or the simulation of relevant climatological variables

⊕: The rainfall is realistic

⊖: (i) There are no extremes larger than the observed; (ii) this relies either upon another model, which will itself have to be adapted for a changed climate; or it uses weather types, of which it is not known whether they cover all relevant future climatological conditions – but the use of analogue locations offers a way round this.

RESAMPLING METHOD

Weather types that are modelled independently determine the *class* of the time-step (e.g. day), or a set of *climate variables* (Temperature, SLP, Relative Humidity, etc.) at that time.

Historical data from days of the same class/from other time-steps that are analogous according to some distance, are sampled

Additional constraints may be introduced to ensure temporal correlation

2. Purely statistical models

<u>Markov models</u> where the dependence upon the past is the key driver of the simulation

<u>Regression models</u> and <u>Generalised Linear Models</u> (e.g. Chandler) which rely chiefly upon establishing relations between the rainfall and climatological/geographical information

Re (i): interactions between variables can be included in a GLM for instance
Re (ii): the Markov assumption can be relaxed to include dependence upon more than one previous time-step

⊖: (i) Regressions and GLMs assume a general form of the dependence (flexible in GLMs) on causing factors and the explanatory variables are not independent.

(ii) The Markov assumption does not represent persistence properly

LINEAR REGRESSION

$$R = \sum_{i=1}^{n} \gamma_i C_i + Z$$

GENERALISED LINEAR MODELS

$$Occurrence:\ \frac{p}{1-p} = exp\{\sum_{i=1}^{n} \alpha_i X_i\} \quad Depth:\ E(R) = exp\{\sum_{i=1}^{n} \beta_i Y_i\}\ \&\ R \sim \text{Gamma}(a, b)$$

3. Phenomenological models focussing upon the scaling features of rainfall

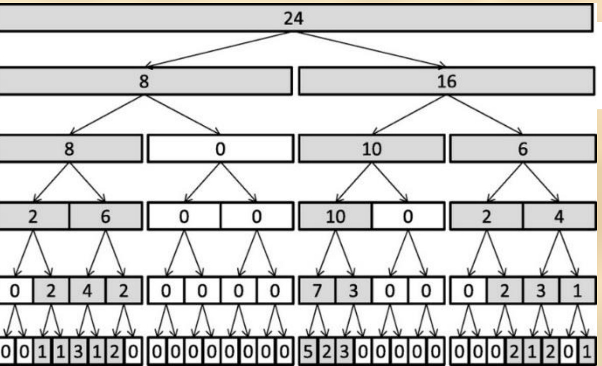
Characteristic rainfall features such as the non-centred moments of rainfall intensities or the probabilities of exceeding scale related thresholds exhibit scaling features which can be captured by a number (fractal) or a function (multifractal)

Rainfall can therefore be directly modelled by a (multi-)fractal cascade. The theory leads ... ively,

> Re (ii): there have however been studies in which the cascade generator has been allowed to have a distribution that changes with the temporal scale, and with the intensity at the next scale below.

⊕: (i) The rainfall is easy to generate; (ii) the model is parsimonious

⊖: (i) A cascade process requires coarse-scale information as starting point, so the most obvious application of the model is in downscaling mode; (ii) There are breaks in the scaling so that scale independence only holds over certain ranges. Where the breaks occur may however depend upon the type of rainfall.



MICRO-CANONICAL CASCADE

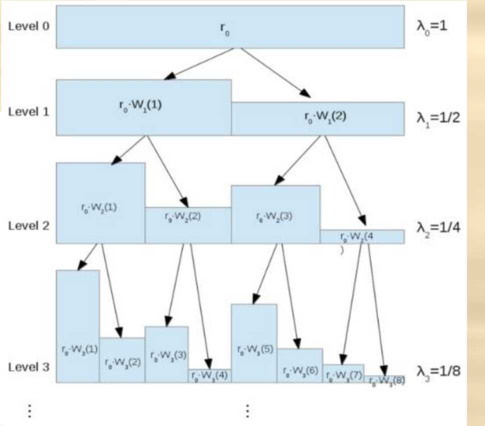
MACRO-CANONICAL CASCADE

MULTI-SCALING:

$E(R_{(h)}^{q}) \propto h^{-K(q)}$

$P(R_{(h)} > h^{-\gamma}) \propto h^{c(\gamma)}$

$E(f) \propto f^{-\beta}$



4. Phenomenological models focussing upon scale-dependent features of rainfall

Characteristic rainfall features such as the clustering of rainfall cells (convective or sub-mesoscale) within storms (mesoscale), the intermittency patterns can be captured by driving the generation of short events (instantaneous or rectangular pulses of rainfall) by a process that

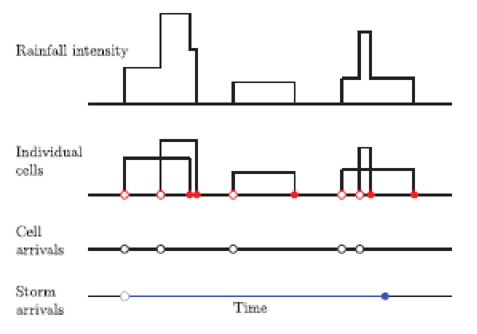
Re (i): a better understanding of the reliability of parameter estimates can now be obtained through confidence intervals
Re (ii): a number of approaches have been proposed to tackle this problem

(Cox) *models*

⊕: (i) The model parameters are easily interpretable; (ii) a number of properties (for the first) or the likelihood function (for the second) can be derived mathematically

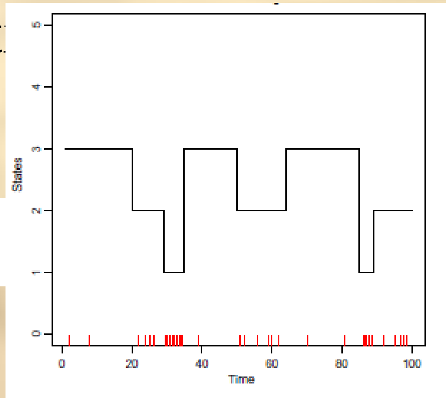
⊖: (i) Parameter identification (for the first) is difficult; (ii) The models tend to overestimate daily extremes and underestimate hourly and sub-hourly extremes (part

e second)



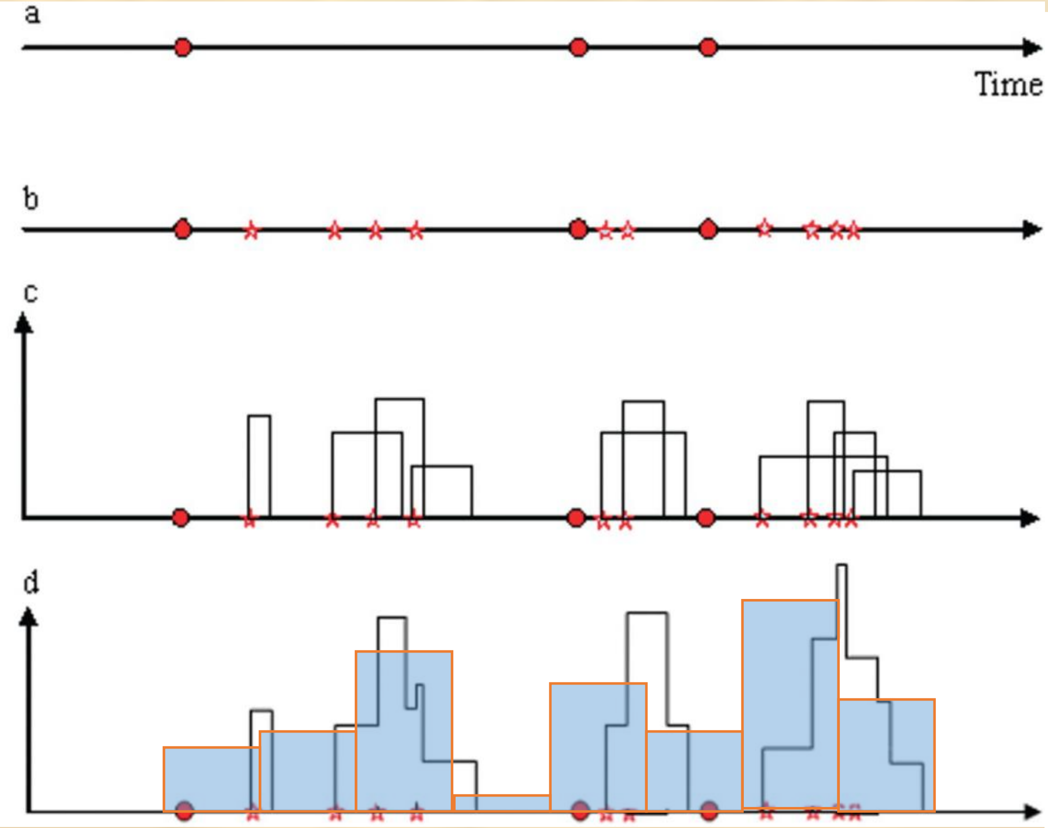
POISSON-CLUSTER MODEL

DOUBLY-STOCHASTIC POISSON MODEL



III – The Poisson cluster process approach

1. Model description



Time t

Poisson-cluster process

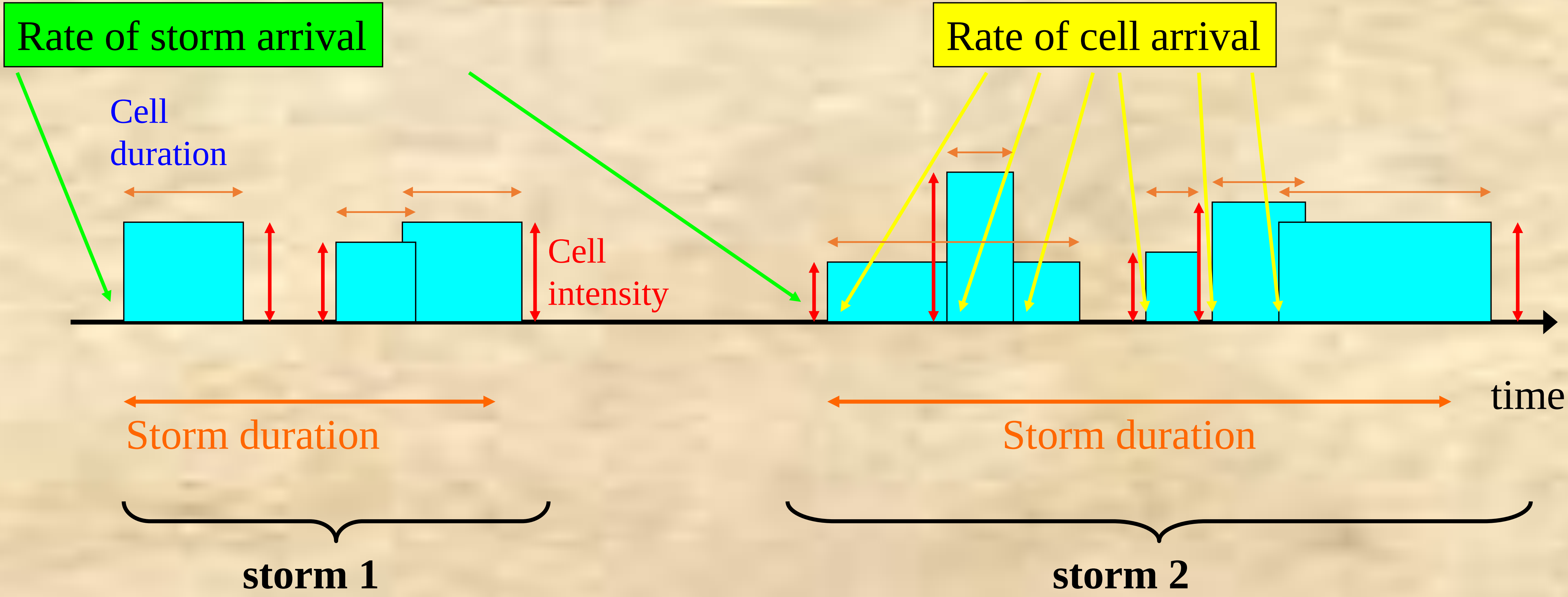
Cells clustered in storms

Blue Line: Continuous rainfall Y(t)

Red Line: Aggregated rainfall $Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(t)dt$

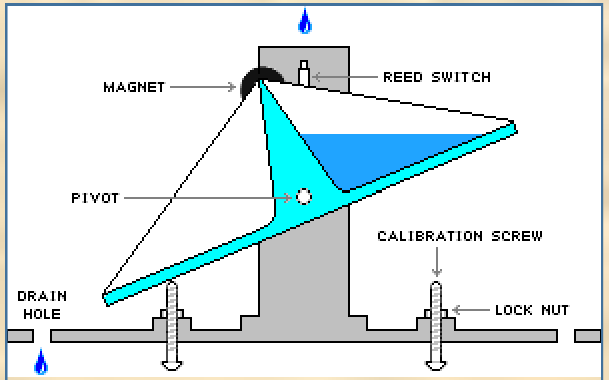
Random Parameter (or Modified) Bartlett-Lewis Rectangular Pulse Model

Arrival of storms according to a *Poisson* process	**(λ)**
Mean cell durations *gamma* distributed	**(α,ν)**
Arrival of cells in a storm according to a *Poisson* process	**($\kappa\eta$)**
Storm duration *exponentially* distributed	**($\varphi\eta$)**
Cell intensity *exponentially* (or other) distributed	**(μ_x) (and μ_x^2)**
Cell duration *exponentially* distributed	(η)



2. Model fitting

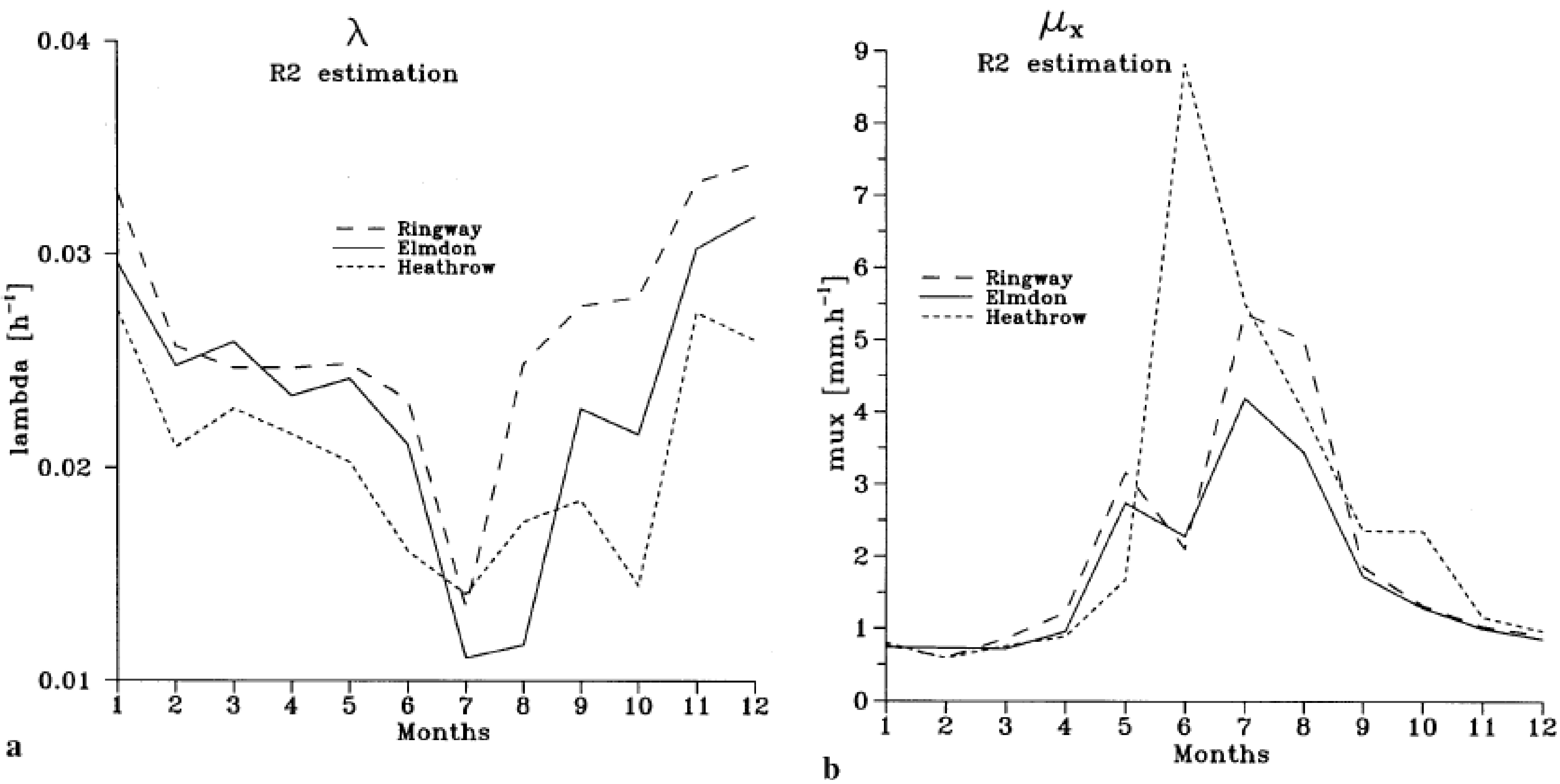
- $A_1(\lambda,\mu_X,\ldots)$: Mean rainfall intensity
- $A_2(\lambda,\mu_X,\ldots)$: Standard deviation of rainfall intensities
- $A_3(\lambda,\mu_X,\ldots)$: Autocorrelation (lag 1) structure of rainfall intensities
- $A_5(\lambda,\mu_X,\ldots)$: Skewness of rainfall intensities
- $A_5(\lambda,\mu_X,\ldots)$: Proportion of dry intervals
- …



Rainfall data set (typically hourly data) from which sample statistics are estimated: $\{\Omega_1, \Omega_2, \ldots, \Omega_n\}$

$$F(\lambda,\mu_X,\ldots) = \sum_{i=1}^{p} \omega_i \left(A_i(\lambda,\mu_X,\ldots) - \Omega_i\right)^2$$

One set of parameters is obtained for each calendar month of the year, so that the model is able to reproduce seasonality

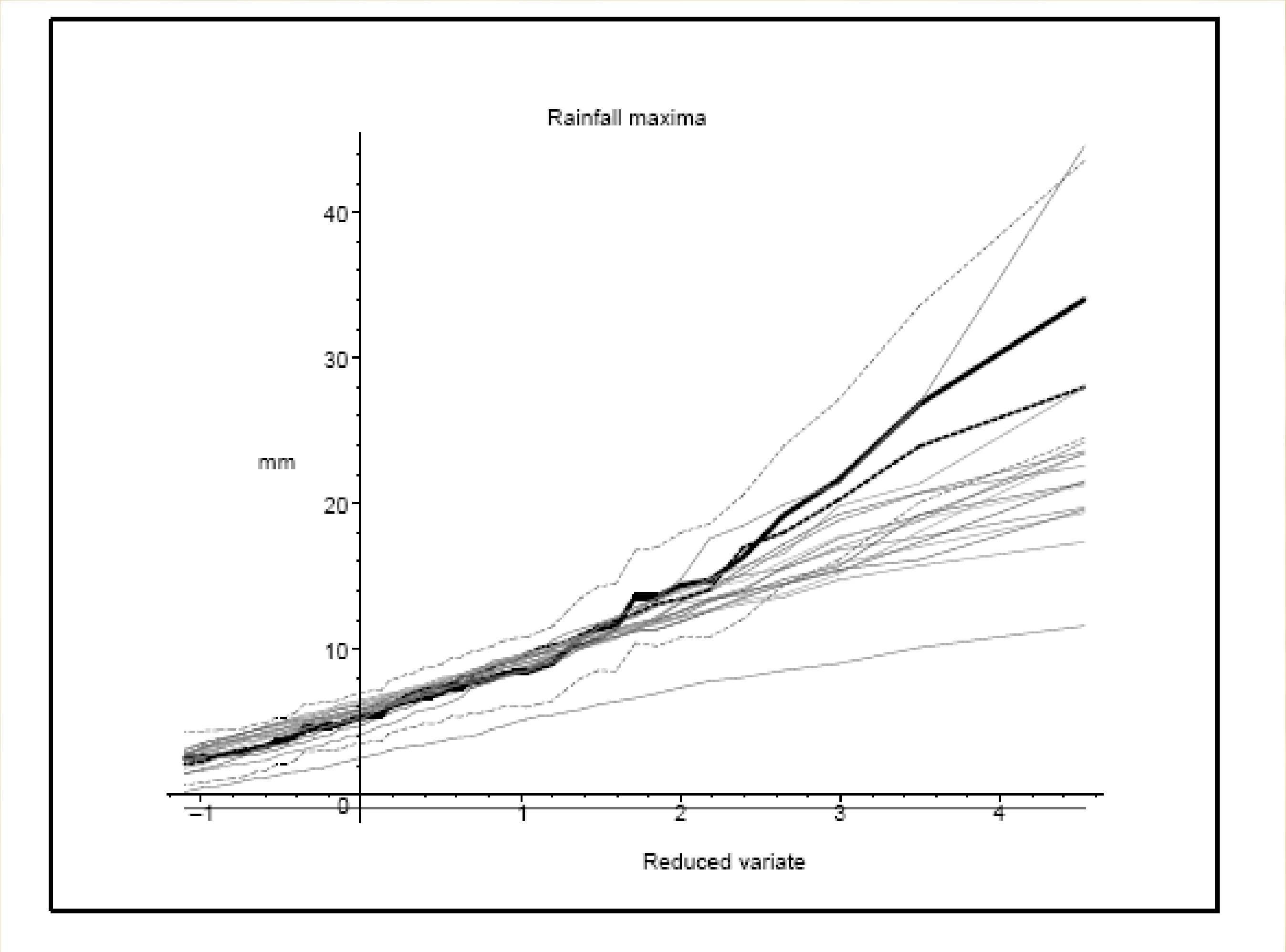


III. The Poisson cluster process approach

IV – Main challenges

1. Impact of parameter identifiability issues

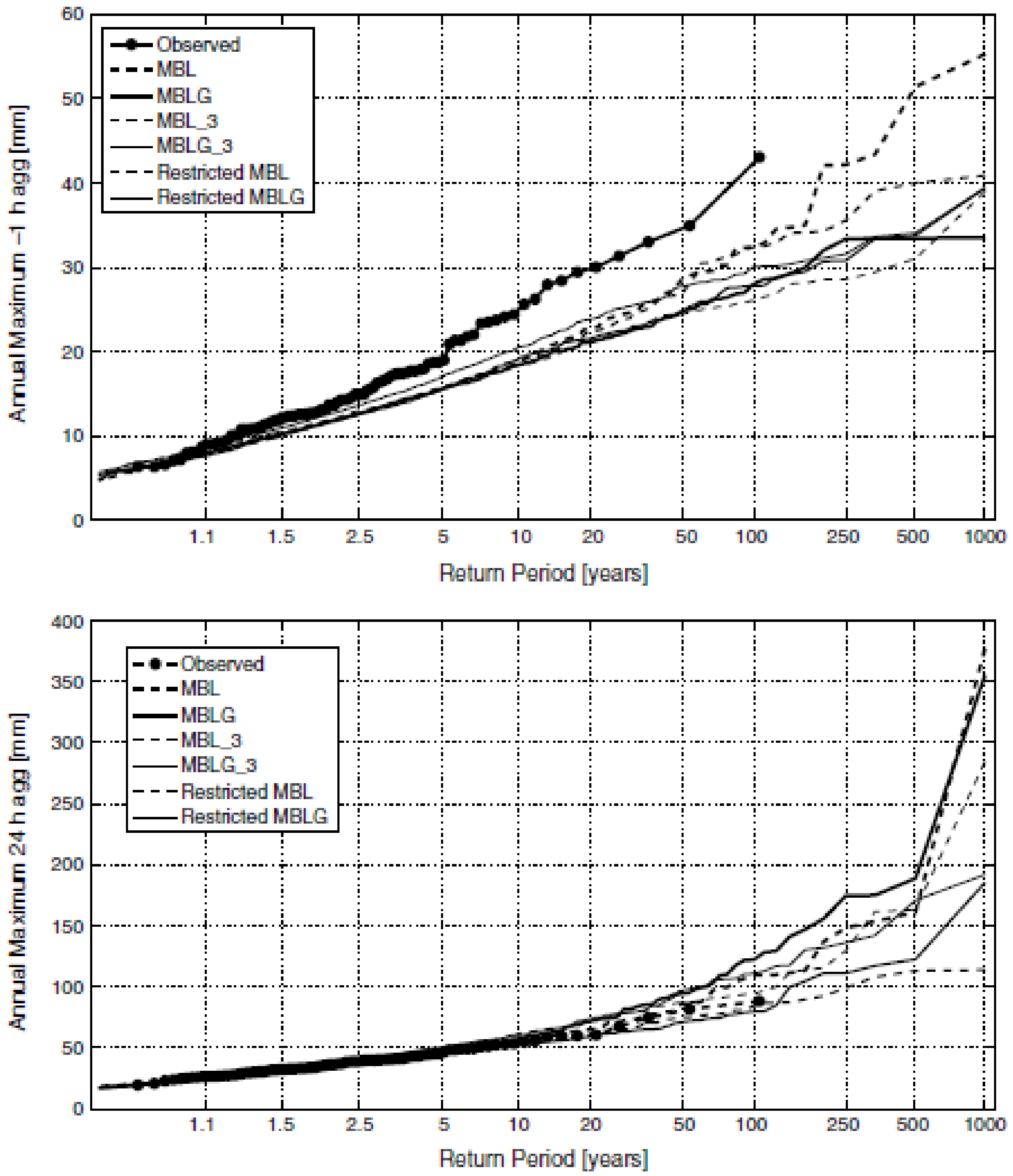
There are many near-optimal parameter sets that produce significantly different types of rainfall, as can be seen, e.g. from the annual extreme rainfall depths



T = 3.2 8 20 55 149 years

2. Problems in reproducing extreme rainfalls

There are many studies (e.g. Verhoest et al., 2010) showing a tendency for many variants of these models to underestimate hourly (and sub-hourly) extreme rainfall depths and overestimate daily extreme depths. This is not necessarily easy to address by changing the distribution of cell depths



3. Problems in reproducing the variance function

As Marani (2003) claimed, the function describing how the variance changes with the time-scale is a very useful indicator of the nature of the stochastic process driving the precipitation.

$$var(R_T) = 2\sigma^2 \int_0^T (T-\tau)\rho(\tau)d\tau$$

Assuming $\rho(\tau)$ is finite, when $T \to \infty$:

if $var(R_T) \propto T$, the process has a finite memory

if $var(R_T) \propto T^{\omega}$, with $1 < \omega < 2$, the process exhibits *long-range dependence*, characterised by the Hurst exponent: $H = \frac{\omega}{2} > 0.5$.This is often because the process has *infinite memory*. It also exhibits scaling behaviour (see also Koutsoyiannis, 2011, 2016),

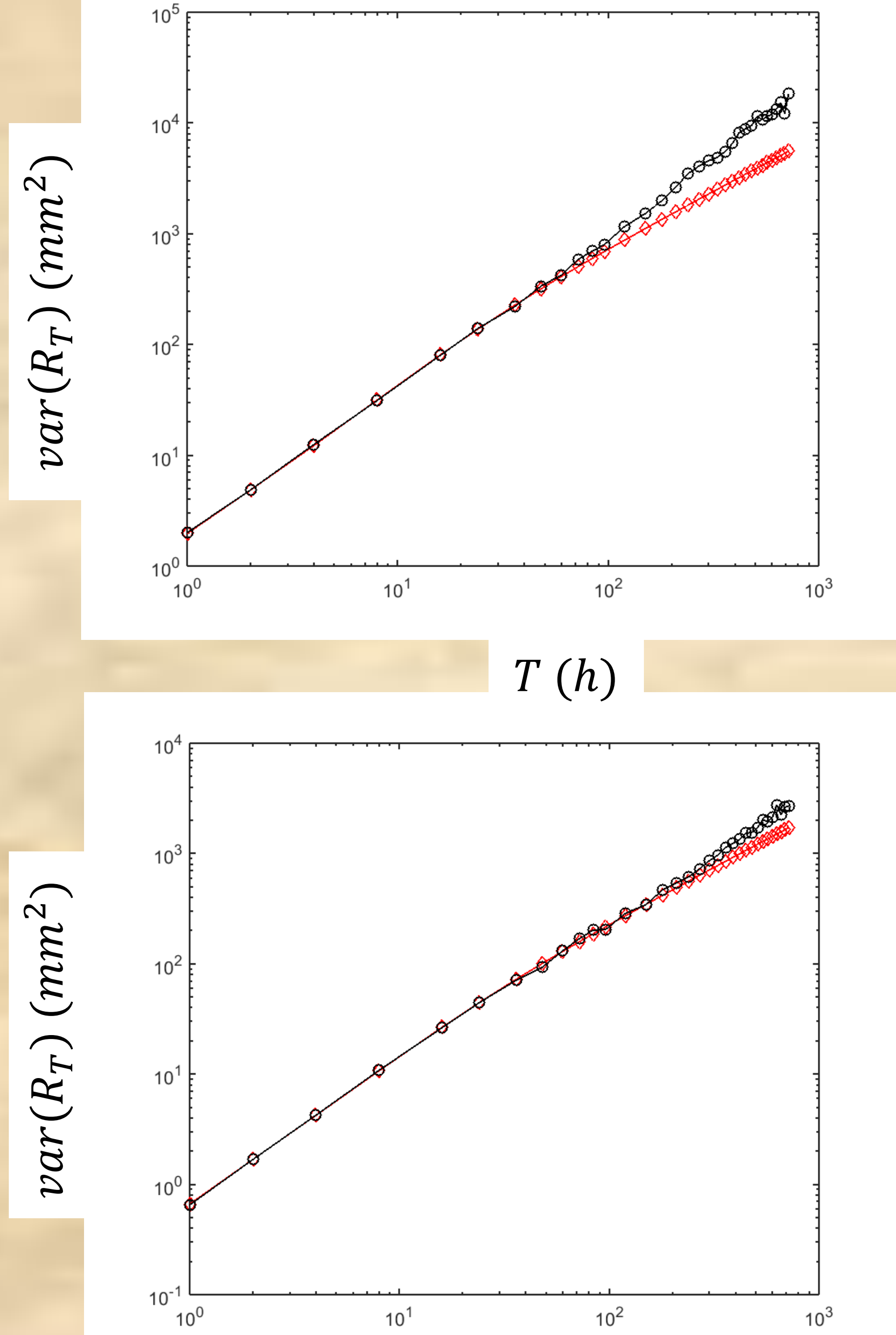
For the Random Parameter Bartlett-Lewis model,

$$\mathrm{d}(\mathrm{var}(R_T))/\mathrm{dT} = \mathrm{a} + \mathrm{b}\,(\mathrm{T}+\mathrm{c})^{2-\alpha} + g\,(\mathrm{fT}+\mathrm{c})^{2-\alpha}$$

and generally $\alpha > 2$, which entails that this model has a finite memory ($\omega = 1$), but the data indicate infinite memory.

Note that *non-stationarity* is reflected in large-scale variability, so the model's limited ability to deal with non-stationarity (through different monthly parameters) is also at stake here.

IV. Main challenges

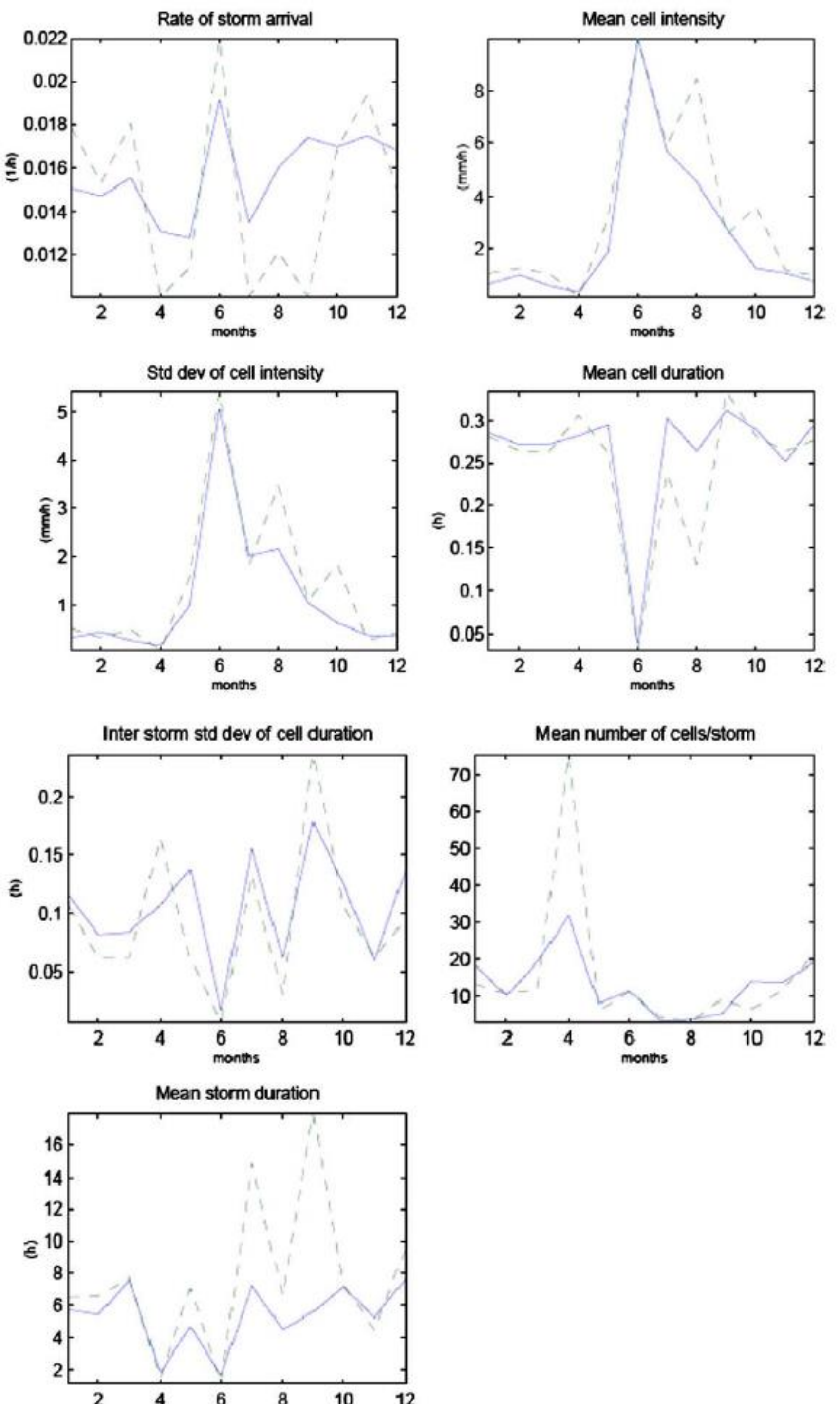


4. Simple methodology for applying such models in the context of a changed climate

Currently, the standard methodology for applying Poisson-cluster models to a changed climate is to infer the change in the p (point) statistics $T(Y_{t,i})$ $(i = 1, \dots, p)$ used for the fit to each calendar month $(t = 1, \dots, 12)$ from the change in the corresponding spatial statistics $S(Y_{t,i})$ $(i = 1, \dots, p;\ t = 1, \dots, 12)$ as they are predicted from runs of a Regional Circulation Model (Kilsby et al., 2007; Onof and Arnbjerg-Nielsen, 2009):

$$\frac{T_{control}(Y_{t,i})}{S_{control}(Y_{t,i})} = \frac{T_{projection}(Y_{t,i})}{S_{projection}(Y_{t,i})}$$

This methodology is widely applied (e.g. UKCP09) although it lacks empirical justification



These problems are interconnected

A - Poor parameter identifiability suggests that there is not enough information in the rainfall signal as it is summarised through moments of different orders of the intensity and wet-dry process to identify a single set of parameters (e.g. even when the skewness is included).

In particular, there is not enough information about extreme intensities so that the different parameter sets lead to a range of extreme value behaviour.

Further information could be obtained from climatological variables; this would give a physical basis to implementations of the model in a changed climate

1. We can seek to include additional information that is not directly about precipitation but its drivers (temperature, SLP, etc.)

2. If we are interested in extreme values only, we can aim to be less ambitious and model a small portion of the depth distribution of rainfall, ignoring low intensities

B – Insufficient variability at very large time-scales (1 month and above) can entail an inability to reproduce intensities that are as extreme (both high and low) as those observed in the data set even at much finer time-scales.

3. We can combine the point process model with another model that generates
a – either coarse-scale (daily+) rainfall
b - or we can use scaling properties of statistics above a certain scale to model the statistics required to fit a BL model

V – Some recent and contemporary work

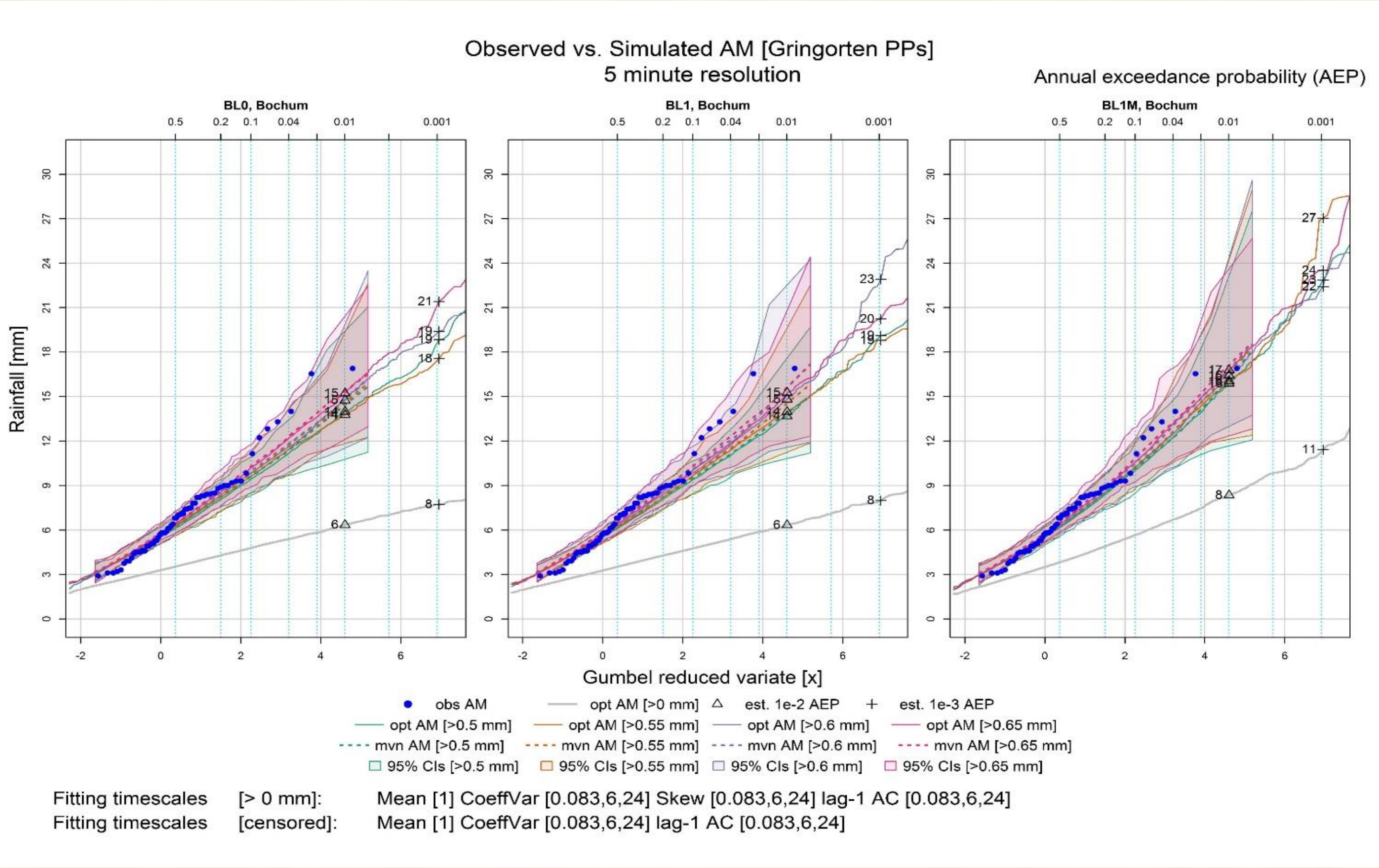
a. Improving the realism of the generated rainfall

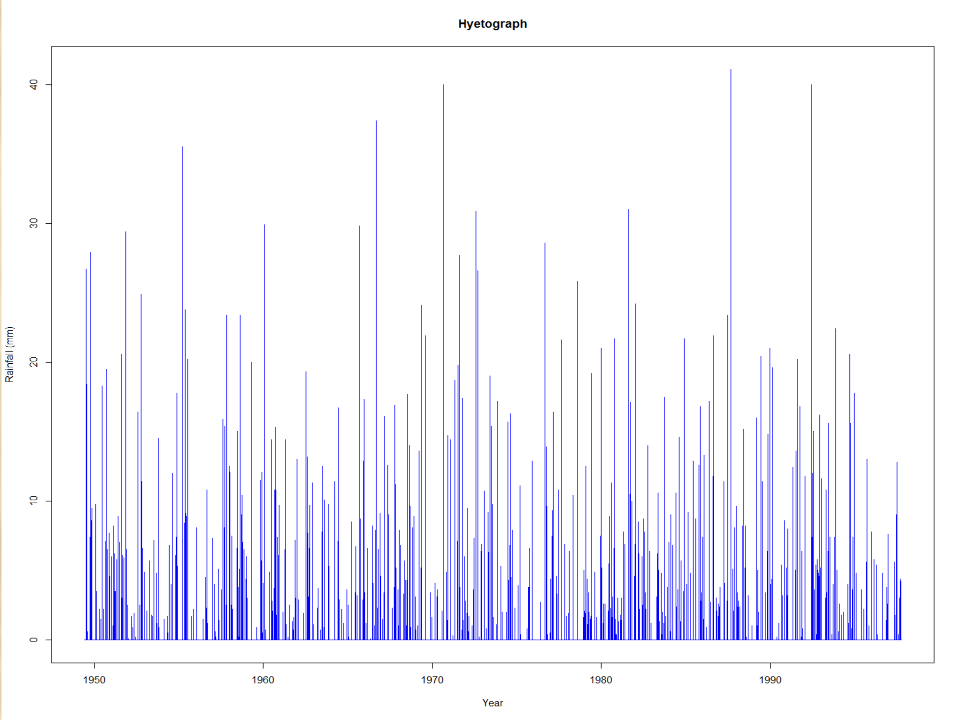
There have been a number of developments in the structure of Poisson-cluster models, some of which are:

- the introduction of two types of cells (Cowpertwait, 1994)
- the replacement of the rectangular cell by another Poisson process, of instantaneous pulses (Cowpertwait, Isham and Onof, 2007)
- the introduction of dependence between cell intensity and duration (e.g. using copulas, Evin & Favre, 2008)
- the development of non-stationary Neyman-Scott models drawing upon the projected changes in rainfall statistics according to Regional Circulation Models (Burton et al., 2010)
- the combination with a random cascade model (Paschalis, Molnar, Fatichi and Burlando, 2014) to improve fine-scale behaviour
- the inclusion of cell intensity in the randomisation of the Bartlett-Lewis model (Kaczmarska, Isham and Onof, 2014) which improves fine-scale behaviour

These papers also make contributions to addressing the main challenges

b. Censoring approach





Questions: 1. How should the censor be chosen?
2. How is the performance at other time-scales?
3. Can the lower intensity rainfall be modelled?
(e.g. use simulated annealing – Bardossy, 1998)

c. Using nearest climatological neighbour data to fit

Standardly, a Bartlett-Lewis model is fitted by separating out calendar months. This is a little arbitrary and the change in the climate suggests a shift of seasons, so it may not be optimal to use only January rainfall to obtain parameters for future January rainfall for instance.

In Kaczmarska et al. (2015), a Local Generalised Method of Moments is applied so that the parameters for a given month are obtained by using data from all months, weighted by the “proximity” which their climatology exhibits to the month of interest. This has been tested for climatologies defined by Temperature and SLP.

So far, for a data set of n months, and with p properties, the optimal parameters were estimated as:

$$\hat{\theta}_m = argmin_{\theta_m}\left\{\sum_{i=1}^{p} w_i\left[\frac{1}{\sum_{t=1}^{n} I(m_t=m)}\sum_{t=1}^{n} I(m_t=m)\big(A_i(\theta_m)-\Omega_i(Y_t)\big)\right]^2\right\}$$

Consider a data set of n months of rainfall and with p statistics for each month $\Omega_i\ (Y_t)\ (t = 1, \dots, n; i = 1, \dots, p)$, and consider now the use of weights $w_i\ (i = 1, \dots, p)$ chosen as the inverse of the variance of property i (Jesus and Chandler, 2011)

Standard generalised method of moments with optimal weights:

An optimal parameter set is obtained for each calendar month $m\ (m = 1, \dots, 12)$ by minimising an objective function based in which only the properties for month m in the data set will play a role

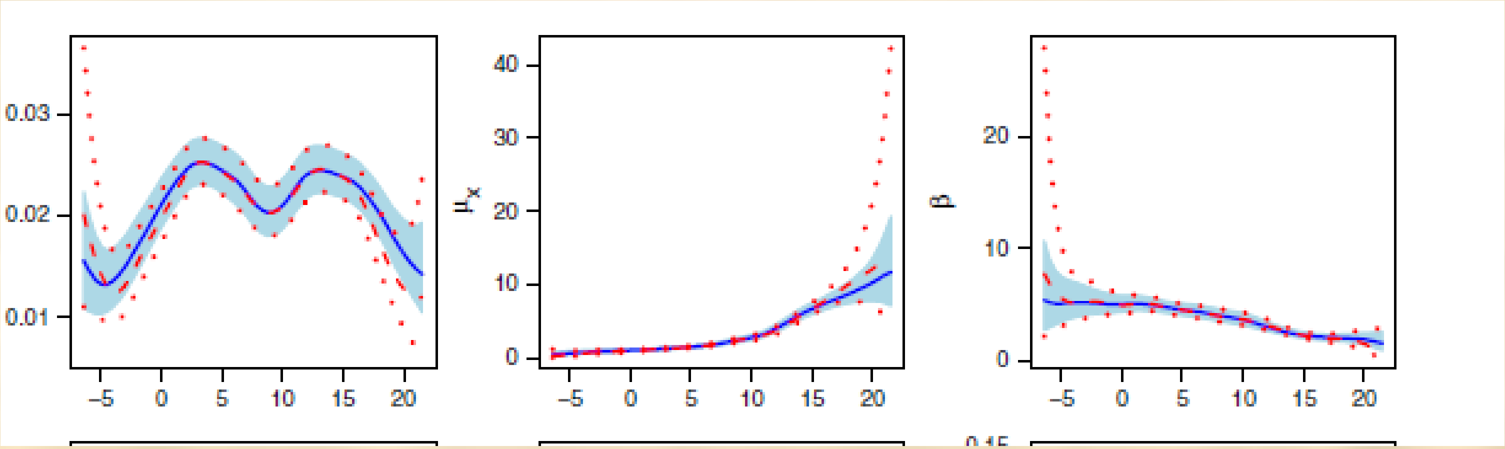
$$\hat{\theta}_m = argmin_{\theta_m}\left\{\sum_{i=1}^{p}\frac{1}{var(\Omega_i)}\left[\frac{1}{\sum_{t=1}^{n} I(m_t = m)}\sum_{t=1}^{n} I(m_t = m)\big(A_i(\theta_m) - \Omega_i(Y_t)\big)\right]^2\right\}$$

Local generalised method of moments with optimal weights:

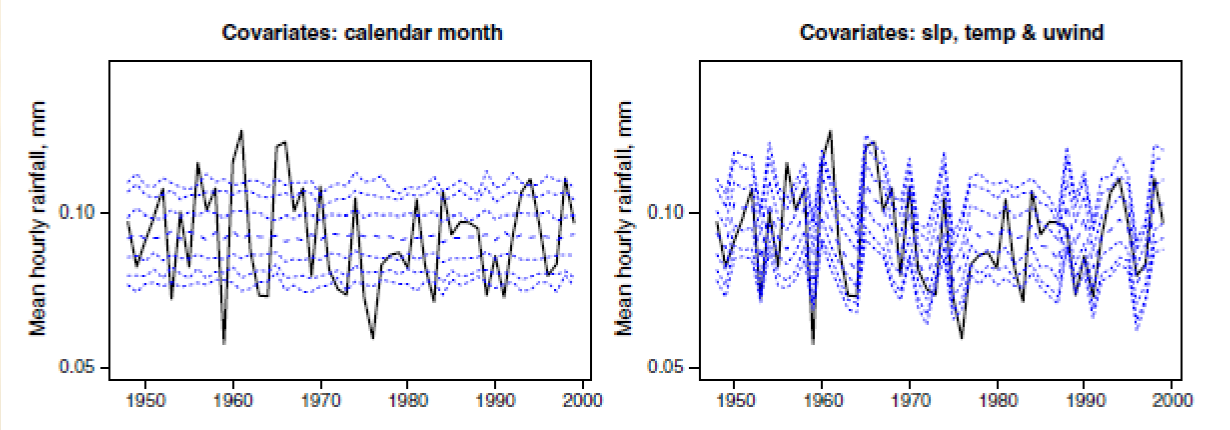
Additionally, we now consider the availability of data for a vector of monthly covariates $\boldsymbol{X}$: $\{\boldsymbol{X_t}\}\ (t = 1, \dots, n)$
With this, we can now estimate a parameter value for any possible value x_0 of covariate $\boldsymbol{X}$
To that end, we introduce a measure of the closeness of each month t of the data set to a month with covariate value $\boldsymbol{x_0}$: $K_h(\boldsymbol{X_t} - \boldsymbol{x_0})$.
This is defined as $K_h(\boldsymbol{X_t} - \boldsymbol{x_0}) = \frac{K[(\boldsymbol{X_t} - \boldsymbol{x_0})/h]}{h}$ where K(.) is a kernel function (e.g. a Gaussian kernel)

$$\hat{\theta}(x_0) = argmin_{\theta_m}\left\{\sum_{i=1}^{p}\frac{1}{var(\Omega_i)}\left[\frac{1}{\sum_{t=1}^{n} K_h(\boldsymbol{X_t} - \boldsymbol{x_0})}\sum_{t=1}^{n} K_h(\boldsymbol{X_t} - \boldsymbol{x_0})\big(A_i(\theta_m) - \Omega_i(Y_t)\big)\right]^2\right\}$$

Parameter identifiability is improved:



Model is able to display more variability thus improving extreme value performance:



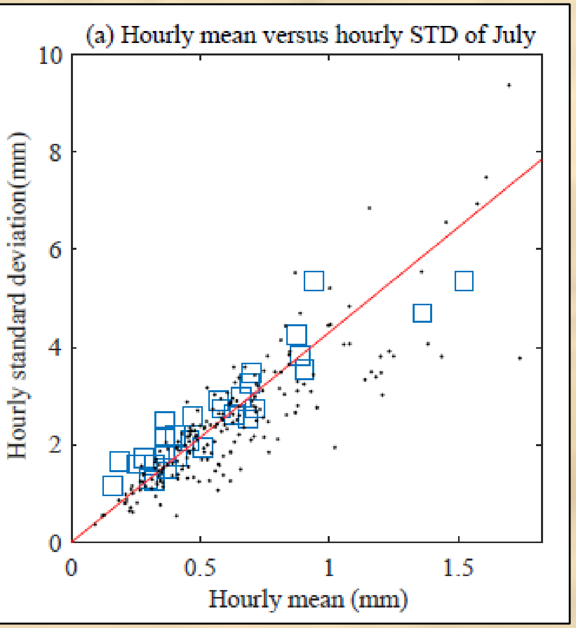
<u>Questions</u>: 1. How should the bandwidth *h* be chosen?
2. Is a monthly value sufficient to take on board the role of climatological variables in generating precipitation?

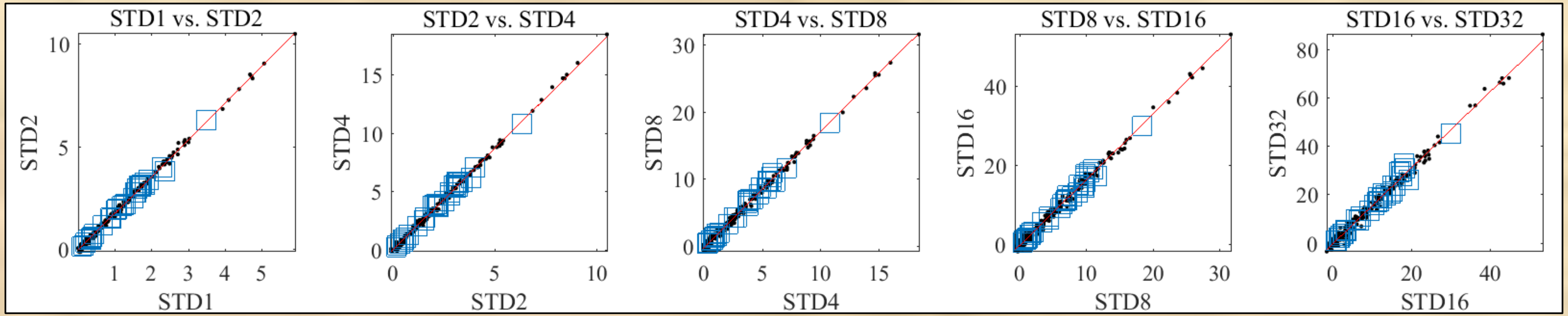
d. Combining Poisson cluster model with a model for coarse scale statistics

The idea here is to fit different models to each month of the data set (Kim et al., 2016). If a suitable model can be found to represent the set of statistics $\{T(Y_t)\}$ required in the fit for month t, then larger scale variability can be conserved by this model.

The scaling properties of these statistics make such a coarse-scale model

Questions: 1. Is a Markov structure for these coarse-scale statistics sufficient?
2. How parsimonious is the final model?



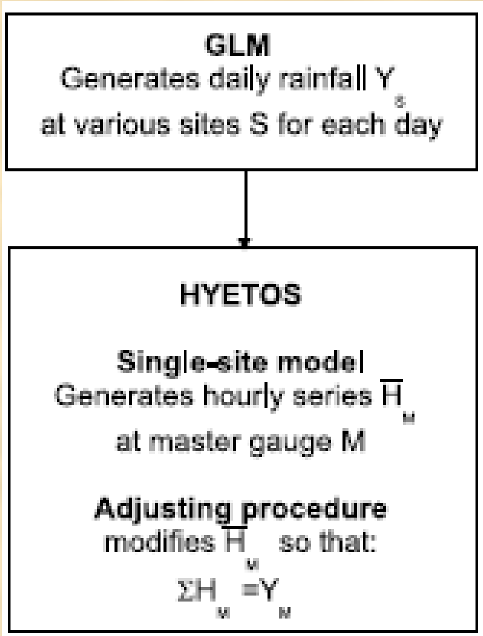


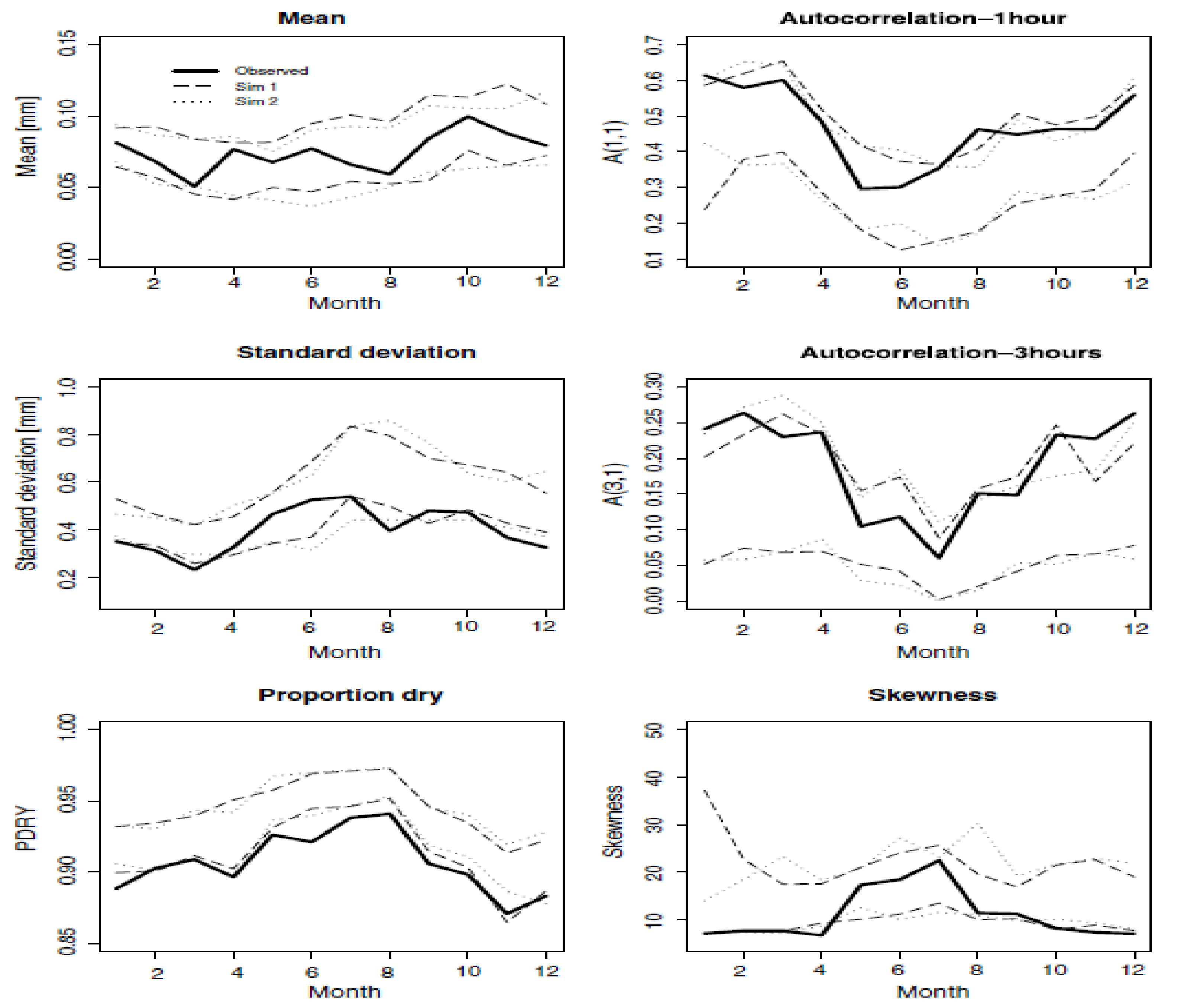
e. Combining Poisson cluster model with a model for the coarse scale rainfall depths

As with the previous option, the idea is to use two models, one of which will ensure the appropriate coarse-scale variability is reproduced. Here, however, this coarse-scale is daily, which will enable us to control rainfall generation using daily climatological information.

There are many candidates for such a daily weather generator. A GLM is chosen, and it is combined with a Bartlett-Lewis model in disaggregation mode (see Koutsoyiannis and Onof, 2001) using the HYETOS software.

This approach has the advantages of (i) allowing the incorporation of any relevant explanatory variable (climatological or other); (ii) allowing rainfall for a changed climate to be generated by incorporating dependence upon climatological variables at the daily time-scale.

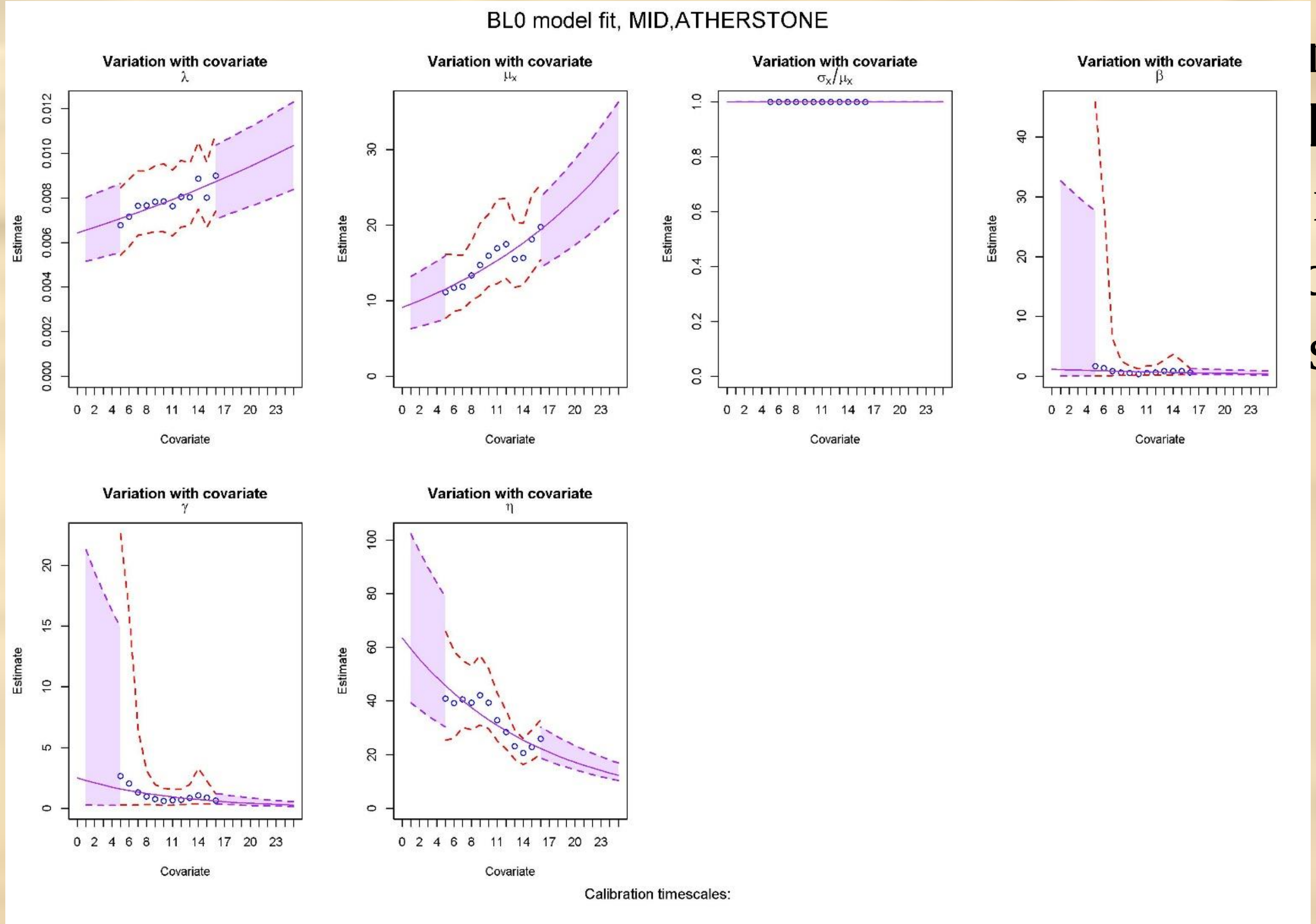




Questions: 1. Does the combination of these two models not introduce too much variability?

2. Does this improve the reproduction of extremes?

f. Combining Poisson cluster model with a model for weather types



nthly time-scale in
logical dependence by
rainfall characteristics or
d Hayhoe, 2007), and
s can then be generated by

Questions: 1. How easily can the Bartlett-Lewis be fitted to rainfall from a given weather type?
2. Might climate change applications not require that new weather types be introduced?

VI. Conclusion

From the above recent and ongoing work, we can see that there are a number of promising areas of research centred upon the use of Poisson-cluster Rectangular Pulse models.

Having chosen to use a Poisson-cluster model as tool for a weather generator, the following choices need to be made:

- A choice of method/model to combine it with so as to reproduce large-scale variability and/or include climatological information
- A decision as to how best to use the data, ideally including climate information to guide the use of historical information
- A choice of time-scales at which the Poisson-cluster model is to be fitted or constrained
- A decision as to whether to divide the data into periods of defining fixed seasons (e.g. months), or periods defined by weather-types.