Building models for spatial-temporal rainfall fields.

Valerie Isham

Department of Statistical Science
University College London

valerie@stats.ucl.ac.uk, *http://www.ucl.ac.uk/stats/*

Cherry Bud Workshop, Keio University, March 2006

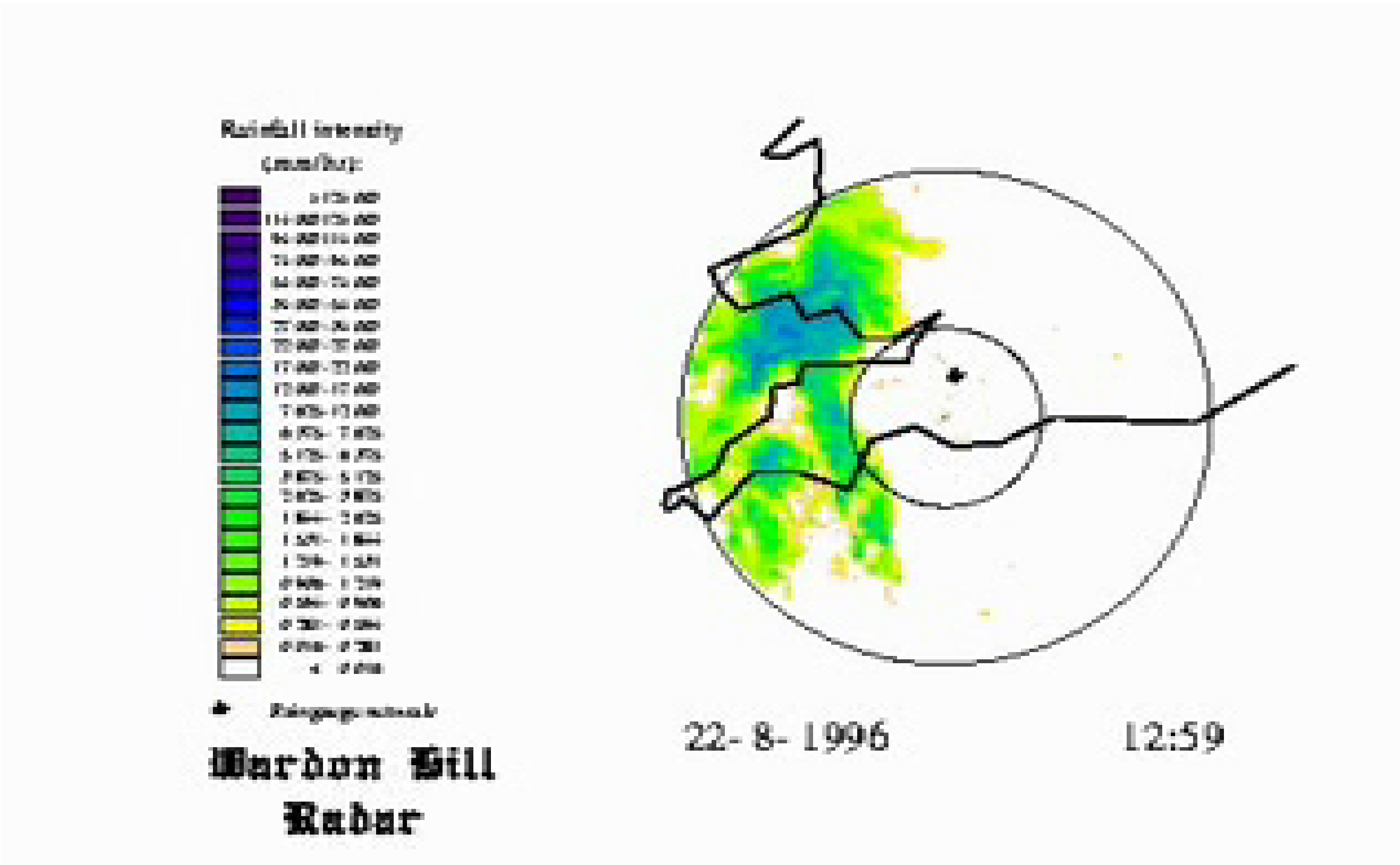
Collaborators:

Richard Chandler[†], Paul Northrop[†]
Howard Wheater*, Christian Onof*
Enrica Bellone[†]*, Chi Yang [†]*

[†] Department of Statistical Science, University College London
*** Department of Civil and Environmental Engineering, Imperial College London**

Acknowledgements: DEFRA

Towards continuous simulation rainfall-runoff modelling for flood risk assessment



Goal: Flood management and assessment of risk, allowing for impacts of changing climate and land use.

Tools: Rainfall-runoff models

- *inputs:* precipitation and evaporation data

- *output:* flows

But: Historical data lack length, temporal resolution and spatial coverage

Need: Long continuous simulations of the input that preserve

- extreme value properties
- spatial structure

Overview

Stochastic (point process-based) models

- within an event
- radar data, model-fitting and assessment
- the advection process
- stationary continuous simulation

Statistical models (GLMs)

- model construction

Combination for nonstationary continuous simulation

Conclusions and future directions

To achieve continuous nonstationary simulation…

… combine

stochastic (point process-based) models

- in *continuous space-time* (aggregate as necessary for fitting) enabling required space-time resolution
- *parsimonious* parameterisation
- parameters relate directly to physical phenomena

$\Rightarrow$ stationary simulation

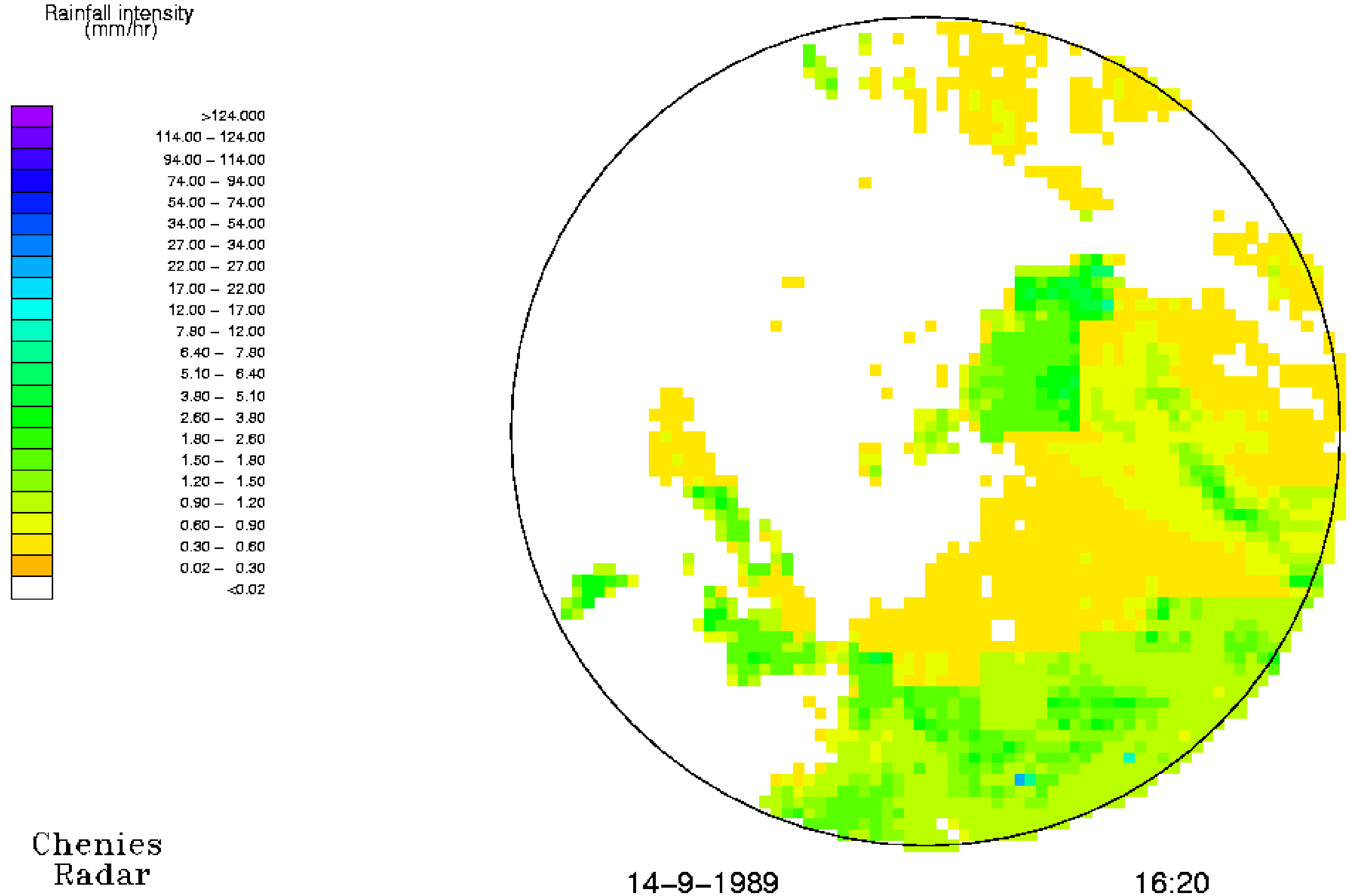
statistical models (GLMs)

- in *discrete space and time*
- dependence on *explanatory variables (eg season, orography)* enabling required space-time nonstationarities

Radar data:

- used for fitting stochastic models
- indirect – reflected energy converted to rainfall intensity, Chenies radar: data from 1990 onwards
- discretised intensities (0.03mm/hr min. non-zero value)
- images at 5 minute intervals, assumed *instantaneous*
- *spatially averaged* rainfall intensities over a rectangular grid of 2×2 km^2 pixels, over circular region, radius 76 km centred on radar
- model fitted to inscribed square of 52 x 52 pixels
- data calibrated on site to ground truth (using rain gauge data)

Chenies radar data - 2km resolution - calibrated on site



Rain gauge data:

- 122 tipping bucket gauges under Chenies radar
- pre-processed to 15 minute rainfall *totals* (0.2mm resolution)
- (spatial) *point* data

Stochastic point process-based rainfall models

Point process-based models with hierarchical structure:
rain *cells* cluster within *storms* within rain *events*….

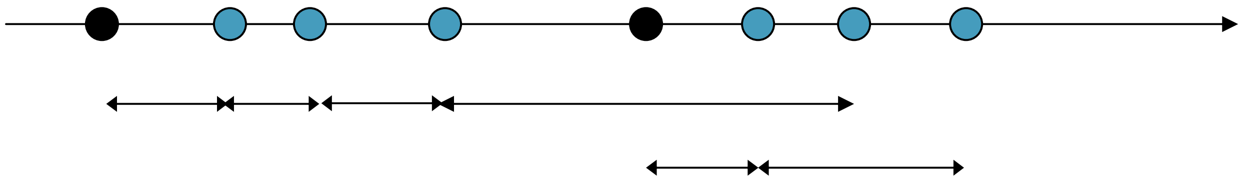
Two stages: the *within-event* structure
the *sequence of rain events* *(advection)*

First…..

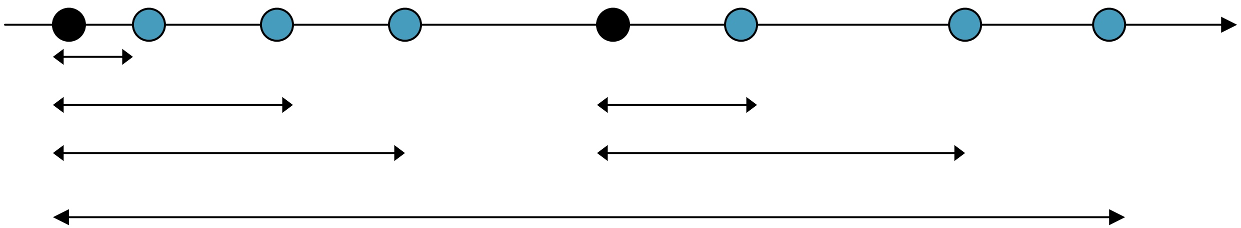
the within rain-event structure

All storms and cells within a rain event have a common *velocity* v and characteristic elliptical ‘*shape*’ (e, θ)

Poisson cluster processes



Bartlett-Lewis process



Neyman-Scott process

● Cluster centre ● Event arrival

⟷ Independent and identically distributed intervals

A snapshot of one storm cluster.

A rain event is the superposition of lots of such clusters

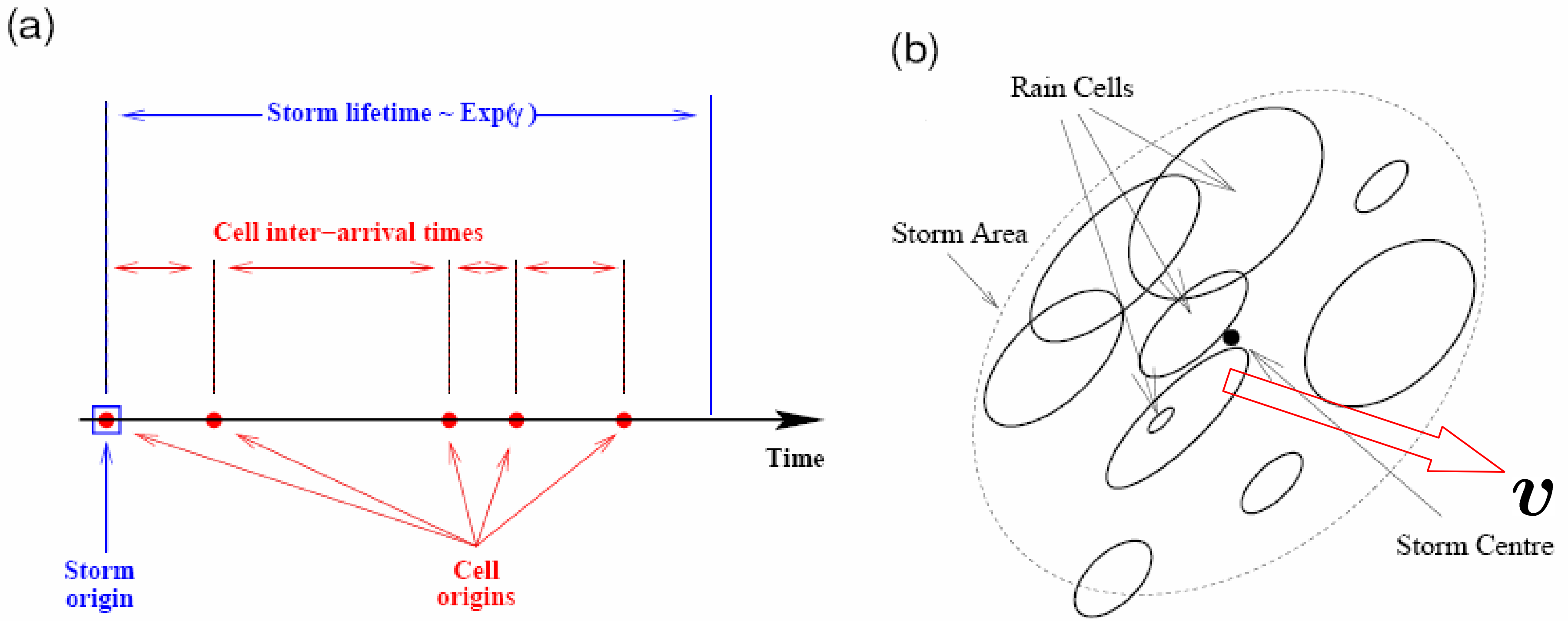


Figure 6.1 Schematic diagram of a storm in the space-time model for the interior of a rain event. (a) Temporal structure: cell origins occur in a Poisson process during the lifetime of a storm, with a cell at the storm origin. (b) Spatial structure: each cell is elliptical, and is displaced from the (moving) storm centre according to a bivariate Gaussian distribution with the same elliptical shape.

Cluster structure

- Storm origins occur in a Poisson process in space and time
- Each storm has a random (exponential) *lifetime*
- Within the storm lifetime,
  - cells occur in a Poisson process in time

    $\Rightarrow$ no of cells has geometric distribution
  - their spatial displacements *relative to moving storm origin* are *i.i.d* Gaussian variates with random *scale* and elliptical contours

$\Rightarrow$

storm has Bartlett-Lewis structure in *time* and Neyman-Scott structure in *space*

Rain cells within clusters

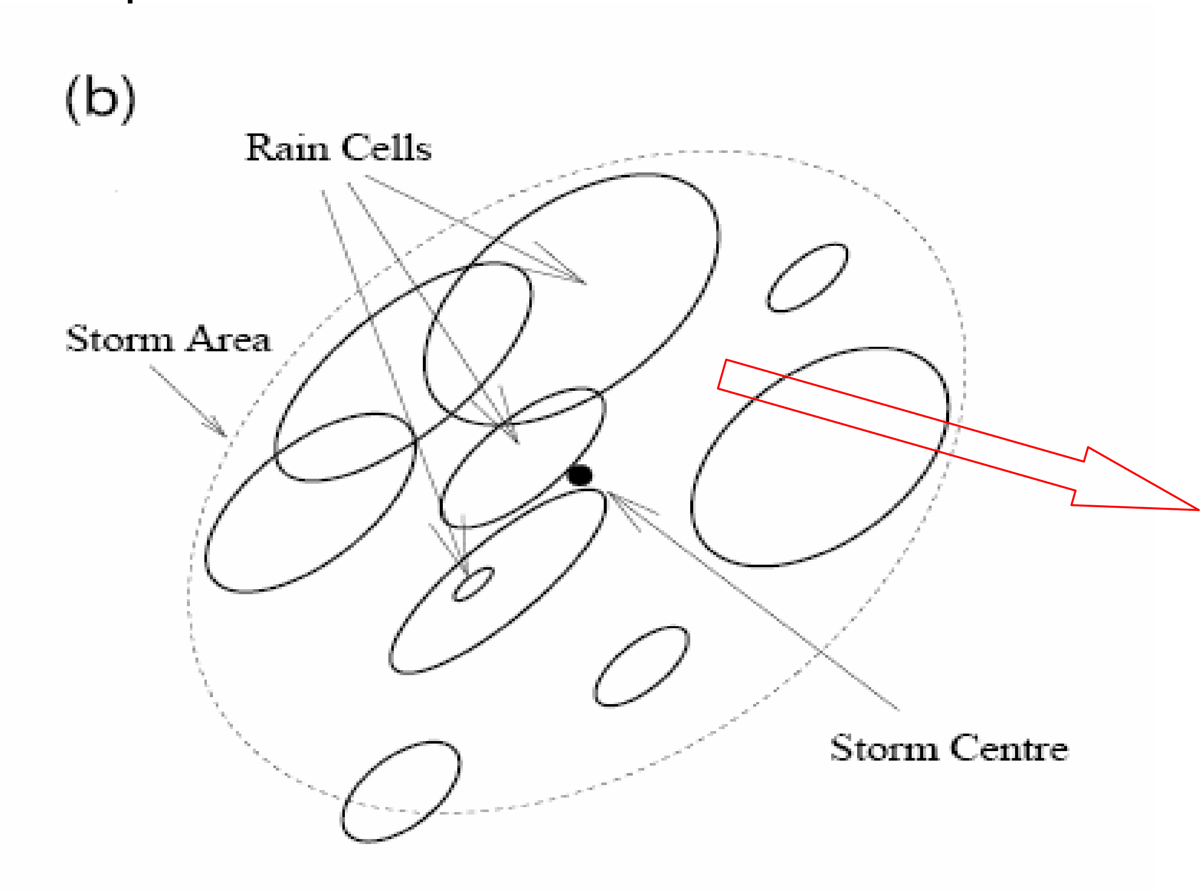
- Cell origins form a *Poisson cluster process* $N(w, \tau)$ in space and time
- Within a cluster, the cells are *i.i.d.* with
  - the *same* elliptical shape (e, θ)
  - random cell *scales*
  - exponential cell *lifetimes*
  - random cell *intensities* X that are constant over a cell's spatial and temporal extent

Effectively each cell is a random elliptical cylinder

A snapshot of one storm cluster.

A rain event is the superposition of lots of such clusters

All clusters have the same velocity, and same characteristic elliptical shape but different spatial scales.



- Cells and storms can overlap.
- The model has 11 (or more) parameters

  - the *distribution* of the cell intensity is not used in model fitting

  - the storm and cell scales are each modelled using a gamma distribution with one parameter fixed
- The temporal structure of the process is Markov

  $\Rightarrow$ availability of explicit expressions for 2nd order properties
- The within-event model is spatially and temporally stationary
- The assumption of a constant cell intensity does not appear statistically important, but *e.g.* a multiplicative random noise could be applied

The *total* cell intensity at u at time t is the sum of the intensities of all the cells covering u at t, which can be written as

$$Y(\boldsymbol{u},t) = \int_{\tau=-\infty}^{t} \int_{\boldsymbol{w}\in\mathbb{R}^2} I(\boldsymbol{w},\tau;\boldsymbol{u},t) X(\boldsymbol{w},\tau) dN(\boldsymbol{w},\tau)$$

and we *observe* the average intensity over a pixel of side h

$$Y_{i\,j}^{(h)}(t) = \frac{1}{h^2} \int_{(i-1)h}^{ih} \int_{(j-1)h}^{jh} Y(\boldsymbol{u},t) du_2 du_1$$

Explicit expressions for the mean intensity and second order properties

$$\rho^{(h)}(\boldsymbol{k},\tau) = \text{corr}\left(Y_{i\,j}^{(h)}(t), Y_{i+k_1\,j+k_2}^{(h)}(t+\tau)\right)$$

Method of model fitting to event interiors

- essentially a method of moments fit using second order properties at a range of spatial and temporal scales
- rain events must have a 15+% coverage, and be spatially and temporally stationary, for at least an hour

Assessment of model fit

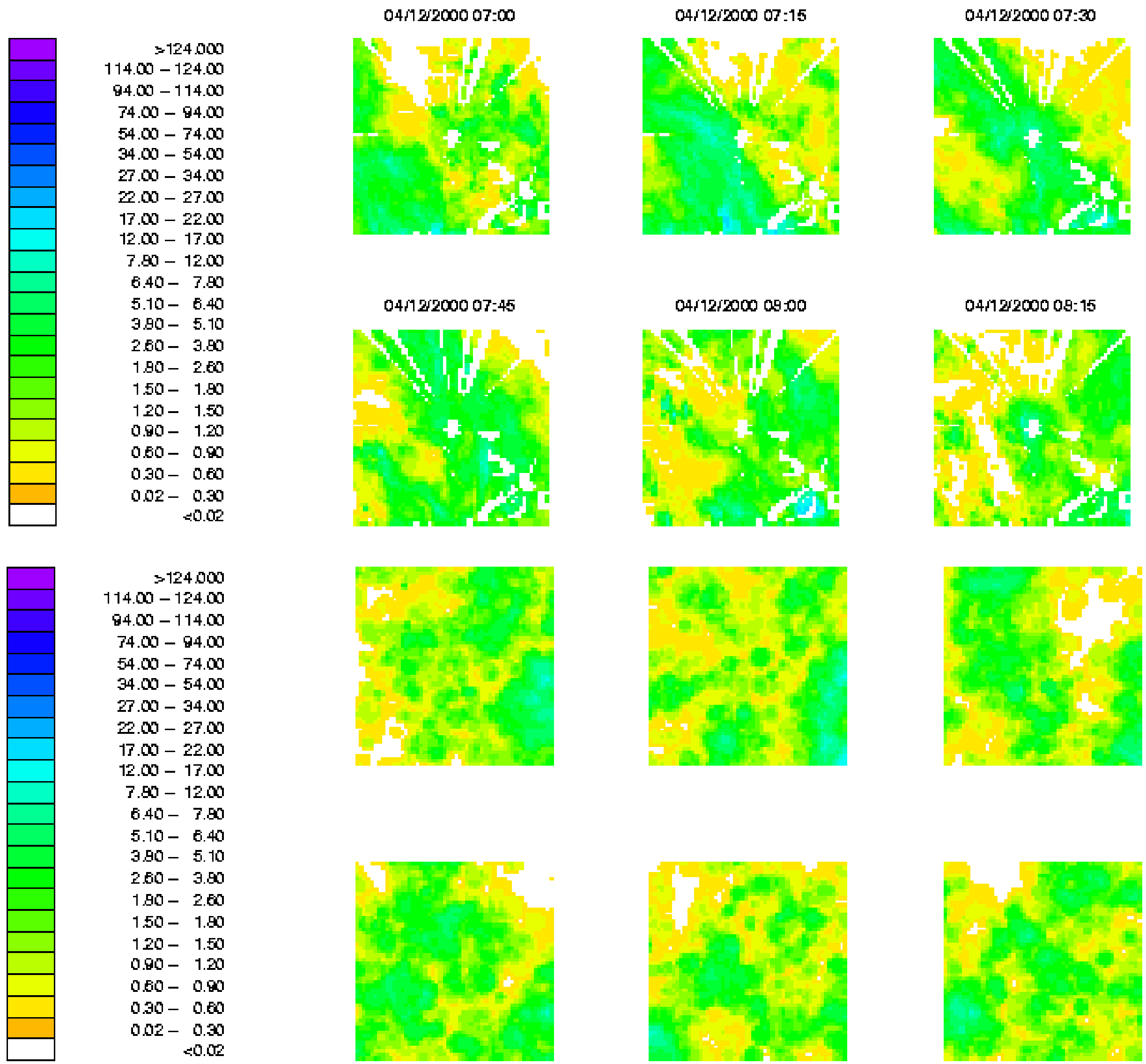
- visual appearance of images **
- comparison of the fitted properties (both explicit and simulated) with empirical sample properties, including
  * moments
  * wet/dry pattern and coverage
  * cumulative rainfall intensities
  * effect of thresholding on the coverage and intensity

at a range of spatial and temporal scales

Event of 4 Dec. 2000

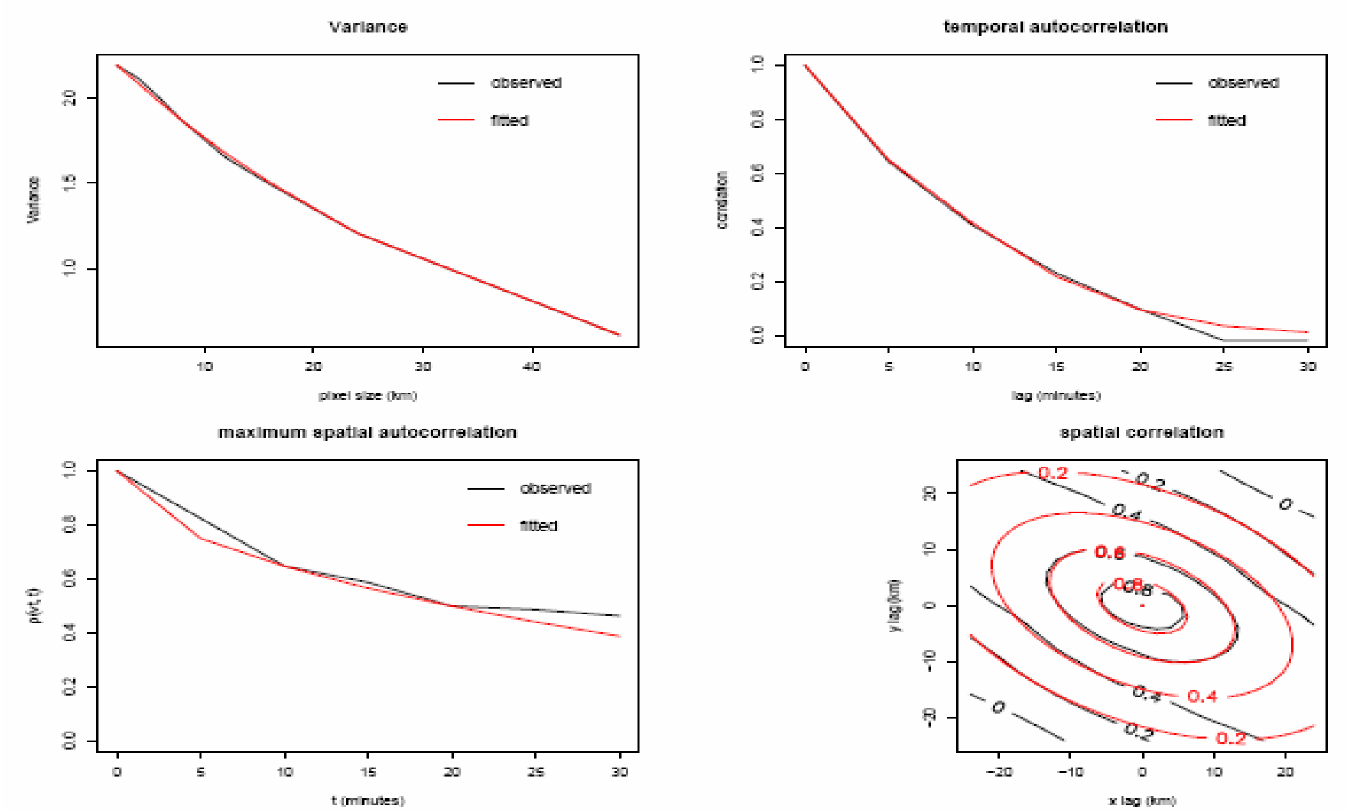
top: 75 min sequence of radar images;

bottom: 75 min sequence of simulated images



Observed and fitted properties: event of 4 Dec. 2000

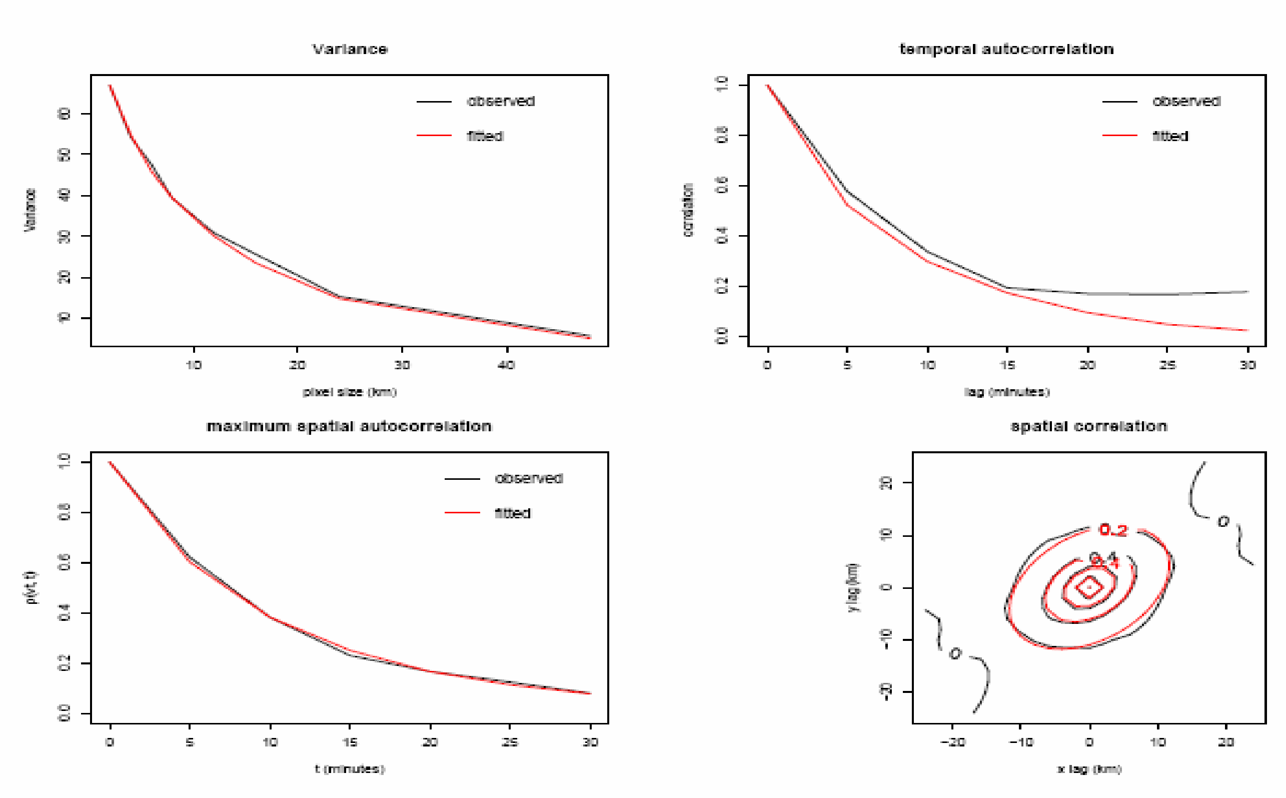
left: variance as a function of spatial scale
right: temporal correlation



left: autocorrelation along estimated direction of event movement
right: spatial correlation

Observed and fitted properties: event of 19 Nov. 1991

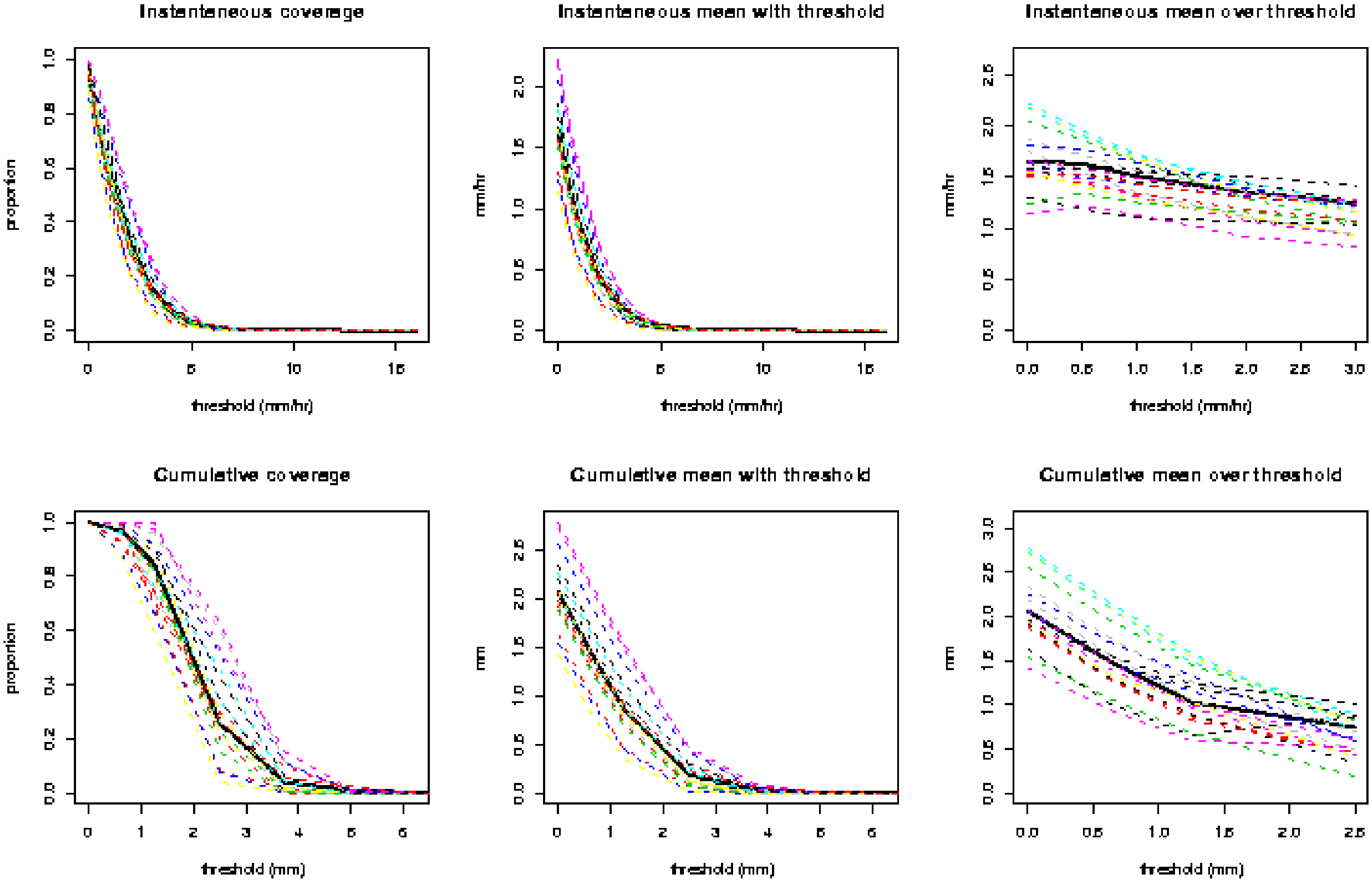
left: variance as a function of spatial scale
right: temporal correlation



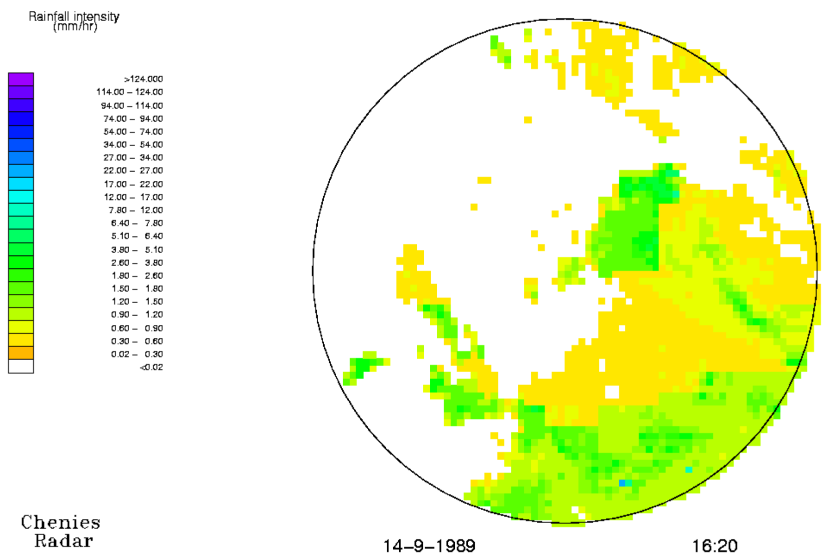
left: autocorrelation along estimated direction of event movement
right: spatial correlation

Threshold analysis for event of 4 Dec. 2000

coverage (lhs), mean (centre), mean over threshold (rhs)
upper: single images; lower: cumulative images over 75 mins
solid lines: data; dashed lines: individual simulations



Calibration issues



Chenies radar data:
2km resolution
calibrated on site

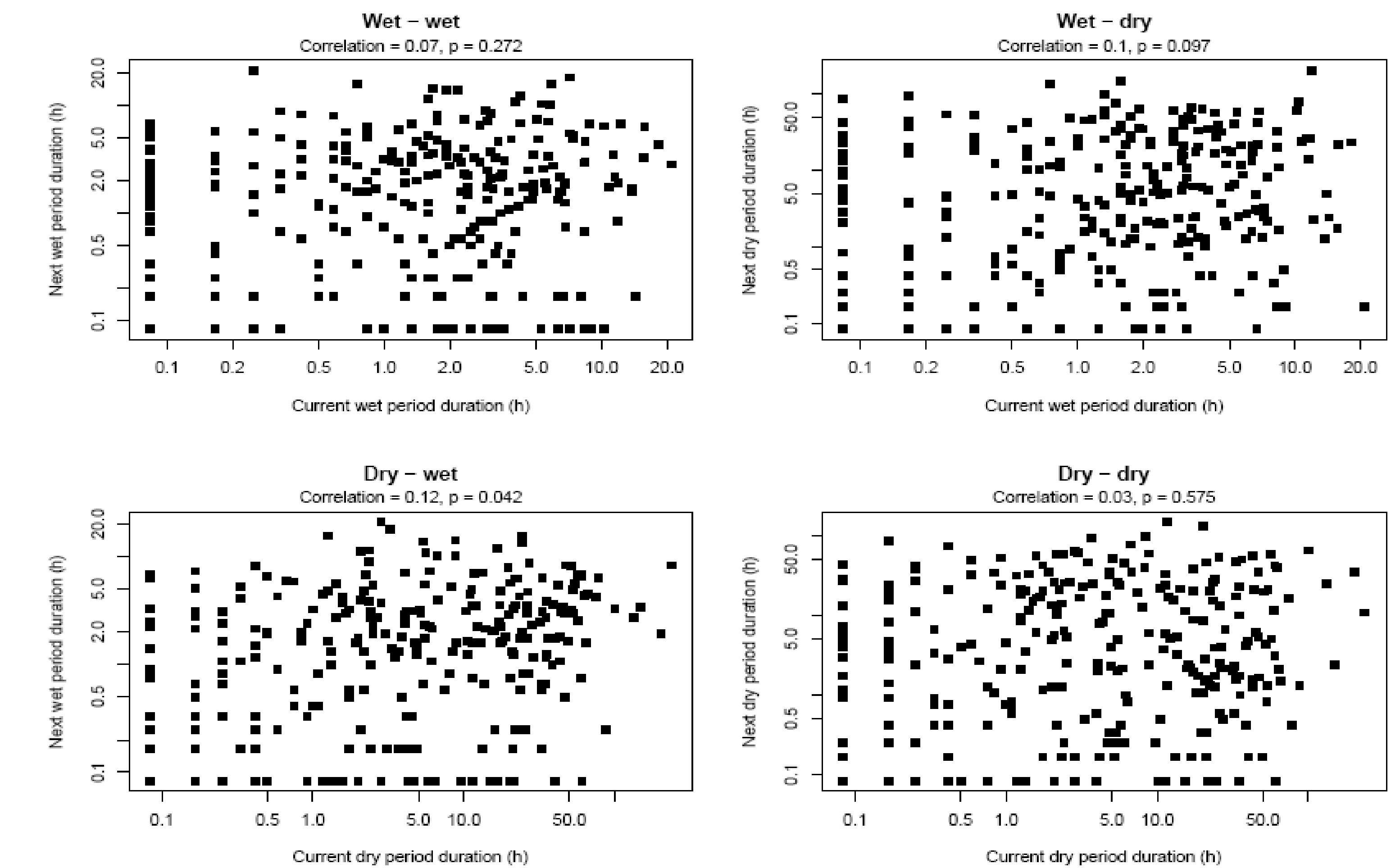
Calibration needs to be smooth in space and time
$\Rightarrow$ recalibration

Issue …. sensitivity of summary statistics used for model fitting to calibration

Modelling arrivals and departures of rain events over catchment (advection)

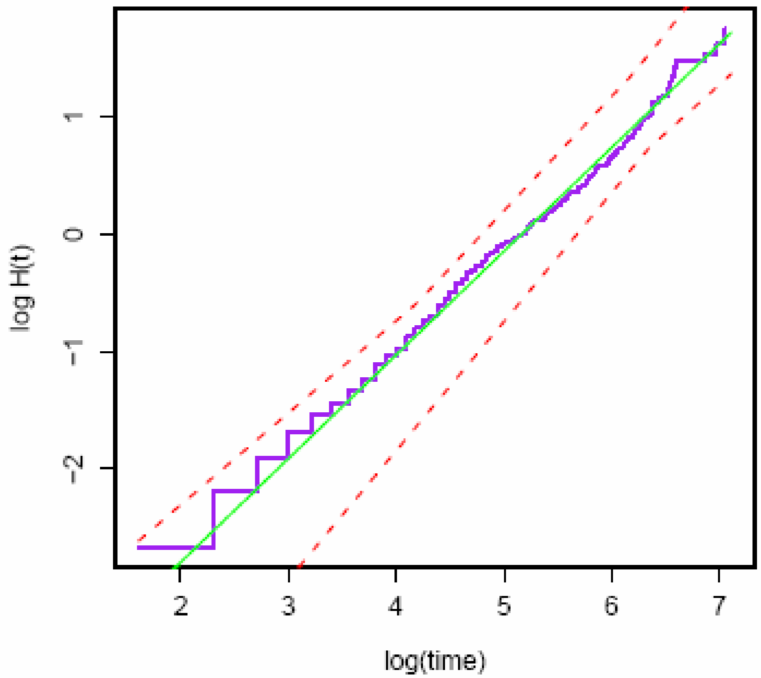
- identify start and end times of rain events by *up* - and *down-crossings* of time series of window *coverage* (proportions of pixels that are wet) across a threshold
- model sequences of durations of ‘dry’ and ‘wet’ intervals as independent *i.i.d.* sequences of Weibull variables (tractable likelihood for censored observations)
- fit orientations of leading and trailing edges of events via linear discriminant analysis

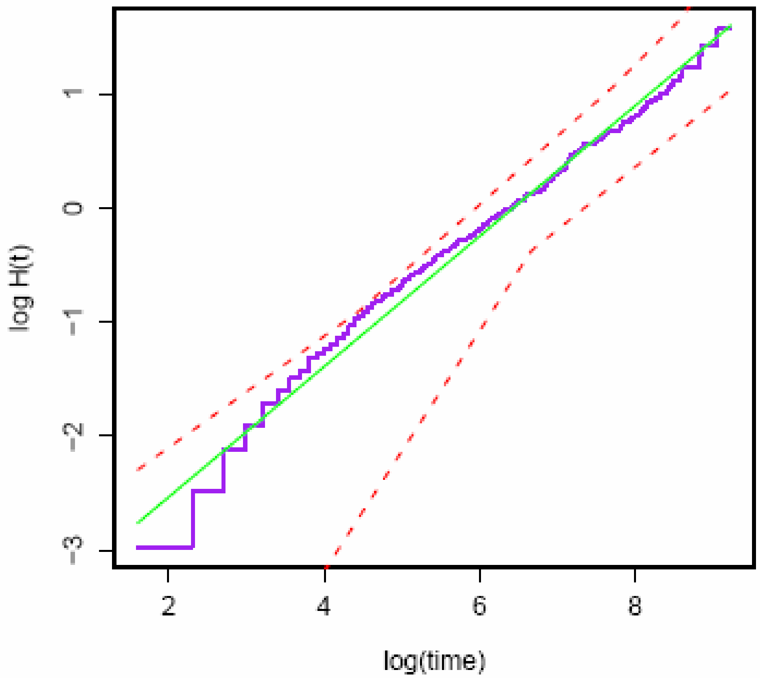
Durations - in log(hours) - of successive wet and dry period durations in July.



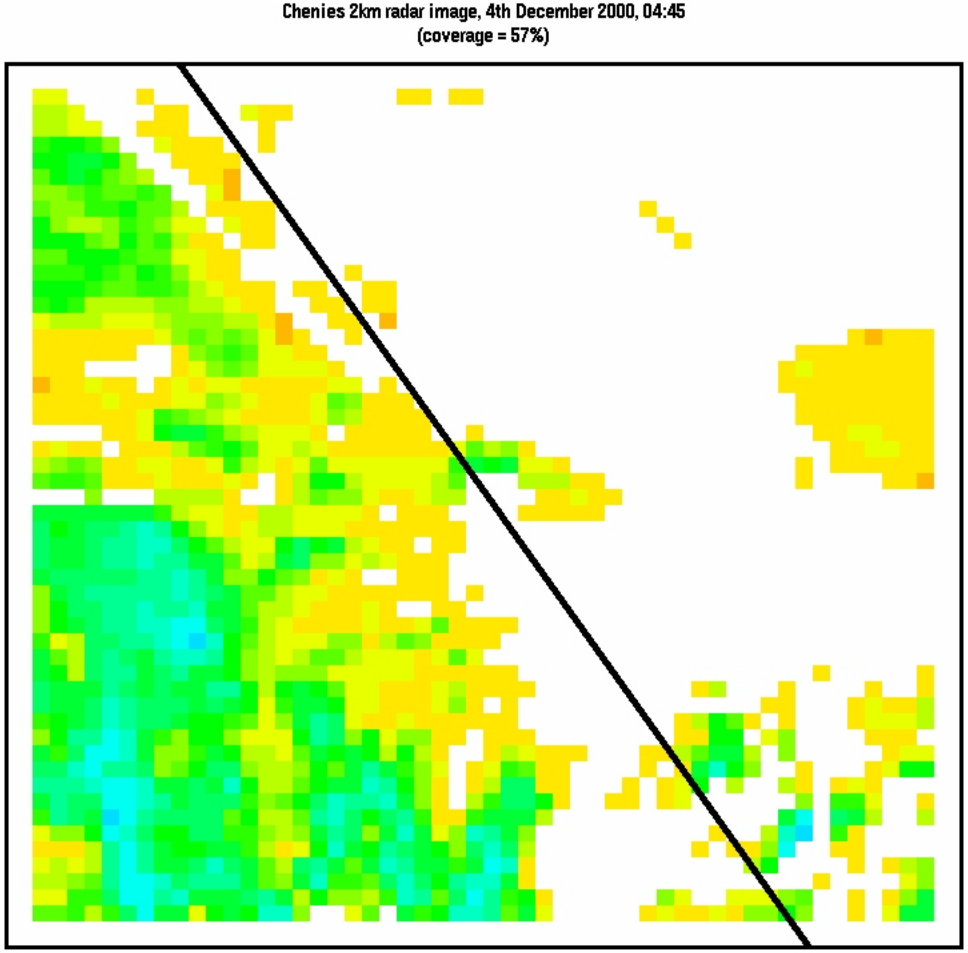
Assessment of Weibull fit:

log-log plots of empirical (step) and theoretical (straight line) cumulative hazard functions for wet and dry period durations in January. Dashed lines indicate the envelopes from 10 simulations.





Leading edge for event of 4 Dec. 2000



Continuous simulation of rainfall

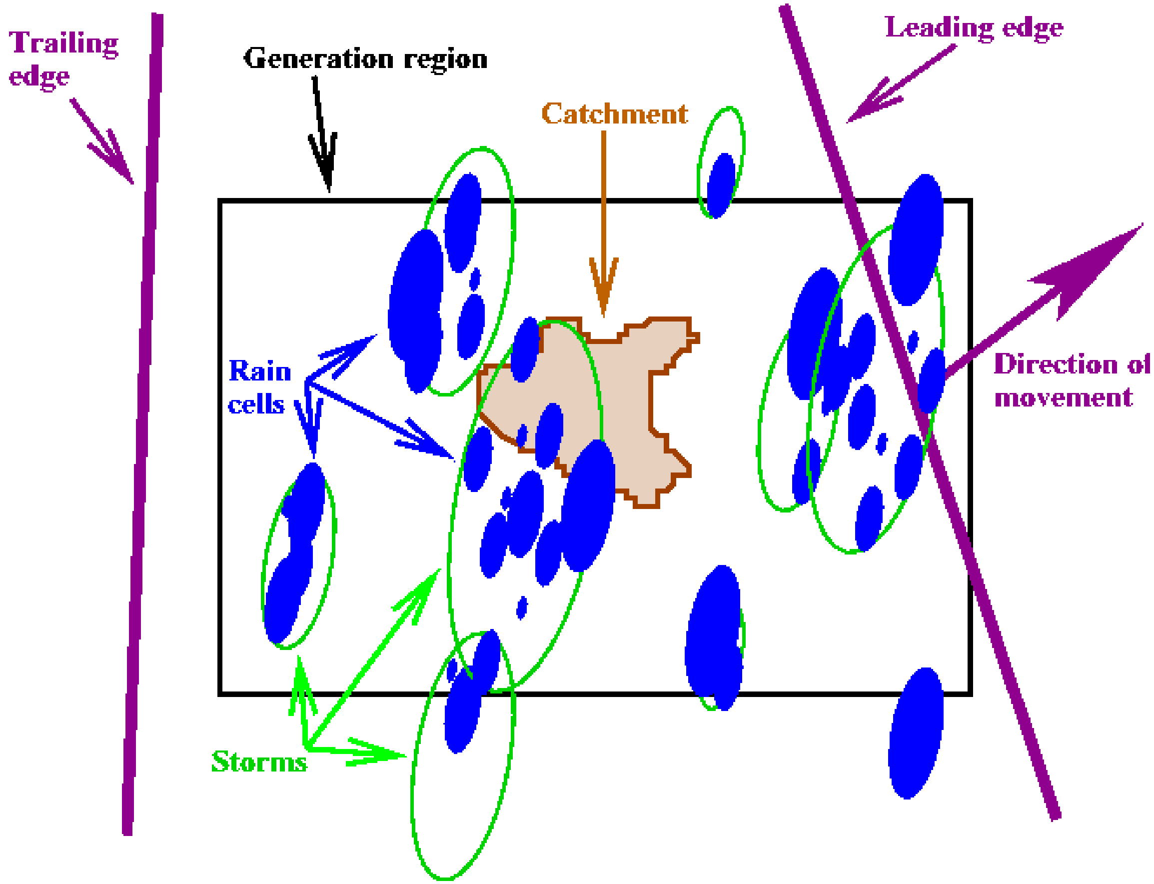
Assemble

- library of parameter sets for large number of fitted events including orientations of leading and trailing edges of each event

Generate - from fitted distributions (on month-by-month basis to allow for seasonality)

- sequence of durations of rain events and inter-event dry periods in alternating renewal process
- other event parameters (including velocity) sampled from library of fitted sets for events *with similar durations*
- rain band wide enough to cover catchment for given event duration, moving at given velocity

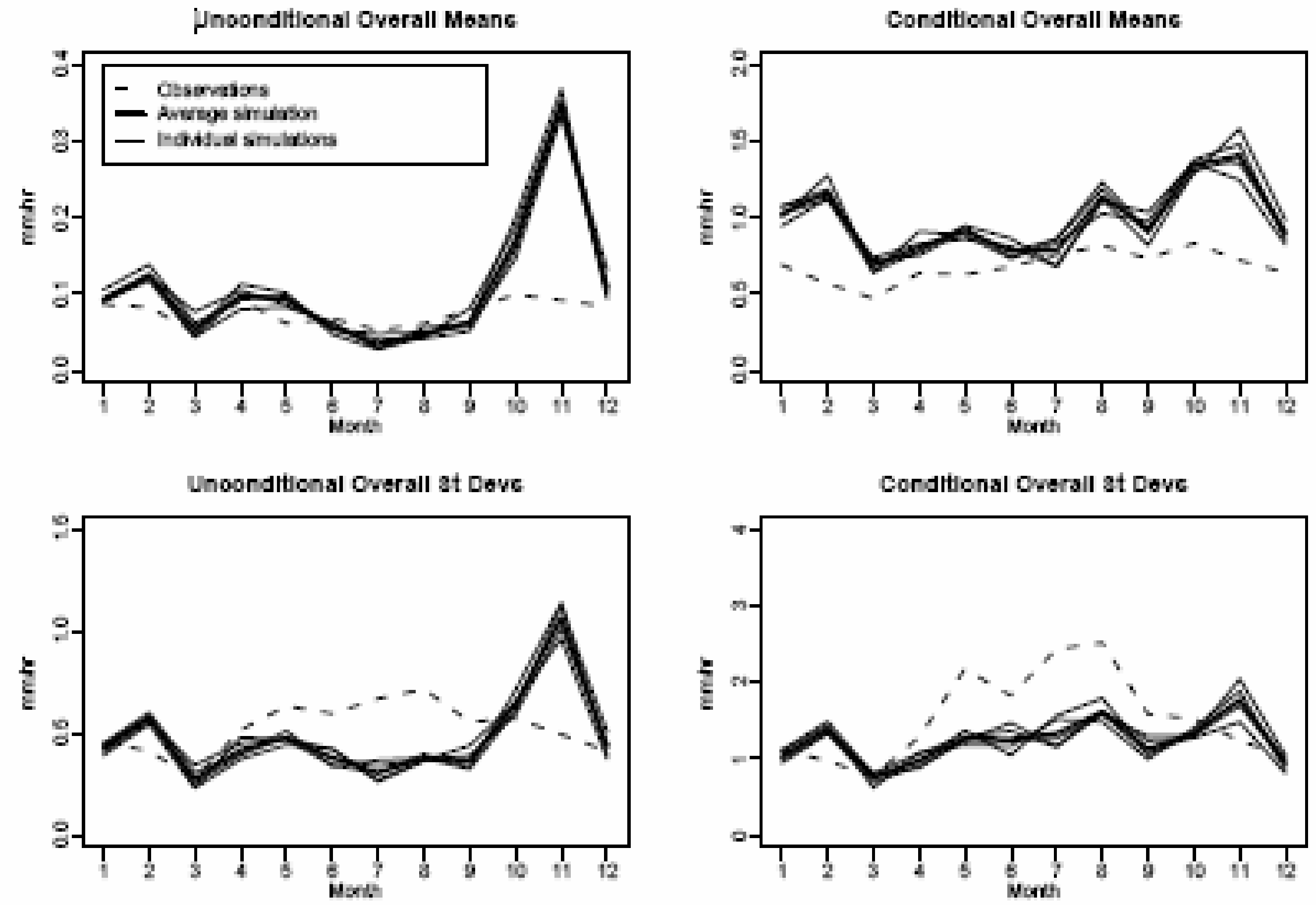
Simulate 'within-event' model within rain band



Schematic of rain band moving across catchment.

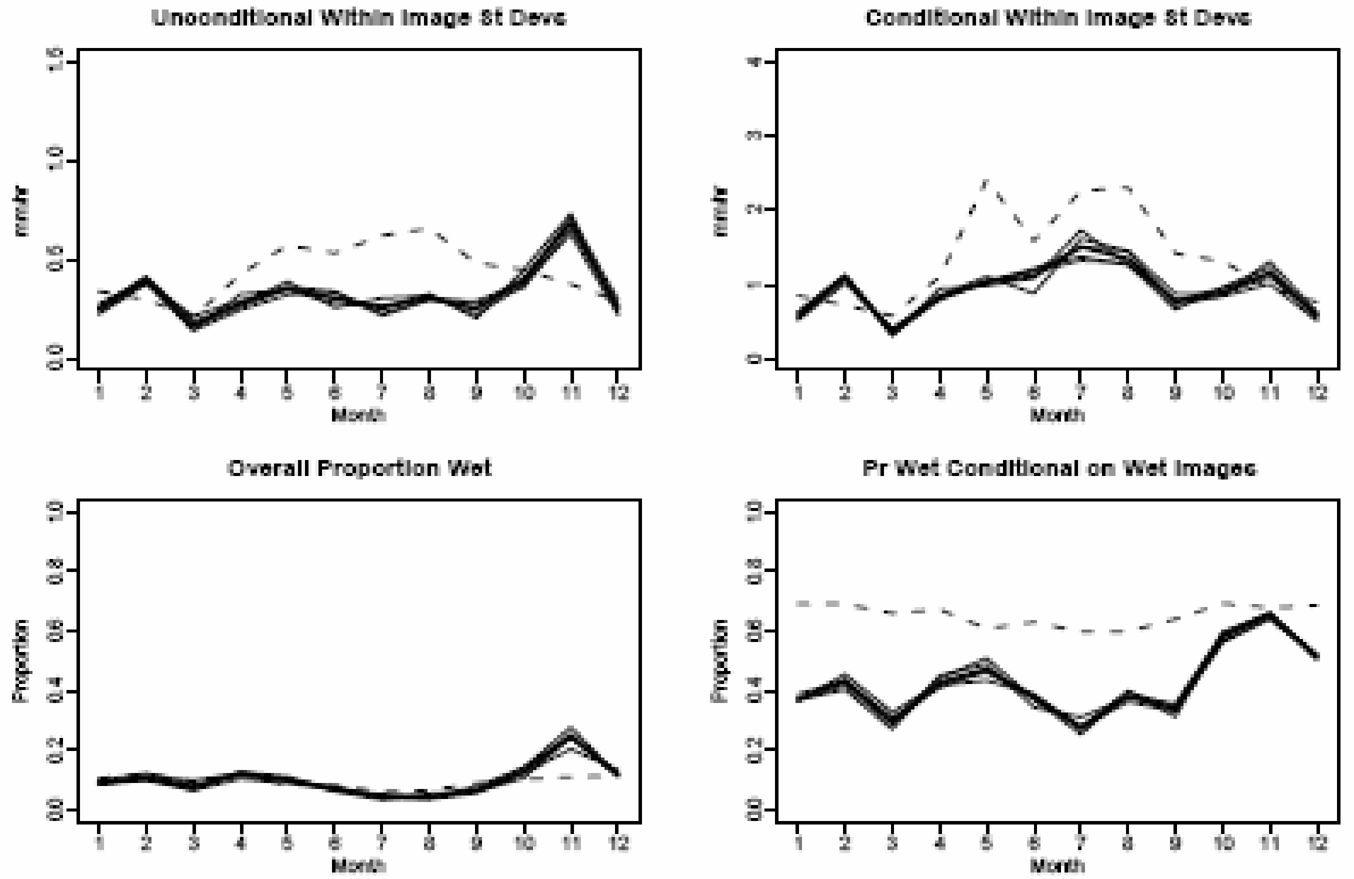
Summary statistics for simulated and empirical data

(hourly & 4 km^2): unconditional and conditional means and standard deviations



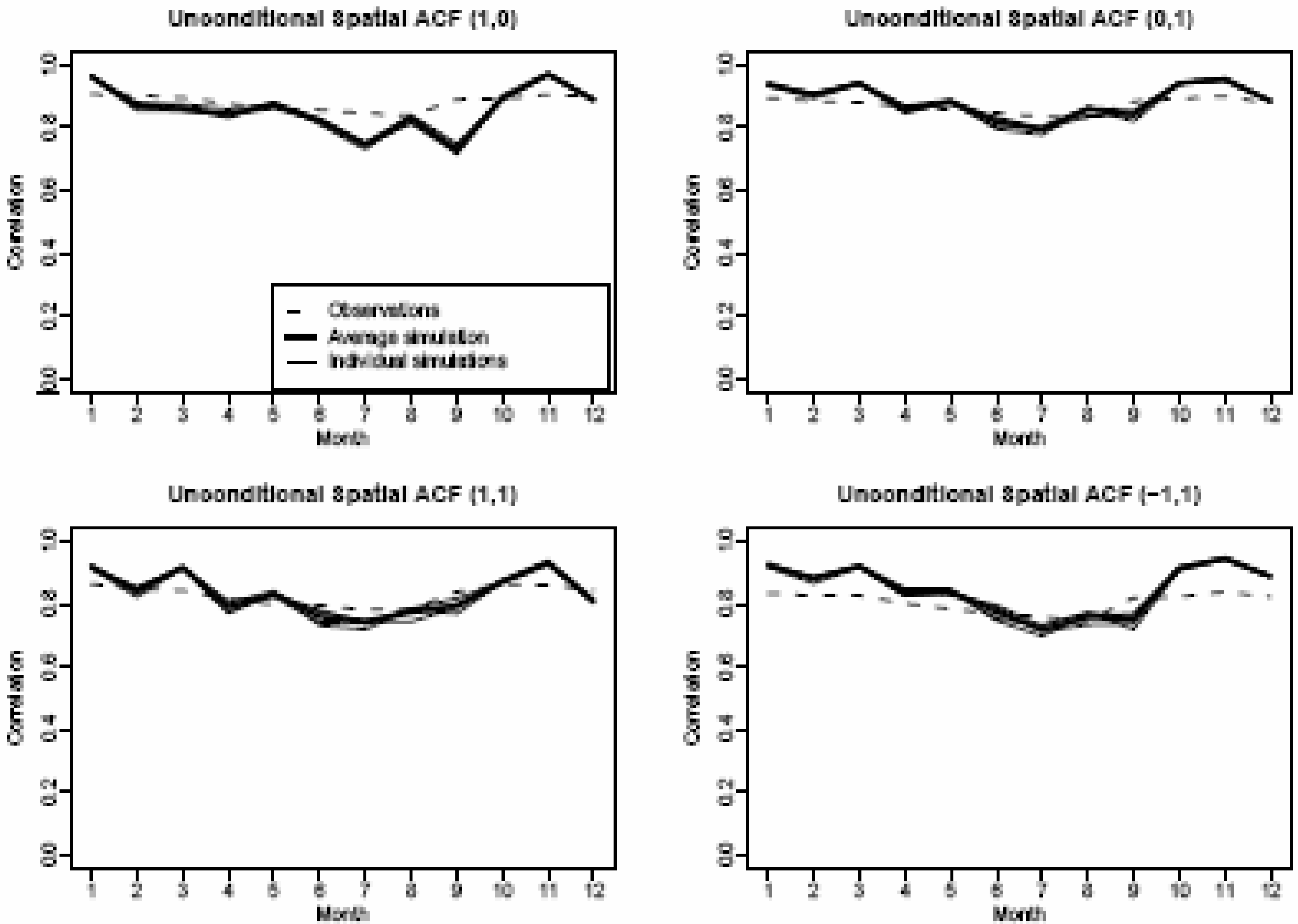
Summary statistics for simulated and empirical data

(hourly & 4 km^2): unconditional and conditional within image standard deviations and proportion wet.



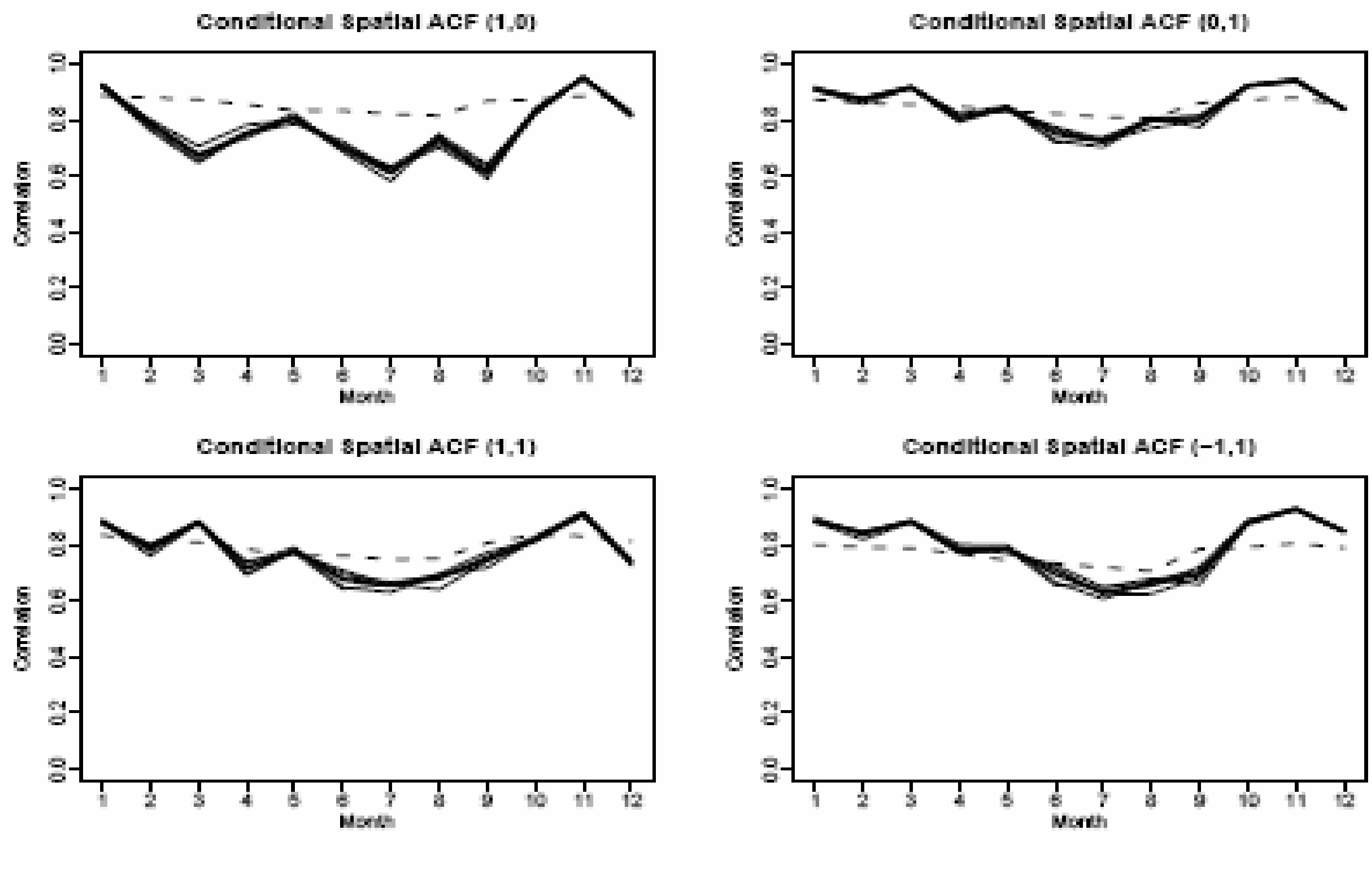
Summary statistics for simulated and empirical data

(hourly & 4 km^2): spatial autocorrelations, lags (1,0), (0,1), (1,1), (-1,1)



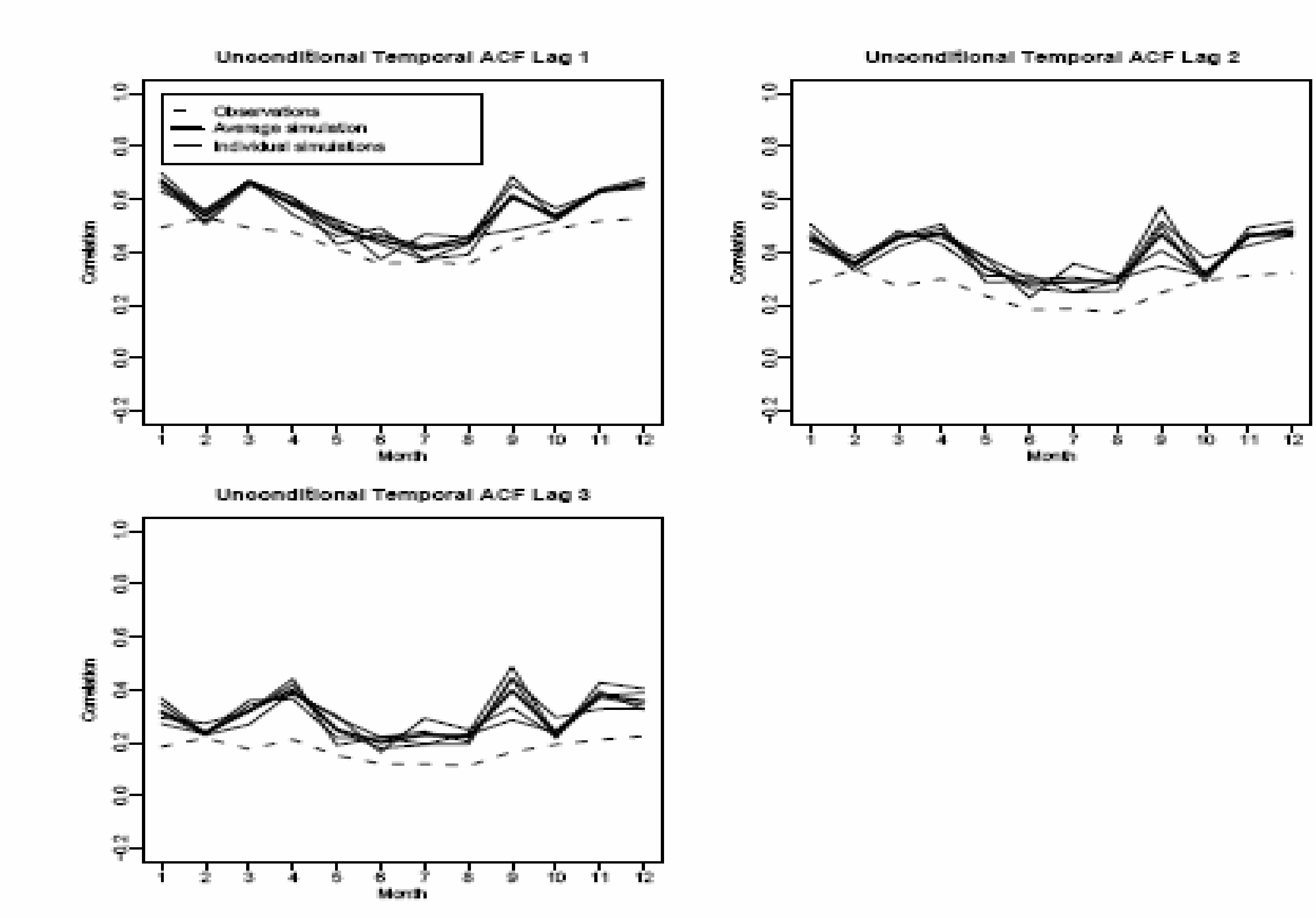
Summary statistics for simulated and empirical data

(hourly & 4 km^2): conditional spatial autocorrelations, lags (1,0), (0,1), (1,1), (-1,1)



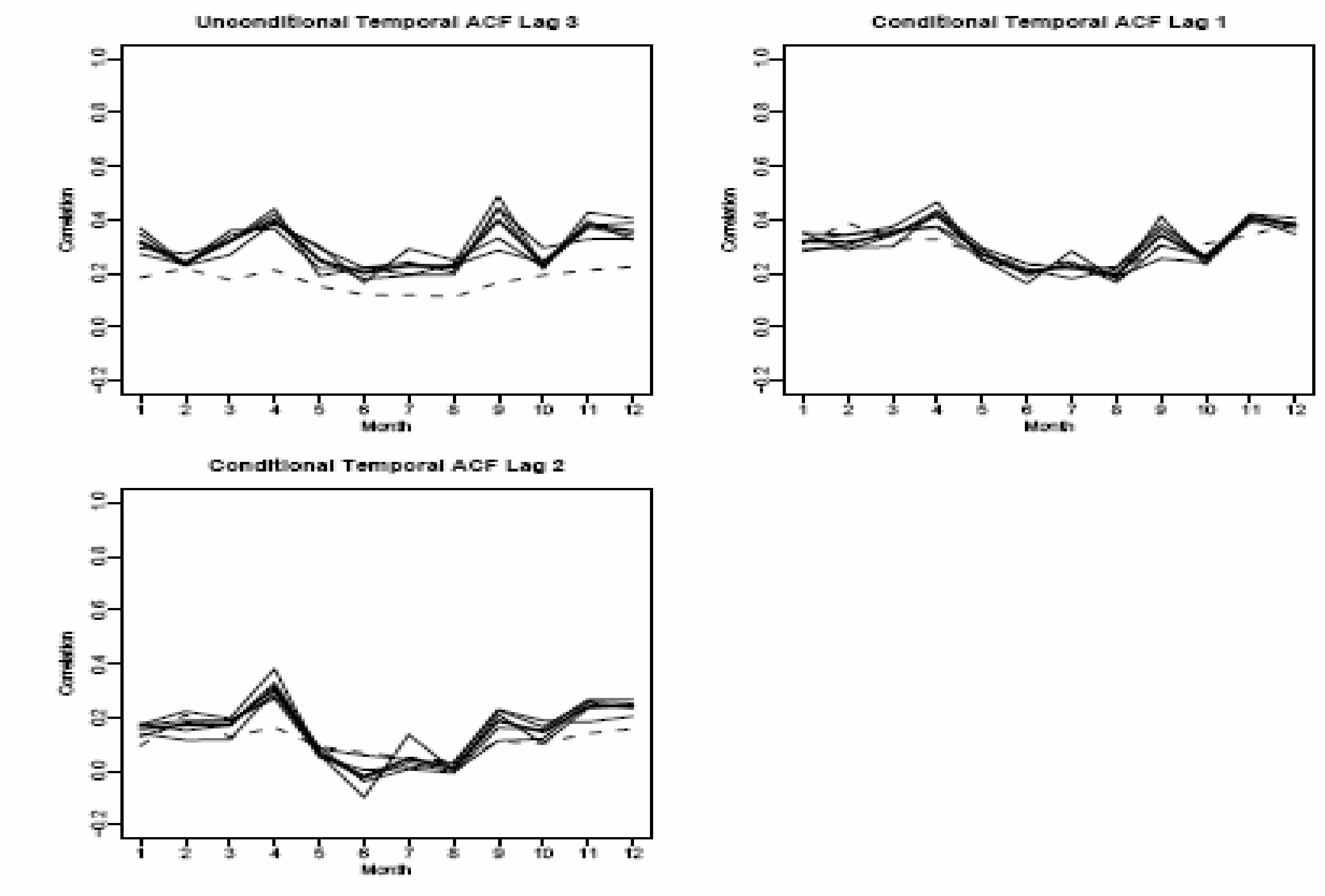
Summary statistics for simulated and empirical data

(hourly & 4 km2): temporal autocorrelations, lags 1, 2, 3.

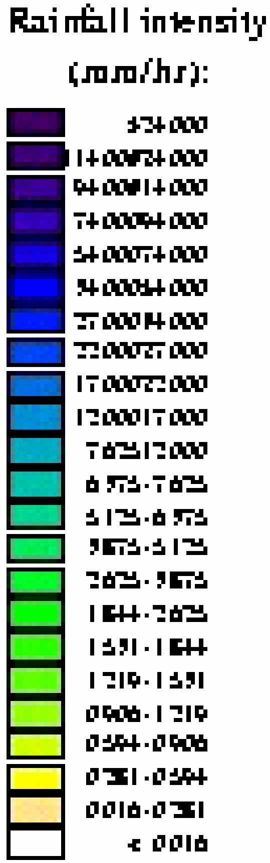


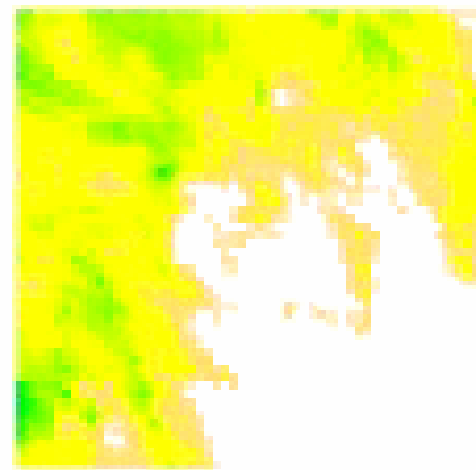
Summary statistics for simulated and empirical data

(hourly & 4 km2): conditional temporal autocorrelations, lags 1, 2, 3.



Data - winter



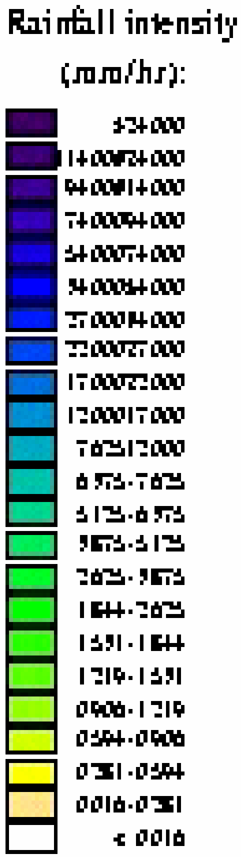


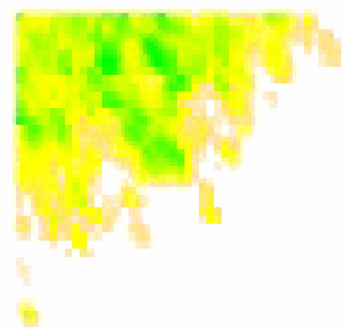
1-02-1994

2200

Wardon Hill Radar

Simulation - winter



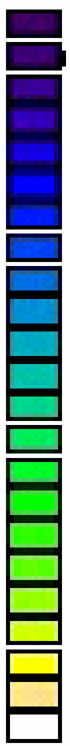


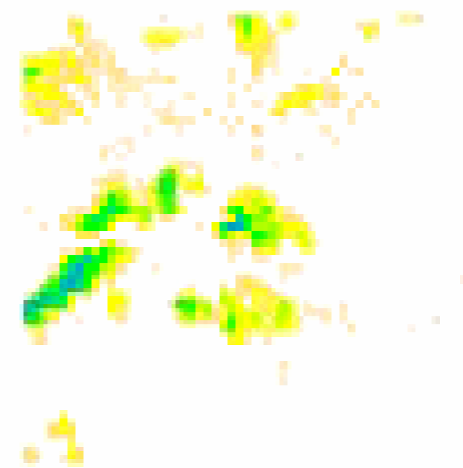
232-1995

2300

Simulated data

Data - summer



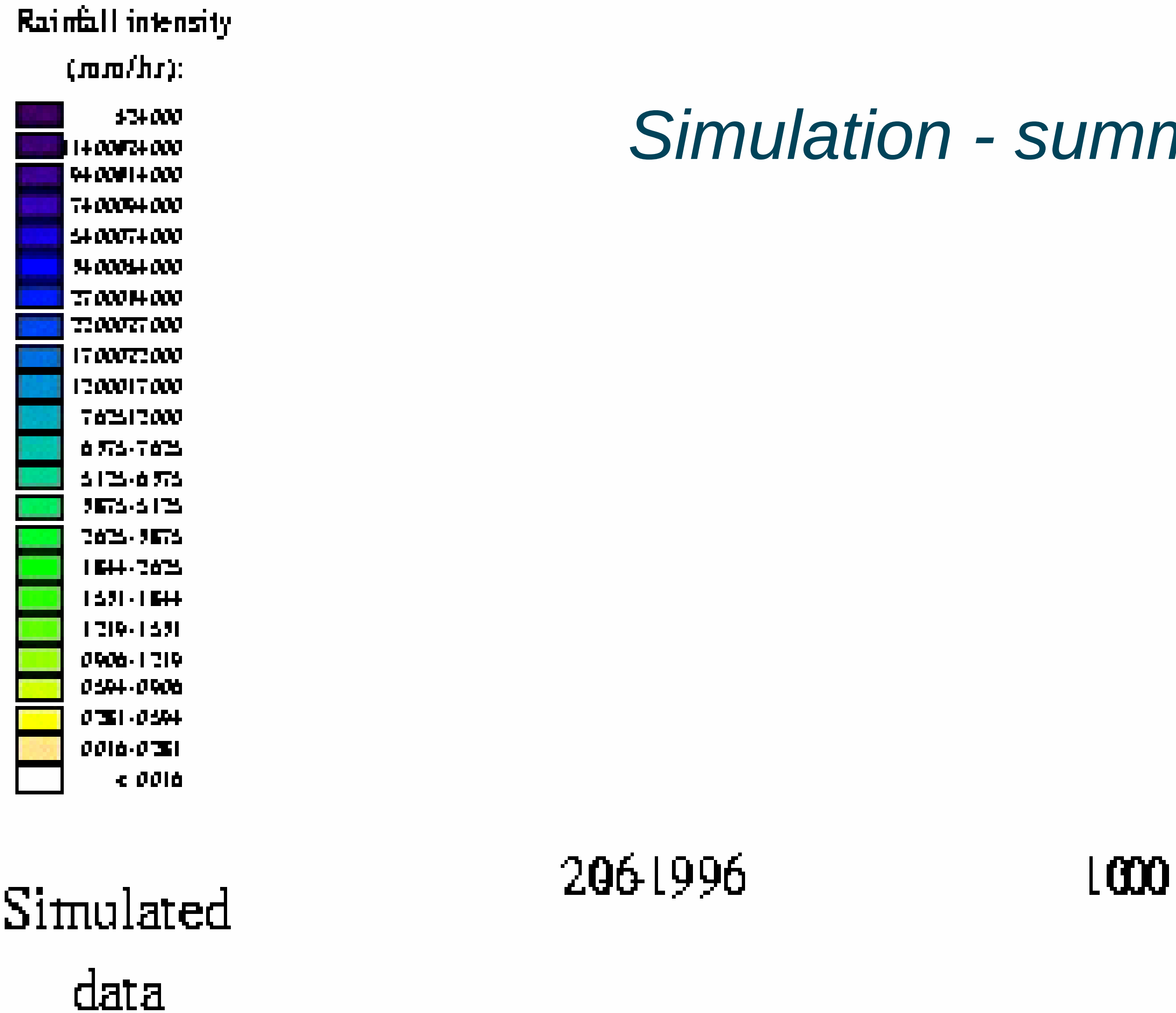


3-6-1994

0700

Wardon Hill
Radar

Simulation - summer



Statistical rainfall models (GLMs)

- Model the temporal sequences (discrete time) at a set of discrete spatial locations
- Incorporate spatial and temporal nonstationarities
  - hydrologically significant even for small catchments
- Forecast the *probability distribution* for daily rainfall at each site, conditionally upon *predictors* (fit using daily raingauge data)
- Predictors can include the site location, time of year, previous days' rainfall, teleconnections (eg El Nino, NAO) and temporal nonstationarities (eg climate change scenarios)
- Allows interpolation at sites not used in fitting

Model rainfall occurrences and amounts separately….

Occurrence model: logistic regression $\ln\left(\frac{p_{st}}{1-p_{st}}\right) = \mathbf{x}'_{st}\boldsymbol{\beta}$

where p_{st} is the probability that site s is wet on day t

Amounts model: the intensity Y_{st} *given* that the site is wet, has a *gamma* distribution with mean μ_{st} such that $\ln \mu_{st} = \boldsymbol{\xi}'_{st}\gamma$ and constant variance/mean ratio

Predictors $\{\mathbf{x}_{st}, \boldsymbol{\xi}_{st}\}$ include: location, elevation, seasonal effects, rainfall autocorrelation and persistence, teleconnections eg NAO, climate-change scenarios via GCM/RCM output, and their interactions

Effects can be assessed formally (likelihood ratio tests) or informally (residual analysis)

Dependence issues

Temporal: functions of past values *at all sites* incorporated in predictors *e.g.* persistence

Spatial: fit models as if sites were spatially independent given predictors and adjust afterwards for inter-site dependence

- simultaneous dependence of *occurrences* is modelled by a *beta-binomial* distribution for the number of sites simultaneously wet (respecting proportions of wet days at each site)
- simultaneous dependence of *amounts* (for simulation) captured through correlation structure of *Anscombe residuals*: for gamma d^{n} $(Y_{st}/\mu_{st})^{1/3}$ $\sim_{\text{approx}}$ Normal

Combination of stochastic and statistical models for nonstationary continuous simulation

Idea: use the point process model to provide the finescale spatial and temporal resolution, and the GLM to drive spatial and temporal nonstationarities.

- Use GLM to generate ensembles of multi-site daily sequences at a grid of sites

  $\Rightarrow$ properties of predictive distribution incorporating spatial dependence (nonstationary in space and in time)

- Condition continuous simulation of stochastic model to respect these properties

Simple multiplicative rescaling

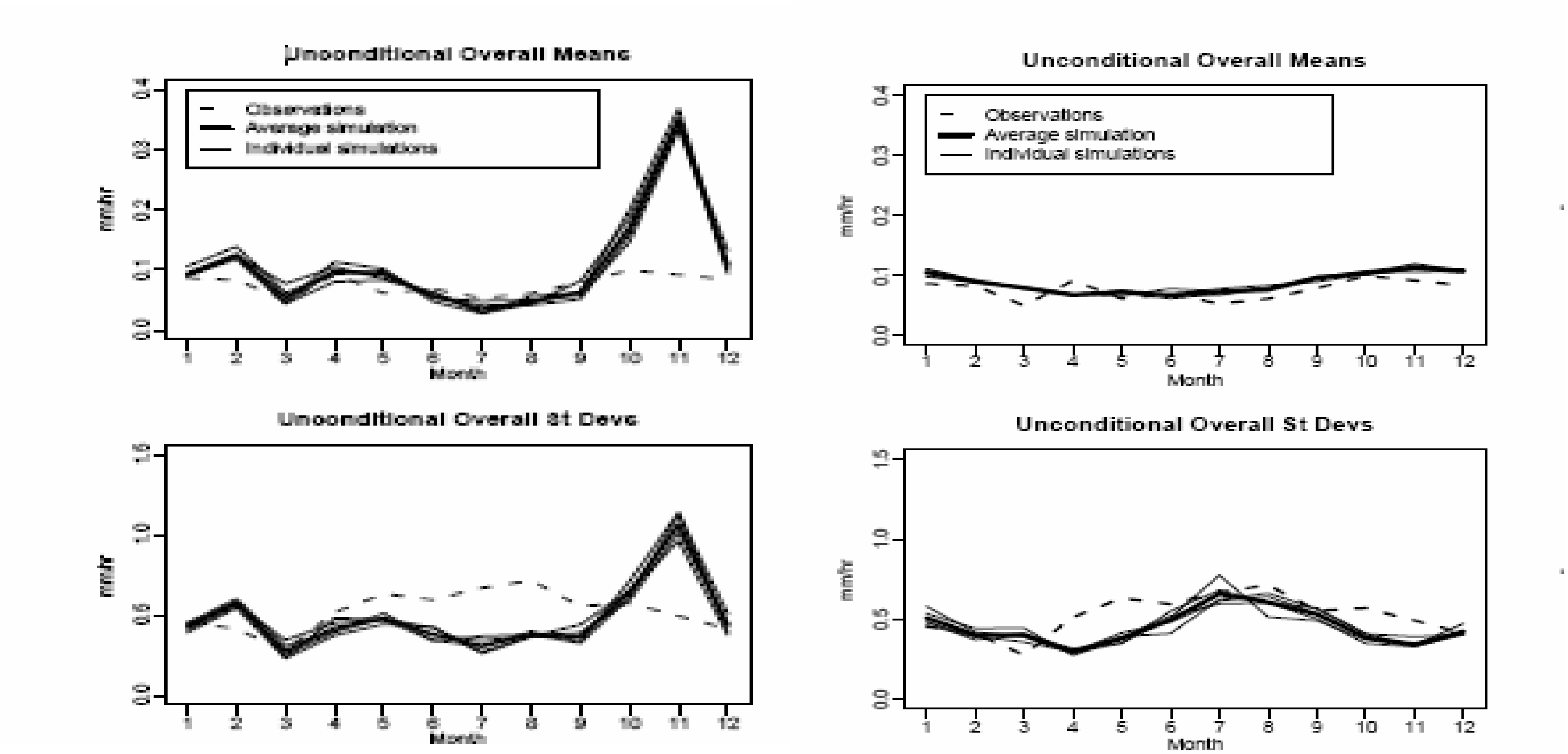
Rescale images using smoothed monthly means (per site per year) for GLM.

Incorporates spatial heterogeneity plus some temporal effects into rainfall amounts, but does not change wet/dry patterns.

Rescaling has effect of calibrating radar values to ground truth (GLMs are fitted to raingauge data)

Summary statistics for simulated and empirical data

(hourly & 4 km²): unconditional means and standard deviations, (lhs) stationary and (rhs) nonstationary simulation



Temporal nonstationarity

1. Match GLM statistics to (temporal) sequences in historical record. Statistics should include properties such as means, proportions of dry days, cross correlations, durations of dry intervals etc

   Use parameters from fits to those data for simulation.

2. Fit stochastic model to statistics derived from GLM simulation - downscaling daily statistics to subdaily values using fitted relationships between time scales as necessary.

NB Continuous simulation enables disaggregation of daily totals to subdaily values.

Conclusions and future directions

- Stochastic model for individual rain events is well-developed for representation of fine-scale spatial-temporal rainfall
- Continuous simulation over longer time periods preserves main features of empirical data
- Combination of stochastic models with GLMs allows incorporation of topographical features and temporal nonstationarities, and calibration to ground truth
- Combination of GLMs with GCM/RCM output will allow incorporation of climate change scenarios.

Further work

- Model extension to allow for light and spatially and/or temporally intermittent rainfall – values below the threshold contribute a minor proportion of the total intensity but may have more significant effect on run-off
- Investigation of the effect on run-off of the sensitivity of summary statistics used for model fitting to calibration method
- Investigation of spatial extremes and the effect on run-off