Local generalised method of moments: an application to point process-based rainfall models

Environmetrics **26 (4)**, 312-325

Jo Kaczmarska, Valerie Isham, Paul Northrop

Department of Statistical Science, University College London

26th Conference of The International Environmetrics Society
Edinburgh
22 July 2016

Main aim

The simulation of realistic artificial rainfall series at a given site

Requirements:

- reproduce statistical properties of real data (mean, variance, skewness, autocorrelations; wet-dry properties; extreme values) at a range of temporal scales
- sub-daily, perhaps sub-hourly, resolution
- can reflect potential impacts of climate change

Such series are required for hydrological studies, e.g. urban drainage design needs sub-hourly data

Statistical downscaling

GCMs: main tools for predicting future climate impacts from the increase of greenhouse gases in the atmosphere.

Problem: cannot reliably model local precipitation:

Solution: “statistical downscaling”. Model relationships between large-scale climate variables of the GCM (or RCM) and local climate

Problem: Structure depends on specific division of time scale, usually daily (and not suitable at subdaily scales)

Solution: Models based on point processes operate in continuous time, allowing aggregation to any timescales

Poisson-cluster process rainfall models

Represent the rainfall process in a simplified way with interpretable parameters

Problem: Models are stationary.

Current practice:

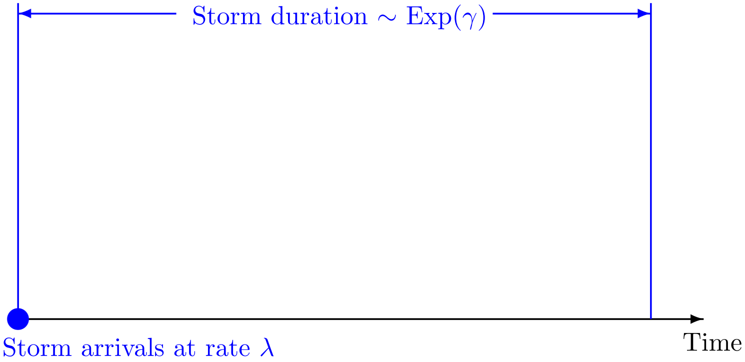
- Seasonality: fit model separately within each month
- Month just a proxy for climate variables. Seasonal patterns changing?

This work:

- allow parameters to vary smoothly with covariates
- ... month, GCM variables: surface temperature, sea-level pressure (slp), zonal (west-east) wind, ...
- expect complicated relationships: use local modelling
- enables simulation of future rainfall for use in climate impact studies

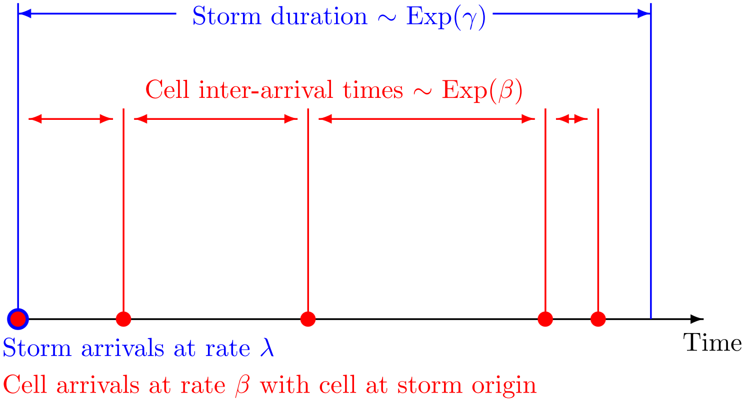
Bartlett-Lewis Rectangular Pulse model

Bartlett-Lewis Poisson cluster process of cell origins:



Bartlett-Lewis Rectangular Pulse model

Bartlett-Lewis Poisson cluster process of cell origins:

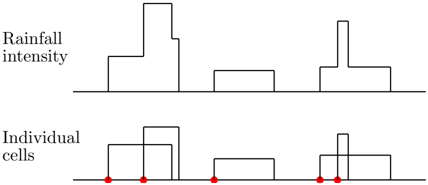


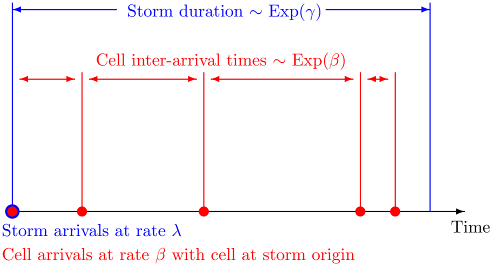
Bartlett-Lewis Rectangular Pulse model

The cells have random durations and intensities.

η rate of cell death

μ_X mean cell intensity





λ rate of storm arrival

γ rate of storm death

β rate of cell arrival

$$\theta = (\lambda, \gamma, \beta, \eta, \mu_X)$$

Data

- 5-min rainfall totals from Bochum, Germany, 1931–1999
- Y_t: data in month t, $t = 1, \ldots, 624$
- From Jan 1948 : monthly NCEP reanalysis data (52.5N, 7.5E) plus NAO index
- 624 monthly fitting properties and covariate values
  - $T(Y_t) = (T_1(Y_t), \ldots, T_k(Y_t))$ sample fitting properties
    [1h mean; 5 min, (1, 6, 24)h coeff. of variation, lag 1 autocorr., skewness]
  - $X_t = (X_{t_1}, X_{t_2}, \ldots, m_t)$: monthly temperature, slp . . . , month

Generalised Method of Moments (GMM)

- $\widehat{\theta}$ minimises a (quadratic form) measure of discrepancy between sample $T(Y_t)$ and model $\boldsymbol{\tau}(\theta)$ values
- Weight sample properties to account for their (co)variances
- Optimal weight matrix W is $\{\mathrm{cov}[T(Y_t)]\}^{-1}$, but in practice $\mathrm{diag}\{\mathrm{var}[T_i(Y_t)]^{-1}\}$ is close to optimal and easier to estimate reliably (Jesus and Chandler, 2011)

Local mean GMM

Estimator for month m

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

$$\left. \widehat{W}_m \left\{ \frac{1}{\sum_{t=1}^{n} I(m_t = m)} \sum_{t=1}^{n} I(m_t = m) \left[T(Y_t) - \tau(\theta_m) \right] \right\} \right]$$

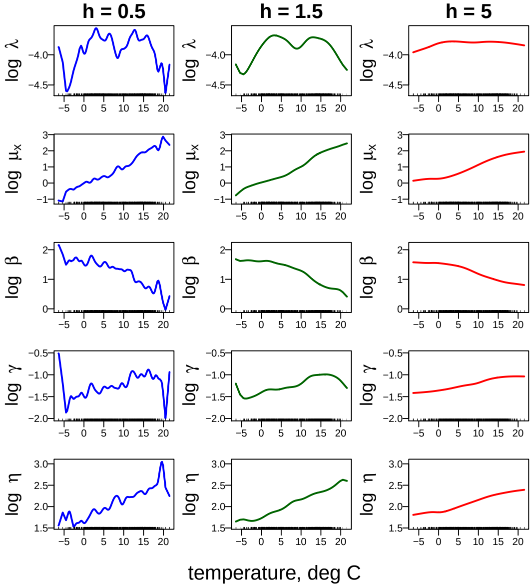
Estimator at covariate $X = x_0$

$$\widehat{\theta}(x_0) = \text{argmin}_{\theta_{x_0}} \left[\left\{ \frac{1}{n} \sum_{t=1}^{n} K_h(X_t - x_0) \left[T(Y_t) - \tau(\theta_{x0}) \right] \right\}^{\mathrm{T}} \right.$$

$$\left. \widehat{W}_{x_0} \left\{ \frac{1}{n} \sum_{t=1}^{n} K_h(X_t - x_0) \left[T(Y_t) - \tau(\theta_{x0}) \right] \right\} \right]$$

- $K_h(X_t - x_0) = \dfrac{1}{h} K\left(\dfrac{X_t - x_0}{h} \right)$, for a kernel function $K()$
- bandwidth h

Single covariate: effect of bandwidth



Bias-variance trade-off: a larger *h* gives a smoother, flatter curve, with lower variance, but higher bias

We use a *global* bandwidth: i.e. the same *h* across the whole curve

Choice of h; multiple covariates

Cross-validation (CV), using repeated random sub-sampling

For each covariate

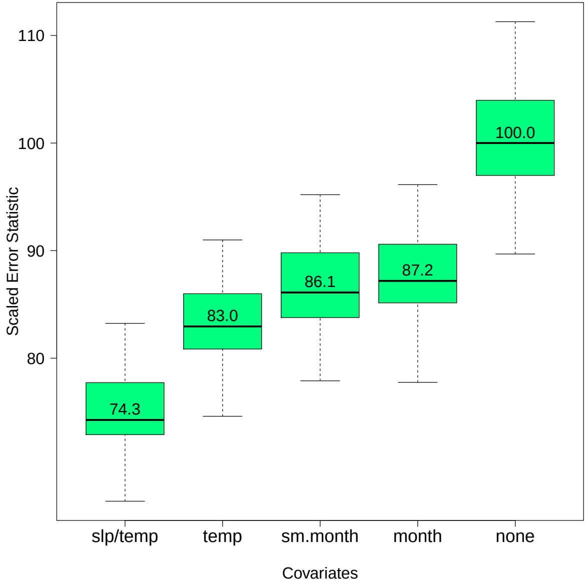
- For each of a set of candidate values of h ...
- Split 624 months into 399 (training) and 225 (test)
- Fit using training set then calculate mean weighted squared error (MWSE) on test set
- Repeat 25 times; find the best h (lowest MWSE) each time

Set 'optimal' h as an average of the best values

Multiple covariates

- use diagonal bandwidth matrix H: product kernels
- use CV to find choose an optimal multiple of a diagonal matrix of the optimal univariate bandwidths
- curse of dimensionality restricts the number of covariates

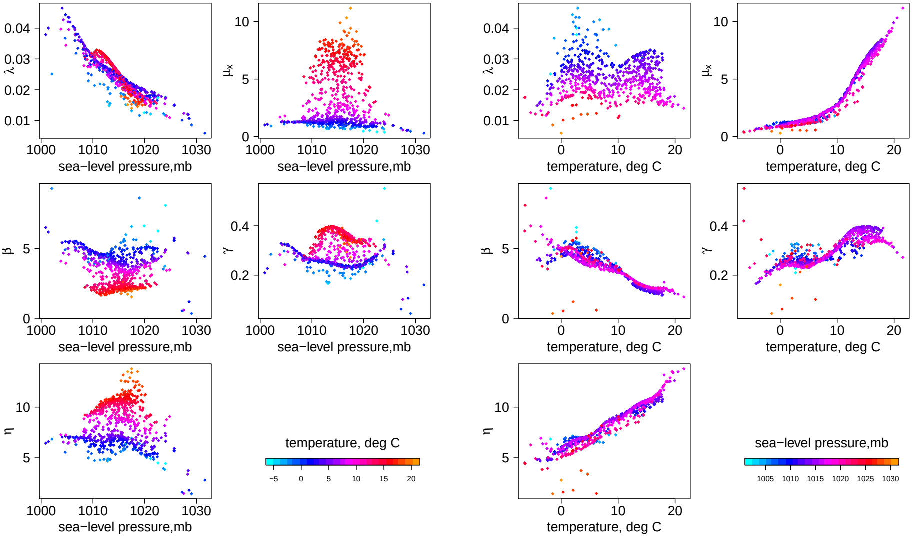
Model Comparison: optimal covariates



We compare prediction errors with different covariates

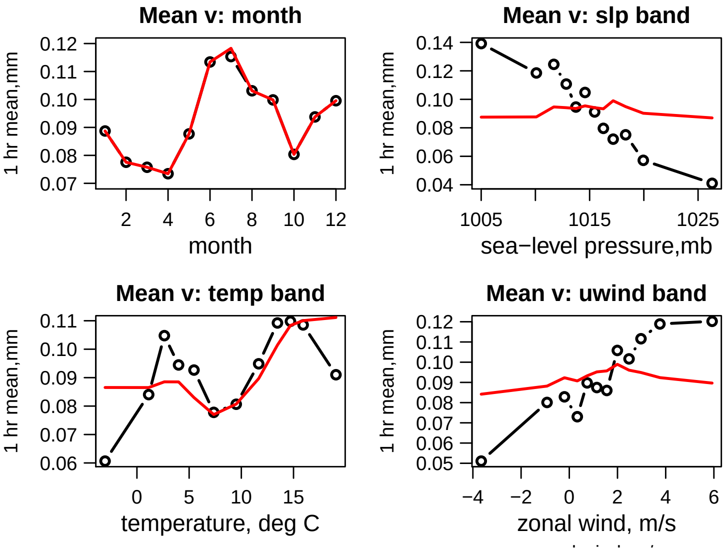
sm.month : local kernel smoothing based on month

Optimal pair of covariates: slp and temp



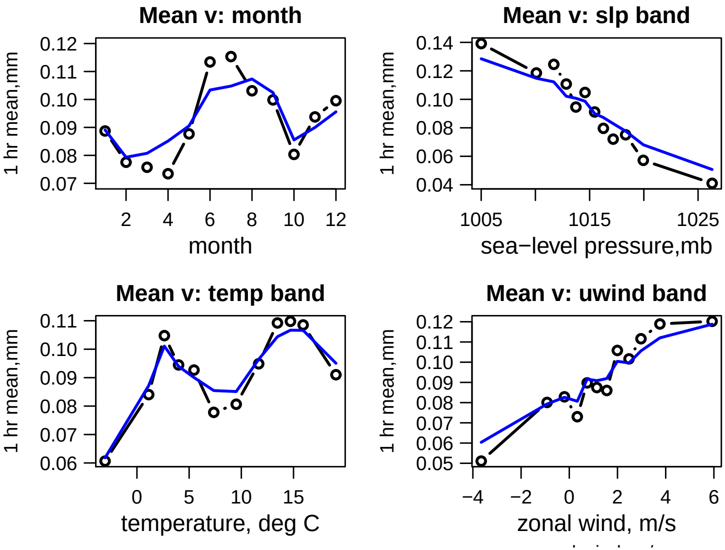
Bandwidths: sea-level pressure: 2.0; temperature: 1.75

Comparison of fit v current approach



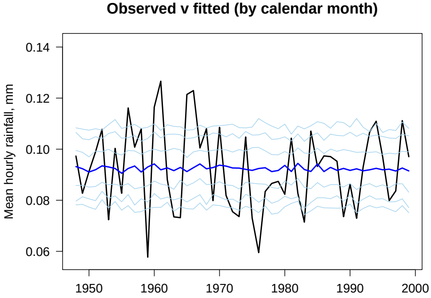
Covariate: Month

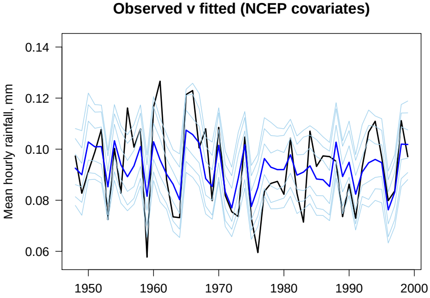
Comparison of fit v current approach



Covariates: Temperature, sea-level pressure and zonal wind

Interannual variability at Bochum





Simulated distributions of mean annual rainfall

- blue bands: (5, 10, 25, 50, 75, 90, 95) percentiles
- black line: sample values

Using covariate information has improved the representation of interannual variability

Concluding remarks

Local smoothing used to relate Poisson-cluster rainfall model parameters to covariates

Advantages

- covariates can replace, and improve, on month : once temperature is included, month adds no further benefit
- improved representation of interannual variability
- simulations can reflect future climate change scenarios

Concluding remarks

Local smoothing used to relate Poisson-cluster rainfall model parameters to covariates

Advantages

- covariates can replace, and improve, on month : once temperature is included, month adds no further benefit
- improved representation of interannual variability
- simulations can reflect future climate change scenarios

Thank you for your attention

References

Research presented here

- Kaczmarska, J. M., Isham, V. S. & Northrop, P. J. (2015), 'Local generalised method of moments: an application to point process-based rainfall models', *Environmetrics* **26 (4)**, 312-325
- Kaczmarska, J. M. (2013), 'Single-site point process-based rainfall models in a nonstationary climate', PhD thesis, University College London

Point process based models/GMM

- 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
- Jesus, J. & Chandler, R. E. (2011), 'Estimating functions and the generalized method of moments', *Interface Focus*, **1(6)**, 871-885

Local fitting

- Fan & Gijbels, L. (1996), *Local Polynomial Modelling and its Applications*, Chapman and Hall.
- Lewbel, A. (2007), 'A local generalized method of moments estimator', *Economics Letters* **94**.
- Carroll, R. J., Ruppert, D. & Welsh, A. H. (1998), 'Local estimating equations', *Journal of the American Statistical Association* **93** (441), 214-227.