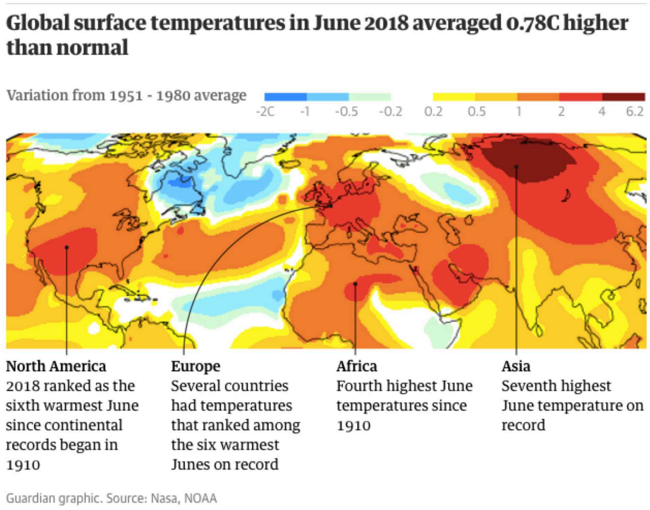
Spatial Extremes

Anthony Davison

https://www.epfl.ch/labs/stat/

Summer 2018

- ☐ Motivation: heatwaves, wildfires, drought, heavy rainfall . . .



Motivations for spatial modelling

- ☐ **Assessment** of significance of ‘hot spots’.
- ☐ **Attribution** of events to possible causes.
- ☐ **Estimation** of changes in extremes of time series, accounting for dependence between related series.
- ☐ **Risk assessment** at a single important site, borrowing strength from sites nearby.
- ☐ **Risk estimation** for large spatial events.

Why specialised models?

- ☐ Basic problem is generally **extrapolation** to rare(r) events.
- ☐ Spatial statistics is mostly based on multivariate normal distributions, inappropriate for modelling tails of distributions, rare events, etc.
- ☐ Extrapolation from a fit to the entire distribution can be misleading:
  - different mechanisms may apply in the extremes
  - different fits to the bulk may give very different tail estimates—in particular, the light tails of the Gaussian density can grossly underestimate probabilities of rare events
  - Gaussian models for multivariate data predict independence of very rare events (‘the formula that killed Wall Street’)
- ☐ Use of standard copulas can deal with transformations to marginal distributions, but not with joint dependence.

Setup

- ☐ Focus on extremes of $Y(x)$ for x in some space or space/time domain $\mathcal{X}$.
- ☐ Aim to estimate probabilities of the form

$$\mathrm{P}\left\{Y(x) \in \mathcal{R}\right\},$$

  where $\mathcal{R}$ represents rare event of interest.
- ☐ Data are available at only a finite subset $\mathcal{X}'$:
  - a few long series (long-term observations, space-poor/time rich)
  - many short series (satellite data, space rich/time poor)
  - many longer series (15-minute radar data, space rich/time rich)

5

Extreme-Value Theory 6

Poisson process

- ☐ Random point pattern $\mathcal{P}$ in a state space $\mathcal{E}$ defined by properties of counts

$$N(\mathcal{A}) = |\{x : x \in \mathcal{P} \cap \mathcal{A}\}|, \quad \mathcal{A} \subset \mathcal{E}:$$

  - $N(\mathcal{A}_1), \ldots, N(\mathcal{A}_k)$ independent for disjoint $\mathcal{A}_1, \ldots, \mathcal{A}_k$,
  - $N(\mathcal{A}) \sim \mathrm{Poiss}\{\mu(\mathcal{A})\}$,

  where the measure μ is non-atomic (diffuse), and often has an **intensity** $\dot{\mu}$.
- ☐ **Mapping theorem**: if $g : \mathcal{E} \to \mathcal{E}^*$ does not create atoms, then $\mathcal{P}^* = g(\mathcal{P})$ is also a Poisson process.
- ☐ Restriction of $\mathcal{P}$ to $\mathcal{E}' \subset \mathcal{E}$ is also Poisson.

7

Classical extremal models

- Use random sample $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$ and for $b_n \in \mathbb{R}$ and $a_n > 0$ define point process

$$\mathcal{P}_n = \{(X_j - b_n)/a_n : j = 1, \ldots, n\}, \quad \mathcal{E} = \mathbb{R}.$$

- Then the rescaled maximum $\{\max(X_1, \ldots, X_n) - b_n\}/a_n$ has a non-degenerate limiting distribution iff $\mathcal{P}_n$ converges to a Poisson process with mean measure

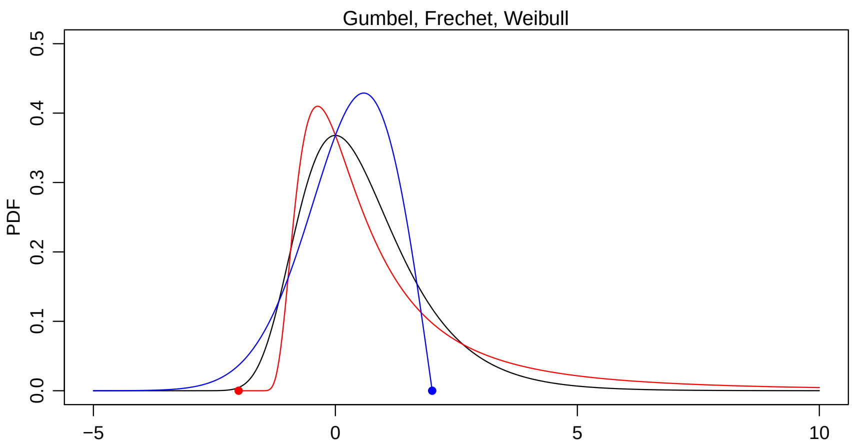
$$\Lambda\{(y,\infty)\} = \left(1 + \xi \frac{y - \eta}{\tau}\right)_+^{-1/\xi}, \quad y \in \mathbb{R},$$

where $u_+ = \max(u, 0)$, and η and τ are location and scale parameters.

- The shape parameter ξ determines the rate of tail decay, with
  - $\xi > 0$ giving the heavy-tailed (Fréchet) case,
  - $\xi = 0$ giving the light-tailed (Gumbel) case—corresponds to Gaussian data,
  - $\xi < 0$ giving the short-tailed (reverse Weibull) case.
- Limiting distributions:
  - for maxima, **generalized extreme-value (GEV)**, $G(y) = \exp\{-\Lambda(y)\}$;
  - for excesses over threshold u, **generalized Pareto (GPD)**, $H(y) = 1 - \Lambda(y + u)/\Lambda(u)$.

8

GEV and shape parameter ξ



- PDFs of **Gumbel** ($\xi = 0$), **Fréchet** ($\xi = 0$) and (reverse) **Weibull** ($\xi < 0$).
- The Fréchet is bounded below, and the reverse Weibull is bounded above.
- The standard Weibull is a distribution for minima.

9

Extrapolation

- ☐ Extreme value theory gives **limiting** models:
  - – GEV applies for maxima of an infinite sample,
  - – GPD applies for exceedances of an 'infinite' threshold.
- ☐ Extrapolation to high levels is based on the fact that the GEV is **max-stable**:

$$G(y)^t = G(b_t + a_t y), \quad t > 0,$$

or equivalently

$$\max(X_1, \ldots, X_t) \stackrel{\mathrm{D}}{=} b_t + a_t X_1$$

for known functions $a_t > 0$ and b_t.

- ☐ For the standard Fréchet, GEV(1,1,1), distribution, $e^{-1/z}$, $(z > 0)$, we have $b_t \equiv 0$, $a_t = t$.
- ☐ Likewise the GPD is **threshold-stable**.
- ☐ Could fit other models, but with weaker mathematical justification.
- ☐ In practice we have finite samples, so the extremal models are approximate and extrapolation may be vulnerable.
- ☐ Now generalize the above **extremal paradigm** to complex settings ...

10

Max-stable processes

- ☐ Can transform maxima to have limiting standard Fréchet distribution, so

$$\max\{Z_1, \ldots, Z_n\} \stackrel{\mathrm{D}}{=} nZ, \quad n = 1, 2, \ldots$$

- ☐ Want processes $Z(x)$ with standard Fréchet margins such that if $Z_1(x), \ldots, Z_n(x) \stackrel{\mathrm{iid}}{\sim} Z(x)$, we can base extrapolation on **max-stability**

$$\max\{Z_1(x), \ldots, Z_n(x)\} \stackrel{\mathrm{D}}{=} nZ(x), \quad x \in \mathcal{X}.$$

- ☐ Let $W(x) \geq 0$ be a random process with $\mathrm{E}\{W(x)\} = 1$ $(x \in \mathcal{X})$, and consider the Poisson process on $\mathbb{R}_+ \times \mathcal{C}_+(\mathcal{X})$:

$$\{(R_j, W_j(x)) : j = 1, 2, \ldots, \}, \quad R_j = (E_1 + \cdots + E_j)^{-1}, \quad E_i \stackrel{\mathrm{iid}}{\sim} \exp(1) \perp\!\!\!\perp W_j \stackrel{\mathrm{iid}}{\sim} W.$$

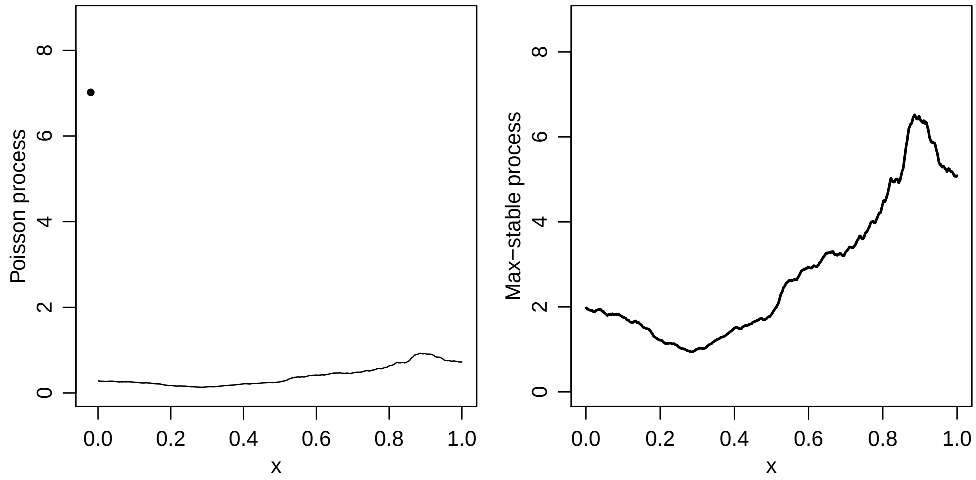
- ☐ Setting $Q_j(x) = R_j W_j(x)$ gives a Poisson process on $\mathcal{C}_+(\mathcal{X})$, and any **max-stable process** has a **spectral representation** (de Haan, 1984)

$$Z(x) = \sup_{j=1}^{\infty} Q_j(x), \quad x \in \mathcal{X}, \tag{1}$$

with $Q_j(x)$ interpreted as the jth event, with overall size R_j and profile $W_j(x)$.

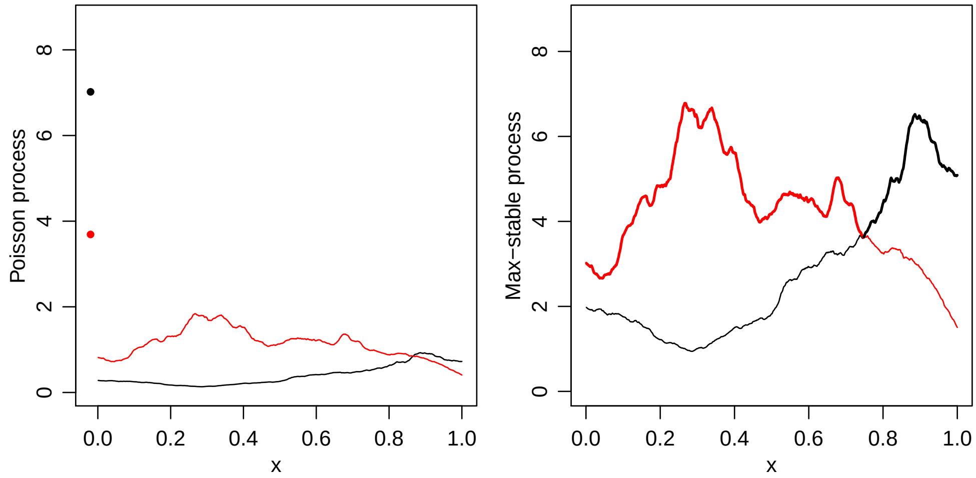
11

Spectral representation



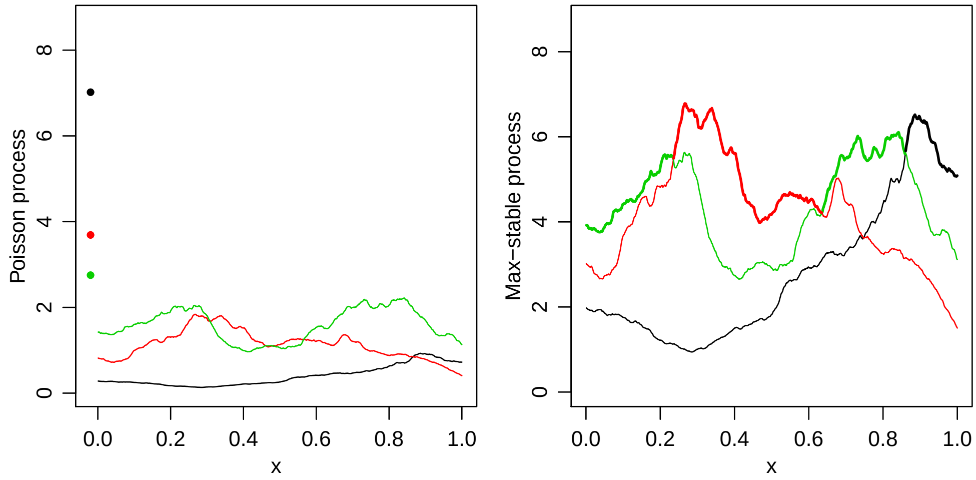
12

Spectral representation



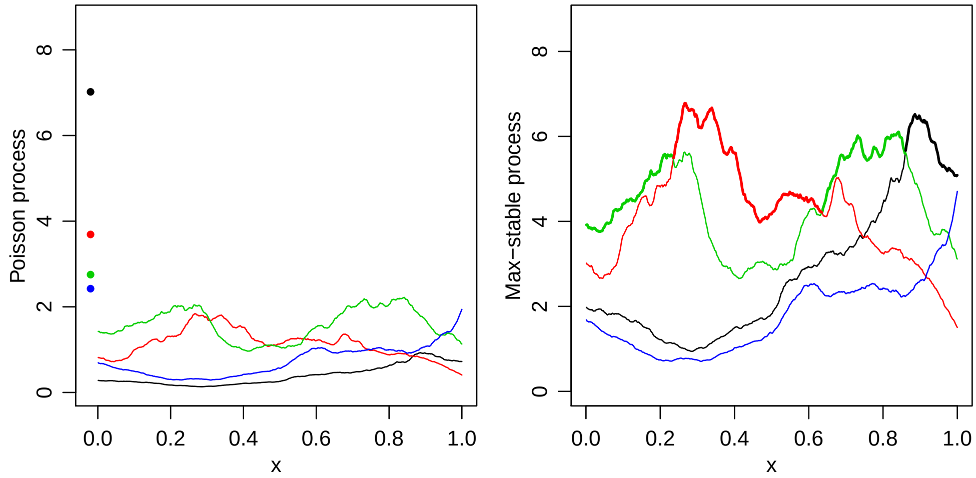
13

Spectral representation



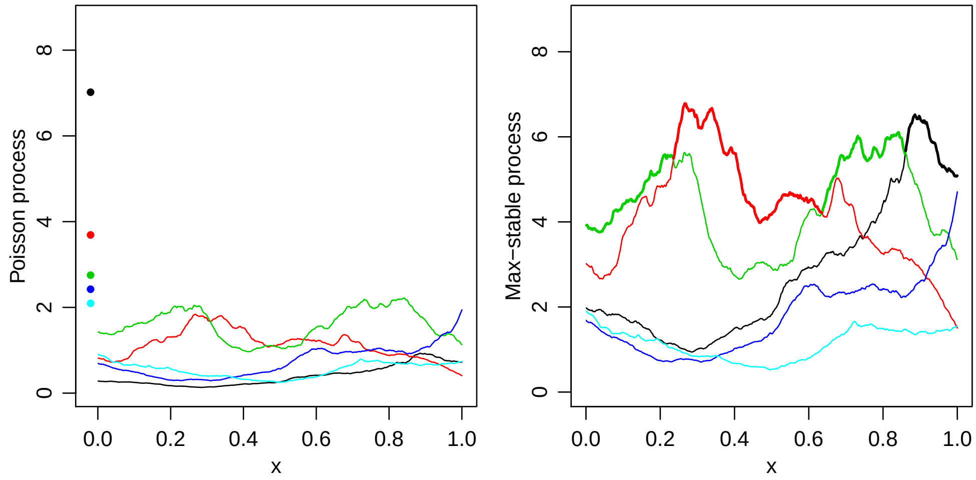
14

Spectral representation



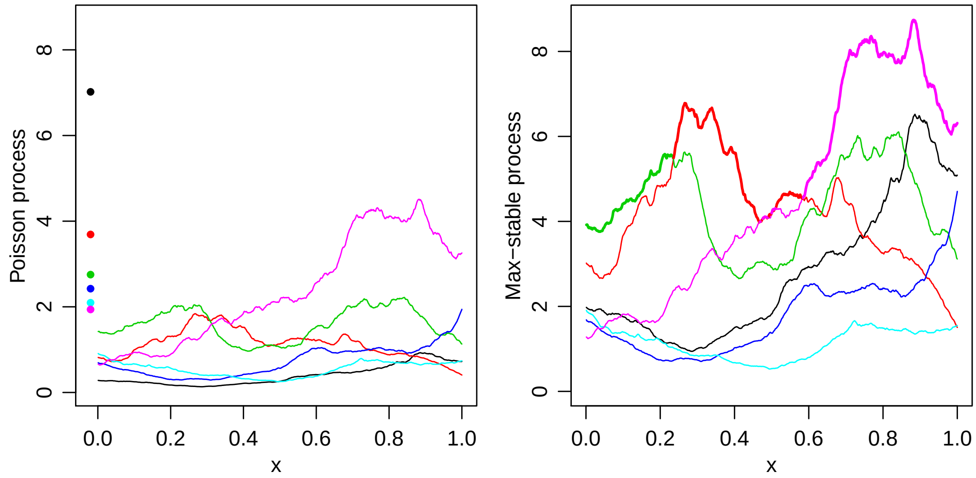
15

Spectral representation



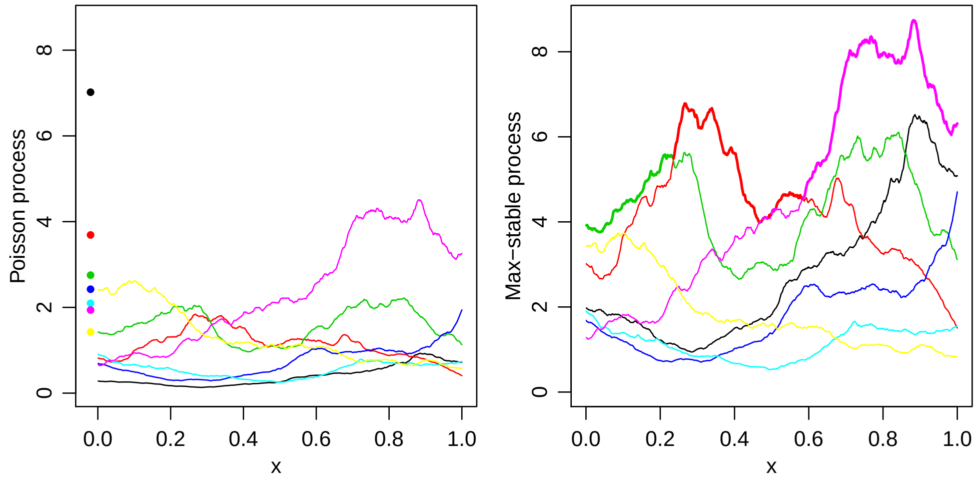
16

Spectral representation



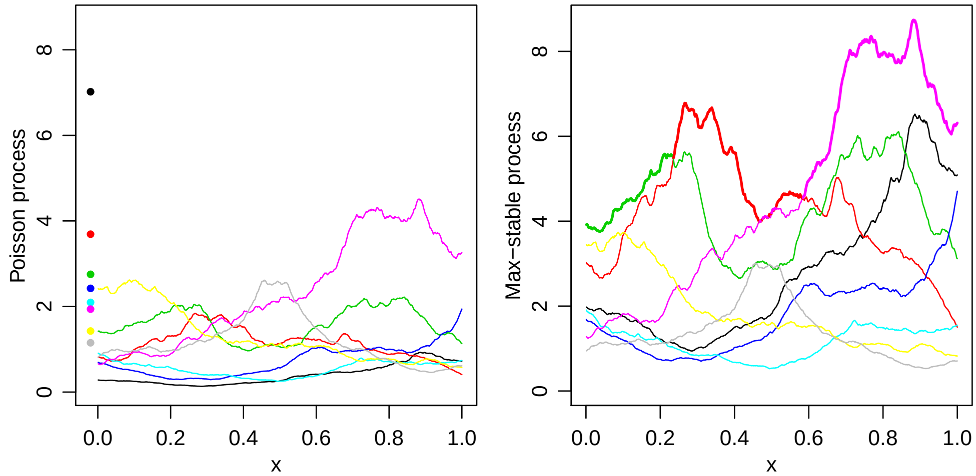
17

Spectral representation



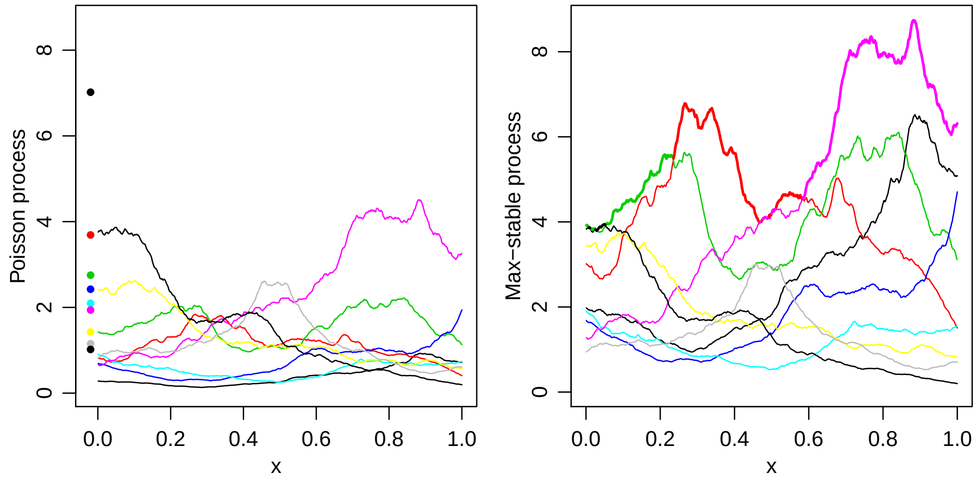
18

Spectral representation



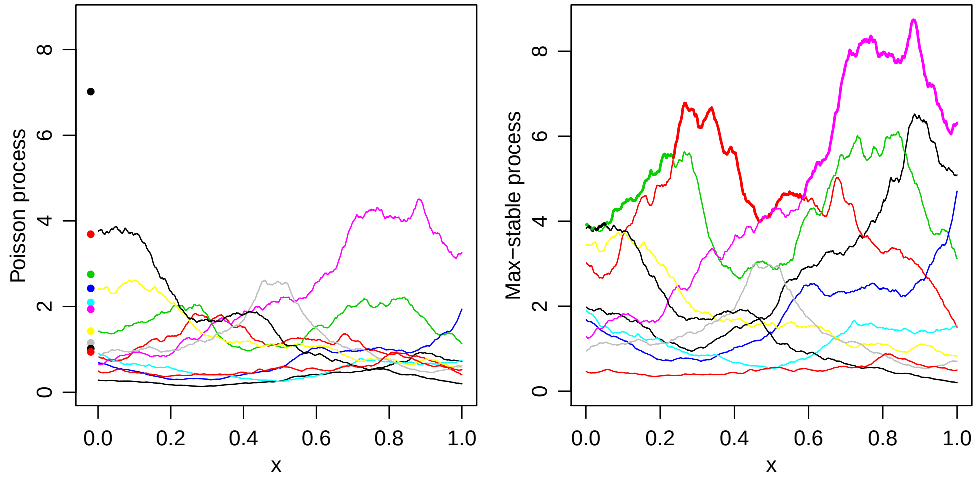
19

Spectral representation



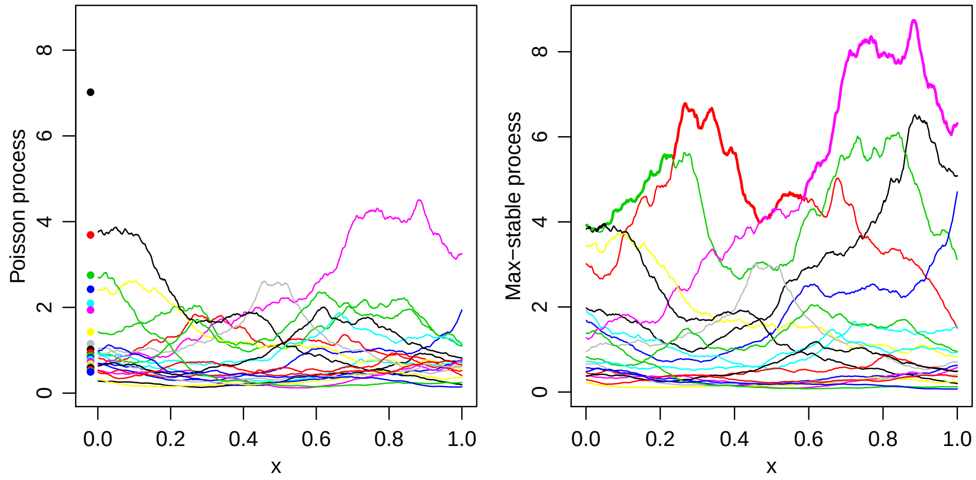
20

Spectral representation



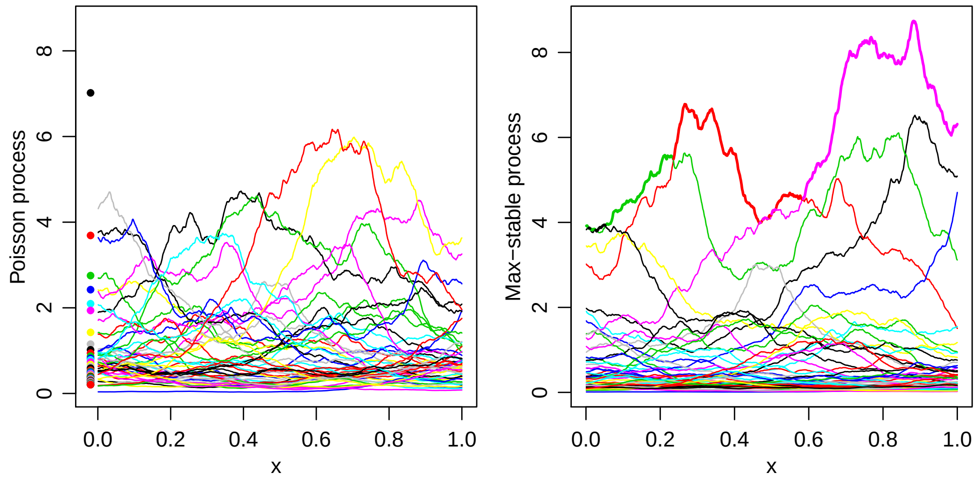
21

Spectral representation



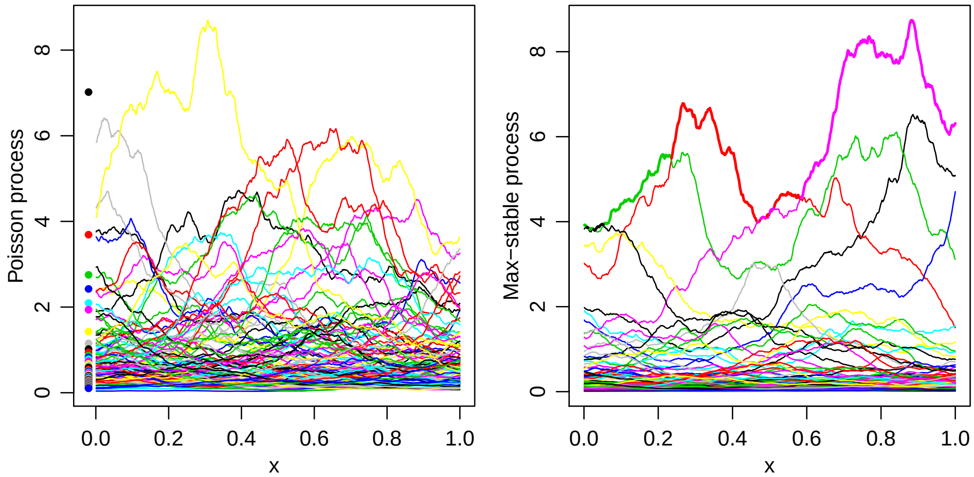
22

Spectral representation



23

Spectral representation



24

Exponent function

- For a function $z(x)$, one can show that

$$\mathrm{P}\left\{Z(x) \le z(x), x \in \mathcal{D}\right\} = \exp\left[-V\{z(x) : x \in \mathcal{D}\}\right], \quad \mathcal{D} \subset \mathcal{X},$$

  where the **exponent function**

$$V\{z(x) : x \in \mathcal{D}\} = \mathrm{E}\left[\sup_{x\in\mathcal{D}}\left\{\frac{W(x)}{z(x)}\right\}\right] = \mu[\{q : q(x) \le z(x), x \in \mathcal{D}\}^c]$$

  is derived from the mean measure μ of the Poisson process $\{Q_j\}$, and expectation is over the '**angular measure**' of W.

- The case $\mathcal{D} = \{x_1, \ldots, x_D\}$ is key to inference, because data are observed on finite sets, and then we write $z_d = z(x_d)$,

$$V(z_1, \ldots, z_D) = \mu(\mathcal{A}_z), \quad \mathcal{A}_z = ([0, z_1] \times \cdots \times [0, z_D])^c \subset \mathcal{E}' = [0, \infty)^D \setminus \{0\}.$$

- μ and V are **homogeneous of order -1**, i.e.,

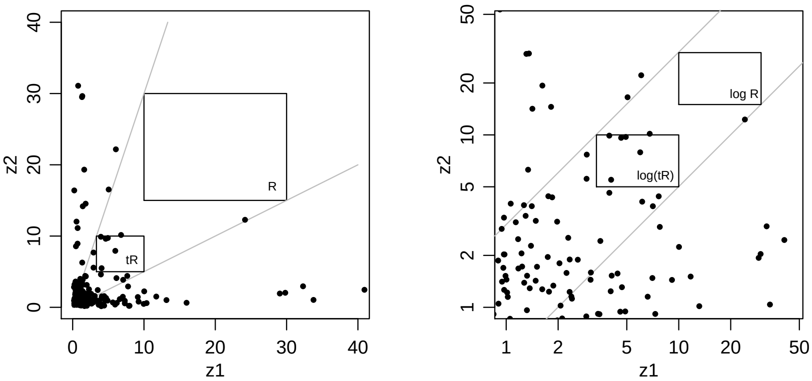
$$\mu(\mathcal{R}) = t \times \mu(t\mathcal{R}), \quad \mathcal{R} \subset \mathcal{E}, t > 0,$$

  which enables extrapolation by 'pulling down' extreme risk sets $\mathcal{R}$ to observable levels.

25

Pulling R down to the origin

Left: point process on unit Fréchet scale, with set $\mathcal{R}$ and its scaled version $\mathcal{R}$.
Right: same, but on Gumbel scale, with logarithmic axes, corresponding to translation of $\log \mathcal{R}$ by $\log t$ towards the origin.



26

Extremal coefficient

- ☐ Homogeneity of V yields

$$\mathrm{P}\left\{Z(x) \le z, x \in \mathcal{D}\right\} = \exp\left\{-V_{\mathcal{D}}(z)\right\} = \exp\left\{-V_{\mathcal{D}}(1)/z\right\} = \left(e^{-1/z}\right)^{V_{\mathcal{D}}(1)}, \quad z > 0,$$

  and the **extremal coefficient**

$$\theta_{\mathcal{D}} = V_{\mathcal{D}}(1)$$

  summarises the degree of dependence of extremes within $\mathcal{D}$.

- ☐ The pairwise version,

$$\theta(x, x') = \mathrm{E}\left[\max\left\{W(x), W(x')\right\}\right], \quad x, x' \in \mathcal{X},$$

  can be regarded as an analogue of the correlation coefficient, with

$$\text{(total dependence)} \quad 1 \le \theta(x, x') \le 2 \quad \text{(independence)},$$

  and the conditional probability interpretation

$$\mathrm{P}\left\{Z(x') > z \mid Z(x) > z\right\} \sim 2 - \theta(x, x'), \quad z \to \infty.$$

- ☐ $\theta(x, x')$ is estimated nonparametrically by the **F-madogram** (Cooley *et al.*, 2006).

27

Maxima and exceedances

- ☐ Inference may be based on
  - replicates of $\{Z(x) : x \in \mathcal{D}\}$, e.g., annual maximum temperatures at sites in $\mathcal{D}$,
  - individual events $\{Q_j(x) : x \in \mathcal{D}\}$, e.g., hurricanes or droughts.
- ☐ Extremal approximations may be better for maxima, but more detailed modelling is possible based on individual events.
- ☐ Choose 'extreme' events using **risk functional** ρ and retaining only events that fall into

$$\mathcal{E}' = \{q : \rho(q) > 1\}.$$

- ☐ Examples involving threshold function $u(x)$:

$$\rho_1(Q) = \sup_{x\in\mathcal{D}} Q(x)/u(x), \quad \rho_2(Q) = \inf_{x\in\mathcal{D}} Q(x)/u(x), \quad \rho_3(Q) = \int_{\mathcal{D}} Q(x)/u(x)\,\mathrm{d}x.$$

- ☐ Inference based on Poisson process likelihood for $\{q_j : q_j \in \mathcal{E}'\}$ involves $\mu(\mathcal{E}')$, which must be finite and computable.
- ☐ If $\rho(aQ) = a\rho(Q)$ for $a > 0$, then $\rho(Q) > 1$ gives $R\rho(W) > 1$; then $\mu(\mathcal{E}') = \mathrm{E}\{\rho(W)\}$ depends only on the distribution of W.

28

Models

29

Models

- ☐ Choice of W determines event size, orientation, smoothness, etc., with weak constraints $W \geq 0$ and $\mathrm{E}\{W(x)\} = 1$ for all $x \in \mathcal{X}$.
- ☐ Several choices are based on zero-mean Gaussian process $\varepsilon(x)$ for $x \in \mathcal{X}$ with variogram

$$\gamma(x, x') = \mathrm{var}\{\varepsilon(x) - \varepsilon(x')\}, \quad x, x' \in \mathcal{X},$$

with $\varepsilon(x)$ either intrinsically stationary or stationary; if stationary

$$0 \leq \gamma(x, x') \leq 2\mathrm{var}\{\varepsilon(x)\} = 2\sigma^2,$$

and if intrinsically stationary, then γ is unbounded but

$$\mathrm{cov}\{\varepsilon'(x_1), \varepsilon'(x_2)\} = \tfrac{1}{2}\left\{\gamma(x_1, x') + \gamma(x_2, x') - \gamma(x_1, x_2)\right\},$$

where $\varepsilon'(x) = \varepsilon(x) - \varepsilon(x')$ for some $x' \in \mathcal{X}$.

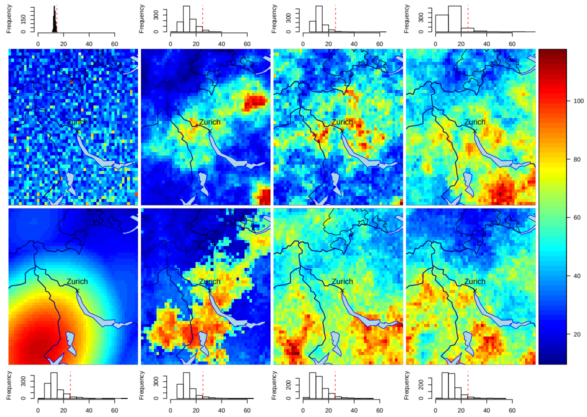
- ☐ Popular examples are the **Brown–Resnick** and **extremal** t processes (Brown and Resnick, 1977; Kabluchko *et al.*, 2009; Thibaud and Opitz, 2015),

$$W(x) = \exp\left[\varepsilon'(x) - \mathrm{var}\{\varepsilon'(x)\}/2\right], \quad W(x) \propto \varepsilon(x)_+^{\alpha}, \quad \alpha > 0,$$

but skew-Gaussian, skew-t, and hierarchical processes can be constructed (Tawn, 1990; Reich and Shaby, 2012; Reich *et al.*, 2014).

30

Realisations from spatial models



Top: latent variable, Student t copula, Hüsler–Reiss copula and extremal-t copula models. Bottom: Smith, Schlather, geometric Gaussian and Brown–Resnick models. The histograms are of 1000 realisations of a summary of rainfall centred on Zürich, and the vertical lines correspond to the realizations shown.
(Davison *et al.*, 2012)

31

Asymptotic dependence and independence

- ☐ The models above are **asymptotically dependent (AD)**:

$$\mathrm{P}\left\{Z(x') > z \mid Z(x) > z\right\} \sim p(x, x') + O(1/z), \quad z \to \infty,$$

so dependence between extremes persists at all levels: extreme events retain the same properties at all levels.

- ☐ Many applications show **asymptotic independence (AI)**, whereby extreme events become more concentrated as they become rarer.
- ☐ Can model this through an **inverted max-stable process** (Wadsworth and Tawn, 2012)

$$Z'(x) = -1/\log\left[1 - \exp\left\{-1/Z(x)\right\}\right],$$

for which

$$\mathrm{P}\left\{Z'(x') > z \mid Z'(x) > z\right\} = \mathcal{L}(z) z^{1-\theta(x,x')}, \quad z \to \infty,$$

for some slowly-varying function $\mathcal{L}$: the rate of approach to the limiting zero probability can vary.

- ☐ Recent work (Huser and Wadsworth, 2019) combines AD and AI, looks very useful.

32

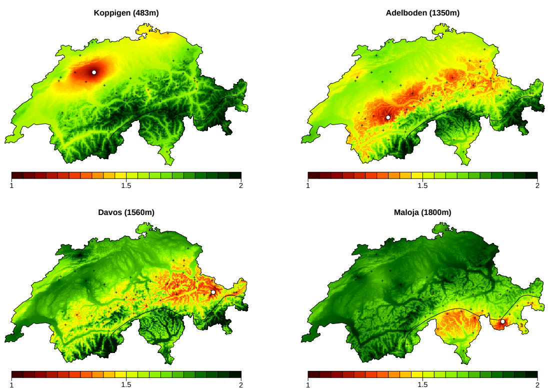
Generalities

- ☐ Extremal models are always mis-specified—inferences likely biased.
- ☐ Must check stability of inferences and possible presence of AI, so vary rarity of chosen events (threshold, ...).
- ☐ Extremal index useful for
  - exploratory and confirmatory analyses based on sub-groups (especially pairs) of observation sites $\mathcal{D}$,
  - simple estimates for model parameters, e.g., by least squares (Buhl and Klüppelberg, 2017).
- ☐ Mainly focus on likelihood-based inferences for parametric models, but also use gradient score.
- ☐ Semiparametric inference preferable, but
  - models are already quite flexible;
  - low power for falsifying models, because data necessarily limited;
  - simulation often needed for risk assessment.

34

Exploratory procedures

- ☐ Exploratory procedures are mostly based on estimates of $\theta(x_1, x_2)$, with the
  - **extremogram** $2 - \widehat{\theta}(x_1, x_2)$ in time series (Davis and Mikosch, 2009),
  - **F-madogram** in spatial cases.



Blanchet and Davison (2011)

35

Margins

- ☐ Extremes $Y(x)$ of original data at $\mathcal{D} = \{x_1, \ldots, x_D\}$ will have GEV/GPD distributions.
- ☐ For maxima, use **marginal transformation**

$$Z(x) = \left\{1 + \xi(x;\vartheta)\frac{Y(x) - \eta(x;\vartheta)}{\tau(x;\vartheta)}\right\}_+^{1/\xi(x;\vartheta)}$$

  to the unit Fréchet scale for inclusion in joint model, with
  - splines for space-varying location, scale and shape parameters,
  - and often constant shape, $\xi(x;\vartheta) \equiv \xi$.
- ☐ Similar transformation for threshold exceedances, with EDF/GPD below/above threshold.
- ☐ Either
  - estimate ϑ first, using **independence likelihood**, and then treat marginal transformation as fixed, or
  - perform **joint estimation** of margins and dependence structure.
- ☐ Balancing good marginal and joint fits can be tricky (easier in some Bayesian formulations).

36

Likelihood for maxima

- ☐ Given independent annual maxima observed at $\mathcal{D} = \{x_1, \ldots, x_D\}$ for n years, the maxima for each year have joint distribution

$$\mathrm{P}\{Z(x_1) \le z_1, \ldots, Z(x_D) \le z_D\} = \exp\{-V(z_1, \ldots, z_D)\}, \quad z_1, \ldots, z_D > 0.$$

- ☐ The form of the CDF means that to compute the likelihood we must differentiate e^{-V} with respect to $z_1, \ldots, z_D$, leading to combinatorial explosion:

$$-V_1 e^{-V}, \quad (V_1V_2 - V_{12})e^{-V}, \quad (-V_1V_2V_3 + V_{12}V_3[3] - V_{123})e^{-V}, \quad \ldots,$$

  with about 10^5 terms for $D = 10$. Clearly this is infeasible for realistic applications, so we try to avoid this, by
  - using a composite (usually a pairwise) likelihood;
  - using event timings to determine the required term, e.g., with $D = 3$,

$$(-V_1V_2V_3 \quad + \quad V_{12}V_3 + V_{13}V_2 + V_{23}V_1 \quad - V_{123})e^{-V};$$

  - or using threshold exceedances.
- ☐ We need V and its derivatives, or Poisson process intensity $\dot{\mu}$ and its integrals ...

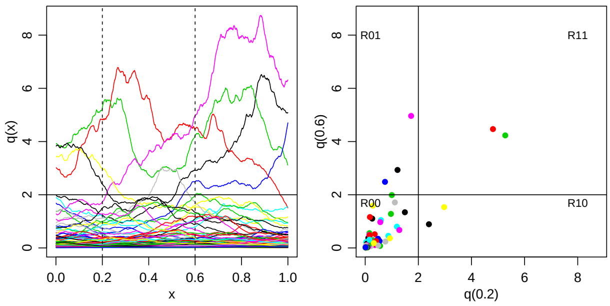
37

Likelihood for events

- Base extremal modelling on those individual events $q(x)$ falling into $\mathcal{E}' = \{q : \rho(q) > 1\}$, where ρ only uses $q(x)$ for $x \in \mathcal{D}$:
  - allows more detailed modelling and may include more data,
  - if $\mu(\mathcal{E}')$ is readily computed, in principle has simpler likelihood,

$$\exp\left\{-\mu(\mathcal{E}')\right\} \times \prod_{q \in \mathcal{E}'} \dot{\mu}(q), \quad \dot{\mu}(q) = -\frac{\partial^D V(z_1, \ldots, z_D)}{\partial z_1 \cdots \partial z_D},$$

  - but components of some q may be non-extreme, so use a **censored likelihood**.



38

Brown–Resnick likelihood

- If $z_d > u$ for $d = 1, \ldots, C$ and $z_d < u$ for $d \in \mathcal{C}' = \{C+1, \ldots, D\}$, and $\mathcal{C} = \{2, \ldots, C\}$, the censored likelihood contribution has form

$$\frac{1}{z_1^2 z_2 \cdots z_C} \times \phi_{C-1}(\log \tilde{z}_{\mathcal{C}}; \tilde{\Omega}_{\mathcal{C},\mathcal{C}}) \times \Phi_{D-C}\left(\tilde{\mu}_{\mathcal{C}'|\mathcal{C}}; \tilde{\Omega}_{\mathcal{C}'|\mathcal{C}}\right),$$

  where ϕ_k and Φ_k denote the k-dimensional normal density and distribution functions, Ω is defined in terms of the variogram γ, and

$$\begin{aligned}
\log \tilde{z}_d &= \log z_d - \log z_1 + \Omega_{d,1}/2, \quad d = 2, \ldots, C, \\
\tilde{\Omega}_{c,d} &= \tfrac{1}{2}\{\Omega_{c,1} + \Omega_{1,d} - \Omega_{c,d}\}, \quad c, d \in \{2, \ldots, D\}, \\
\mu_{\mathcal{C}'|\mathcal{C}} &= (\log u - \log z_1 + \tfrac{1}{2}\Omega_{1,\mathcal{C}'}) - \tilde{\Omega}_{\mathcal{C}',\mathcal{C}}\tilde{\Omega}_{\mathcal{C},\mathcal{C}}^{-1} \log \tilde{z}_{\mathcal{C}}, \\
\tilde{\Omega}_{\mathcal{C}'|\mathcal{C}} &= \tilde{\Omega}_{\mathcal{C}',\mathcal{C}'} - \tilde{\Omega}_{\mathcal{C}',\mathcal{C}}\tilde{\Omega}_{\mathcal{C},\mathcal{C}}^{-1}\tilde{\Omega}_{\mathcal{C},\mathcal{C}'}.
\end{aligned}$$

- Feasible for $D \le 100$, with modified R function for Φ (de Fondeville and Davison, 2018).
- Gradient score needed for higher D:
  - differentiate with respect to data, so normalising constants not needed;
  - use weight function to downweight effects of observations near thresholds.
- Similar computations are possible for extremal-t processes.

39

Closing

- ☐ Basic ideas on maxima and point processes extend to spatial and space-time settings.
- ☐ Max-stable processes give asymptotic dependence models—asymptotic independence also seen.
- ☐ Can fit such models using
  - composite (especially pairwise) likelihood,
  - full likelihood (needs additional information, difficult with large D),
  - Bayesian methods, or
  - gradient score methods.
- ☐ Model-checking possible, using simulation from fitted models and other techniques—but difficult to validate far into tails, because of lack of data.
- ☐ Currently much research in area (e.g., threshold models, non-stationarity, downscaling, semiparametric inference, networks, , . . .).

41

Some reading

- ☐ Coles (2001), *Introduction to the Statistical Modeling of Extreme Values*, Springer
- ☐ de Haan and Ferreira (2006) *Extreme Value Theory: An Introduction*, Springer
- ☐ Davison and Huser (2015) *Annual Review of Statistics and its Applications*
- ☐ Davison, Huser and Thibaud (2019) In *Handbook of Environmental and Ecological Statistics*, CRC Press
- ☐ de Haan (1984) *Annals of Probability*
- ☐ Schlather (2002) *Extremes*
- ☐ Cooley, Naveau, Poncet (2006) In *Lecture Notes in Statistics*, vol 187
- ☐ Cooley, Naveau, Nychka (2007) *JASA*
- ☐ Davison, Padoan and Ribatet (2012) *Statistical Science*
- ☐ Wadsworth and Tawn (2012) *Biometrika*
- ☐ Reich and Shaby (2012) *Annals of Applied Statistics*
- ☐ Wadsworth and Tawn (2015) *Biometrika*
- ☐ Thibaud and Opitz (2015) *Biometrika*
- ☐ Asadi, Davison and Engelke (2015) *Annals of Applied Statistics*
- ☐ de Fondeville and Davison (2018) *Biometrika*
- ☐ Huser and Wadsworth (2019) *JASA*

42

Some reading

Blanchet, J. and Davison, A. C. (2011) Spatial modelling of extreme snow depth. *Annals of Applied Statistics* **5**, 1699–1725.

Brown, B. M. and Resnick, S. I. (1977) Extreme values of independent stochastic processes. *Journal of Applied Probability* **14**, 732—739.

Buhl, S. and Klüppelberg, C. (2017) Generalised least squares estimation of regularly varying space-time processes based on flexible observation schemes. arXiv:1704.05656.

Cooley, D., Naveau, P. and Poncet, P. (2006) Variograms for max-stable random fields. In *Dependence in Probability and Statistics*, eds P. Bertail, P. Doukhan and P. Soulier, volume 187 of *Lecture Notes in Statistics*, pp. 373–390. New York: Springer.

Cooley, D., Nychka, D. and Naveau, P. (2007) Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**, 824–840.

Davis, R. A. and Mikosch, T. (2009) The extremogram: A correlogram for extreme events. *Bernoulli* **15**, 977–1009.

Davison, A. C., Padoan, S. A. and Ribatet, M. (2012) Statistical modelling of spatial extremes (with Discussion). *Statistical Science* **27**, 161–186.

de Fondeville, R. and Davison, A. C. (2018) High-dimensional peaks-over-threshold inference. *Biometrika* **105**, 575–592.

de Haan, L. (1984) A spectral representation for max-stable processes. *Annals of Probability* **12**, 1194–1204.

Fuentes, M., Henry, J. and Reich, B. (2013) Nonparametric spatial models for extremes: Application to extreme temperature data. *Extremes* **16**, 75–101.

Huser, R. and Wadsworth, J. (2019) Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association* **114**(DOI: 10.1080/01621459.2017.1411813), 434–444.

Kabluchko, Z., Schlather, M. and de Haan, L. (2009) Stationary max-stable fields associated to negative definite functions. *Annals of Probability* **37**, 2042–2065.

Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *Annals of Applied Statistics* **6**, 1430–1451.

Reich, B. J., Shaby, B. A. and Cooley, D. (2014) A hierarchical model for serially-dependent extremes: A study of heat waves in the western US. *Journal of Agricultural, Biological and Environmental Statistics* **19**, 119–135.

Sang, H. and Gelfand, A. E. (2010) Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological and Environmental Statistics* **15**, 49–65.

Tawn, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika* **77**, 245–253.

Thibaud, E. and Opitz, T. (2015) Efficient inference and simulation for elliptical Pareto processes. *Biometrika* **102**, 855–870.

Wadsworth, J. L. and Tawn, J. A. (2012) Dependence modelling for spatial extremes. *Biometrika* **99**, 253–272.