Spatial Extremes in Atmospheric Problems

Eric Gilleland
Research Applications Laboratory (RAL)
National Center for Atmospheric Research (NCAR),
Boulder, Colorado, U.S.A.
http://www.ral.ucar.edu/staff/ericg

The 19th TIES Conference, Kelowna, BC Canada, 12 June 2008

Brief Overview of Spatial Extremes Methods

- Multivariate Extremes and max-stable processes
- Copulas (e.g., Renard and Lang, 2006; Mikosch, 2006, and ensuing discussions).
- Regional Frequency Analysis (RFA)
- Loss function approach (IWQSEL), e.g. Craigmile et al. (2006)
- Upcrossings (e.g., Åberg and Guttorp, 2008, and references therein)
- Bayesian (e.g., Casson and Coles, 1999; Cooley et al., 2007)
- Spatio-temporal Models (e.g., Davis and Mikosch, 2006; Wikle and Cressie, 1999)
- EVD with Spatial Model on Parameters

Brief Overview: Multivariate Extremes

Coles (2001, chapter 8)

Schlather and Tawn (2003)

Max-stable processes
Smith (1990, and references therein)
Schlather (2002, and references therein)
Cooley et al. (2006)

Conditional approach
Heffernan and Tawn (2004, and ensuing discussions),
see also Mendes and Pericchi (2008)

Brief Overview: Regional Frequency Analysis (RFA)

Hosking and Wallis (1997)

Flood maps

Daly et al. (2002)
Daly et al. (1994)
Sveinsson et al. (2002)

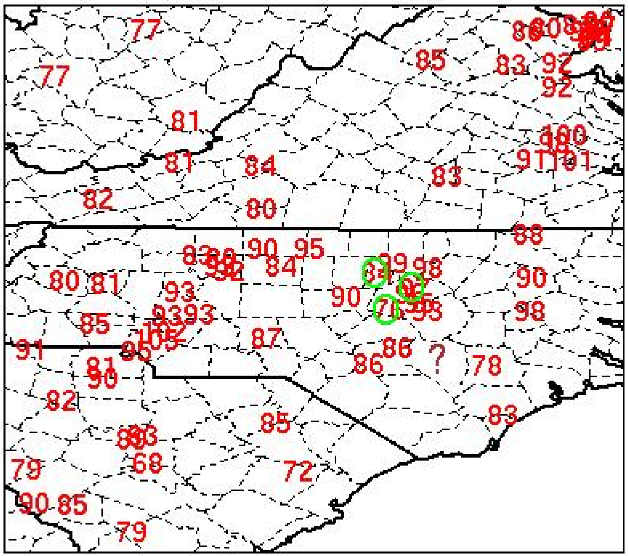
Precipitation

Schaefer (1990)
Fowler et al. (2005)
Fowler and Kilsby (2003)

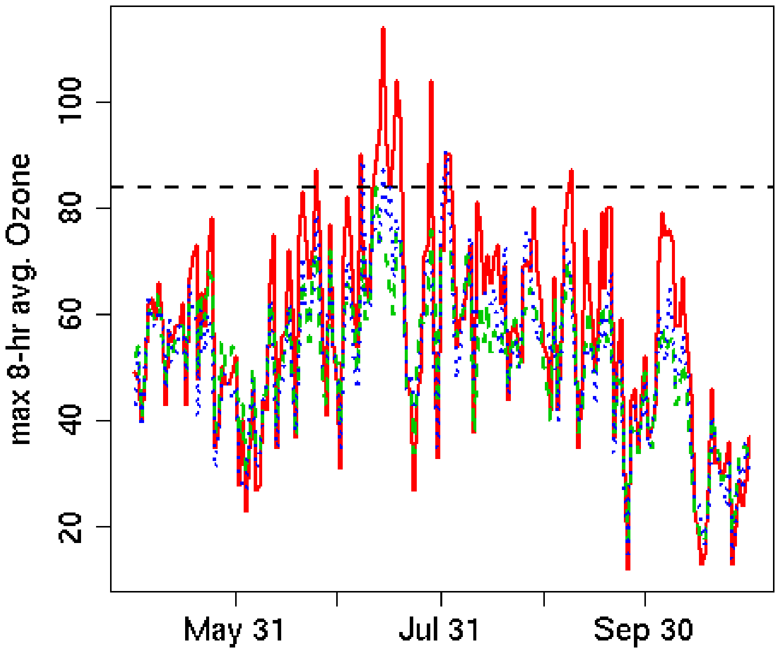
A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

North Carolina



Three sites



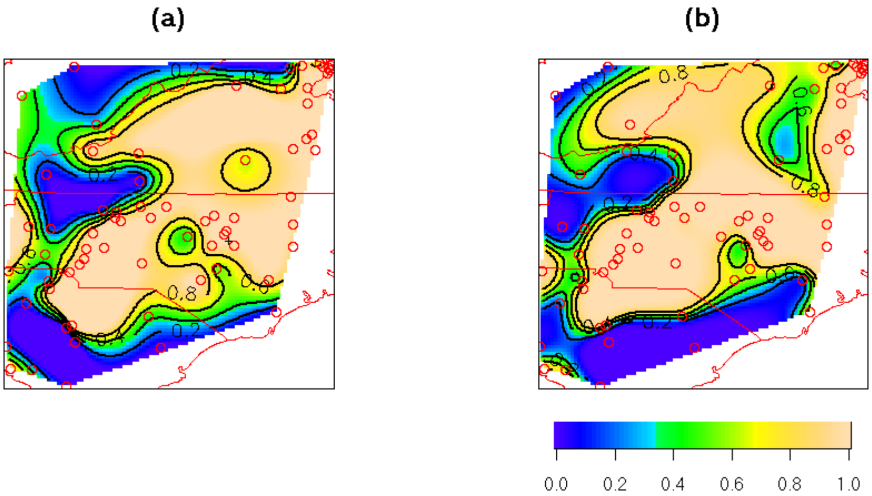
FHDA – Fourth Highest Daily maximum 8-hr Average ozone

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo (Gilleland and Nychka, 2005)

GPD with spatial model on scale parameter (Gilleland et al., 2006)



$$\Pr\{\text{FHDA}_{1997} > 80\}$$

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo

$$y(\boldsymbol{s};t) = \sigma(\boldsymbol{s})u(\boldsymbol{s};t) + \mu(\boldsymbol{s};t)$$

$$u(\boldsymbol{s};t) = \rho(\boldsymbol{s})u(\boldsymbol{s};t-1) + \varepsilon(\boldsymbol{s};t)$$

$$\boldsymbol{\varepsilon}(\cdot;t) \sim \mathrm{Gau}\left(\mathbf{0},\, \boldsymbol{\Sigma}\right)$$

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo

$$\boldsymbol{\Sigma} = \left[\mathrm{cov}(\boldsymbol{s}_i,\, \boldsymbol{s}_j)\right],$$

$\mathrm{cov}(\boldsymbol{s}_i,\, \boldsymbol{s}_j) = \psi(h)$ (Stationary).

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo

$$\operatorname{cov}\left(u(\boldsymbol{s};t), u(\boldsymbol{s}';t-\tau)\right) = \frac{(\rho(\boldsymbol{s}))^{\tau}\sqrt{1-\rho^2(\boldsymbol{s})}\sqrt{1-\rho^2(\boldsymbol{s}')}}{1-\rho(\boldsymbol{s})\rho(\boldsymbol{s}')}\psi(h),$$

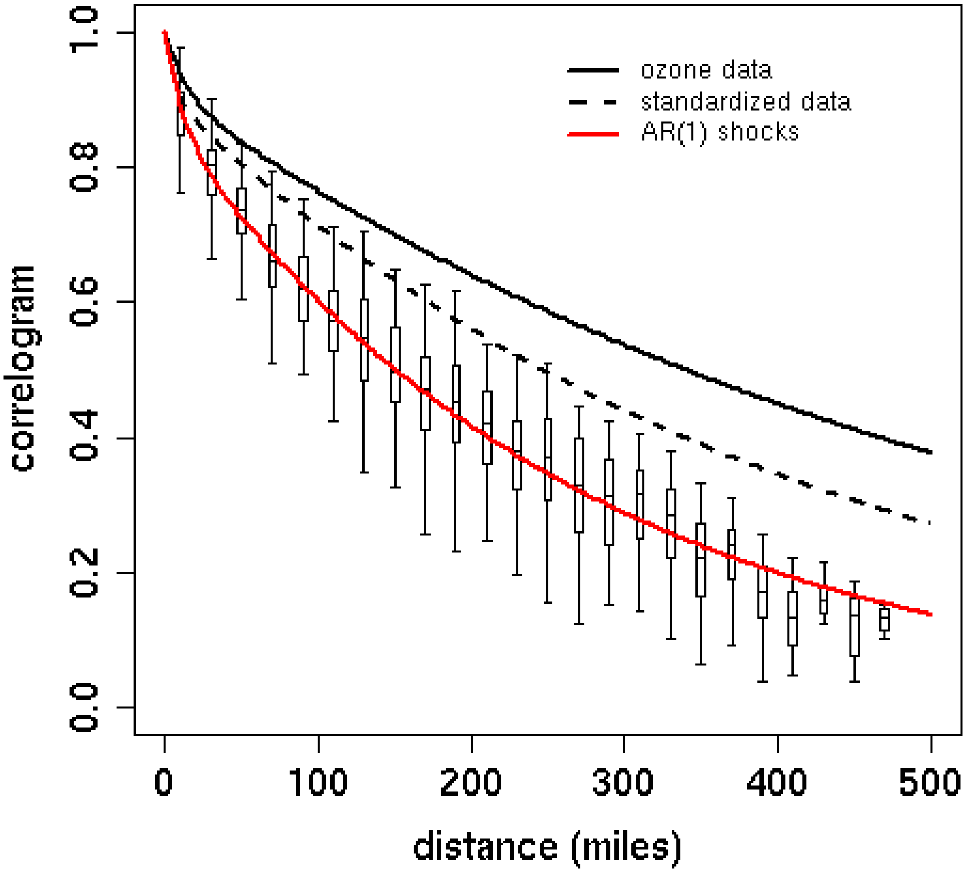
for $\tau \geq 0$.

If $\rho(\boldsymbol{s}) = \rho$, then $\operatorname{cov}\left(u(\boldsymbol{s};t), u(\boldsymbol{s}';t-\tau)\right) = \rho^{\tau}\psi(h)$.

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

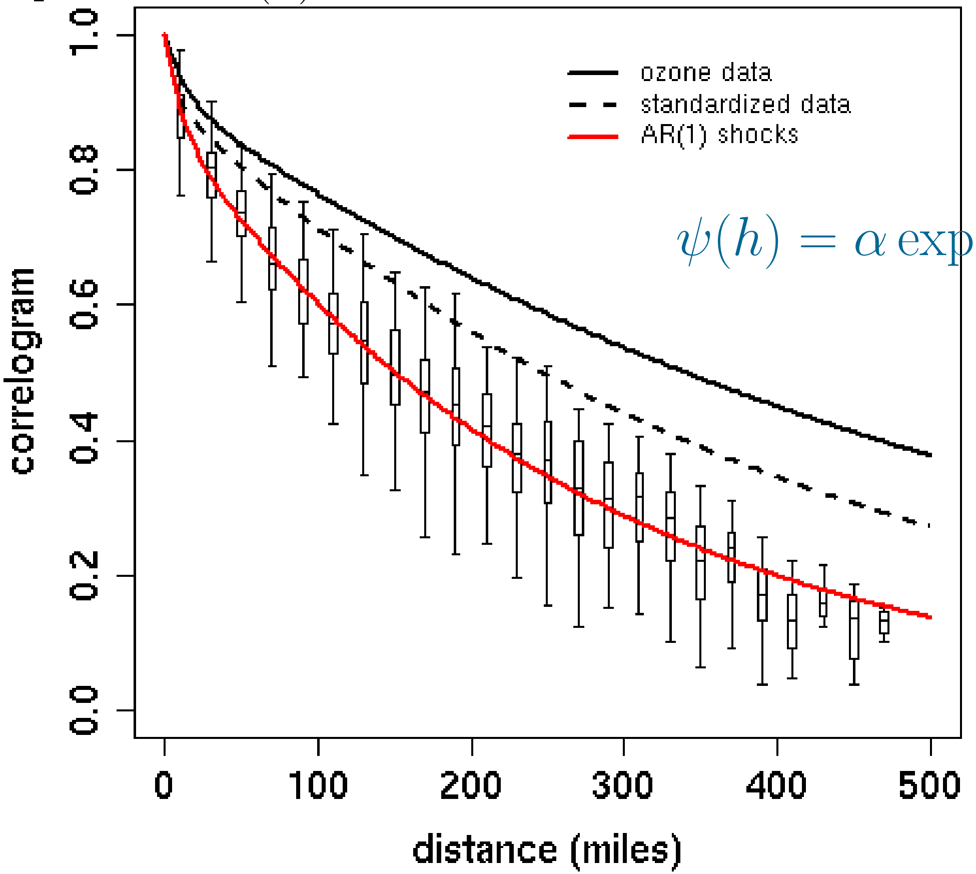
Spatial AR(1) Monte Carlo



A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo



$$\psi(h) = \alpha \exp\left(-\frac{h}{\theta_1}\right) + (1-\alpha)\exp\left(-\frac{h}{\theta_2}\right)$$

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo

$$\psi(h) = \alpha \exp\left(-\frac{h}{\theta_1}\right) + (1-\alpha)\exp\left(-\frac{h}{\theta_2}\right)$$

Here: $\hat{\alpha} \approx 0.13$ (± 0.02), $\hat{\theta}_1 \approx 11$ miles (± 3.37 miles) and $\hat{\theta}_2 \approx 272$ miles (± 16.89 miles).

Uncertainty via parametric bootstrap.

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo

Algorithm to predict FHDA at unobserved location(s), $\boldsymbol{s}_0$.

1. Simulate $u(\boldsymbol{s}_0; t)$ for an entire ozone season
   (a) Interpolate spatially from $u(\boldsymbol{s}, 1)$ to get $\hat{u}(\boldsymbol{s}_0, 1)$.
   (b) Also interpolate spatially to get $\hat{\rho}(\boldsymbol{s}_0)$, $\hat{\mu}(\boldsymbol{s}_0, \cdot)$ and $\hat{\sigma}(\boldsymbol{s}_0)$.
   (c) Sample shocks at time t from $[\varepsilon(\boldsymbol{s}_0, t)|\varepsilon(\boldsymbol{s}, t)]$.
   (d) Propagate AR(1) model.

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo

Algorithm to predict FHDA at unobserved location, $\boldsymbol{s}_0$*.*

1. Simulate $u(\boldsymbol{s}_0; t)$ for an entire ozone season.
2. Back transform $\hat{y}(\boldsymbol{s}_0, t) = \hat{u}(\boldsymbol{s}_0, t)\hat{\sigma}(\boldsymbol{s}_0) + \hat{\mu}(\boldsymbol{s}_0, t)$
3. Take fourth-highest value from Step 2.
4. Repeat Steps 1 through 3 many times to get a sample of FHDA at unobserved location(s).

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo

Distribution for the AR(1) shocks

$[\varepsilon(\boldsymbol{s}_0, t)|\varepsilon(\boldsymbol{s}, t)]$ (Step 1c) given by

$$\mathrm{Gau}(\boldsymbol{M}, \boldsymbol{\Sigma})$$

with

$$\mathbf{M} = \boldsymbol{k}'(\boldsymbol{s}_0, \boldsymbol{s})\boldsymbol{k}^{-1}(\boldsymbol{s}, \boldsymbol{s})\varepsilon(\boldsymbol{s}, t)$$

and

$$\boldsymbol{\Sigma} = \boldsymbol{k}'(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{k}'(\boldsymbol{s}_0, \boldsymbol{s})\boldsymbol{k}^{-1}(\boldsymbol{s}, \boldsymbol{s})\boldsymbol{k}(\boldsymbol{s}, \boldsymbol{s}_0),$$

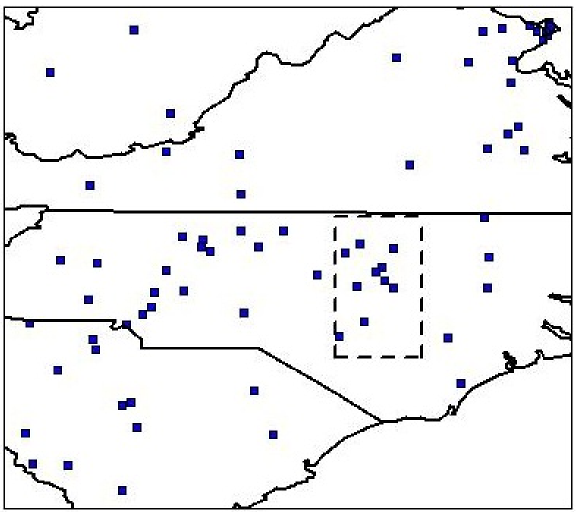
where $\boldsymbol{k}(\boldsymbol{x}, \boldsymbol{y}) = [\psi(\boldsymbol{x}_i, \boldsymbol{y}_j)]$ the covariance matrix for two sets of spatial locations.

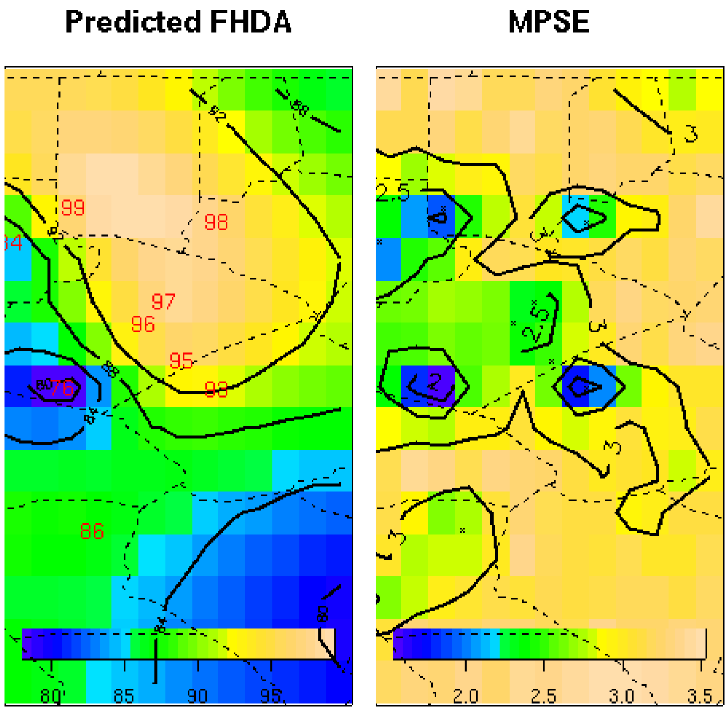
A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo

Results of predicting FHDA spatially with daily model (1997)

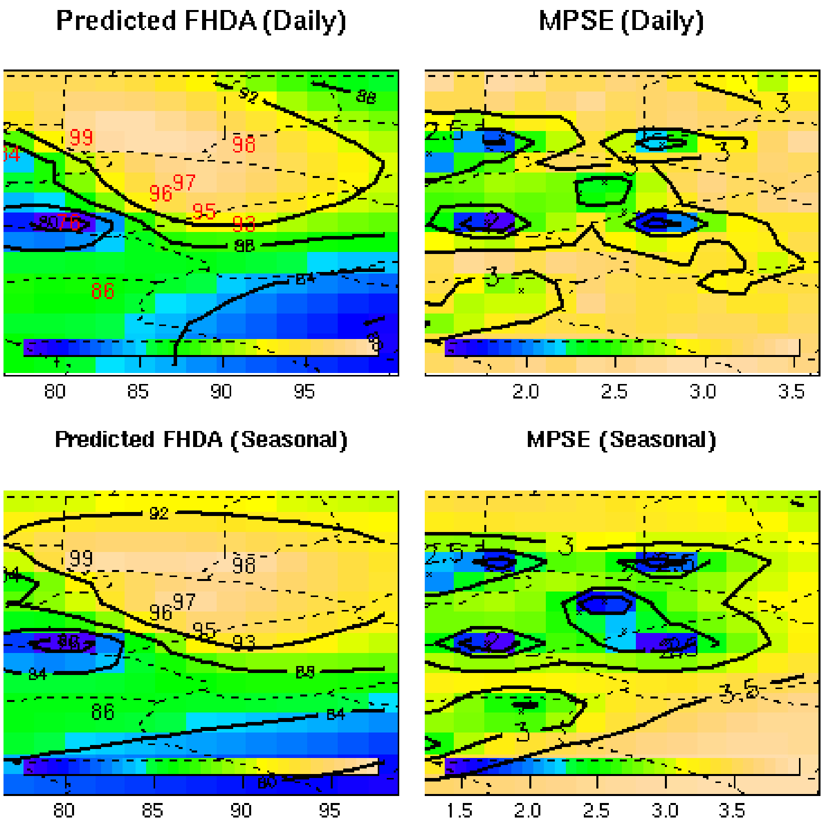




A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

Spatial AR(1) Monte Carlo



A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

GPD with spatial model on scale parameter

Given a spatial process, $Z(\boldsymbol{s})$, what can be said about

$$\Pr\{Z(\boldsymbol{s}) > z\}$$

when z is large?

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

GPD with spatial model on scale parameter

Given a spatial process, $Z(\boldsymbol{s})$, what can be said about

$$\Pr\{Z(\boldsymbol{s}) > z\}$$

when z is large?

Note:
This is not about dependence between $Z(\boldsymbol{s})$ and $Z(\boldsymbol{s}')$–this is another topic!

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

GPD with spatial model on scale parameter

Given a spatial process, $Z(\boldsymbol{s})$, what can be said about

$$\Pr\{Z(\boldsymbol{s}) > z\}$$

when z is large?

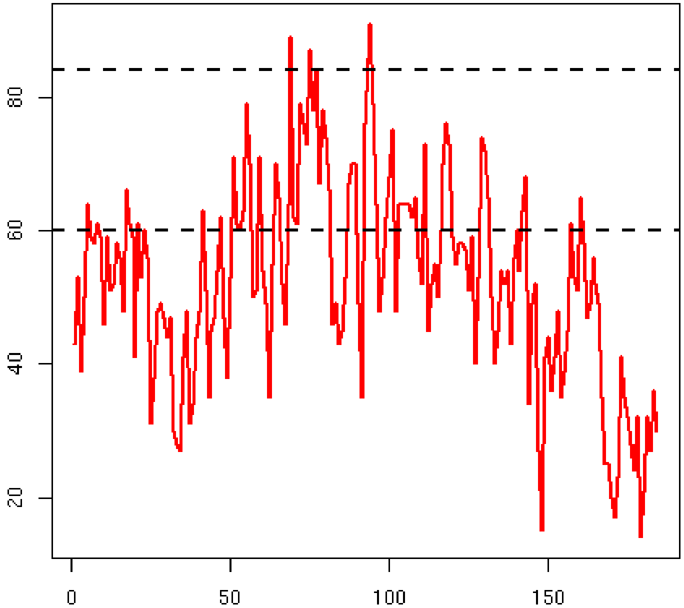
Note:
This is not about dependence between $Z(\boldsymbol{s})$ and $Z(\boldsymbol{s}')$–this is another topic!

Spatial structure on parameters of distribution (not FHDA).

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

GPD with spatial model on scale parameter



For a (large) threshold u, the GPD is given by

$$\Pr\{X > x | X > u\} \approx [1 + \frac{\xi}{\sigma}(x - u)]^{-1/\xi}$$

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

GPD with spatial model on scale parameter

Observation Model:

$y(\boldsymbol{s}, t)$ surface ozone at location $\boldsymbol{s}$ and time t

$$[y(\boldsymbol{s}, t) | \sigma(\boldsymbol{s}), \xi(\boldsymbol{s}), u, y(\boldsymbol{s}, t) > u]$$

Spatial Process Model:

$$[\sigma(\boldsymbol{s}), \xi(\boldsymbol{s}), u | \boldsymbol{\theta}]$$

Prior for hyperparameters:

$$[\boldsymbol{\theta}]$$

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

GPD with spatial model on scale parameter

A Hierarchical Spatial Model

Assume extreme observations to be *conditionally independent* so that the joint pdf for the data and parameters is

$$\prod_{i,t}[y(\boldsymbol{s}_i,t)|\sigma(\boldsymbol{s}),\xi(\boldsymbol{s}),u,y(\boldsymbol{s}_i,t)>u]\ [\sigma(\boldsymbol{s}),\xi(\boldsymbol{s}),u|\boldsymbol{\theta}]\ [\boldsymbol{\theta}]$$

t indexes time and i stations.

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

GPD with spatial model on scale parameter

Shortcuts and Assumptions

- Assume threshold, u, fixed.
- $\xi(\boldsymbol{s}) = \xi$ (i.e., shape is constant over space). *Justified by univariate fits.*
- Assume $\sigma(\boldsymbol{s})$ is a Gaussian process with isotropic Matérn covariance function.
- Fix Matérn smoothness parameter at $\nu = 2$, and let the range be very large–leaving only λ (ratio of variances of nugget and sill).

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

GPD with spatial model on scale parameter

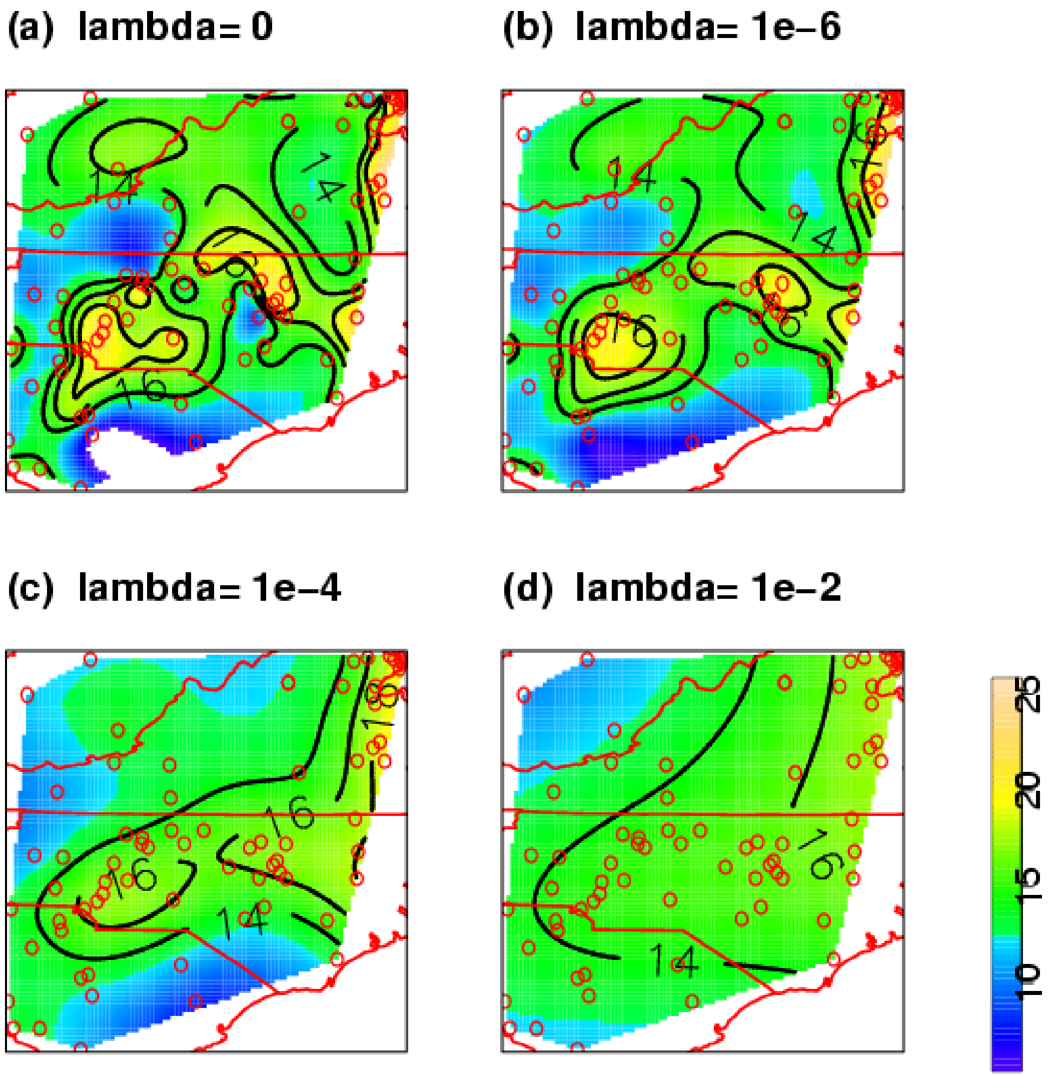
More on $\sigma(\boldsymbol{s})$

$$\sigma(\boldsymbol{s}) = P(\boldsymbol{s}) + e(\boldsymbol{s}) + \eta(\boldsymbol{s})$$

with P a linear function of space, e a smooth spatial process, and η white noise (nugget).

λ is the only hyper-parameter

- As $\lambda \longrightarrow \infty$, the posterior surface tends toward just the linear function.
- As $\lambda \longrightarrow 0$, the posterior surface will fit the data more closely.


(a) lambda= 0
(b) lambda= 1e−6
(c) lambda= 1e−4
(d) lambda= 1e−2
14
16
18
10
15
20
25

A couple of simple approaches

Spatial Distribution for fourth-largest order statistic

GPD with spatial model on scale parameter

log of joint distribution

$$\sum_{i=1}^{n} \ell_{\text{GPD}}(y(\boldsymbol{s}_i, t), \sigma(\boldsymbol{s}_i), \xi) -$$

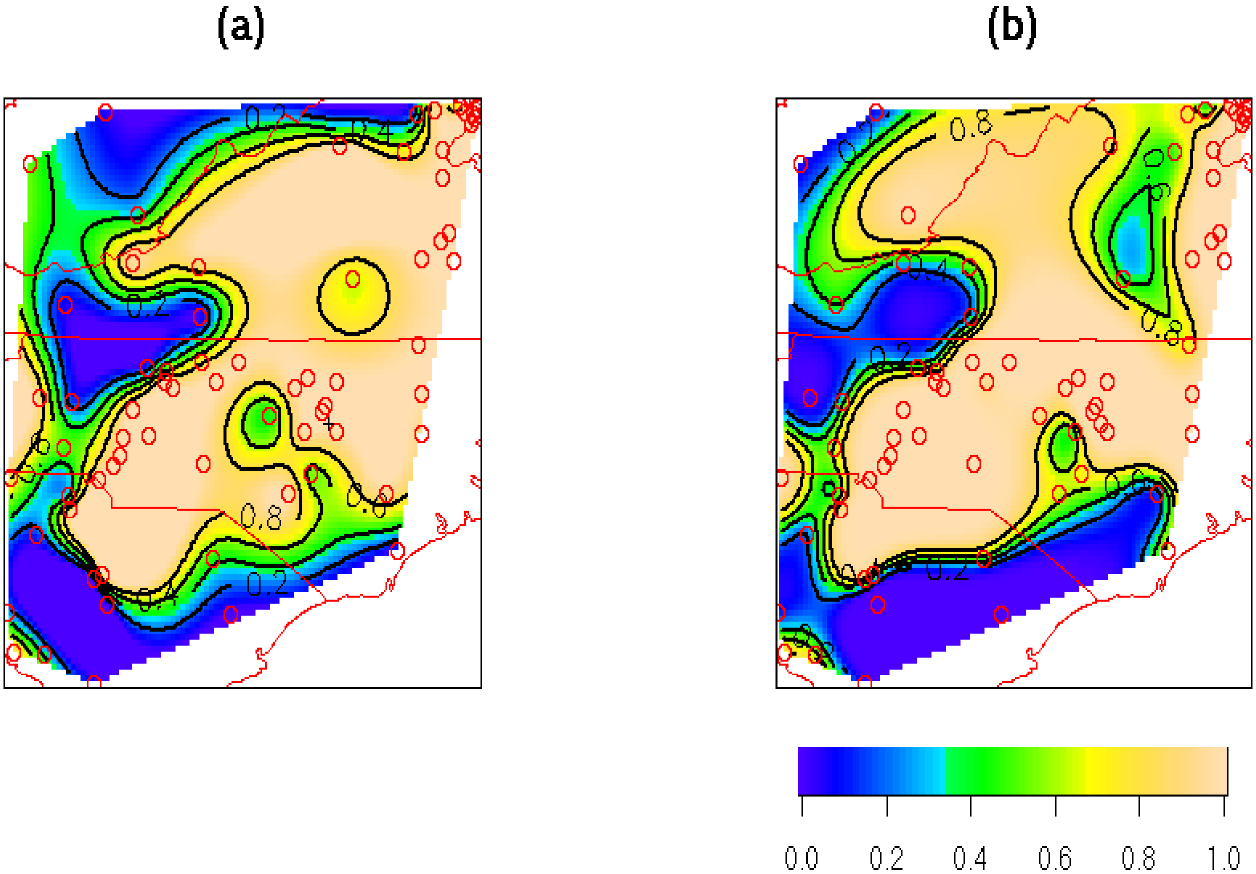
$$\lambda(\boldsymbol{\sigma} - \boldsymbol{X\beta})^T K^{-1}(\boldsymbol{\sigma} - \boldsymbol{X\beta})/2 - \log(|\lambda K|) + C$$

K is the covariance for the prior on $\boldsymbol{\sigma}$ at the observations.

This is a penalized likelihood:

The penalty on $\boldsymbol{\sigma}$ results from the covariance and smoothing parameter λ.

Probability of exceeding the standard



Conclusions and Possible Extensions

Spatial AR(1) Monte Carlo

- Each part simple, but can model relatively complex processes
- Lower CV than naïve approach assuming FHDA$\sim$ Gau$(\boldsymbol{M}, \boldsymbol{V})$
- Computationally intensive
- Prediction standard errors too optimistic?
- Let $u(\boldsymbol{s}; t) = \rho(\boldsymbol{s})u(\boldsymbol{s}; t-1) + \beta\varepsilon_1(\boldsymbol{s}; t) + (1-\beta)\varepsilon_2(\boldsymbol{s}; t)$
- Allow ψ to be nonstationary for larger domains
- Incorporate meteorological covariates

Conclusions and Possible Extensions

GPD with spatial model on scale parameter

- Simple extension to independent fits at each spatial location
- Compares relatively well with spatial AR(1) Monte Carlo approach
- Direct use of EVA
- More difficult if region is not homogeneous
- Does not model the underlying process spatially
- Apply spatial model to the return levels instead of the parameters
- Incorporate meteorological covariates

That's it...

References will be posted on my web page at

http://www.ral.ucar.edu/~ericg

Questions/Comments

References

Åberg, S. and P. Guttorp, 2008: Distribution of the maximum in air pollution fields. *Environmetrics*, **19**, 183–208, doi:10.1002/env.866.

Casson, E. and S. Coles, 1999: Spatial regression models for extremes. *Extremes*, **1**, 449–468.

Coles, S., 2001: *An introduction to statistical modeling of extreme values*. Springer-Verlag, London, UK, 208 pp.

Cooley, D., P. Naveau, and P. Poncet, 2006: *Variograms for spatial max-stable random fields*. In Dependence in Probability and Statistics. Ed. Bertail, P. and Doukhan, P. and Soulier, P. Springer Lecture Notes in Statistics no. 187.

Cooley, D., D. Nychka, and P. Naveau, 2007: Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Stat. Assoc.*, **102**, 824–840.

Craigmile, P., N. Cressie, T. Santner, and Y. Rao, 2006: A loss function approach to identifying environmental exceedances. *Extremes*, **8**, 143–159, doi:10.1007/s10687–006–7964–y.

Daly, C., W. Gibson, G. Taylor, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Research*, **23**, 99–113.

Daly, C., R. Neilson, and D. Phillips, 1994: A statistical topographic model for mapping climatological precipitation in mountainous terrain. *J. Appl. Meteorol.*, **33**, 140–158.

Davis, R. and T. Mikosch, 2006: Extreme value theory for space-time processes with heavy-tailed distributions. *Submitted (Available at: http://www.stat.colostate.edu/%7Erdavis/technical_reports.html)*.

Fowler, H., M. Eskstrom, C. Kilsby, and P. Jones, 2005: New estimates of future changes in extreme rainfall across the uk using regional climate model integrations, part 1: Assessment of control climate. *J. Hydrology*, **300**, 212–233.

Fowler, H. and C. Kilsby, 2003: A regional frequency analysis of united kingdom extreme rainfall from 1961 to 2000. *International J. Climatol.*, **23**, 1313–1334.

Gilleland, E. and D. Nychka, 2005: Statistical models for monitoring and regulating ground-level ozone. *Environmetrics*, **16**, 535–546.

Gilleland, E., D. Nychka, and U. Schneider, 2006: *Hierarchical modelling for the Environmental Sciences: statistical methods and applications*, Eds. JS Clark and A Gelfand. Oxford University Press, New York, chapter Spatial models for the distribution of extremes. 170–183.

Heffernan, J. and J. Tawn, 2004: A conditional approach for multivariate extreme values. *J.R. Statist. Soc. B*, **66**, 497–546.

Hosking, J. and J. Wallis, 1997: *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press, Cambridge, UK, 240 pp.

Mendes, B. and L. Pericchi, 2008: Assessing conditional extremal risk of flooding in puerto rico. *Stoch. Environ. Res. Risk. Assess.*, doi:10.1007/s00477-008-0220-z.

Mikosch, T., 2006: Copulas: Tales and facts. *Extremes*, **9**, 3–20, doi:10.1007/s10687–006–0015–x.

Renard, B. and M. Lang, 2006: Use of a gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, **30**, 897–912, doi:10.1016/j.advwatres.2006.08.001.

Schaefer, M., 1990: Regional analysis of precipitation annual maxima in washington state. *Water Resources Research*, **26**, 119–131.

Schlather, M., 2002: Models for stationary max-stable random fields. *Extremes*, **5**, 33–44.

Schlather, M. and J. Tawn, 2003: A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, **90**, 139–156.

Smith, R., 1990: Max-stable processes and spatial extremes. *Unpublished Manuscript (available at: http://www.stat.unc.edu/postscript/rs/spatex.pdf)*, 1–32.

Sveinsson, O., J. Salas, and D. Boes, 2002: Regional frequency analysis of extreme precipitation in northeastern colorado and fort collins flood of 1997. *J. Hydrologic Engineering*, **7**, 49–63.

Brief Overview: Multivariate Extremes

General Formulation

$$\lim_{n\longrightarrow\infty}\left(\Pr\left[\frac{M_{n1}(\boldsymbol{x}_1)-b_{n1}}{a_{n1}}\leq x_1,\ldots,\frac{M_{nd}(\boldsymbol{x}_d)-b_{n1}}{a_{n1}}\leq x_d\right]\right)=$$
$$G(x_1,\ldots,x_d),$$

where $\boldsymbol{x}_i = (x_{1i},\ldots,x_{ni})$ are iid $d-$dimensional random vectors, $M_{ni}(\boldsymbol{x}_i) = \max_{j=1,\ldots,n}(x_{ji})$, $a_{n1},\ldots,a_{nd} > 0$ and $b_{n1},\ldots,b_{nd}$ are normalizing constants, and G is a non-degenerate $d-$dimensional cdf.

Brief Overview: Multivariate Extremes

General Formulation

Simplify by assuming each $\boldsymbol{x}_i$ has a standard Fréchet marginal cdf. Then,

$$G(x_1, \ldots, x_d) = \exp\{-V(x_1, \ldots, x_d)\},$$

where

$$V(x_1, \ldots, x_d) = \int \max_{j=1,\ldots,d} \left\{ \frac{w_j}{x_j} \right\} dH(w_1, \ldots, w_d)$$

Brief Overview: Multivariate Extremes

General Formulation

Extremal coefficient measures dependence in the tails.

$$\theta = \int \max_{j=1,\ldots,d} w_j dH(w_1, \ldots, w_d)$$

Brief Overview: Copulas

General Formulation

$\boldsymbol{X} = (X_1, \ldots, X_d)$ a $d-$dimensional random vector with cdf

$$F(x_1, \ldots, x_d) = \Pr\left(X_1 \leq x_1, \ldots, X_d \leq x_d\right).$$

A *copula* is a function, c, s.t. $c : [0,1]^d \mapsto [0,1]$ and

$$F(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d))$$

Brief Overview: Copulas

Assuming the marginal distributions F_i, then

- c exists, and
- if the F_i's are continuous, then c is unique.
- The dependence structure of $\boldsymbol{X}$ can be reconstructed from the copula and the F_i's.

Brief Overview: Regional Frequency Analysis (RFA)

General Formulation

Multiple-step procedure:

1. Determine relatively homogeneous regions.
2. Normalize annual maxima series by an index (flood) measure.
3. Fit a distribution (e.g., GEV) to pooled dimensionless sample.
4. Scale distribution from 3 by indexes from 2 to get local distributions.

Brief Overview: Regional Frequency Analysis (RFA)

- L-moments used for parameter estimation.
- Criteria based on L-moments are suggested for selecting homogeneous regions and for choosing a probability distribution.
- Uncertainty obtained via bootstrap methods.