CBMS Lecture 10

Alan E. Gelfand
Duke University

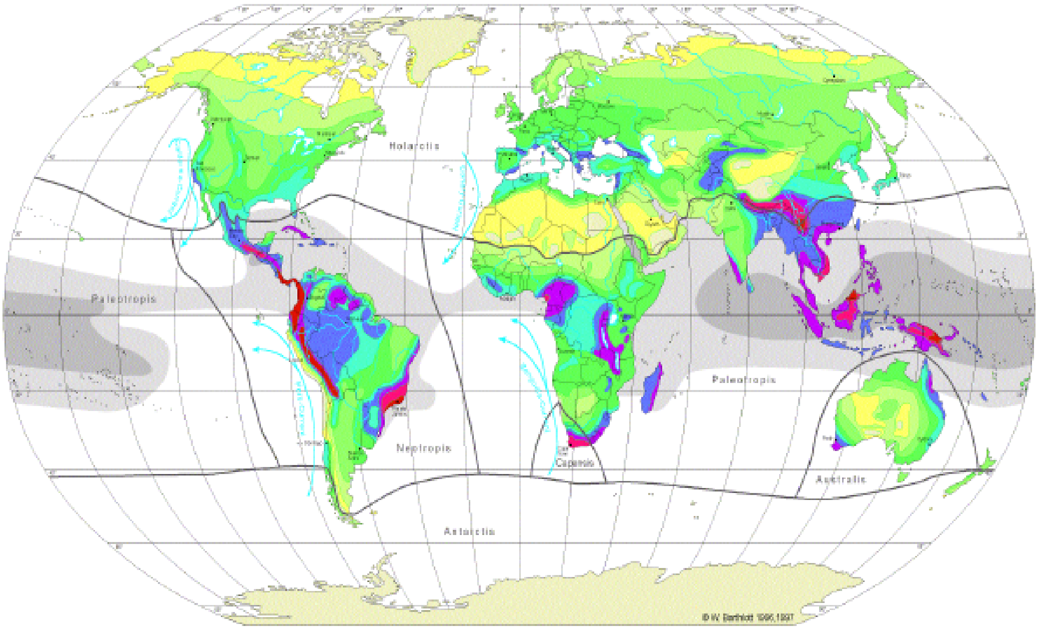
Space-time modeling for extremes

- Extreme value analysis is frequently applied to environmental science data.
- Extremes in exposure to environmental contaminants, extremes in weather.
- For plants, extreme weather events such as drought, heavy rainfall and very high or low temperatures, might be more significant factors in explaining plant performance with regard to survival, growth, reproductivity, etc., than trends in the mean climate.

Cape Floristic Region

- We illustrate the analysis of spatio-temporal weather extremes with precipitation data from the Cape Floristic Region (CFR)
- The smallest but, arguably, the richest of the world's six floral kingdoms, encompassing a region of roughly $90,000 km^2$ in southwestern South Africa.
- The daily precipitation surfaces arise via interpolation to grid cells at 10km resolution based on records reported by up to 3000 stations over South Africa over the period from 1950-1999.
- We have 50 derived surfaces of annual maxima surfaces from daily rainfalls for 1332 grid cells.
- We hold out year 1999 for model validation.

GLOBAL BIODIVERSITY: SPECIES NUMBERS OF VASCULAR PLANTS



Robinson Projection
Standard Parallels 38°N und 38°S
Scale 1: 130000000

Diversity Zones (DZ): Number of species per 10.000km²

DZ 1 (-100)
DZ 2 (100 - 200)
DZ 3 (200 - 500)
DZ 4 (500 - 1000)
DZ 5 (1000 - 1500)
DZ 6 (1500 - 2000)
DZ 7 (2000 - 3000)
DZ 8 (3000 - 4000)
DZ 9 (4000 - 5000)
DZ 10 (>5000)

Capensis floristic regions

sea surface temperature

>29°C
>27°C

cold currents

W. Barthlott, N. Biedinger, G. Braun
F. Feig, G. Kier, W. Lauer & J. Mutke 1997
modified after
W. Barthlott, W. Lauer & A. Placke 1996
Department of Botany and Geography
University of Bonn
German Aerospace Research Establishment, Cologne

Cartography: M. Graf
Department of Geography
University of Bonn

Cape Floristic (Biogeographic) Region

90,000 km^2

9000+ plant species

70% found nowhere else

1400+ threatened or endangered species

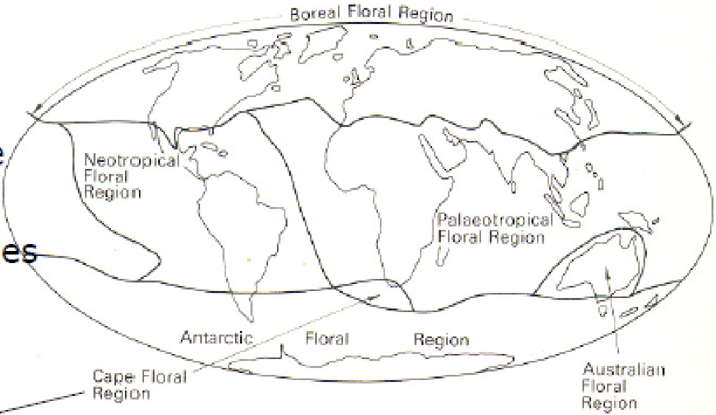
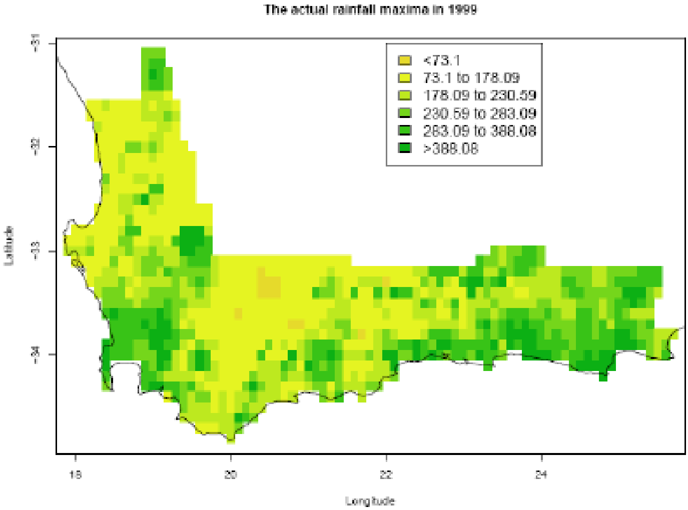


Figure 63. The floral regions of the world today. (After Good, 1964; courtesy of Longmans, Green and Co. Ltd., London.)



The actual rainfall maxima in 1999
<73.1
73.1 to 178.09
178.09 to 230.59
230.59 to 283.09
283.09 to 388.08
>388.08
Latitude
-31
-32
-33
-34
Longitude
18
20
22
24

Spatial work

- An enormous literature on the modeling of extremes.
- The standard approach utilizes Generalized Extreme Value (GEV) distribution families.
- Alternatively, daily precipitation exceedances for a given threshold using the Generalized Pareto Distribution (GPD).
- Here, we consider hierarchical models for extreme rainfall which reflect dependence in space and in time.
- We use the GEV which is characterized by a location, a scale and a shape parameter. Each of these could vary in space and time and they could be mutually dependent.
- So, one can envision a range of such models, fitting them, and comparing them.
- First, grid cell data using Markov random field models.
- Then, a dataset at station (point) level and use Gaussian processes.

Possibilities for modeling maxima

- Suppose at site $\mathbf{s}$, daily data j in year t
- That is, we envision the collection of surfaces $W_{t,j}(\mathbf{s})$ which are observed at $\{s_i, i = 1, 2, ..., n\}$.
- Seek inference regarding $Y_t(s) = \max_j W_{t,j}(s)$, i.e., we are considering maxima over time at a particular location, not maxima over space at a given time.
- Evidently in the setting of multivariate extreme value theory.
- It seems clear that spatial dependence in the maxima will be weaker than spatial dependence in the daily data.
- Inference focuses on parametric issues, e.g., trends in time, dependence in space, as well as on prediction at new locations and at future (one-step ahead) times.
- Also, risk assessment with regard to return time. That is, if $P(Y_t(\mathbf{s}) > y) = p$, the expected return time until an exceedance of y is $1/p$ time units.

cont.

- We could model the W's, which induces inference for the Y's.
- Here, an enormous literature, using hierarchical space-time modeling.
- For a generic model, suppose we obtain the posterior, $[\theta|\{W_{t,j}(\mathbf{s}_i)\}]$.
- And at a new s_0, obtain the predictive distribution for $W_{t,j}(\mathbf{s}_0)$ in the form of posterior samples.
- These posterior samples can be converted into posterior samples of $Y_t(s_0)$, i.e., "derived" quantities.
- We can do similarly for forecasting to time $T+1$.

cont.

- Why is this path not of interest here.
- First, it is not modeling or explaining the maxima
- Second, it is extremely computationally demanding, generating hundreds and hundreds of daily data samples and discarding them all, retaining just the maxima as a summary.
- Most importantly, there will be greater uncertainty in predictions made using derived quantities than in prediction based on direct modeling of the maxima.

cont.

- There are still two paths to directly model the maxima:
  - We can model the maxima or
  - model the "process" that drives the maxima.
- Under the first path, the usual assumptions are that the $Y_t(\mathbf{s})$ are spatially dependent but temporally independent, i.e., they are viewed as replicates.
- The argument here is strong temporal dependence at the scale of days but temporal dependence will be negligible between yearly values. Usually, not an explanatory model, no covariates
- The second path is to provide space-time modeling for the parameters in the GEV's for the $\{Y_t(\mathbf{s})\}$.
- We develop this "process" path

Review of Extreme Value Theory

- For a sequence $Y_1, Y_2, \ldots$ of i.i.d. random variables and, for a given n, models for $M_n = \max(Y_1, \ldots, Y_n)$.
- If the distribution of the Y_i is specified, the exact distribution of M_n is known.
- Extreme value theory considers the existence of $\lim_{n\to\infty} Pr((M_n - b_n)/a_n \leq y) \equiv F(y)$ for two sequences of real numbers $a_n > 0, b_n$.
- If $F(y)$ is nondegenerate, it belongs to either the Gumbel, the Fréchet or the Weibull class of distributions, which can all be usefully expressed under the umbrella of the GEV:

$$G(y; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

  for $\{y : 1 + \xi(y-\mu)/\sigma > 0\}$.
- Here, $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and $\xi \in \mathbb{R}$ is the shape parameter. The "residual" $V = (Y - \mu)/\sigma$ follows a GEV$(0, 1, \xi)$.

cont.

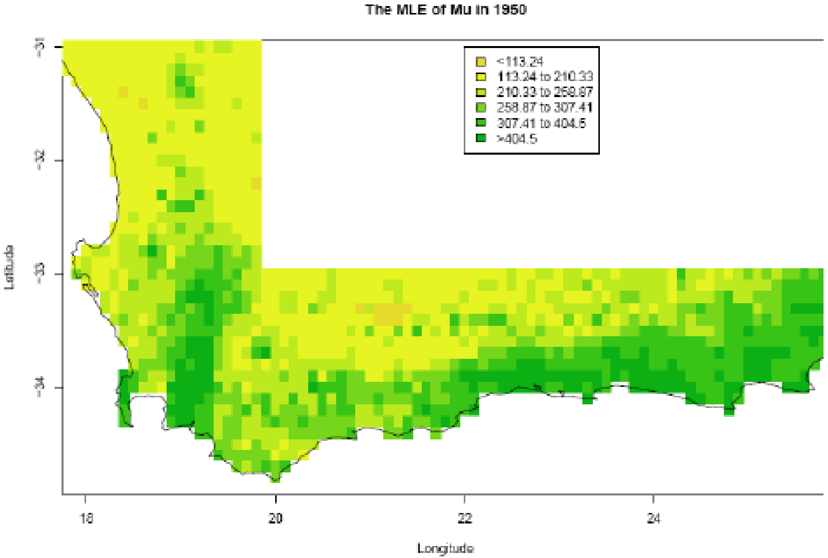
- Let $Z = (1 + \xi V)^{\frac{1}{\xi}} \Leftrightarrow V = \frac{Z^{\frac{1}{\xi}} - 1}{\xi}$. Then Z follows a standard Fréchet distribution, with distribution function $\exp(-z^{-1})$, i.e., an $IG(1, 1)$.)
- Note that n is not specified; the GEV is viewed as an approximate distribution to model the maximum of a sufficiently long sequence of random variables.
- The GEV distribution is a first stage model for annual precipitation maxima
- We specify μ, σ, and ξ at the second stage to reflect underlying spatio-temporal structure.

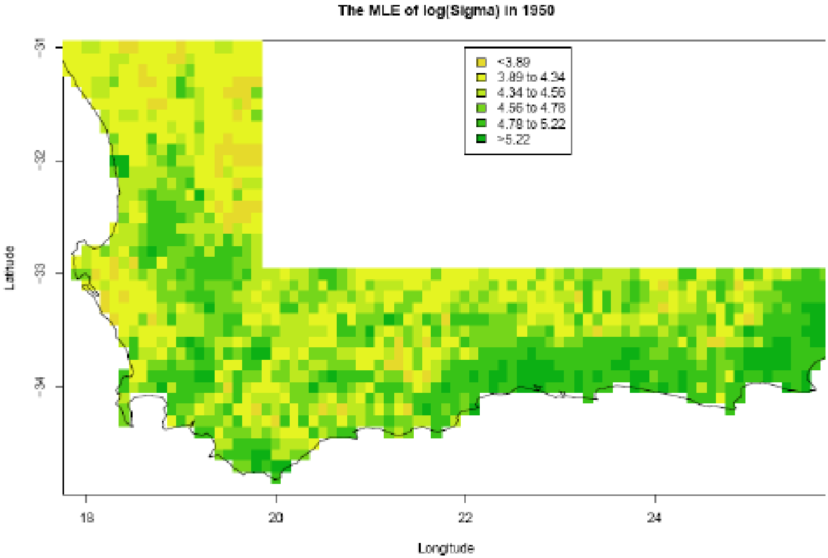
cont.

- Working with grid cells, let $Y_{i,t}$ denote the annual maximum of daily rainfall for cell i in year t.
- We assume the $Y_{i,t}$ are conditionally independent, each following a GEV distribution with parameters $\mu_{i,t}$, $\sigma_{i,t}$, and $\xi_{i,t}$, respectively.
- Focus on specification of the models for $\mu_{i,t}$, $\sigma_{i,t}$ and $\xi_{i,t}$.
- Conditional independence implies interest in smoothing the surfaces around which the interpolated data is centered rather than smoothing the data surface itself.
- Perhaps OK at annual time scale and at 10km grid cell resolution

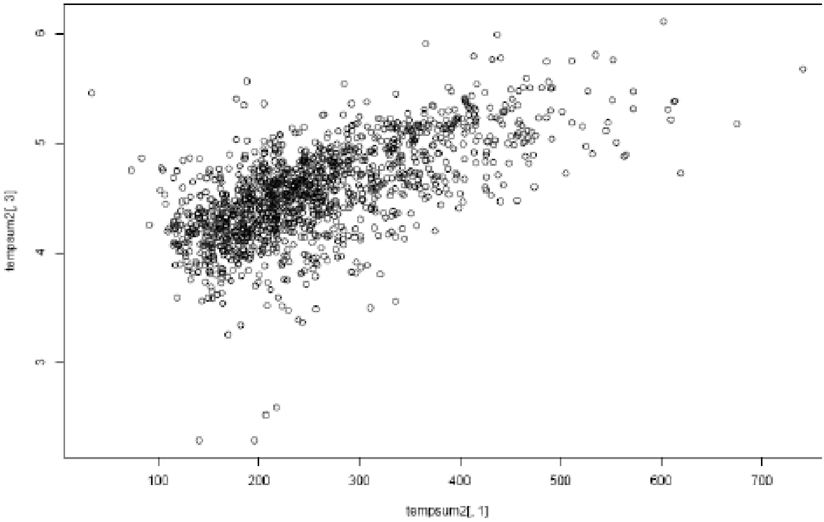
Modeling details

- Exploratory analysis for the CFR precipitation data suggests $\xi_{i,t,} = \xi$ for all i and t and $\sigma_{i,t} = \sigma_i$ for all t
- So, modeling focuses on the $\mu_{i,t}$.
- In addition, we want the $\mu_{i,t}$ and σ_i to be dependent at the same site.
- We adopt the coregionalization approach, making a random linear transformation of independent conditionally autoregressive (CAR) models


The MLE of Mu in 1950
<113.24
113.24 to 210.33
210.33 to 258.87
258.87 to 307.41
307.41 to 404.5
>404.5
-31
-32
-33
-34
Latitude
18
20
22
24
Longitude


The MLE of log(Sigma) in 1950
<3.89
3.89 to 4.34
4.34 to 4.56
4.56 to 4.78
4.78 to 5.22
>5.22
Latitude
-31
-32
-33
-34
Longitude
18
20
22
24

Scatter plot of log(Sigma) against Mu in 1950



Modeling

- For modeling the $\mu_{i,t}$, many options. With spatial covariates $\mathbf{X}_i$, a regression model with random effects, $[\mu_{i,t}|\boldsymbol{\beta}, W_{i,t}, \tau^2] = N(\mathbf{X}_i^T \boldsymbol{\beta} + W_{i,t}, \tau^2)$.
- For example, the $\mathbf{X}_i$ could include altitude or a trend surface while $W_{i,t}$ is a spatio-temporal random effect.
- Possibilities for modeling $W_{i,t}$ include:
  - (i) an additive form, **Model A**: $W_{i,t} = \psi_i + \delta_t$, $\delta_t = \phi\delta_{t-1} + \omega_t$, where $\omega_t \sim N(0, W_0^2) \quad i.i.d$;
  - (ii) a linear form in time with spatial random effects, **Model B**: $W_{i,t} = \psi_i + \rho(t - t_0)$;
  - (iii) a linear form in time with local slope, **Model C**: $W_{i,t} = \psi_i + (\rho + \rho_i)(t - t_0)$;
  - (iv) a multiplicative form in space and time, **Model D**: $W_{i,t} = \psi_i\delta_t$, $\delta_t = \phi\delta_{t-1} + \omega_t$, where $\omega_t \sim N(0, W_0^2) \quad i.i.d.$

Dependence?

- Again, we want dependence between location and scale parameters in the GEV model.
- In models A, B, and D, we do this by specifying $\log \sigma_i$ and ψ_i to be dependent. We work with $\sigma_i = \sigma_0 \exp(\lambda_i)$ and a coregionalized bivariate CAR model.
- In model C, we specify $\log \sigma_i, \psi_i$ and ρ_i to be dependent and use a coregionalized trivariate CAR specification.

cont.

- In the foregoing models, temporal evolution in the extreme rainfalls is taken into account in the model for $W_{i,t}$. This enables prediction for any grid cell for any year.
- Model comparison for the CFR precip data: Posterior medians are adopted as the point estimates of the predicted annual maxima because of the skewness of the predictive distributions.
- Predictive performance by computing the averaged absolute predictive errors (AAPE) for each model. Also empirical vs. nominal (.95%) coverage and the deviance information criterion (DIC).
- Using these criteria, Model C emerged as best

Using 1999 as hold-out data

	$\bar{D}$	p_D	DIC	AAPE	AAD	$\hat{r}$
Model A	788778	9115	797892	84.6	110.5	0.985
Model B	790575	8884	799459	83.7	104.7	0.947
Model C	789392	9212	798604	77.9	103.1	0.945
Model D	803382	9201	812583	81.8	118.9	0.968

Table: DIC, AAPE and $\hat{r}$ for Models (A) - (D).

Model C: $W_{i,t} = \mu_i^* + (\rho + \rho_i)(t - t_0)$

A Continuous spatial process model

- Again, we focus on spatial extremes for precipitation events.
- Now, data are annual maxima of daily precipitations derived from station records at 281 sites in South Africa in 2006.
- With point-referenced station data, spatial interpolation for the predictive distribution of the extreme value at unmonitored locations.
- We modify the previous development to point-referenced extreme values to achieve multi-scale dependence along with spatial smoothing for realizations of the surface of extremes.
- Relax the first stage conditional independence assumption
- A (mean square) continuous spatial process model for the actual extreme values for spatial dependence which is unexplained by the latent spatial specifications for the GEV parameters.
- Range of space-time dependence at second stage is longer than that for first stage spatial smoothing

Modeling details

- Formally, the first stage of the hierarchical model can be written as:

$$Y(\mathbf{s}) \;=\; \mu(\mathbf{s}) + \frac{\sigma(\mathbf{s})}{\xi(\mathbf{s})}(Z(\mathbf{s})^{\xi(\mathbf{s})} - 1)$$

  where $Z(\mathbf{s})$ follows a standard Fréchet distribution.
- We view $Z(\mathbf{s})$ as the "standardized residual" in the first stage GEV model.
- Conditional independence $\Leftrightarrow$ $Z(\mathbf{s})$ are $i.i.d.$
- So, even if the surface for each model parameter is smooth, realizations of the predictive surface will be everywhere discontinuous
- Instead, to create spatially dependent $Z(\mathbf{s})$, transform a GP to a spatial standard Fréchet process

Finally

- We assume that the $Z(\mathbf{s})$ follow a standard Fréchet process
- Specification of the second stage, to facilitate model fitting is to assume there is spatial dependence for the $\mu(\mathbf{s})$ but that $\sigma(\mathbf{s})$ and $\xi(\mathbf{s})$ are constant across the study region.
- The data is not likely to be able to inform about processes for $\sigma(\mathbf{s})$ and for $\xi(\mathbf{s})$.
- Suppose $\mu(\mathbf{s}) = X(\mathbf{s})^T\beta + W(\mathbf{s})$. $X(\mathbf{s})$ is the site-specific vector of potential explanatory variables. The $W(\mathbf{s})$ are spatial random effects, a zero-centered Gaussian process
- Plugging in the model for $\mu(\mathbf{s})$ we obtain

$$Y(\mathbf{s}) = X^T(\mathbf{s})\beta + W(\mathbf{s}) + \frac{\sigma}{\xi}(Z(\mathbf{s})^{\xi} - 1)$$

- Two sources of spatial dependence, the $Z(\mathbf{s})$ and the $W(\mathbf{s})$ with two associated scales of dependence.

Finally

- With the foregoing annual maxima of daily rainfalls from the station data in South Africa
- Fitting 200 monitoring sites, with both the conditionally independent first stage specification and the smoothed first stage
- The smoothed first stage specification was superior, achieving 20% reduction in AAPE and more accurate empirical coverage.

Performance of the model in (6.1) and the model in (6.2) using averaged absolute predictive errors ($AAPE$) and the empirical coverage probability $\hat{r}$.

	$AAPE$	$\hat{r}$
Smoothed GEV	2.6474	0.955
Nonsmoothed GEV	3.5306	0.966