Extreme Value Analysis and Ventures into Space and Time

15 July 2013

Weather and Climate Impacts Assessment Science Program
http://www.assessment.ucar.edu/

Eric Gilleland
Joint Numerical Testbed
Research Applications Laboratory
National Center for Atmospheric Research

Next Generation Climate Data Products Workshop
July 15 - 19, 2013

Hosted by the National Center for Atmospheric Research
Boulder, Colorado

Extreme Value Analysis

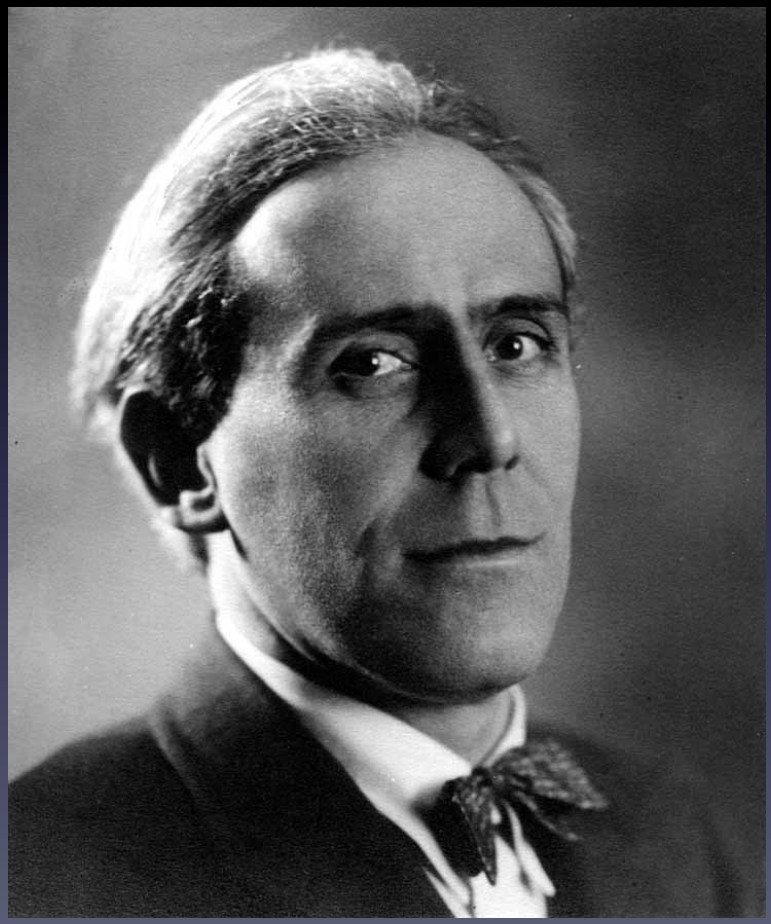
Colorado Lottery

POWER PLAY PAYOUT TABLE

MATCH	PRIZE	X2	X3	X4	X5
●●●●●●	Jackpot	POWER PLAY does not apply.			
●●●●●	$200,000	$1,000,000*			
●●●●●	$10,000	$20,000	$30,000	$40,000	$50,000
●●●●	$100	$200	$300	$400	$500
●●●●	$100	$200	$300	$400	$500
●●●	$7	$14	$21	$28	$35
●●●	$7	$14	$21	$28	$35
●●	$4	$8	$12	$16	$20
●	$3	$6	$9	$12	$15

Extreme Value Analysis

"Il est impossible que l'improbable n'arrive jamais"
--Emil Gumbel



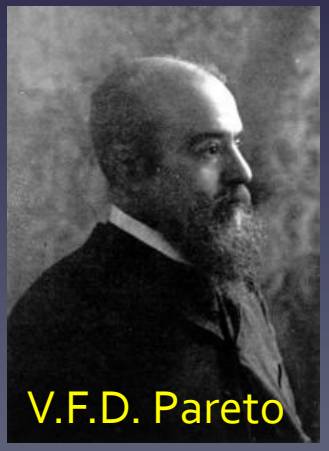
Extreme Value Analysis

$$\Pr\{X - u > x | X > u\}$$

$$\approx \left[1 + \frac{\xi x}{\sigma}\right]^{-1/\xi}, \; x > 0,$$

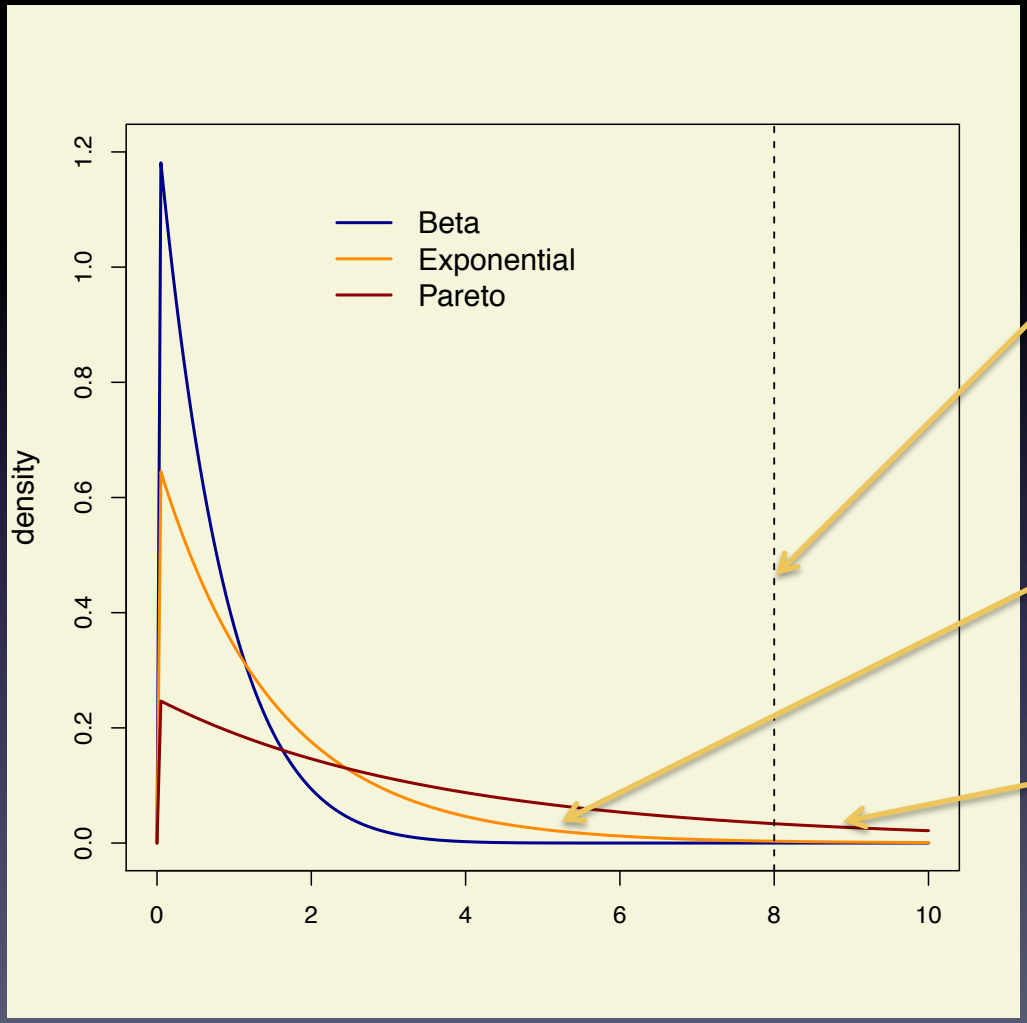
for sufficiently large u

Generalized Pareto Distribution


V.F.D. Pareto

Extreme Value Analysis

Three types of Extreme Value df's



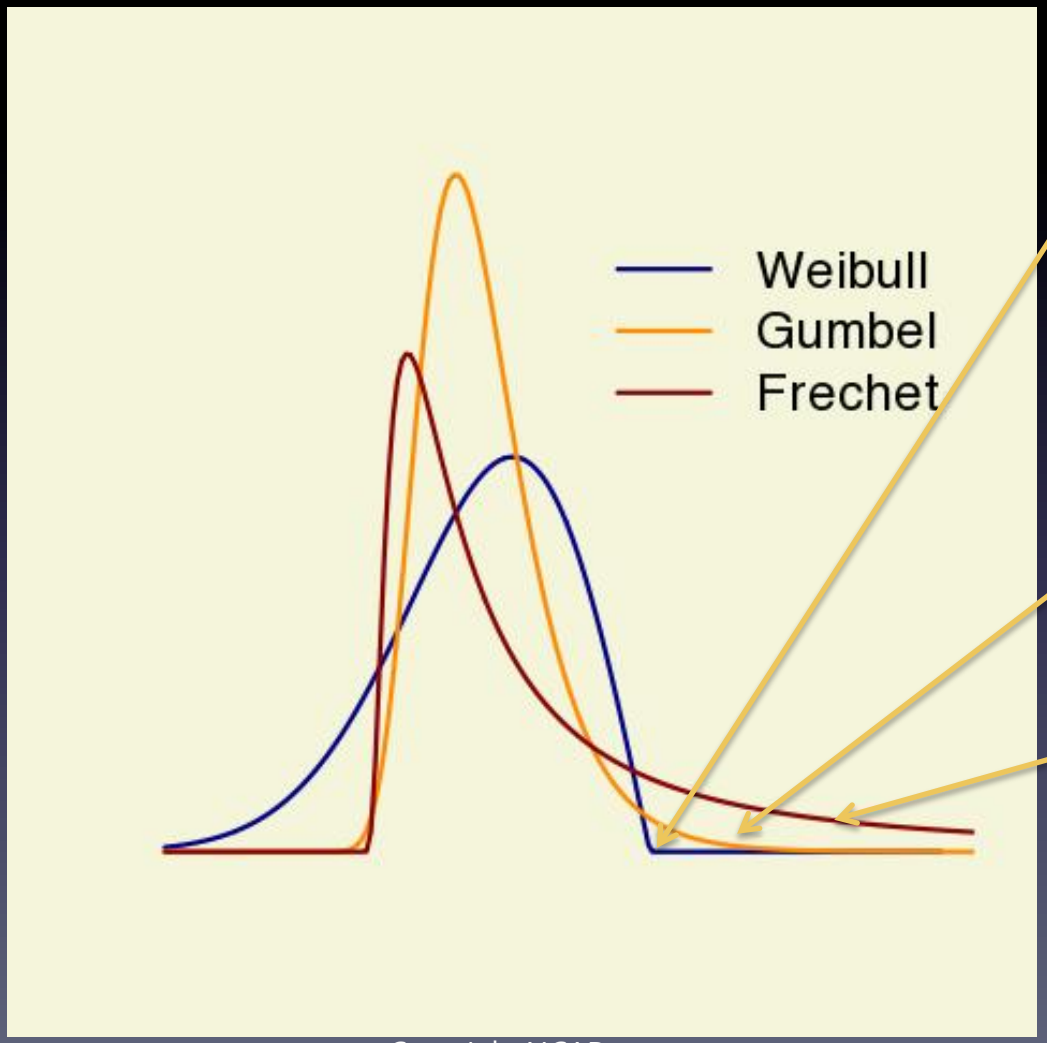
Beta df; bounded upper tail

Exponential; light tail

Pareto; heavy tail

Extreme Value Analysis

Three types of Extreme Value df's



Weibull df; bounded upper tail

Gumbel df; light tail

Fréchet df; heavy tail

Extreme Value Analysis

Paula Radcliffe, 11.6 mph world marathon record
London Marathon, 13 April 2003

Predicted Speed Limits

Thoroughbreds (Kentucky Derby)	**≈ 38 mph**
Greyhounds (English Derby)	≈ 38 mph
Men (100 m distance)	≈ 24 mph
Women (100 m distance)	≈ 22 mph
Women (marathon distance)	≈ 12 mph
Women (marathon distance using a different model)	≈ 11.45 mph

Denny, M.W., 2008: Limits to running speed in dogs, horses and humans. *J. Experim. Biol.*, **211**:3836–3849.

Extreme Value Analysis

- Poisson-GP (frequency of occurrence-intensity above the threshold)
- Point Process (PP) Representation: view as points in two-dimensional space
- Latter subsumes former

Relation of $\mathrm{GEV}(\mu,\sigma,\xi)$ to those of a $\mathrm{PP}(\lambda,\sigma^*,\xi)$

(i) ξ is identical,

(ii) $\ln\lambda = -\frac{1}{\xi}\ln\left(1+\xi\frac{u-\mu}{\sigma}\right),$

(iii) $\sigma^* = \sigma + \xi(u-\mu)$

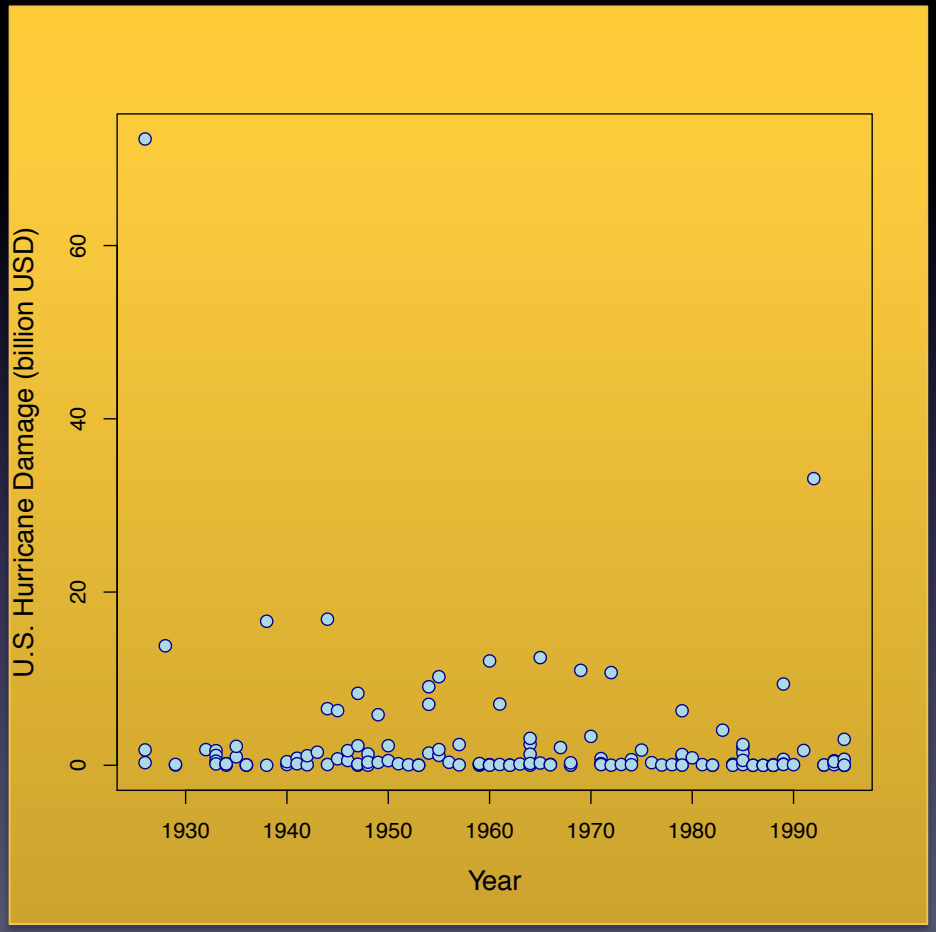
Extreme Value Analysis

- Estimation can be performed in many ways, but maximum likelihood is most common, and must be done numerically
- Gumbel/Exponential types are single points in a continuous parameter space
- Block/Threshold selection is a trade-off between low variance and low bias
- Usually interest is in return levels (often that exceed the range of the data), which can easily be obtained by inverting the GEV or GP df
- Non-stationary series can be modeled by incorporating covariates (e.g., time) in the GEV/GP parameters
- If the threshold excesses of the log transformation of a variable are max-stable (i.e., follow a GP df in the limit) , then the untransformed variable is super-heavy tailed, and its excesses do not have a non-degenerate df in the limit.

Extreme Value Analysis

Example

Economic damage from hurricanes in the United States from 1925 to 1995



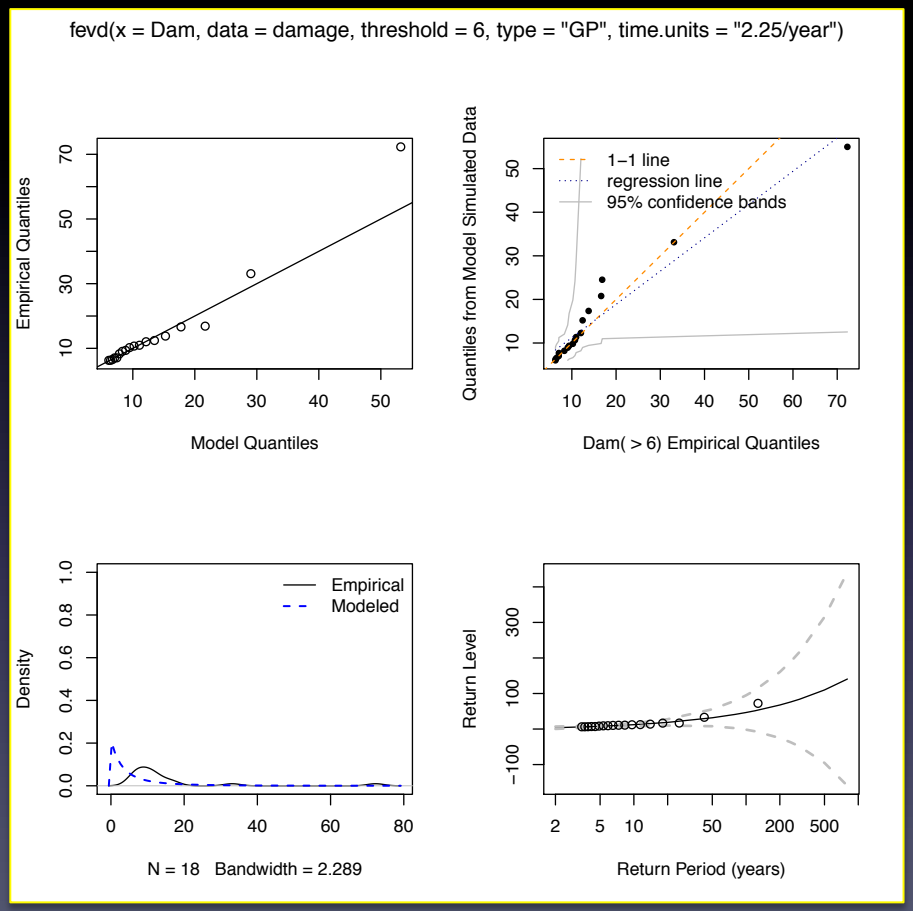
Extreme Value Analysis

Example

Economic damage from hurricanes in the United States from 1925 to 1995

GP df fit to damage data.

Return levels estimated using an approximation of 2.25 hurricanes per year.



Extreme Value Analysis

Example

Economic damage from hurricanes in the United States from 1925 to 1995

$$\Pr\{\text{damage} > 20 \text{ billion USD}\} \approx 0.16$$

$$\Pr\{\text{damage} > 60 \text{ billion USD}\} \approx 0.02$$

$$\Pr\{\text{damage} > 100 \text{ billion USD}\} \approx 0.01$$

Extreme Value Analysis

Multivariate Extreme Value Analysis

Without loss of generality, assume a standard max-stable df (can always transform to one). For simplicity, consider the bivariate case for variables X and Y.

Model is for:

$$\mathbf{M}_n = \left(\max\left\{ X_1, \ldots, X_n \right\} / n, \max\left\{ Y_1, \ldots, Y_n \right\} / n \right)$$

Extreme Value Analysis

Multivariate Extreme Value Analysis

Model is for:

$$\mathbf{M}_n = \left(\max\left\{X_1,\ldots,X_n\right\}/n, \max\left\{Y_1,\ldots,Y_n\right\}/n\right)$$

Analogous to the univariate setting, if there is a non-degenerate limiting df, *G*, then it must have the form:

$$G(x,y) = \exp\left\{-V(x,y)\right\}, x>0, y>0$$

where

$$V(x,y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w), \text{ with}$$

$$\int_0^1 w dH(w) = 1/2$$

Extreme Value Analysis

Copula

$$\Pr\{X_1 \le x_1, \ldots, X_d \le x_d\} = C\{F_1(x_1), \ldots, F_d(x_d)\}$$

- Separates the joint df into *d* marginal dfs and a joint df, *C*, called the copula, for the variables on a common marginal df.
- Dependence is modeled by the copula, *C*.
- Must choose *C*.

Extreme Value Analysis

$$\Pr[T_A<1, T_B<1] = \Phi_2(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)), \gamma)$$

Recipe for Disaster: The formula that Killed Wall Street

Wired Magazine, 2/23/2009, by Feliz Salmon

Spatial EVA

- Incorporate spatial information into parameter estimates
  - Vector Generalized Additive Models (VGAM)
  - Bayesian Hierarchical Modeling (BHM)
- Copula Models
- max-stable processes
- BHM + max-stable process
- More …

Recent Reviews

Turkman et al., 2010, *Extremes*, **13**, 375 – 397.

Davison et al., 2012, *Statistical Science*, **27** (2), 161 – 186.

Conditional EVA Model

Note: This is from the Heffenan and Resnick (2007, *Annals of Applied Probability*, **17**, 537 – 571) formulation of the problem.

$$\Pr\left\{ Y - u > y, \frac{X - a(Y)}{b(Y)} \le z \middle| Y > u \right\} \underset{u \to \infty}{\to} \exp(-y) G(z)$$

Heffernan and Tawn (2004, *JRSS B*, **66**, 497 – 546) found that for a wide class of copula models, the form of the sequences of functions *a* and *b* fell into a simple class for positively associated variables (and a slightly more complicated form for negatively associated random variables). Namely,

$$a(y) = \alpha y \text{ and } b(y) = y^{\beta}$$
$$(\alpha, \beta) \in [0,1] \times (-\infty, 1)$$

Conditional EVA Model

Note: This is from the Heffenan and Resnick (2007, *Annals of Applied Probability*, **17**, 537 – 571) formulation of the problem.

$$\Pr\left\{Y-u>y, \frac{X-a(Y)}{b(Y)} \le z \middle| Y>u\right\} \underset{u\to\infty}{\to} \exp(-y)G(z)$$

Heffernan and Tawn (2004, *JRSS B*, **66**, 497 – 546) found that for a wide class of copula models, the form of the sequences of functions *a* and *b* fell into a simple class for positively associated variables (and a slightly more complicated form for negatively associated random variables). Namely,

$$a(y)=\alpha y \text{ and } b(y)=y^{\beta}$$
$$(\alpha,\beta)\in[-1,1]\times(-\infty,1)$$

Using Laplace marginals! (obviates the added complexity for negatively associated r.v.'s)

Conditional EVA Model

A consequence of the result that leads to a way to estimate the model (semi-parametrically) is that:

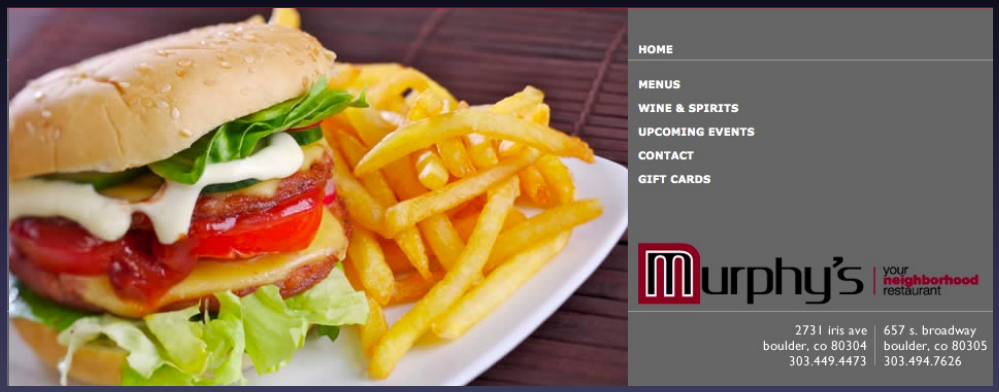
$$X_{|Y>u} = \alpha Y + Y^{\beta} Z_{|Y>u}$$

$$Z \perp Y, Z \sim G$$

No simple closed-form expression for *G*.

Conditional EVA Model

Estimation and Inference



"People who love sausage and respect the law should never watch either one being made."

--Murphy's Law

Conditional EVA Model

"…there is the method of estimation itself. The authors use maximum likelihood for estimating the generalized Pareto margins, Gaussian estimation for the conditional means and standard deviations and pseudolikelihood estimation for combining the various conditional distributions into a multivariate family, a veritable witches' soup of estimation methods, all nicely stirred up with the bootstrap as seemingly the only means of keeping control of all the estimation errors. Although I applaud the authors' eclecticism, would it not be better to have a more coherent estimation strategy?"

--Richard L. Smith
(Discussion on Heffernan and Tawn, 2004)

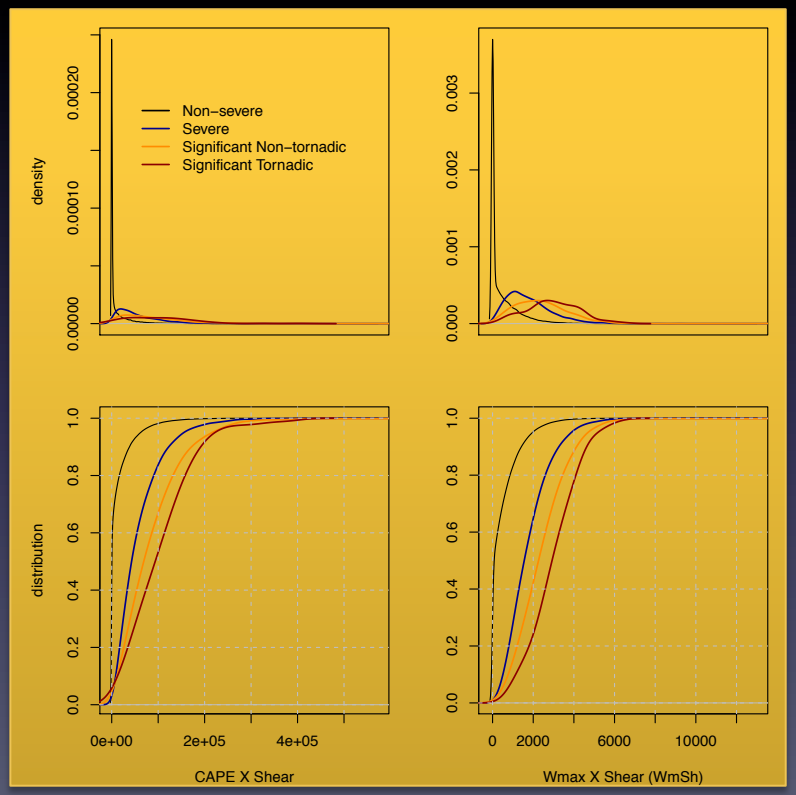
Conditional EVA Model

Example: maximum vertical wind (Wmax, m/s) and 0 – 6 km shear (Shear, m/s)

$$\text{Wmax} = \sqrt{2\text{CAPE}}$$

convective available potential energy (CAPE) J/kg

CAPE * Shear

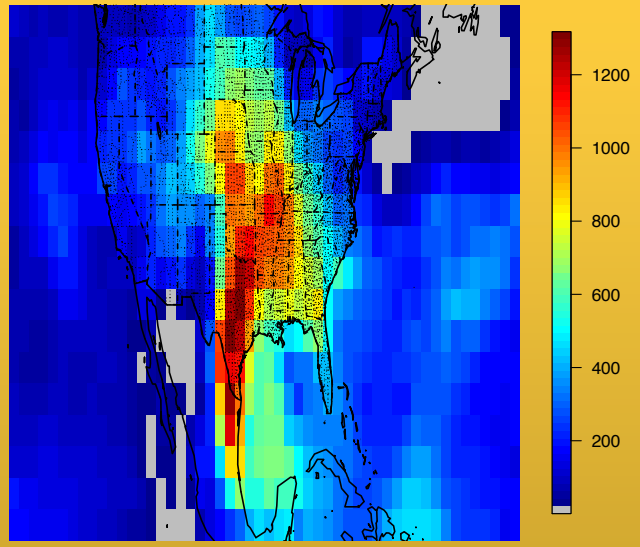


Wmax * Shear = WmSh

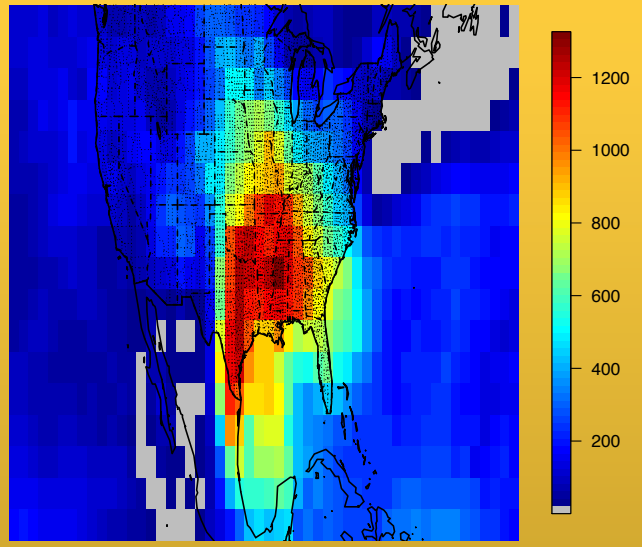
Conditional EVA Model

$$\left[\mathrm{WmSh}_1, \ldots, \mathrm{WmSh}_n \middle| \text{High field energy} \right]$$

Conditioned on Field Sum



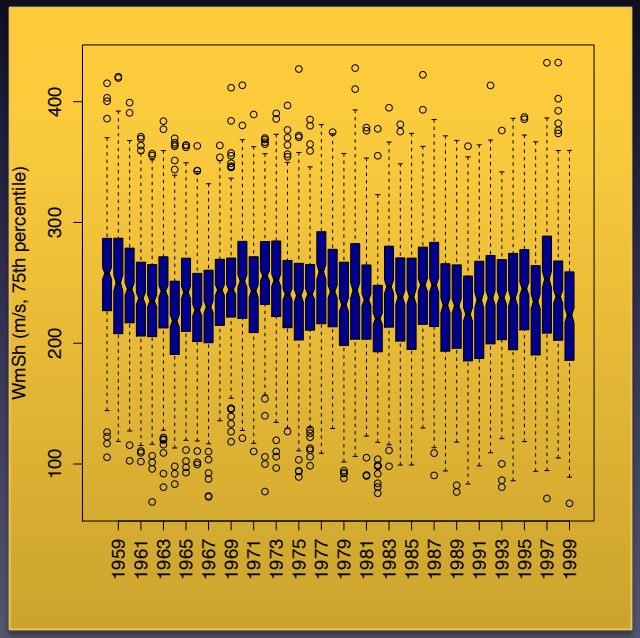
Conditioned on upper quartile

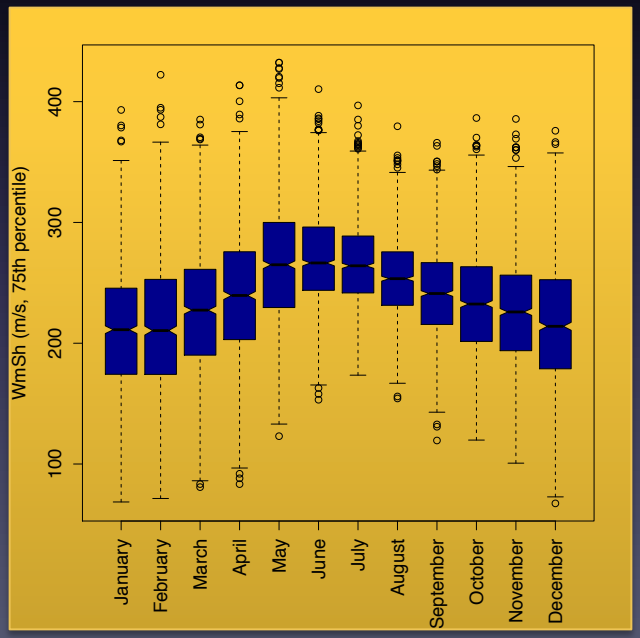


Conditional EVA Model

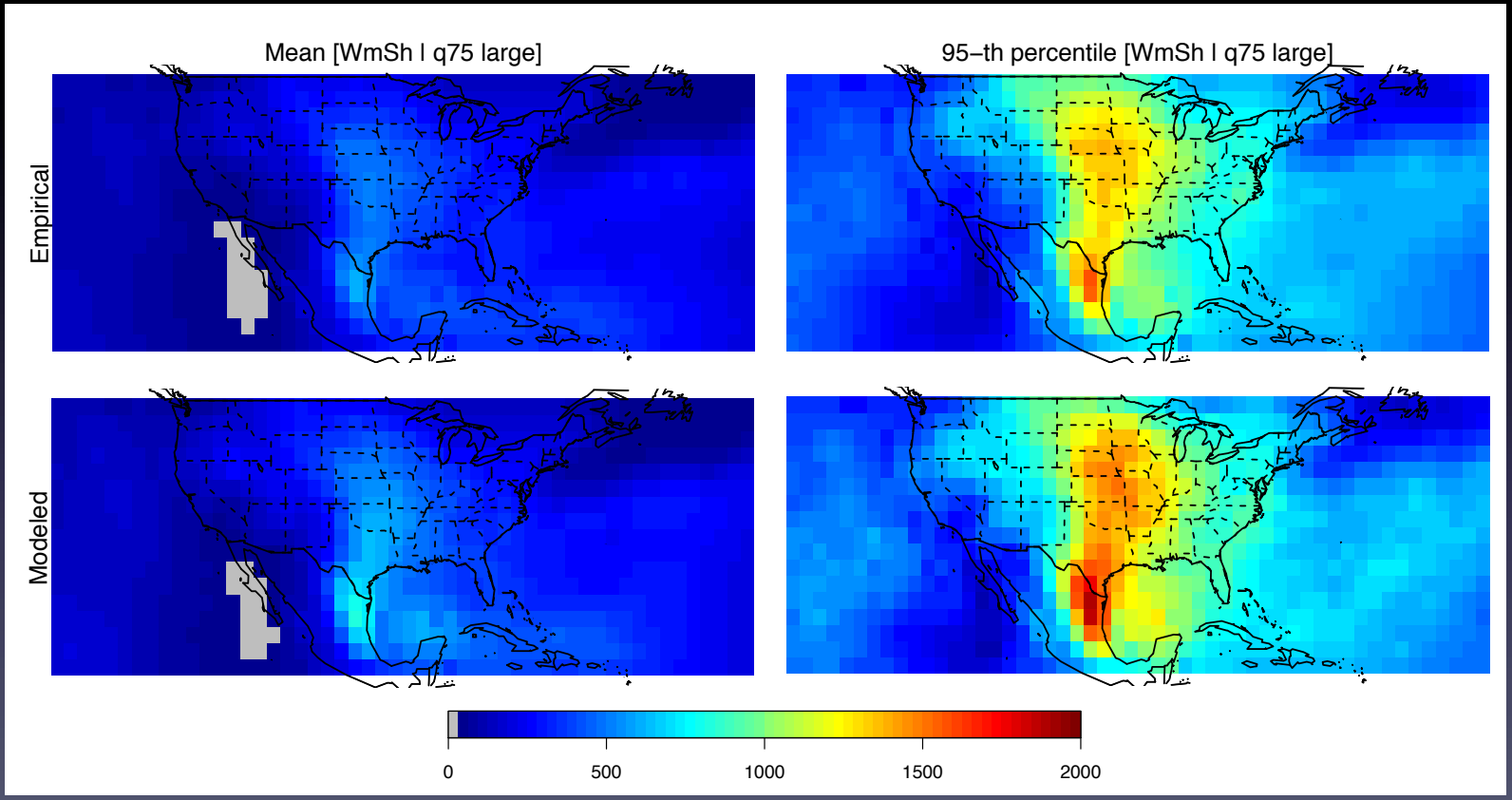
Upper quartile over space

- univariate time series
- measure of high WmSh over, possibly a small area of, space.
- Call it q75.

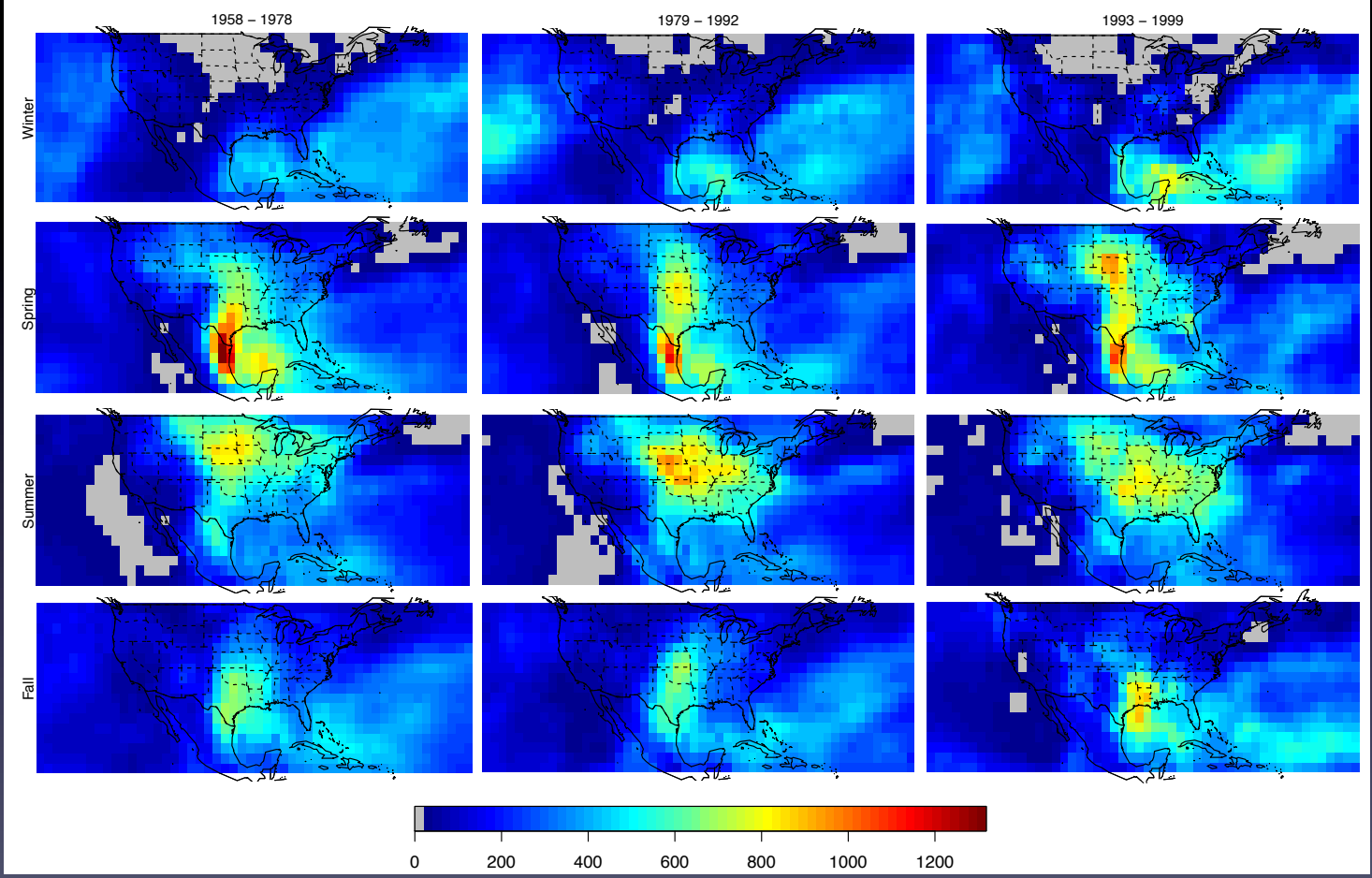




Conditional EVA Model

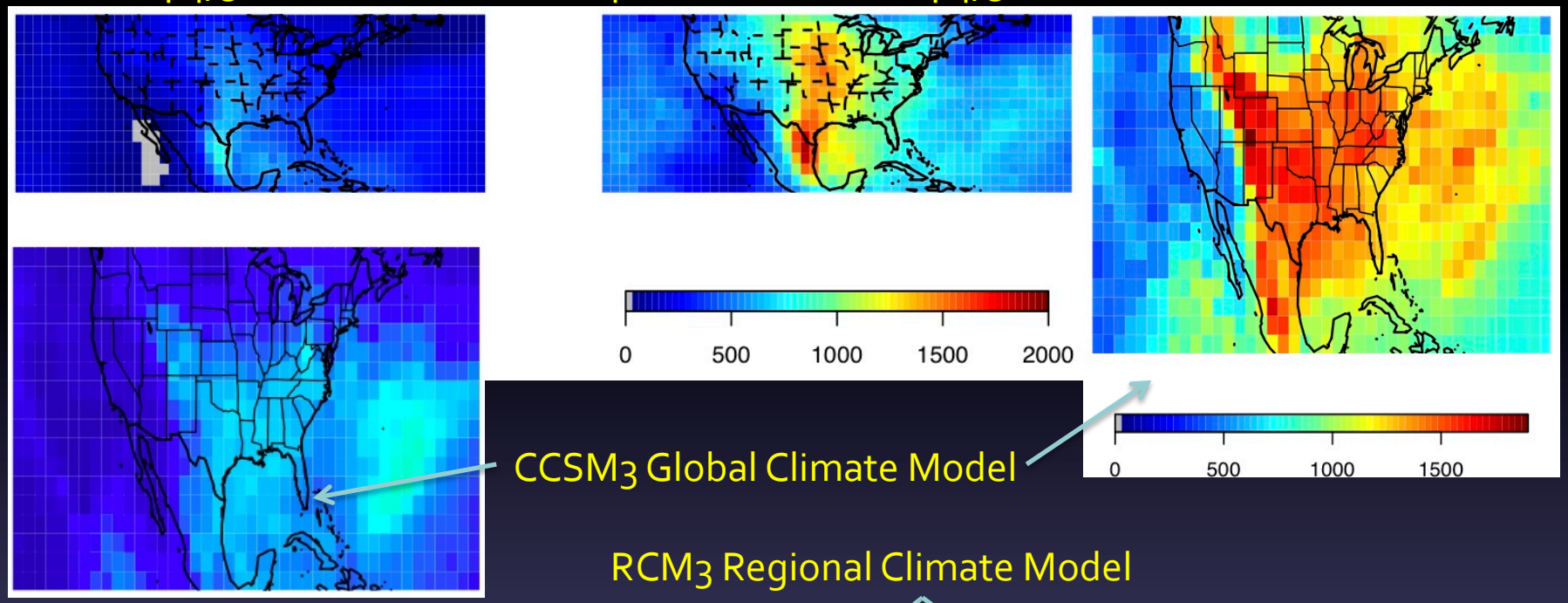


Conditional EVA Model



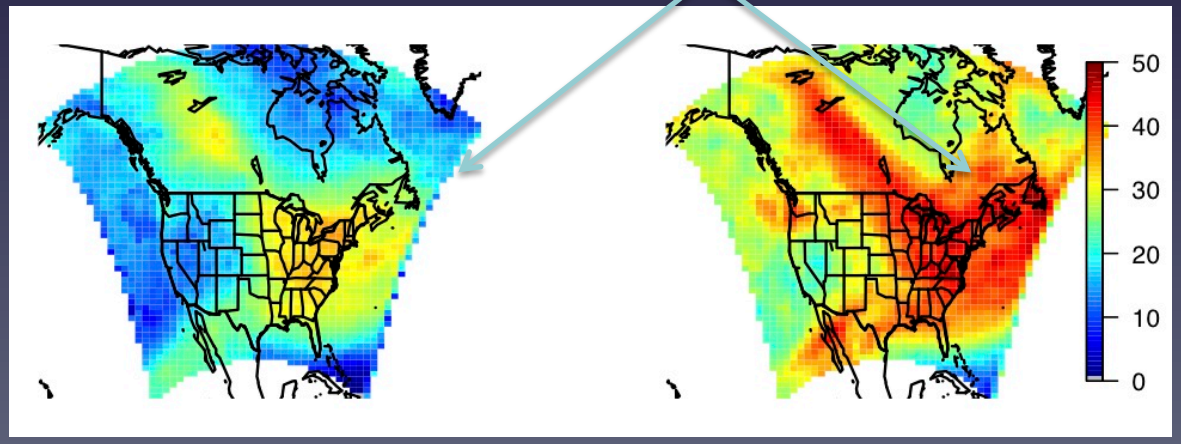
Observed Reanalysis Mean [WmSh | q75 > u]

Observed Reanalysis 95-th percentile [WmSh | q75 > u]



CCSM3 Global Climate Model

RCM3 Regional Climate Model



Data Products / Modeling tools

needed to advance climate/weather data analytics or impact studies

- High resolution!
- More informative data for hurricanes
- Severe storm environment variables
- long historic record
- observational uncertainty information/ensembles for models
- Data base with multiple variables along with location and time information (preferably at the same times)

Questions?

"We are now cruising at a level of two to the power of twenty-five thousand to one against and falling, and we'll be restoring normality just as soon as we are sure what is normal anyway."

--Douglas Adams

Hitchhiker's Guide to the Galaxy

[concerning the Infinite Improbability Drive]