Sparse-GEV: Sparse Latent Space Model for Multivariate Extreme Value Time Series Modeling

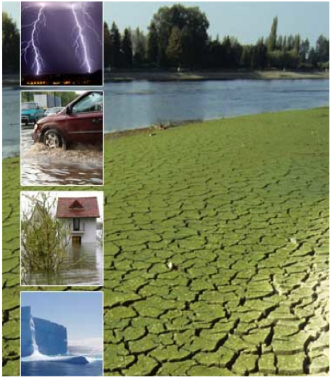
Yan Liu

Computer Science Department
Viterbi School of Engineering
University of Southern California

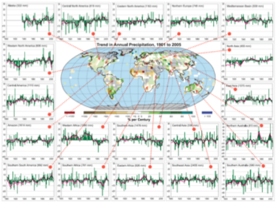
Second Workshop on Understanding Climate Change from Data

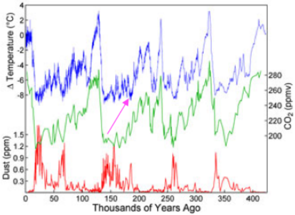
August 6, 2012

Climate Change: One of the Most Critical Issues in the 21st Century

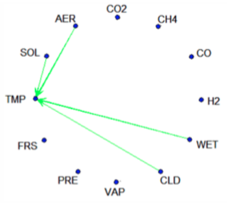


Discovery of Temporal Dependence Relationships





Output: temporal causal graph of climate forcing agents



Granger Causality

Main Idea: Cause is prior to Effect ⇒ Past values of the cause should help prediction of the Future values of the Effect.

$$x_t = \sum_{l=1}^{L} a_l x_{t-l} \tag{1}$$

$$x_t = \sum_{l=1}^{L} a'_l x_{t-l} + \sum_{l=1}^{L} b'_l y_{t-l}, \tag{2}$$

If Eq. (2) is a significantly better model than Eq. (1), we determine that time series y Granger causes time series x.

Lasso-Granger [Arnold et al, KDD 2007]: To achieve superior accuracy and scalability:

$$\min_{\{\mathbf{a}_i\}} \sum_{t=L+1}^{T} \left\| x_t^{(i)} - \sum_{j=1}^{P} \mathbf{a}_{i,j}^\top \mathbf{x}_{t,Lagged}^{(j)} \right\|_2^2 + \lambda \|\mathbf{a}_i\|_1 , \tag{3}$$

$\mathbf{x}_{t,Lagged}^{(j)} = \left[x_{t-L}^{(j)}, \ldots, x_{t-1}^{(j)} \right]$.

Challenges in Practical Applications and Theory

Our work to address practical challenges:

- Non-stationary time series [KDD 2009]
- Natural grouping of time series [KDD 2009]
- Spatial time series [KDD 2009]
- Nonlinear time series [AAAI 2010]
- Relational time series [ICML 2010]
- Irregular time series [SDM 2012]
- Extreme-value time series [ICML 2012]
- Hidden variables [Climate Informatics Workshop 2012]

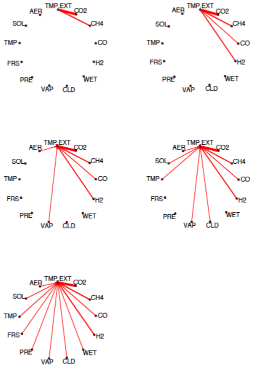
Our work to address theory challenges:

- Granger causality versus true causality [In submission]

Spatial time series: *elastic net* [KDD 2009]

$$\hat{\beta} = \arg\min_{\beta} \sum_{s\in S} \sum_{t=L+1}^{T} \left(x^i_{t,s} - \sum_{j=1}^{N}\sum_{l=1}^{L}\sum_{\omega\in\Omega} \beta^j_{l,\omega} x^j_{t-l,s+\omega}\right)^2 + \lambda_2 \underbrace{\sum_{j=1}^{N}\sum_{l=1}^{L} (\beta^j_{l,:})^T \tilde{\Delta}_j \beta^j_{l,:}}_{\text{SpatialPenalty}} + \lambda_1 \underbrace{\sum_{j=1}^{N} \|\beta^j_{:}\|_{\Delta_j}}_{\text{SparsityPenalty}}$$

Variables (Variable group)	**Type**	**Source**
Methane (CH_4)	Greenhouse Gases	NOAA
Carbon-Dioxide (CO_2)		
Hydrogen (H_2)		
Carbon-Monoxide (CO)		
UV (AER)	Aerosol Index	NASA
Temperature (TMP)	Climate	CRU
Temp Range (TMP)		
Temp Min (TMP)		
Temp Max (TMP)		
Precipitation (PRE)		
Vapor (VAP)		
Cloud Cover (CLD)		
Wet Days (WET)		
Frost Days (FRS)		
Global Horizontal (SOL)	Solar Radiation	NCDC
Direct Normal (SOL)		
Global Extraterrestrial (SOL)		
Direct Extraterrestrial (SOL)		
1-year return level for temperature extreme (TMP.EXT)	Climate	Estimated using temp from CDIAC



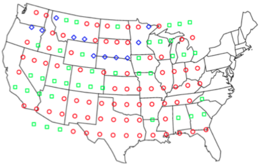
Relational time series: *hidden Markov random fields* [ICML 2010]

$$P_\lambda(X^{(1)}, \ldots, X^{(M)}, s^{(1)}, \ldots, s^{(M)} | \boldsymbol{\beta}, w) = \frac{1}{Z} \prod_{i=1}^{M} \Phi_\lambda(X^{(i)} | \beta^{(s^{(i)})}) \prod_{(i,j) \in G_r} \Phi(s^{(i)}, s^{(j)} | w),$$

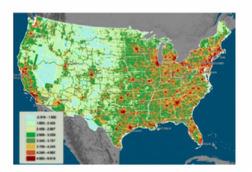
where $\Phi(s^{(i)}, s^{(j)} | w) = \exp(\sum_{s,s'} w_{ss'}, \delta_{ss'}(s^{(i)}, s^{(j)}))$, and

$$\Phi_\lambda(X^{(i)} | \beta^{(s^{(i)})}) \propto \prod_{t=L+1}^{N} \exp(-\frac{1}{2}(\vec{x}_t^{(i)} - \vec{x}_{(t-1)..(t-L)}^{(i)} \cdot \beta^{s^{(i)}})^T \cdot (\vec{x}_t^{(i)} - \vec{x}_{(t-1)..(t-L)}^{(i)} \cdot \beta^{s^{(i)}})) + \lambda \| \beta^{(s)} \|$$

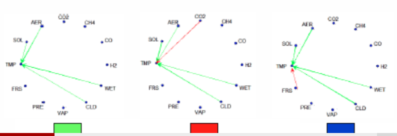
Clusters of US locations by our method
(number of clusters = 3)



Map of US CO2 Concentration
(http://www.purdue.edu/eas/carbon/vulcan/GEarth)



Causal graphs associated with each state

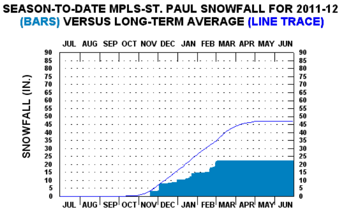


Motivation

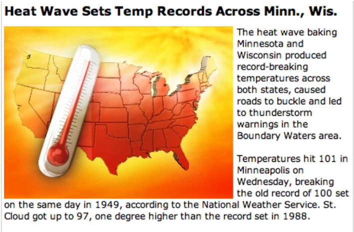
Climate Change: More frequent occurrences of extreme weather

Examples: Minneapolis in 2012



(a) Warm Winter with Little Snow



Heat Wave Sets Temp Records Across Minn., Wis.

The heat wave baking Minnesota and Wisconsin produced record-breaking temperatures across both states, caused roads to buckle and led to thunderstorm warnings in the Boundary Waters area.

Temperatures hit 101 in Minneapolis on Wednesday, breaking the old record of 100 set on the same day in 1949, according to the National Weather Service. St. Cloud got up to 97, one degree higher than the record set in 1988.

(b) Heatwave

Problem Definition

The Problem

Given Multiple extreme-valued time series $\left(x_{1:T}^{1:N}\right)$.

Goal Recover a temporal graph that represents the temporal dependencies between the time-series.

Challenges

- Heavy tail of the data
- Scarcity of the data

Sparse-GEV: Gumbel Distribution for the Observations

Proposed Model - Sparse-GEV

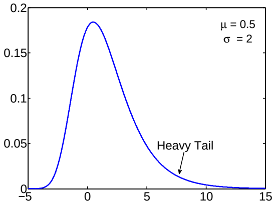
$$\boxed{p(\{x_t^i\}, \{\mu_t^i\} | \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{c}) = \prod_{i=1}^{P} \prod_{t=L+1}^{T} p(x_t^i | \mu_t^i, \sigma^i) p(\mu_t^i | \{\mu_{t-l}^j\}, \boldsymbol{\beta}, \boldsymbol{c}),}$$

We assume a simpler model: $\xi \to 0$. The result is the Gumbel distribution:

$$\mathcal{G}(x|\mu, \sigma) = \exp\left\{-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right\}.$$

Properties of the distribution:

- Connections to Exponential Family
- Heavy Tail ($\propto \exp(-x/\sigma)$)
- Computationally Challenging: Exponential terms in the log-likelihood



Sparse-GEV: The Latent Structure for the Parameters

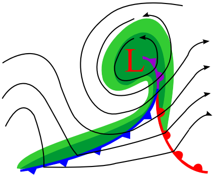
The temporal structure for μ_t^i:

$$\mu_t^i = f\left(\mu_{\text{Lagged}}^{\text{All Neighbours}}\right) + \epsilon.$$

In this work, the linear model:

$$\mu_t^i = c^i + \sum_{\ell=1}^{L} \sum_{j=1}^{P} \beta_{j,\ell}^i \mu_{t-l}^j + \epsilon_t$$

- c^i: Location specific bias
- ϵ_t: White Gaussian Noise
- $\beta_{j,\ell}^i$: *sparse* temporal dependency parameters.



Inference: EM Algorithm

Maximum Likelihood Solution:

$$\{\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{c}}\} = \arg\max \mathcal{L}(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^P; \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{c}) + \sum_{i=1}^{P} \lambda \|\boldsymbol{\beta}^i\|_1,$$

Solved via EM Algorithm.

- E-Step: Expectation computed using Particle Filtering.
- M-Step: Newton-Raphson for $\boldsymbol{\sigma}$ and Lasso solvers for $\boldsymbol{\beta}$ and $\boldsymbol{c}$.

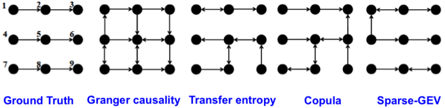
Prediction

$$\hat{x}^i_{T+1} = \bar{\mu}^i_{T+1} + \gamma_E \sigma^i,$$

where $\bar{\mu}^i_{T+1} = c^i + \sum_{l=1}^{L} \sum_{j=1}^{P} \beta^i_{j,l} \bar{\mu}^j_{T-l}$, and $\gamma_E \approx 0.5771$ is the Euler constant.

Experiment Results: Simulation Data

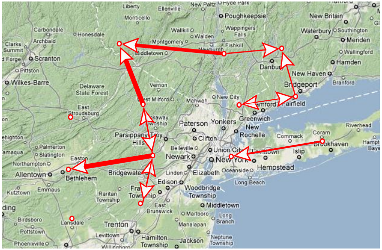
Graph Learning Accuracy (AUC)



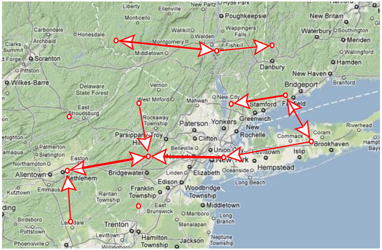
Algorithms	Avg AUC Score
Sparse-GEV	**0.9257**
Granger	0.9046
Transfer Entropy	0.8701
Copula	0.8836

The temporal dependence graph learned by Sparse-GEV on the extreme value time series of Wind in NY and Gust in NY. Thicker edges imply stronger dependency.

Wind Graph:



Gust Graph:



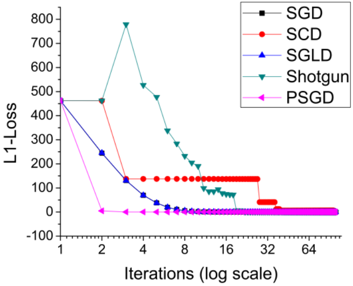
Experiment Results: Prediction Accuracy

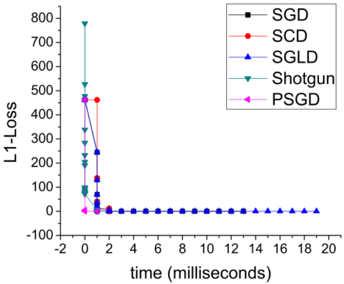
Prediction Accuracy (RMSE)

	Synth.	Wind	Gust
Sparse-GEV	**0.2644**	**0.0660**	**0.0927**
Granger	0.2923	0.0695	0.0943
TE	0.3135	0.0692	0.0983
Copula	0.2987	0.0678	0.0934

Scalability: Parallel Stochastic Optimization Algorithms

- **Stochastic Subgradient Langevin Dynamics** (*SGLD*) [Welling & Teh, 2011]: combining mini-batch stochastic subgradient descent and Langevin Dynamics.
- **Parallel Stochastic Coordinate Descent** (*Shotgun*) [Bradley et al, 2011]: a parallel implementation of Stochastic Coordinate Descent.
- **Parallel Stochastic Gradient Descent** (*PSGD*) [Zinkevich et al, 2010]: randomly partitioning the data, giving one partition to each processor, which sequentially uses each data point of its own partition to update β using a constant step size η.





Summary

One Important Task in Climate Change

- Discovery of Temporal Dependence between Extreme Value Time Series



- Sparse Latent GEV Models: an effective and efficient temporal point process model to capture sparse temporal dependence between extreme value time series

Acknowledgement

USC Melody Lab

Taha Bahadori

Huida Qiu

Yi Chang

IBM Research

Aurelie Lozano

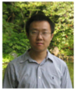
Naoki Abe

Hongfei Li

NEC Research

Alexandru Niculescu-Mizil

CMU



Xi Chen

Princeton

Han Liu

Thank you!