Spatial Temporal Statistical Modeling of Extremes and a Marginal Approach

Jun Yan

Joint work with Zhuo Wang, Xuebin Zhang

Department of Statistics, University of Connecticut

June 16, 2016 @ BIRS

Outline

Issues from Spatial Temporal Modeling of Extremes

- Extreme observations are often spatial temporal by nature as data are collected from monitoring stations or grid boxes over years.
- Spatial temporal extreme modeling has attracted much attention.
- Objectives of statistical analyses: estimation (parameters, return levels); prediction; risk assessment; simulation.
- Spatial temporal dependence is critical in statistical inferences: can be of primary interest in some applications but not all.
- Marginal approaches without fully dependence specification can be useful and efficient when dependence is nuissance.

Methodological Overview

- How do we introduce dependence in models?
  - Markov process (mostly in temporal setting, uni-directional)
  - Latent processes
  - Copula (ev-copula vs. non-ev-copula)
  - Full parametric specification: max-stable processes
  - No dependence specification (marginal approach)
- How do we handle dependence in inferences? This determines the pros and cons of different methods.
  - Likelihood or composite likelihood
  - Bayesian (possible with composite likelihood)
  - Working dependence

Univariate Margin: GEV Distribution

GEV distribution is the limit distribution of properly normalized sample maximum, with density

$$f(y;\, \mu, \sigma, \xi) = \frac{1}{\sigma} t(y)^{\xi+1} e^{-t(y)},$$

where

$$t(y) = \begin{cases} \left(1 + \left(\frac{y-\mu}{\sigma}\right)\xi\right)^{-1/\xi}, & 1 + \left(\frac{y-\mu}{\sigma}\right)\xi > 0, \quad \xi \neq 0; \\ e^{(y-\mu)/\sigma}, & \xi = 0. \end{cases}$$

The cumulative distribution function is

$$F(y;\, \mu, \sigma, \xi) = e^{-t(y)},$$

- μ, σ and ξ are the location, scale and shape parameters, respectively.
- ξ determines the tail behavior.
- covariates are often incorporated in μ.

EVA for Time Series Data: Marginal Approach

(A recent review on temporal extremes is Reick and Shaby (2016).)

- Two major inference goals: return level; serial dependence.
- Marginal approach (when return level is of primary interest): POT needs to account for serial dependence
  - declustering (e.g., Davison and Smith, 1990; Ferro and Segers, 2003)
  - working independence and sandwich variance correction (Smith, 1991; Fawcett and Walshaw, 2007)
- Remark: sandwich variance needs a moderately large sample size ($n = 10,000$ in simulation study of Fawcett and Walshaw (2007)).

Modeling Serial Dependence

- Markov chain model (Smith et al., 1997; Fawcett and Walshaw, 2006): conditional distribution constructed from bivariate ev distribution and univariate GEV distribution.
- Markov-switching model (Shaby et al., 2016) with two states (bulk, extreme)
- Max-stable process
  - Hierarchical Bayesian (Reich et al., 2014)
  - Censored pairwise likelihood (Raillard et al., 2014)
- Copula model (copula may needs to be ev-copula; connection to Markov model and max-stable model)

Auto-Regressive and Moving Average Models

A recent review is Zhang et al. (2016).

- MARMA(p, q) process (Davis and Resnick, 1989)
- Max positive alpha stable process (Naveau et al., 2011)
- Multivariate maxima and moving maxima (M4) model (Smith and Weissman, 1991; Zhang and Smith, 2004, 2010)
- Sparse moving maxima models (Zhang, 2005) with random coefficients (SM3R) (Tang et al., 2013).

Remark: inference is challenging.

Spatial Extremes: Latent Variables

(A recent review on spatial extremes modeling is Davison et al. (2012).)

- Latent Gaussian process (e.g., Casson and Coles, 1999; Cooley et al., 2007; Sang and Gelfand, 2010)
- Latent positive stable variables on a grid of knots (e.g., Stephenson, 2009; Reich and Shaby, 2012) (see more mileage Brian Reich's talk on Friday)

Remarks

- Both facilitate Hierarchical Bayesian inference.
- Latent Gaussian process does not offer extremal dependence, and the marginal distribution is no longer GEV.
- Laten positive stable variables retain marginal GEV distribution.

Spatial Extremes: Copulas

- From Sklar's Theorem, every multivariate continuous distribution can be uniquely represented by its marginal distributions and a *copula* which characterizes the dependence structure.
- Copula has standard uniform margins while simple max-stable process has unit Fréchet margins.
- Limiting distribution of multivariate componentwise sample maxima leads to marginal GEV distributions and an ev copula (max-stable).
- Parametric models
  - Hüsler–Reiss copula (Hüsler and Reiss, 1989) (connection to Gaussian extreme value process (Smith, 1990))
  - *t* copula (Demarta and McNeil, 2005) (connection to *t* extremal *t* process (Opitz, 2013))

Remark: A slightly different view but shares the same difficulties faced by max-stable process models.

Spatial Extremes: Max-Stable Process

- Extension of multivariate extreme value distribution to processes (de Haan, 1984; Schlather, 2002).
- Elegant in theory and popular in recent applications (e.g., Padoan et al., 2010; Davison and Gholamrezaee, 2012).
- Several parametric families implemented in R packages `SpatialExtremes` and `RandomFields`.
- Fully specification of the dependence structure allows prediction and simulation.
- Estimation often with composite likelihood because the full likelihood is unavailable.
- Full likelihood and Bayesian inference freshly tackled (Raphael Huser's talk)

Spatial Temporal Extremes

- A rising field.
- Models based on the spectral representation (Schlather, 2002; Kabluchko et al., 2009)
  - Constructions: spatial components and a single time coponent. (Davis et al., 2013; Huser and Davison, 2014; Buhl and Klüppelberg, 2016)
  - Composite likelihood inference.
  - Continous instead of discrete time; Time has no specific role from spatial dimension; No explicit temporal dynamics.
- Space-time max-stable models with spectral separability (Embrechts et al., 2015)
  - Partly decouple the influence of time and space, but such that time influences space through an operator on space.
  - Both continuous-time and discrete-time versions.

More Remarks

- Full dependence specification allows inference for jointly defined events and simulation.
- When the spatial dependence is misspecified, full specification can lead to bias (Wang et al., 2014) (e.g., mixture dependence).
- Goodness-of-fit test can provide some guard but can be difficult to do in high dimension (Kojadinovic et al., 2015).
- In some applications where the primary interest is inference about marginal parameters, spatial dependence is a nuisance.
- Efficient marginal inference is possibly by somewhat accounting for spatial dependence without exact modeling; no specification beyond the univariate GEV distribution.

Motivating Application: Detection and Attribution (Fingerprinting)

- Changes in climate extremes have more influential environmental and societal impacts than changes in mean climate states.
- Detection and attribution of changes in climate extremes have gained sharpened focus.
- This is a challenging problem: low signal-noise ratio, sparsity of data, distributional properties of extremes.
- No fully satisfactory analog of the optimal fingerprinting method has been established for changes in climate extremes with the GEV distribution.
  - Apply standard fingerprinting to measures of extremes or to estimated parameters from extreme value distributions.
  - The state-of-the-art is Zwiers et al. (2011).
- Can we increase the efficiency in Zwiers et al. (2011)?

Combined Score Equations

Idea: *Combine the score equation of the marginal GEV distribution at each monitoring sites in some optimal way to improve efficiency by accounting the spatial correlation among them.*

- Y_{ts}: extreme observation of interest at site s in year t with density $f(\cdot;\theta_{ts})$, $s = 1, \ldots, m$, $t = 1, \ldots, n$.
- X_{ts}: $p \times 1$ covariate vector for θ_{ts}.
- $g(\theta_{ts}) = \eta_{ts} = X_{ts}^{\top}\beta$, where g is a known link function.
- Assume data from year to year are independent while spatial dependence exists within the same year.
- Only assume marginal distribution f is the correctly specified GEV distribution.

Combined Score Equations

- Score function: $S_{ts} = \mathrm{d} \log f(Y_{ts}; \theta_{ts}) / \mathrm{d}\theta_{ts}$.
- Score equation for β at site s:

$$\sum_{t=1}^{n} X_{ts} \frac{\mathrm{d}\theta_{ts}}{\mathrm{d}\eta_{ts}} S_{ts} = 0.$$

- Combined score equation:

$$\sum_{t=1}^{n} X_t^\top A_t W_t^{-1} S_t = 0,$$

  where $X_t^\top = (X_{t1}, \ldots, X_{tm})$, $A_t = \mathrm{diag}(\mathrm{d}\theta_{t1}/\mathrm{d}\eta_{t1}, \ldots, \mathrm{d}\theta_{tm}/\mathrm{d}\eta_{tm})$, W_t^{-1} is the weight matrix, and $S_t = (S_{t1}, \ldots, S_{tm})^\top$.
- When W_t is the identity matrix, it reduces to the derivative of the independence likelihood.

Optimal Weight

- Optimal W_t (Nikoloulopoulos et al., 2011):

$$W_t = \Omega_t \Delta_t^{-1},$$

where $\Omega_t = \mathrm{cov}(S_t)$ and

$$\Delta_t = -\mathrm{diag}\left\{ \mathrm{E}\left(\frac{\mathrm{d}^2 \log f_{t1}(y_{t1}, \theta_{t1})}{\mathrm{d}\theta_{t1}^2} \right), \ldots, \mathrm{E}\left(\frac{\mathrm{d}^2 \log f_{tm}(y_{tm}, \theta_{tm})}{\mathrm{d}\theta_{tm}^2} \right) \right\}.$$

- Approximate the covariance matrix Ω_t of the score functions S_t:
  - Apply the idea of GEE (Liang and Zeger, 1986) with simple form of working spatial correlation structure.
  - Assume all the clusters (years) share a same correlation matrix, R, of the score function:

$$\Omega_t = \Delta_t^{1/2} R \Delta_t^{1/2}.$$

Approximation of Optimal Weight

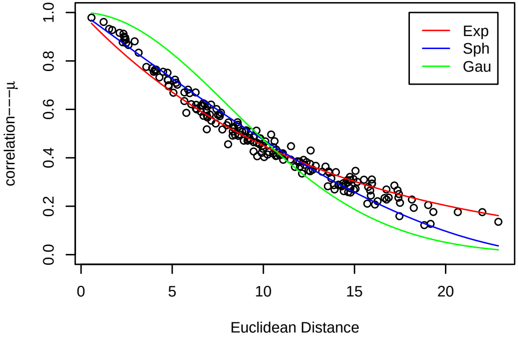

Figure: The empirical correlation of the standardized score function of μ (points), and the corresponding non-linear least square fitted correlation curves from exponential (red), spherical (blue) and gaussian (green) correlation function. Data generated from an isotropic Smith model with $m = 20$, $n = 1000$, and moderate dependence level in region $[-10, 10]$.

Approximation of Optimal Weight

- It would be nice to know the pairwise correlation between site j and site k, ρ_{jk}, but approximation is good too.
- Exponential correlation

$$\rho_{jk} = \exp(-d_{jk}/r),$$

where d_{jk} is the pairwise distance and r is the parameter to be estimated through the empirical correlation of the standardized score function.

- Spherical correlation

$$\rho_{jk} = \left[1 - 1.5(r/d_{jk}) + 0.5(r/d_{jk})^3\right] I_{d_{ij}<r},$$

which leads to sparse correlation matrix and can be exploited computationally when the number of sites is big.

Combined Score Equation of GEV Distribution

- μ_{ts}, σ_{ts}, and ξ_{ts}: the location, scale, and shape parameter, respectively, for the GEV distribution of Y_{ts}.
- $X_{\mu,ts}$, $X_{\sigma,ts}$, and $X_{\xi,ts}$: corresponding covariate vector.
- GEV parameters connect to covariates through known link functions:

$$g_\mu(\mu_{ts}) = X_{\mu,ts}^\top \beta_\mu, \qquad g_\sigma(\sigma_{ts}) = X_{\sigma,ts}^\top \beta_\sigma, \qquad g_\xi(\xi_{ts}) = X_{\xi,ts}^\top \beta_\xi,$$

- Regression parameter: $\Theta = (\beta_\mu^\top, \beta_\sigma^\top, \beta_\xi^\top)^\top$.
- Iteratively estimate one set of the parameters at a time while the other two sets are being fixed until convergence.

Inferences

- Sandwich variance estimator
  - May underestimate variation with shorter records
  - Need adjustment under temporal dependence
- Semiparametric bootstrap with no dependence structure specified
  - Need to preserve spatial (and temporal) dependence (Heffernan and Tawn, 2004)
  - Can be used in detection and attribution analysis

Simulation Study

- Number of sites m.
- Study region $[-c, c]^2$.
- Three combinations of (c, m): $(10, 20)$, $(10, 80)$, and $(20, 80)$.
- Number of years $n = 100$.
- Marginal model at each site s, $s = 1, \ldots, m$: GEV model with covariates latitude $X_1(s)$ and longitude $X_2(s)$:

$$\begin{cases} \mu_s = \beta_{\mu,0} + \beta_{\mu,1} X_1(s) + \beta_{\mu,2} X_2(s), \\ \sigma_s = \beta_{\sigma,0}, \\ \xi_s = \beta_{\xi,0}, \end{cases}$$

where $\beta_{\mu,0} = 15, \beta_{\mu,1} = -0.2, \beta_{\mu,2} = 0.25, \beta_{\sigma,0} = 4$ and $\beta_{\xi,0} = 0.2$.

Simulation Study

- Data between different years were independent, while within the same year spatial dependence was imposed.
- Data generating scenarios:
  1. An isotropic SM max-stable model.
  2. A mixture of a GG max-stable model and a GA model.
- Dependence level: moderate (M) and strong (S).
- Estimation methods:
  - Combined score equation (CSE).
  - Independence likelihood (IL).
  - Pairwise likelihood (PL).
- 1000 replications for each case.
- Report: average point estimate, root mean square error (RMSE), and the relative efficiency (RE) in terms of MSE using the IL method as the reference.

Scenario 1: SM model

Table: Simulation results for S dependence level.

			Estimate			RMSE			RE	
(c, m)	Par	True	IL	PL	CSE	IL	PL	CSE	PL	CSE
(10, 20)	$\beta_{\mu,0}$	15	15.034	15.032	15.030	0.353	0.347	0.317	1.039	1.243
	$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.026	0.025	0.023	1.062	1.255
	$\beta_{\mu,2}$	0.25	0.248	0.248	0.249	0.025	0.025	0.023	1.042	1.226
	$\beta_{\sigma,0}$	4	4.001	4.002	3.998	0.268	0.268	0.245	1.000	1.193
	$\beta_{\xi,0}$	0.2	0.198	0.199	0.199	0.061	0.045	0.056	1.804	1.202
(10, 80)	$\beta_{\mu,0}$	15	15.041	15.038	15.028	0.353	0.344	0.296	1.058	1.421
	$\beta_{\mu,1}$	−0.2	−0.199	−0.199	−0.199	0.023	0.023	0.019	1.010	1.439
	$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.023	0.023	0.019	1.006	1.477
	$\beta_{\sigma,0}$	4	3.993	3.994	3.983	0.259	0.257	0.222	1.014	1.360
	$\beta_{\xi,0}$	0.2	0.196	0.197	0.199	0.059	0.043	0.052	1.844	1.296
(20, 80)	$\beta_{\mu,0}$	15	15.024	15.022	15.022	0.241	0.238	0.208	1.030	1.350
	$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.010	0.010	0.009	1.022	1.437
	$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.011	0.011	0.009	1.024	1.463
	$\beta_{\sigma,0}$	4	3.993	3.992	3.993	0.183	0.182	0.156	1.015	1.379
	$\beta_{\xi,0}$	0.2	0.196	0.197	0.196	0.041	0.036	0.035	1.337	1.374

Scenario 2: Mixed model of GG and GA

- Proportion p from GG max-stable model with Gaussian correlation function $\rho(h) = \exp[-(\|h\|/\phi)^2]$.
- Proportion $1 - p$ from GA model with exponential correlation function $\rho(h) = \exp(-h/\tau)$.
- Contamination rate $p \in \{0, 0.1, 0.25, 0.5, 0.75, 1\}$.
- PL method was applied with the dependence model specified as GG.

Scenario 2: Mixed model of GG and GC

Table: Simulation results for S dependence level when $p = 0.25$ and 0.75.

				Estimate			RMSE			RE	
p	(c, m)	Par	True	IL	PL	CSE	IL	PL	CSE	PL	CSE
0.25	(10, 20)	$\beta_{\mu,0}$	15	14.739	14.694	14.730	0.438	0.459	0.419	0.907	1.091
		$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.024	0.023	0.022	1.053	1.194
		$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.023	0.022	0.021	1.076	1.227
		$\beta_{\sigma,0}$	4	4.031	4.013	4.029	0.257	0.251	0.231	1.049	1.233
		$\beta_{\xi,0}$	0.2	0.195	0.225	0.196	0.055	0.050	0.052	1.199	1.123
	(20, 80)	$\beta_{\mu,0}$	15	14.789	14.769	14.785	0.318	0.330	0.307	0.931	1.073
		$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.010	0.010	0.008	1.039	1.335
		$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.010	0.010	0.009	1.040	1.353
		$\beta_{\sigma,0}$	4	4.070	4.061	4.069	0.190	0.186	0.180	1.052	1.123
		$\beta_{\xi,0}$	0.2	0.192	0.204	0.192	0.039	0.033	0.037	1.431	1.141
0.75	(10, 20)	$\beta_{\mu,0}$	15	14.920	14.842	14.914	0.352	0.371	0.337	0.902	1.095
		$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.022	0.022	0.020	1.059	1.184
		$\beta_{\mu,2}$	0.25	0.250	0.250	0.250	0.023	0.022	0.021	1.058	1.186
		$\beta_{\sigma,0}$	4	3.998	3.971	3.998	0.251	0.248	0.239	1.025	1.106
		$\beta_{\xi,0}$	0.2	0.200	0.255	0.201	0.048	0.067	0.045	0.504	1.108
	(20, 80)	$\beta_{\mu,0}$	15	14.932	14.893	14.924	0.273	0.283	0.253	0.935	1.170
		$\beta_{\mu,1}$	−0.2	−0.200	−0.200	−0.200	0.010	0.009	0.008	1.042	1.334
		$\beta_{\mu,2}$	0.25	0.250	0.249	0.250	0.010	0.010	0.009	1.041	1.252
		$\beta_{\sigma,0}$	4	4.016	3.998	4.014	0.185	0.182	0.169	1.037	1.207
		$\beta_{\xi,0}$	0.2	0.197	0.221	0.197	0.031	0.035	0.029	0.806	1.141

Some Discussion

- Comparison between CSE and PL
  - What if PL is correctly specified?
    PL does better in shape but not in location parameters.
- Simulation study under what spatial dependence
  - max-stable (ev copula)
  - non-ev copula (e.g. Gaussian)
  - mixture of ev and non-ev copula
  - unknown dependence (from observed data)
- Handling large data with spherical working correlation and sparse matrix operation.
- Non-invertible working correlation matrix?
- Performance of bootstrap confidence interval

Optimal Fingerprinting

- The fingerprint refers to the pattern of change in the climate that is expected in response to external forcing of the climate system, and is typically estimated from climate model simulations.
- Regresses observations onto the fingerprints to determine whether they are present in the observations.
- Classical optimal fingerprinting: scaling factors are estimated by a generalized least squares approach, with the optimal weight chosen to be the inverse variance matrix of the residuals. The method gives the smallest variance of the scaling factor estimator.
- No fully satisfactory analog of the optimal fingerprinting method has been established for changes in climate extremes with the GEV distribution.

Fingerprint Method

- Signal Estimation
- Detection Analysis
- Uncertainty Assessment
- Goodness-of-fit Test

Signal Estimation

- Estimated from the ensembles of climate model simulation data of the external forcing.
- GEV location parameters, which vary every h-yr ($h = 5$ or 10), are used to represent the model-simulated changes in the climate extremes.

Signal Estimation

Suppose

- Number of year n is multiple times of h, then the number of h-yr block is $B = n/h$.
- Totally l ensembles for one individual climate model.
- Z_{tsu}: the model output at grid box s in year t from ensemble u for $s = 1, \ldots, m$, $t = 1, \ldots, n$, and $u = 1, \ldots, l$.

A GEV model is assumed for Z_{tsu} with parameters

$$\begin{cases} \mu_{tsu} = \mu_{b(t),s}, \\ \sigma_{tsu} = \sigma_s, \\ \xi_{tsu} = \xi_s, \end{cases} \tag{1}$$

where $b(t) = \text{ceiling}(t/h)$, $b = 1, \ldots, B = n/h$.

Detection Analysis

- The observed annual extreme climate variable Y_{ts} at grid box s is modeled by a GEV distribution with

$$\begin{cases} \mu_{ts} = \alpha_s + X_{ts}^\top \beta, \\ \sigma_{ts} = \sigma_s, \\ \xi_{ts} = \xi_s, \end{cases} \tag{2}$$

  - X_{ts}: $p \times 1$ vector of the relative signals of external forcing of interest estimated from climate model data.
  - β: scaling vector.
  - α_s, σ_s, and ξ_s: grid box specific location, scale, and shape parameters, respectively.
- CSE reduces to Zwiers et al. (2011) with working independence.
- Bootstrap procedures in Zwiers et al. (2011) remain valid.

Inference about β

The confidence interval of β:

- Above 0: a response to external forcing is said to be "detected".
- Above 0 and also contains 1: no evidence that observation and model simulations are inconsistent; conclusion of attribution is made.
- Above 1: observation is underestimated by climate model.
- Between 0 and 1: observation is overestimated by climate model.

CSE Method with a Coordinate Descent Approach

Totally $3m + p$ unknown parameters.

Coordinate descent approach: a two-step iterative process.

1. Given current estimate $\hat{\beta}$ of β, obtain the likelihood estimate $\hat{\zeta}_s$ of $\zeta_s = (\alpha_s, \sigma_s, \xi_s)$ separately at each grid box $s \in \{1, \ldots, m\}$.
2. Given current estimate $\hat{\zeta}_s$, obtain the CSE estimate $\hat{\beta}$ of β from solving the estimating equation with an appropriately chosen working correlation structure.

The two steps iterate until $\hat{\beta}$ converges.

Uncertainty Assessment

A 32×32 block bootstrap that preserves both temporal and spatial dependence is performed.

- 32 bootstrap samples of signal: account for the effects of signal uncertainty that arise from climate-model-simulated internal variability.
- 32 bootstrap samples of the observational data: account for the natural internal variability in the climate system.

5% quantile and 95% quantile leads to an approximate 90% confidence interval.

Goodness-of-fit Test

- Test for each grid box separately.
- Remove the nonstationary component by the fitted GEV parameters.
- Account for the uncertainty in parameter estimation.
- Combine the site-wise p-values into one region level p-value to alleviate the multiple testing issue.

Simulation Study in Fingerprinting Setting

- Mimic the daily maximum temperature setting in Australia ($n = 140$, $m = 29$).
- Recall detection model:

$$\mu_{ts} = \alpha_s + X_{ts}^\top \beta, \qquad \sigma_{ts} = \sigma_s, \qquad \xi_{ts} = \xi_s.$$

- Estimated signals $\tilde{\mu}_{d(t),s}$ were used as input X_{ts} to generate data. Parameters α, σ and ξ are the estimates based on Australia data.
- $\beta \in \{0, 0.5, 1\}$.
- Dependence model: a mixture of a GG model (proportion p) and a GA model (proportion $1 - p$).
- CSE method with an exponential correlation structure.

			Estimate			RMSE			RE	
p	Dep	True	IL	PL	CSE	IL	PL	CSE	PL	CSE
0	M	0	−0.001	−0.001	0.001	0.120	0.114	0.103	1.10	1.37
		0.5	0.503	0.503	0.502	0.118	0.111	0.097	1.12	1.49
		1	1.005	1.005	1.005	0.119	0.112	0.098	1.13	1.48
	S	0	−0.001	−0.002	−0.004	0.153	0.140	0.104	1.19	2.14
		0.5	0.503	0.502	0.501	0.147	0.133	0.102	1.21	2.05
		1	1.007	1.007	1.002	0.146	0.134	0.103	1.19	1.99
0.5	M	0	0.004	0.003	0.001	0.116	0.112	0.094	1.08	1.51
		0.5	0.507	0.507	0.502	0.115	0.111	0.096	1.07	1.42
		1	0.997	0.997	1.000	0.115	0.112	0.097	1.06	1.40
	S	0	−0.004	−0.004	0.000	0.138	0.131	0.098	1.12	2.00
		0.5	0.500	0.500	0.499	0.144	0.136	0.100	1.13	2.08
		1	1.005	1.005	1.006	0.138	0.131	0.097	1.12	2.02
1	M	0	−0.001	−0.001	−0.002	0.110	0.108	0.091	1.04	1.46
		0.5	0.504	0.504	0.502	0.110	0.108	0.092	1.04	1.44
		1	0.997	0.997	1.000	0.112	0.110	0.093	1.04	1.44
	S	0	−0.001	0.000	−0.002	0.132	0.128	0.098	1.07	1.83
		0.5	0.505	0.505	0.502	0.133	0.129	0.099	1.07	1.80
		1	0.996	0.996	1.000	0.135	0.131	0.100	1.07	1.82

(The relative efficiency (RE) was based on the MSE, with the IL estimate as reference.)

Applications on Extreme Temperatures

- Perfect model detection for Australia
- Extreme temperatures in Northern Europe (NEU)
  - Annual maximum of daily maximum (TXx) — warmest day
  - Annual maximum of daily minimum (TNx) — warmest night
  - Annual minimum of daily maximum (TXn) — coldest day
  - Annual minimum of daily minimum (TNn) — coldest night
- Data period 1951–2010 ($n = 60$, $m = 67$).
- CSE method with an exponential correlation structure.

Perfect Model Detection for Australia

- Annual maxima of daily minimum temperature (TNx) in Australia during 1861–2000 ($n = 140$, $m = 29$).
- Both the "observational" data and signals come from simulations of climate models.
- Climate simulation data with 10 ensembles under combined effect of anthropogenic and natural forcings. Each ensemble was treated as an observation dataset and the rest nine ensembles as model simulations.
- If the signal is strong enough to be detectable, β should be around 1 (which may be affected by data resolution).

Table: Summaries of the estimate of the scaling factor β, the corresponding 90% confidence interval and the interval length for the 10 perfect model detection analyses.

	Estimate		90% Confidence Interval		Interval Length	
Ensemble	IL	CSE	IL	CSE	IL	CSE
1	1.03	0.91	(0.78, 1.28)	(0.75, 1.09)	0.49	0.34
2	0.93	0.79	(0.68, 1.19)	(0.60, 0.99)	0.51	0.39
3	1.11	1.06	(0.90, 1.32)	(0.83, 1.30)	0.42	0.46
4	1.10	0.94	(0.86, 1.35)	(0.76, 1.15)	0.49	0.39
5	0.99	0.93	(0.75, 1.26)	(0.71, 1.15)	0.51	0.44
6	0.98	0.79	(0.73, 1.24)	(0.55, 1.03)	0.51	0.47
7	0.95	0.88	(0.77, 1.16)	(0.67, 1.10)	0.39	0.42
8	1.16	0.89	(0.94, 1.38)	(0.67, 1.09)	0.44	0.42
9	0.98	0.81	(0.67, 1.27)	(0.54, 1.07)	0.60	0.53
10	1.02	1.01	(0.81, 1.24)	(0.83, 1.20)	0.43	0.36

Detection and Attribution Analysis of TNx in NEU

Forcing	Me	Par	est	90% CI	len	gof
ALL	IL	β	1.10	(0.73, 1.48)	0.75	2
	CSE	β	0.69	(0.46, 0.95)	0.49	1
ANT	IL	β	1.19	(0.77, 1.62)	0.85	2
	CSE	β	0.52	(0.31, 0.74)	0.43	2
ANT&NAT	IL	β_A	1.12	(0.75, 1.50)	0.76	2
		β_N	0.91	(-0.28, 2.07)	2.35	
	CSE	β_A	0.70	(0.47, 0.95)	0.48	1
		β_N	0.59	(0.17, 1.01)	0.84	

Summary

- CSE improves efficiency without specifying spatial dependence
- Application to climate extremes increases power of detection and attribution analysis.

Future Work

- Applications
  - Regional frequency analysis
  - Detection and attribution of changes in extreme precipitation
- Methodological development.
  - Correct bias from measurement error (signals are not known but estimated): ongoing with Yujing Jiang.
  - Avoid correlation parameter estimation (Stoner and Leroux, 2002).
  - Minimize the objective function (Qu et al., 2000; Bai et al., 2012).

Bivariate Density

The bivariate density of site i and j, $f_{i,j}$, is

$$f_{i,j}(y_i, y_j; \beta, \alpha) = g_{i,j}\,(z_i, z_j; \alpha)\,|J(y_i, y_j; \beta)|,$$

where $g_{i,j}(z_i, z_j; \alpha)$ is the bivariate marginal density of the max-stable process model with unit Fréchet margins, $z_i = G^{-1}\{F_i(y_i; \beta)\}$, and

$$|J(y_i, y_j; \beta)| = \Big|\frac{\mathrm{d}}{\mathrm{d}y_i} G^{-1}\{F_i(y_i; \beta)\}\frac{\mathrm{d}}{\mathrm{d}y_j} G^{-1}\{F_j(y_j; \beta)\}\Big|.$$

Optimal Weight

From Cauchy–Schwarz inequality (Chaganty and Joe, 2004), the optimal choice of W_t satisfies

$$X_t^\top W_t^{-1} = \Psi_t^\top \Omega_t^{-1}, \tag{3}$$

where $\Psi_t = -\mathrm{E}(\mathrm{d}S_t/\mathrm{d}\beta^\top)$, and $S_t = (S_{t1}, \ldots, S_{tm})^\top$. Since

$$-\mathrm{E}(\mathrm{d}S_{ts}/\mathrm{d}\beta^\top) = -\mathrm{E}\left(\frac{\mathrm{d}^2 \log f_{ts}(y_{ts}, \theta_{ts})}{\mathrm{d}\theta_{ts}^2}\right) x_{ts}^\top = \mathcal{I}_{ts} x_{ts}^\top,$$

We have

$$\Psi_t = -\mathrm{E}(\mathrm{d}S_t/\mathrm{d}\beta^\top) = \Delta_t X_t,$$

where $\Delta_t = \mathrm{diag}(\mathcal{I}_{t1}, \ldots, \mathcal{I}_{tm})$. Substitute Ψ_t into equation (3),

$$X_t^\top W_t^{-1} = \Psi_t^\top \Omega_t^{-1} = X_t^\top \Delta_t^\top \Omega_t^{-1} \quad \Rightarrow \quad W_t = \Omega_t \Delta_t^{-1}.$$

Bootstrap Sampling of Signal when $h = 10$

(A) For each individual grid box of each ensemble, divide the n-year simulation data into $n/10$ nonoverlapping 10-yr blocks. Randomly sample 5-yr blocks data with replacement within the 10-yr blocks. All the grid boxes share the same sample order to keep the spatial dependence.

(B) All the ensembles of all climate models share the same sample order.

(C) Estimate the signals from the reordered data in step A.

(D) Repeat steps A to C 32 times.

Bootstrap Sampling of Signal when $h = 5$

(A) For each ensemble, transform the simulation data into Gumbel residuals by the fitted GEV parameters in model (1) at each grid box (Kharin and Zwiers, 2005).

(B) For each individual grid box, divide the data into nonoverlapping 5-yr blocks. Randomly sample 5-yr blocks data with replacement. All the grid boxes share the same sample order to keep the spatial dependence.

(C) Transform the reordered Gumbel residuals back into GEV distribution by the fitted GEV parameters in model (1), then estimate the signals from the transformed data.

(D) Repeat steps A to C 32 times.

Bootstrap Sampling of Observational Data

(a) Subtracting the scaled signal $X_{ts}^{\top}\hat{\beta}$ from Y_{ts} in Model (2) to obtain the residuals.

(b) For each individual grid box, divide the residuals into nonoverlapping 5-yr blocks. Randomly reorder the 5-yr blocks residuals. All the grid boxes share the same sample order to keep the spatial dependence.

(c) Adding the scaled signal $X_{ts}^{\top}\hat{\beta}$ back to the reordered residuals, and denote it as $\tilde{Y}_{ts}$.

(d) Estimate the scaling factor from $\tilde{Y}_{ts}$ with the given signal.

(e) Repeat steps b to d 32 times for each of the 32 bootstrap samples of signal.

Goodness-of-fit Test

i. Transform the observational data into Gumbel residuals by the fitted GEV parameters in Model (2) at each grid box. So testing the goodness-of-fit of the GEV distribution of observed data equals to testing the goodness-of-fit of the standard Gumbel distribution of the Gumbel residuals.
ii. Calculate the Kolmogorov–Smirnov test statistic at each grid box.
iii. Apply the semiparametric bootstrap algorithm in Heffernan and Tawn (2004) on the observational data to generate a bootstrap sample of observed data. Apply steps i to ii on the sample data. Repeat this step 1000 times to obtain 1000 bootstrap samples of the Kolmogorov–Smirnov statistics at each grid box.

References I

Bai, Y., P. X.-K. Song, and T. E. Raghunathan (2012). Joint composite estimating functions in spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(5), 799–824.

Buhl, S. and C. Klüppelberg (2016). Anisotropic Brown–Resnick space-time processes: Estimation and model assessment. *Extremes*. Forthcoming.

Casson, E. and S. Coles (1999). Spatial regression models for extremes. *Extremes 1*(4), 449–468.

Chaganty, N. R. and H. Joe (2004). Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 66*(4), 851–860.

Cooley, D., D. Nychka, and P. Naveau (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association 102*(479), 824–840.

Davis, R. A., C. Klüppelberg, and C. Steinkohl (2013). Max-stable processes for modeling extremes observed in space and time. *Journal of the Korean Statistical Society 42*(3), 399–414.

Davis, R. A. and S. I. Resnick (1989). Basic properties and prediction of max-ARMA processes. *Advances in applied probability 21*(4), 781–803.

Davison, A. C. and M. M. Gholamrezaee (2012). Geostatistics of extremes. *Proceedings of the Royal Society A 468*, 581–608.

Davison, A. C., S. A. Padoan, and M. Ribatet (2012). Statistical modeling of spatial extremes. *Statistical Science 27*(2), 161–186.

Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological) 52*, 393–442.

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

Demarta, S. and A. McNeil (2005). The t copula and related copulas. *International Statistical Review 73*(1), 111–129.

Embrechts, P., E. Koch, and C. Robert (2015). Space-time max-stable models with spectral separability. *arXiv preprint arXiv:1507.07750*.

Fawcett, L. and D. Walshaw (2006). Markov chain models for extreme wind speeds. *Environmetrics 17*(8), 795–809.

References II

Fawcett, L. and D. Walshaw (2007). Improved estimation for temporally clustered extremes. *Environmetrics 18*(2), 173–188.

Ferro, C. A. and J. Segers (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 545–556.

Heffernan, J. E. and J. A. Tawn (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(3), 497–546.

Huser, R. and A. Davison (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(2), 439–461.

Hüsler, J. and R.-D. Reiss (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters 7*, 283–286.

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

Kharin, V. and F. Zwiers (2005). Estimating extremes in transient climate change simulations. *Journal of Climate 18*(8), 1156–1173.

Kojadinovic, I., H. Shang, and J. Yan (2015). A class of goodness-of-fit tests for spatial extremes models based on max-stable processes. *Statistics and Its Interfaces 8*(1), 45–62.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Naveau, P., Z. Zhang, and B. Zhu (2011). An extension of max autoregressive models. *Statistics and Its Interfaces 4*, 253–266.

Nikoloulopoulos, A. K., H. Joe, and N. R. Chaganty (2011). Weighted scores method for regression models with dependent data. *Biostatistics 12*, 653–665.

Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis 122*, 409–413.

Padoan, S. A., M. Ribatet, and S. A. Sisson (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association 105*(489), 263–277.

References III

Qu, A., B. G. Lindsay, and B. Li (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika 87*(4), 823–836.

Raillard, N., P. Ailliot, and J. Yao (2014). Modeling extreme values of processes observed at irregular time steps: Application to significant wave height. *The Annals of Applied Statistics 8*(1), 622–647.

Reich, B. J. and B. A. Shaby (2012). A hierarchical max-stable spatial model for extreme precipitation. *The annals of applied statistics 6*(4), 1430–1451.

Reich, B. J., B. A. Shaby, and D. Cooley (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*(1), 119–135.

Reick, B. J. and B. A. Shaby (2016). Time sereis of extremes. In D. K. Dey and J. Yan (Eds.), *Extreme Value Modeling and Risk Analysis: Methods and Applications*, pp. 131–151. CRC Press.

Sang, H. and A. E. Gelfand (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics 15*(1), 49–65.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes 5*(1), 33–44.

Shaby, B. A., B. J. Reich, D. Cooley, and C. G. Kaufman (2016). A Markov-switching model for heat waves. *The Annals of Applied Statistics 10*(1), 74–93.

Smith, R. L. (1990). Max-stable processes and spatial extremes. Unpublished Manuscript.

Smith, R. L. (1991). Regional estimation from spatially dependent data. Preprint.

Smith, R. L., J. A. Tawn, and S. G. Coles (1997). Markov chain models for threshold exceedances. *Biometrika 84*(2), 249–268.

Smith, R. L. and I. Weissman (1991). Characterization and estimation of the multivariate extremal index. Preprint.

Stephenson, A. G. (2009). High-dimensional parametric modelling of multivariate extreme events. *Australian & New Zealand Journal of Statistics 51*(1), 77–88.

Stoner, J. A. and B. G. Leroux (2002). Analysis of clustered data: A combined estimating equations approach. *Biometrika 89*(3), 567–578.

References IV

Tang, R., J. Shao, and Z. Zhang (2013). Sparse moving maxima models for tail dependence in multivariate financial time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 143*, 882–895.

Wang, Z., J. Yan, and X. Zhang (2014). Incorporating spatial dependence in regional frequency analysis. *Water Resources Research 50*(12), 9570–9585.

Zhang, Z. (2005). A new class of tail-dependent time series models and its applications in financial time series. *Advances in Econometrics 20(B)*, 323–358.

Zhang, Z., L. Peng, and T. Idowu (2016). Max-autoregressive and moving maxima models for extremes. In D. K. Dey and J. Yan (Eds.), *Extreme Value Modeling and Risk Analysis: Methods and Applications*, pp. 153–178. CRC Press.

Zhang, Z. and R. L. Smith (2004). The behavior of multivariate maxima of moving maxima processes. *Journal of Applied Probability 41*, 1113–1123.

Zhang, Z. and R. L. Smith (2010). On the estimation and application of max-stable processes. *Journal of Statistical Planning and Inference 140*, 1135–1153.

Zwiers, F. W., X. Zhang, and Y. Feng (2011). Anthropogenic influence on long return period daily temperature extremes at regional scales. *Journal of Climate 24*(3), 881–892.