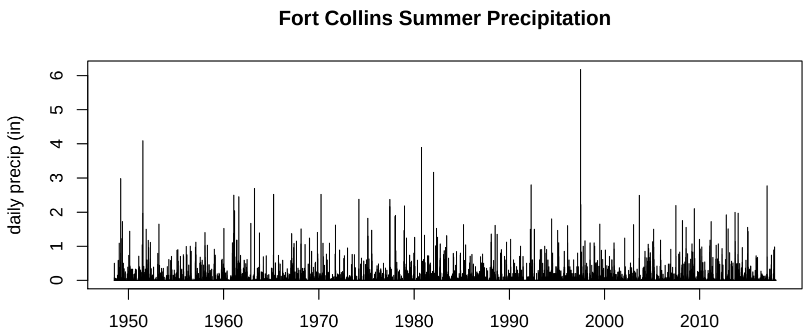
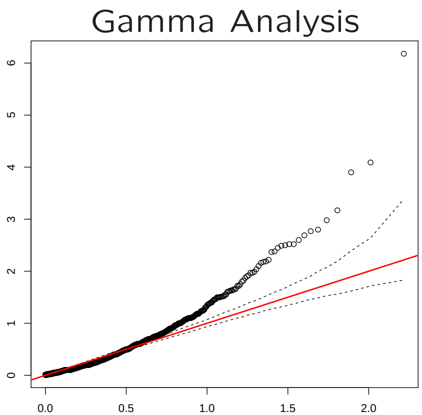
Transformed-Linear Methods for Multivariate Extremes and Application to Climate

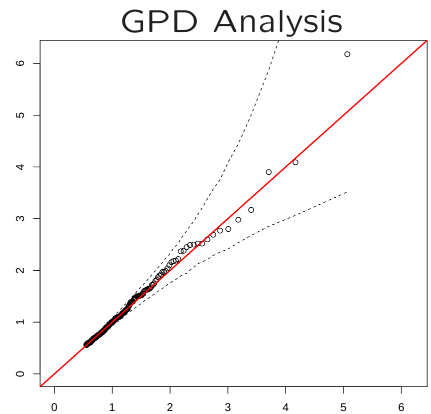
Dan Cooley
Department of Statistics
Colorado State University

Joint work with:
Yujing Jiang, Jeongjin Lee,
Nehali Mhatre, Troy Wixson, CSU

Extremes Mantra: ‘Let the tail speak for itself’



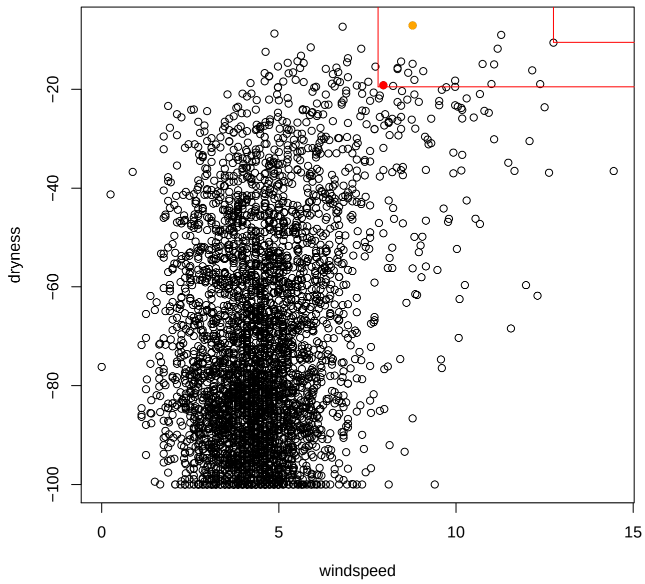




Rtn period est: 35 million years 146 years

- Fit only a subset of extreme data.
- Use a model appropriate for extremes.

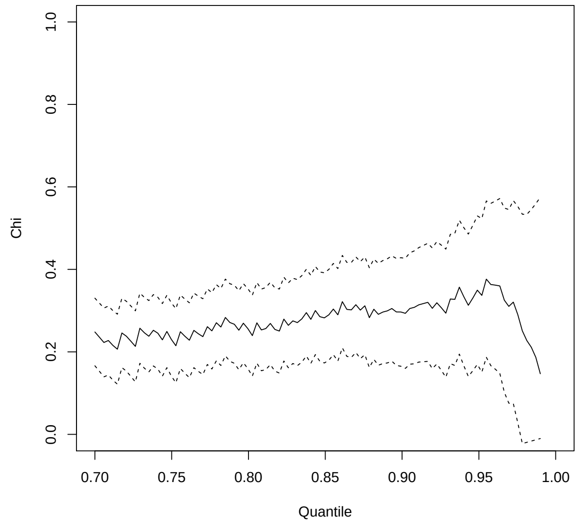
Bivariate Extremes: S. Calif. Fire Conditions



- Points in upper right correspond to explosive fire risk.
- Santa Ana Winds: simultaneous extreme dry/wind.
- Red: Cedar Fire (2003), Witch Fire (2007).
- What is probability of conditions in regions $\mathcal{R}_1, \mathcal{R}_2$?
- Correlation does not describe upper tail.

How strong is tail dependence?

$$\chi = \lim_{u \to 1} P(X_2 > F_{X_2}^{-1}(u) | X_1 > F_{X_1}^{-1}(u))$$



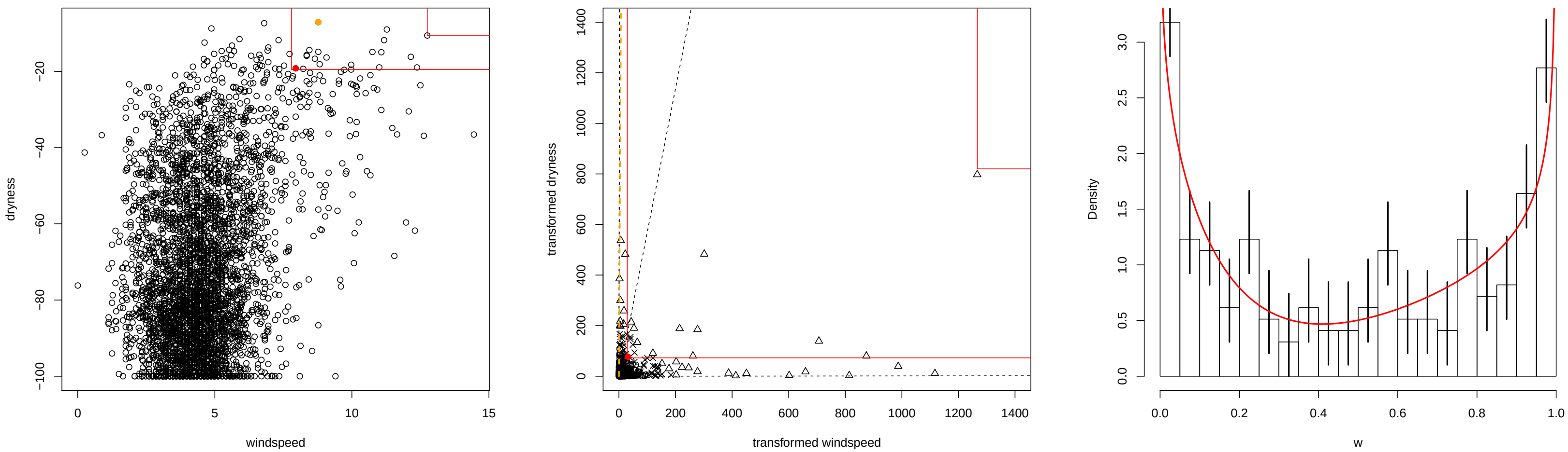
- χ summarizes tail dependence with one number.
- $\chi \approx .25$ is pretty strong!
- We will use a different tail dependence summary later.
- Not enough to assess $P(\boldsymbol{X} \in \mathcal{R}_i)$.

Regular Variation: Modeling Tail Dependence

Estimating $P(\boldsymbol{X} \in \mathcal{R}_i)$ requires us to *model* tail dependence.

Multivariate Regular Variation

- Framework suggested by classical MV extremes.
- Assumes data are heavy-tailed. (Below $\rightarrow \alpha = 1$.)
- Dist'n of large points given by radial/angular components.

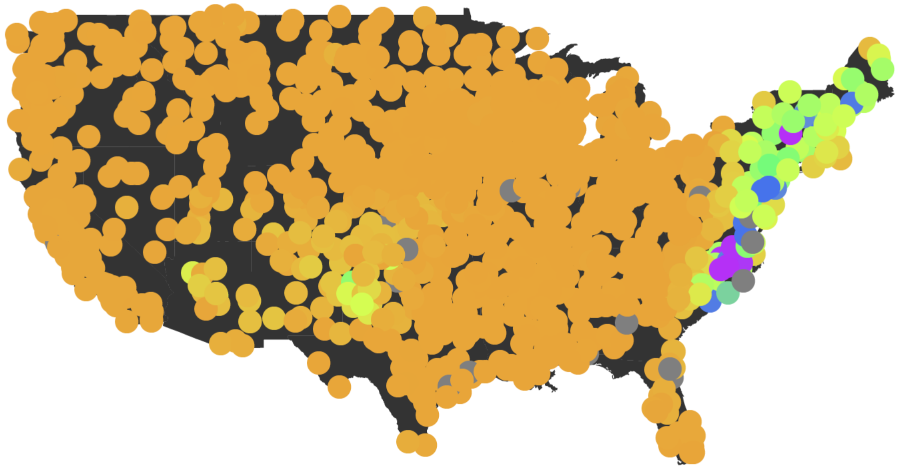


For $\mathcal{R}_2$: $\widehat{p}_2 = 1.5 \times 10^{-4}$, Empirical: $\tilde{p}_2 < 2.5 \times 10^{-4}$

95% CI: $(0.8, 5.5) \times 10^{-4}$.

What about high dimensional data?

Hurricane Floyd, Precip Sept 16-18, 1999



- Modeling angular measure H is hopeless.
- Maybe we don't need to...
  - Questions are likely more simple.
  - In non-extreme analysis, use covariance as summary measure.
  - Many useful methods are linear: PCA, linear time series, linear prediction, spatial methods.

Tail Pairwise Dependence Matrix

Assume $\boldsymbol{X} \in [\mathbf{0}, \infty)^p$ is regularly varying with index $\alpha = 2$, and has angular measure H on L_2 unit ball Θ_{d-1}^+.

Define TPDM $\Sigma_{\boldsymbol{X}} := [\sigma_{ik}]_{i,k=1,\ldots,p}$, and

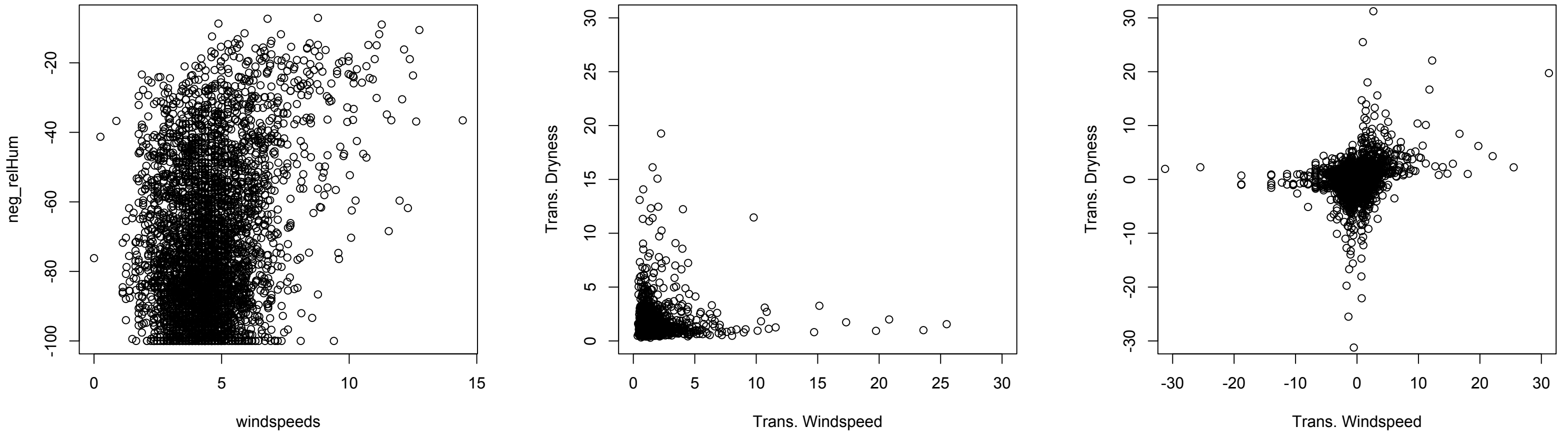
$$\sigma_{ik} = \int_{\Theta_{d-1}^+} w_i w_k dH_{\boldsymbol{X}}(\boldsymbol{w}).$$

- $\Sigma_{\boldsymbol{X}}$ is *non-negative definite*!
- A matrix of χ_{ik}'s isn't.
- Diagonals describe 'scale' of elements.
- Asymptotic independence iff $\sigma_{ik} = 0$.
- Also is completely positive: $\exists A \in \mathbb{R}^{p \times q} \geq 0$ s.t. $\Sigma_{\mathbf{X}} = AA^T$.

Linear operations in the positive orthant?

Q: Why model in positive orthant?

A: Allows us to focus on the *upper* tail.



How do you perform linear operations and restrict to positive orthant?

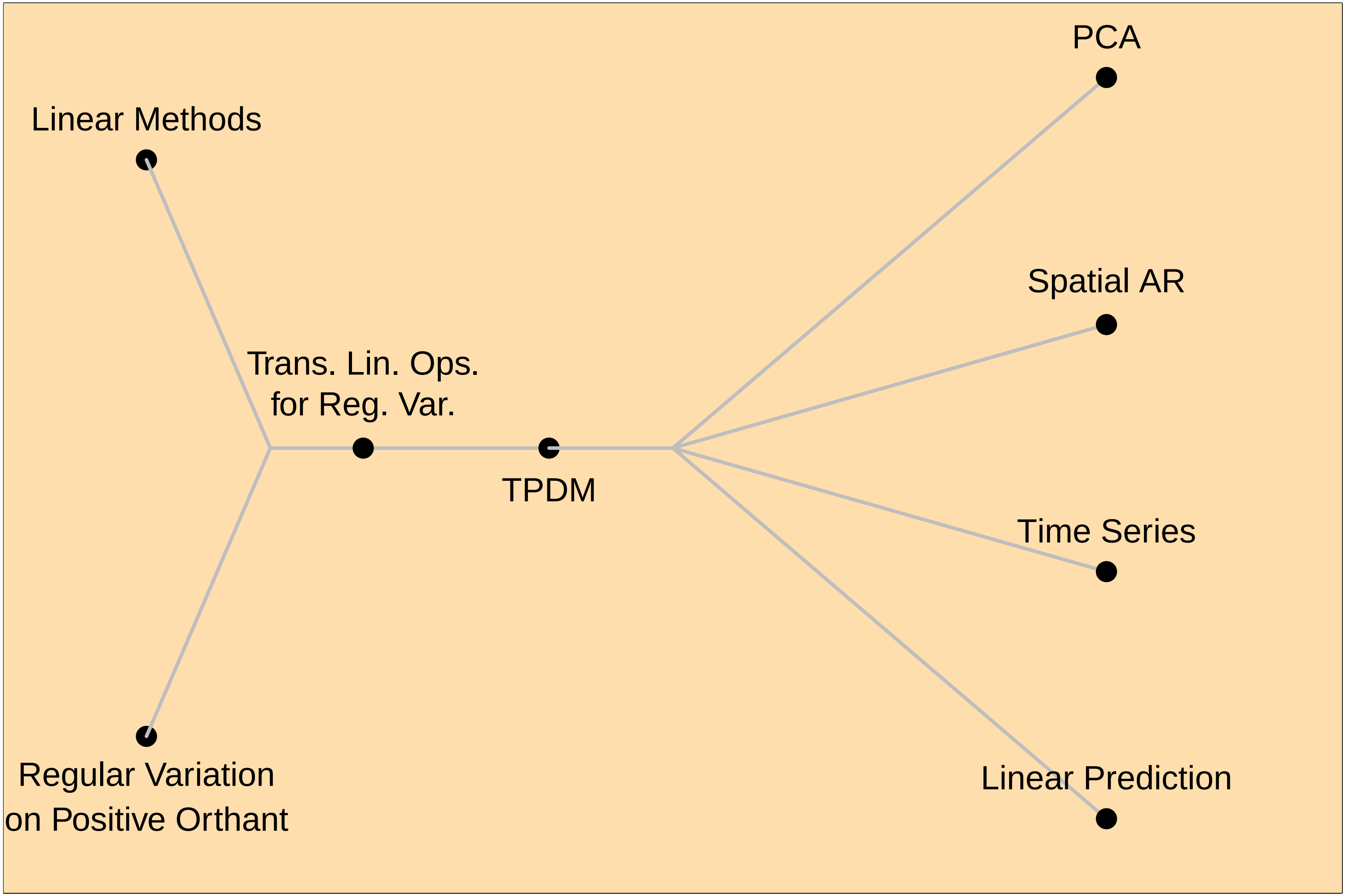
Transformed-Linear Operations

Goal: define "linear" operations in $\mathbb{R}_+^p$:

- $\boldsymbol{y} \in \mathbb{R}^p$
- t: 'transform', monotone function $\mathbb{R} \mapsto \mathbb{R}_+$
- Define $\boldsymbol{x} = t(\boldsymbol{y}), \boldsymbol{x} \in \mathbb{R}_+^p$, componentwise
- $\boldsymbol{x}_1 \oplus \boldsymbol{x}_2 := t\Big(t^{-1}(\boldsymbol{x}_1) + t^{-1}(\boldsymbol{x}_2)\Big)$
- $c \circ \boldsymbol{x} = t(ct^{-1}(\boldsymbol{x}))$ for $c \in \mathbb{R}$
- Can show creates a vector space in $\mathbb{R}_+^p$.

...applied to regularly-varying random vectors

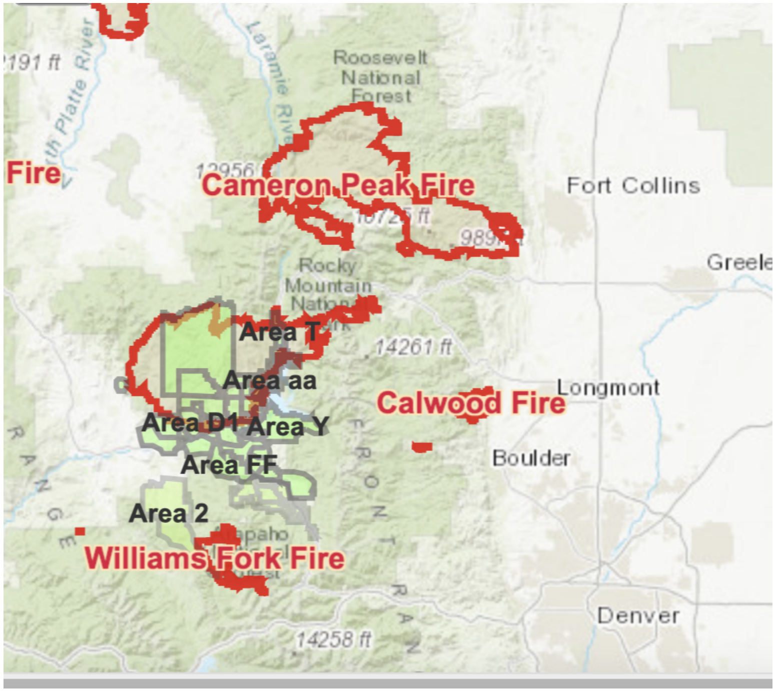
- Use $t(y) = \log(1 + \exp(y))$; $\lim_{y\to\infty} \frac{t(y)}{y} = \lim_{x\to\infty} \frac{t^{-1}(x)}{x} = 1$.
- Regular variation 'preserved':
  - $\boldsymbol{X}_1, \boldsymbol{X}_2 \in RV_+(\alpha) \to \boldsymbol{X}_1 \oplus \boldsymbol{X}_2 \in RV_+(\alpha)$.
  - $\boldsymbol{X} \in RV_+(\alpha), a \in \mathbb{R} \to a \circ \boldsymbol{X} \in RV_+(\alpha)$.



PCA
Linear Methods
Spatial AR
Trans. Lin. Ops.
for Reg. Var.
TPDM
Time Series
Regular Variation
on Positive Orthant
Linear Prediction

Colorado Wildfire Season 2020

- Three largest fires in recorded history occurred in 2020.
- We focus on Northern Colorado.

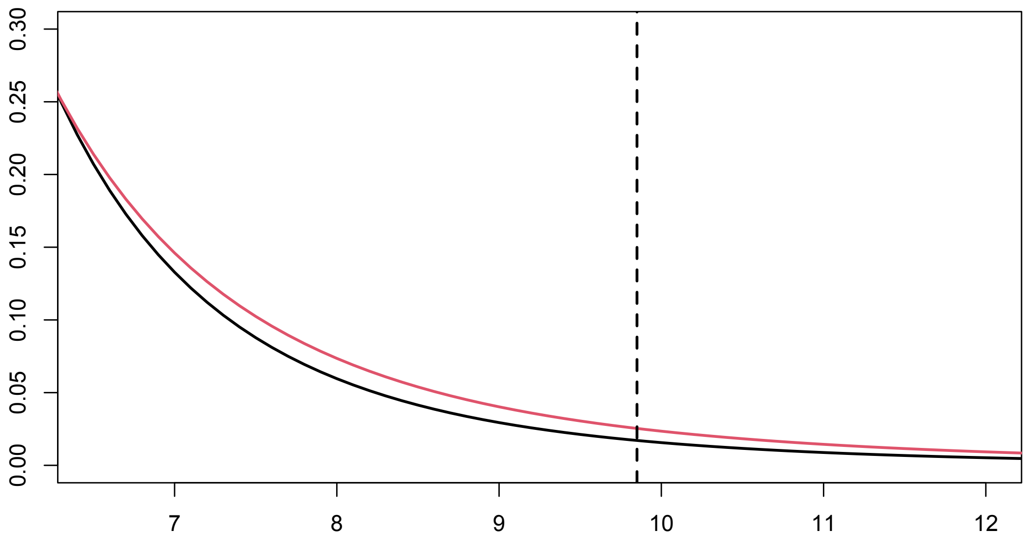


- Cameron Peak: 209K acres, Aug 13 - Dec 2.
- East Troublesome: 193K acres, Oct 14 - Nov 30.

Extreme Event Attribution

Public: "Is an extreme event caused by climate change?"

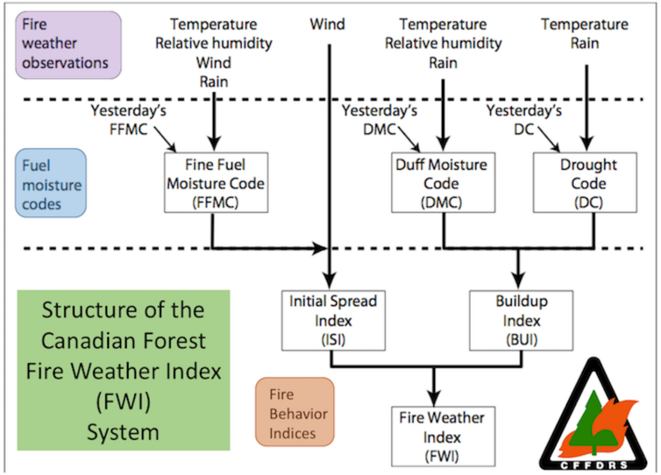
Goal: Compare event probabilities under current and reference climates.



Attribution done largely for "single spike" events.
Q: How to do attribution for *seasonal* risk?
A: Time series analysis.

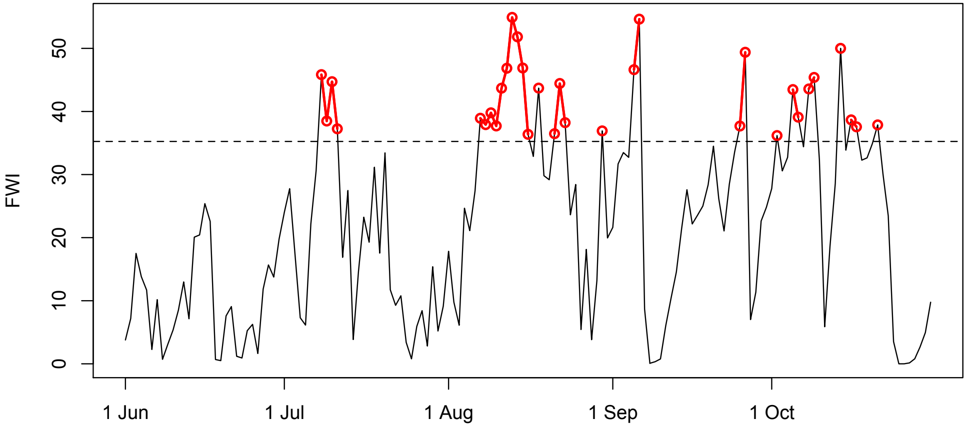
Fire Weather Index

Takes weather input (temp, relative humidity, wind, rain), outputs fire *weather* severity.



- Short term (ISI) and cumulative (BUI) effects.
- Does not tell whole story about fire risk:
  - – Does not account for management, ignition.
  - + Focuses on *climate signal* associated with fire risk.
- FWI computed from ERA5 Reanalysis data (for now).

Characteristics of FWI Time Series



- Goal: Model the upper tail.
- Exploratory analysis:
  - Strong tail dependence:
    - Lag 1: $\widehat{\chi}(.98) = 0.45$
    - Lag 5: $\widehat{\chi}(.98) = 0.23$.
  - Bounded tail: $\widehat{\xi} \approx -.3$.
- Positive-valued, interest is one-directional.
- Non-stationary due to seasonality.

Approach for Attribution

1. Fit a *simple* time series model which captures the *extreme* behavior of the Fire Weather Index (FWI).
2. Simulate many fire seasons under current and past climate.
3. Compare 2020 fire season to the simulated fire seasons in both periods.

Classical Time Series Analysis

- Can be done assuming only *second-order stationarity*.
  - TS characterized by ACVF (lagged pairwise dependence)
- Familiar models are *linear*.
  - $X_t = \sum_{j=0}^{\infty} a_j Z_{t-j}; \quad \{Z_t\} \sim WN(0, \sigma^2)$
  - ARMA models

Tail Stationarity and TPDF

Let $\{X_t\}$ be *non-negative* regularly-varying TS with $\alpha = 2$.

- Tail Pairwise Dependence Function (TPDF)

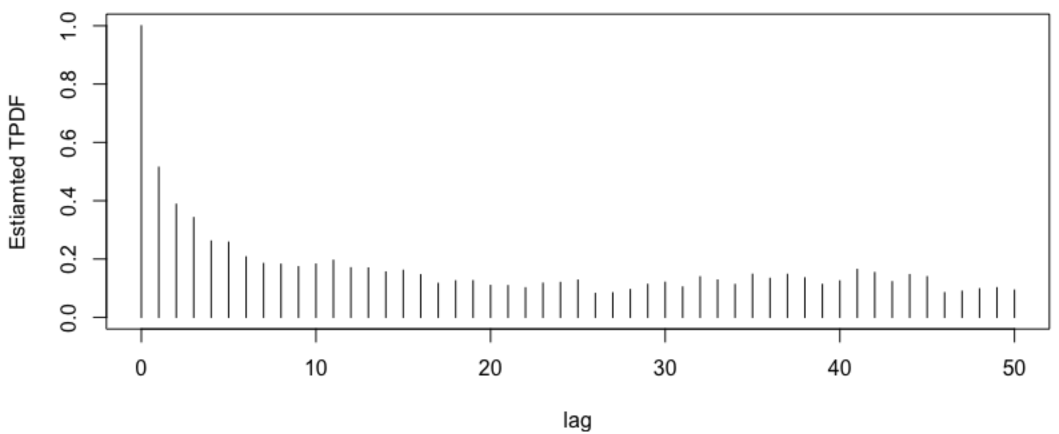
$$\sigma(X_t, X_{t+h}) = \int_{\Theta_1^+} w_t w_{t+h} \mathrm{d}H_{X_t, X_{t+h}}(\boldsymbol{w}).$$

  - H is angular measure, using L_2 norm.
  - Shown to be a positive definite function (like ACVF).

- Tail stationarity:

$$\sigma(X_t, X_{t+h}) = \sigma(h); \ \forall t.$$

$\rightarrow$ Analysis focuses on capturing TPDF.



Transformed-Linear Models for Time Series

Consider TS models of the form

$$X_t = \bigoplus_{j=-\infty}^{\infty} \psi_j \circ Z_{t-j},$$

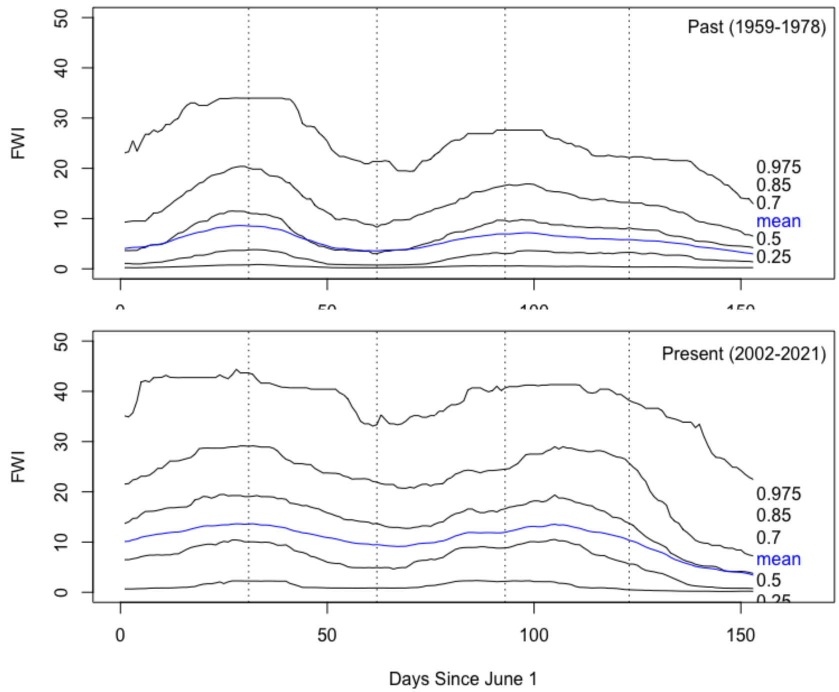
where $\{Z_t\}$ are indep, "tail stationary", $\in RV_+(\alpha = 2)$.

- X_t is non-negative, coefficients $\psi_j \in \mathbb{R}$.
- convergence wp 1 if $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

Trans-linear ARMA models

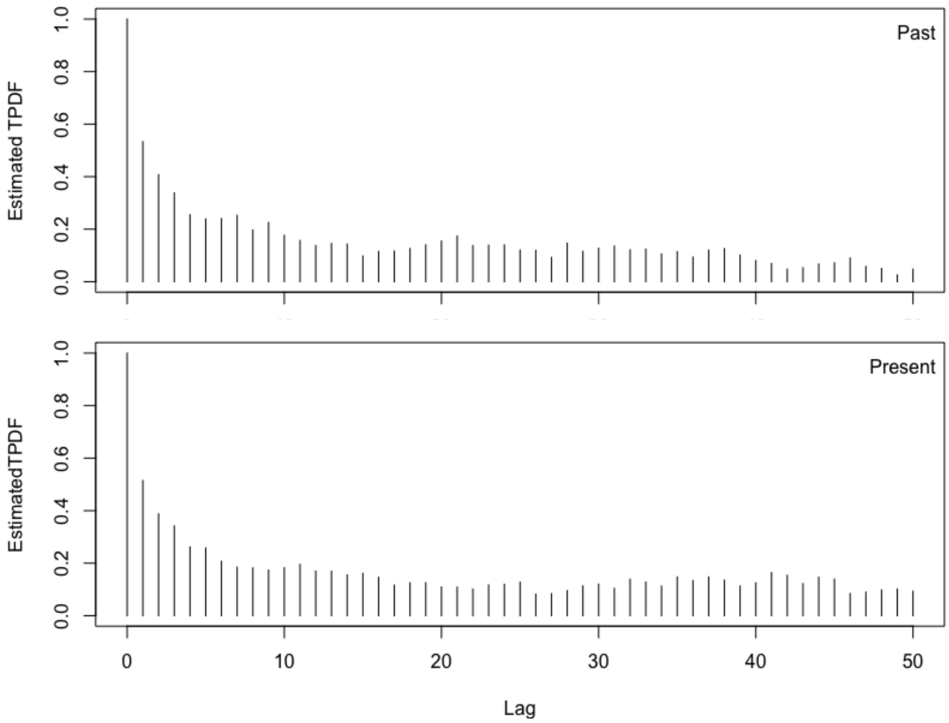
- MA(q)
- MA(∞); existence via tail analogue to MS convergence.
- AR(1); defined via trans-linear backshift operator, show staty soln.
- ARMA(1,1)
- ARMA(p,q)

Estimating and Transforming the Marginal



- Apparent shift upward under present climate vs. past.
- Unexpected bimodal seasonality. Arises from ISI (wind).
- Marginal estimated and data converted to stationary regularly varying $\alpha = 2$.

TPDF Estimation



- Similar tail dependence past and present.
- Stronger short-range dependence (ISI).
- Some mid-range dependence (BUI).
- Long-range dependence likely due to estimator bias.

Fitting a Trans-linear RV Model

Approaches

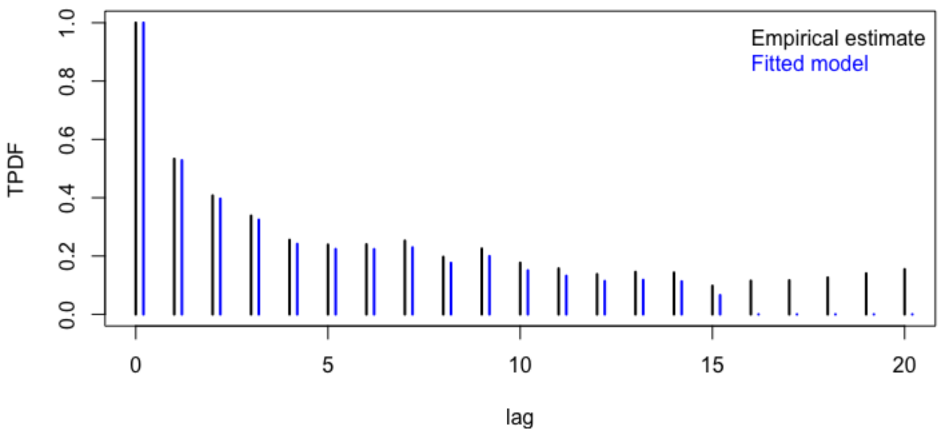
1. Minimize model/empirical TPDF squared error difference.
   - ◦ Closed-form TPDFs for MA(q), AR(1), ARMA(1,1).
   - ◦ Numerically fit higher order, ARMA(3,3)
   - – Natural questions about lags and weights.
2. Use innovations algorithm to fit an MA(q) of any order.
   - \+ Method-of-moments-like use of empirical TPDF.
   - – Larger number (q) of model parameters.

- Both methods have need to perform model selection.

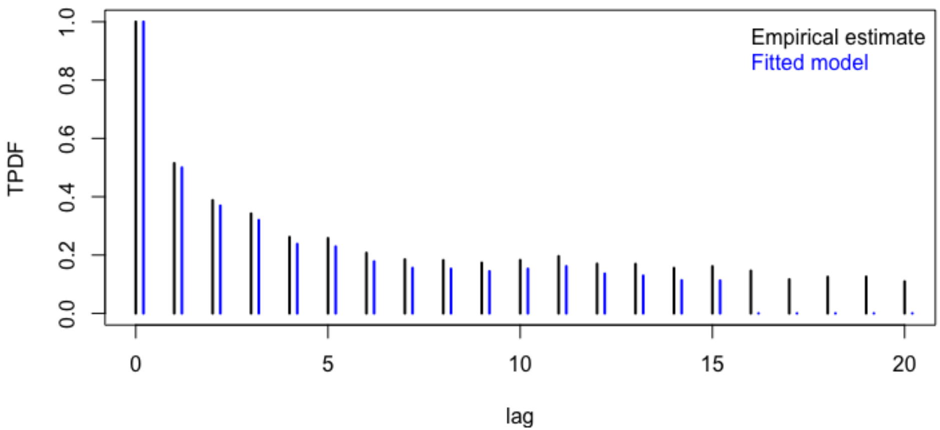
Choosing the order of the MA(q)

Select MA(15); (run-length and high quantiles of sums)

Past



Current



Probability Est. and Attribution for 2020 Season

Quantile	T-hold	# 2020	Past	Present	Ratio
95%	34.4	22	0.003 (0.000, 0.010)	0.037 (0.005, 0.077)	12.4 (1.3, ∞)
97.5%	39.5	14	0.013 (0.005, 0.0054)	0.054 (0.012, 0.099)	4.1 (0.36, 8.7)
99%	44.9	6	0.020 (0.012, 0.185)	0.193 (0.128, 0.303)	9.7 (0.82, 16.5)

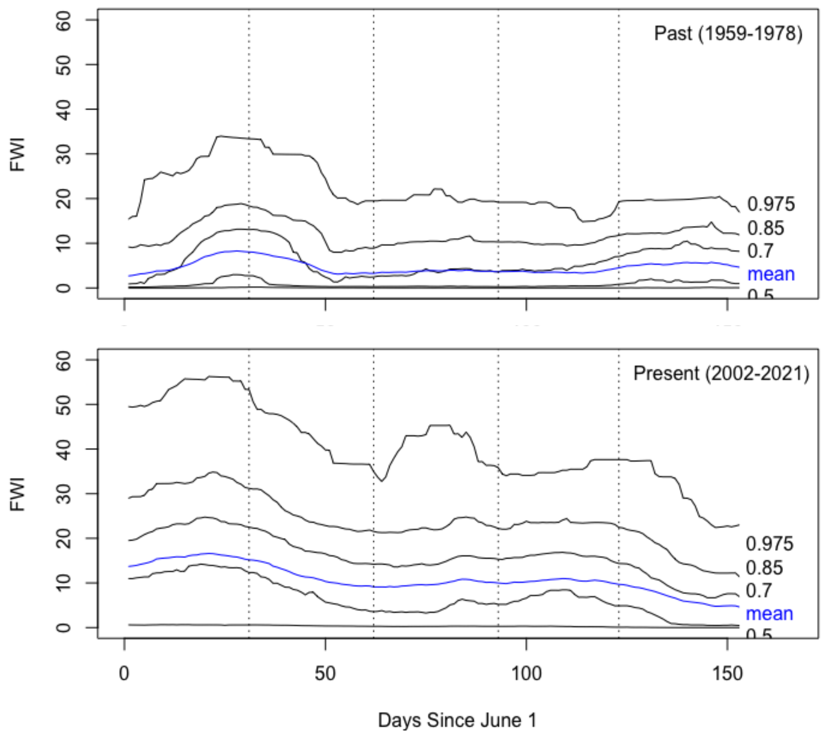
- Estimates not sensitive to MA(q) order ($q = 20$).
- Difference arising mostly b/c change in marginal.
- Lots of modeling choices.

Attribution Statement:

According to our model fitted to the ERA Reanalysis data, the probability of seeing N. Colorado fire weather as intense as 2020 is at least four times more likely under the current climate as compared to climate of fifty years ago.

But wait! There's more...

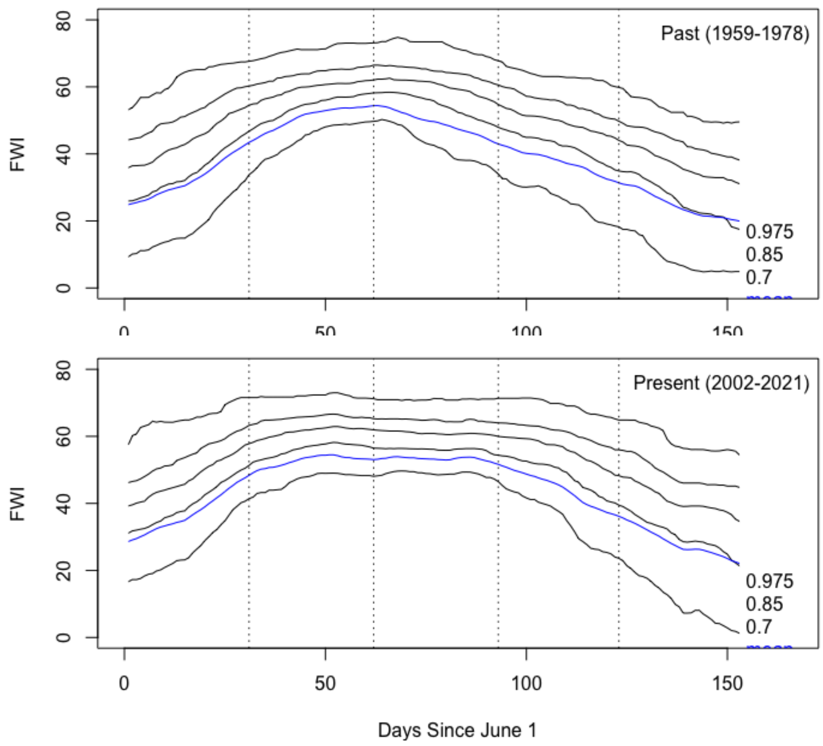
Northern Colorado Weather Station Data



- Higher FWI values. Larger shift in marginal.
- Also, stronger tail dependence in present.

Attribution: 2020 a 0% event under earlier climate.

Northern California ERA Data

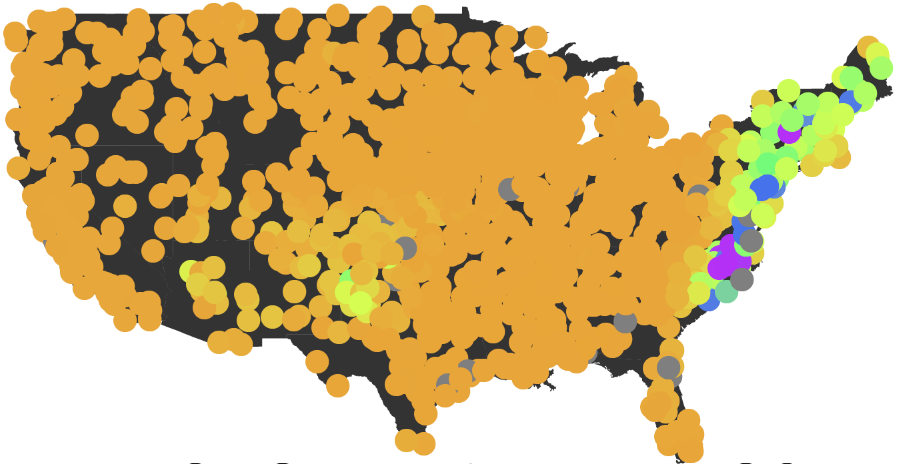


- FWI values much higher than in Colorado.
- Less apparent shift in FWI climate.

Extremal PCA and Precipitation Trends

“Scientist: Let’s investigate trends in extreme precipitation.”

Hurricane Floyd, Precip Sept 16-18, 1999



- Data from 1140 GHCN stations over CONUS, 1950-2016.
- Focus on hurricane season (August to October).
- 6162 total days.
- 3-day moving average to account for asynchronicity.
- ENSO index also collected for ASO, 1950-2016.

PCA: Eigen-decompostion of Σ

How? Σ is positive definite: $\Sigma = UDU^T$,
$U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)$ is unitary, D is diagonal w/ $\lambda_1 \geq \ldots \geq \lambda_p > 0$.

Why? $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ is a basis for $\mathbb{R}^p$, $\boldsymbol{e}_i = t(\boldsymbol{u}_i)$ are basis for $\mathbb{X}^p$
$\Rightarrow$ for any realization $\boldsymbol{x} \in \mathbb{X}^p$ $\exists$ representation

$$\begin{aligned} \boldsymbol{x} &= v_1 \circ \boldsymbol{e}_1 \oplus \ldots \oplus v_p \circ \boldsymbol{e}_p \\ &= U \circ t(\boldsymbol{v}) \end{aligned}$$

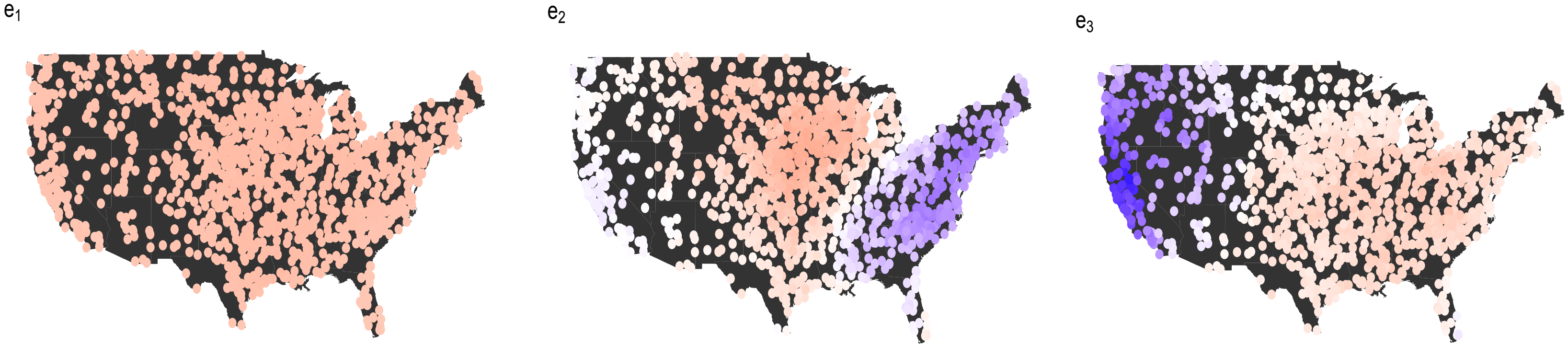
So? Eigenvectors u_i are ordered in importance.

Define PC's $\boldsymbol{V} := U^T t^{-1}(\boldsymbol{X})$.

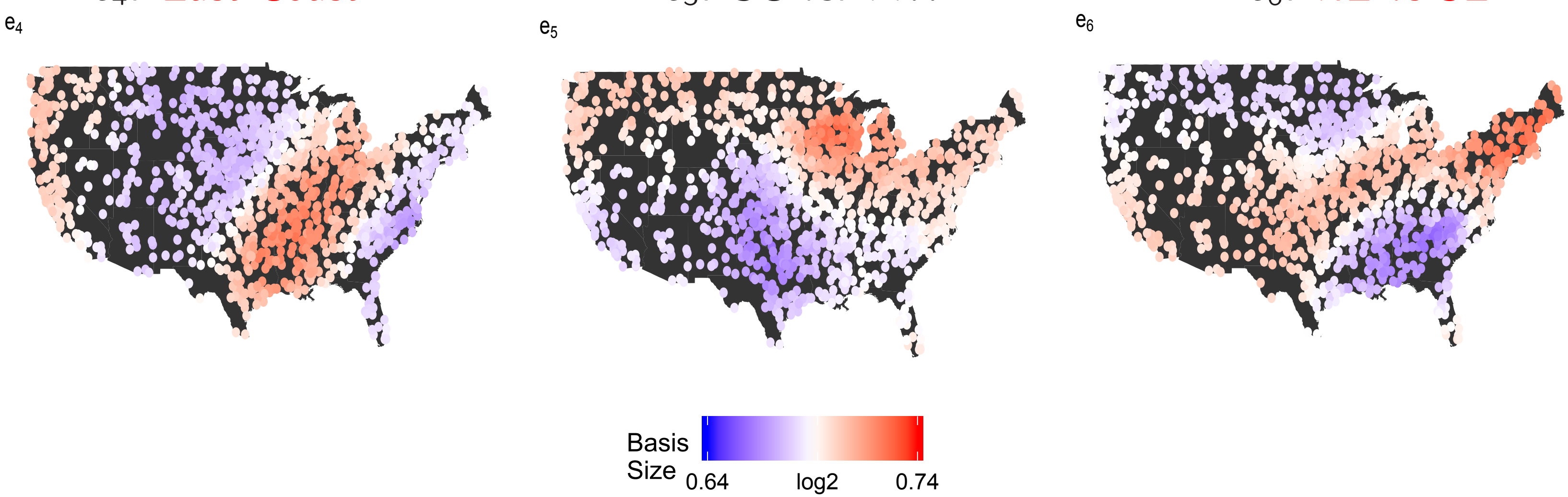
- $\boldsymbol{V} \in RV^p(\alpha = 2)$
- Can observe behavior through the lens of the basis, by looking at time series of PC's.

Eigenvectors

e_1: Continental — e_2: East Third — e_3: West Coast

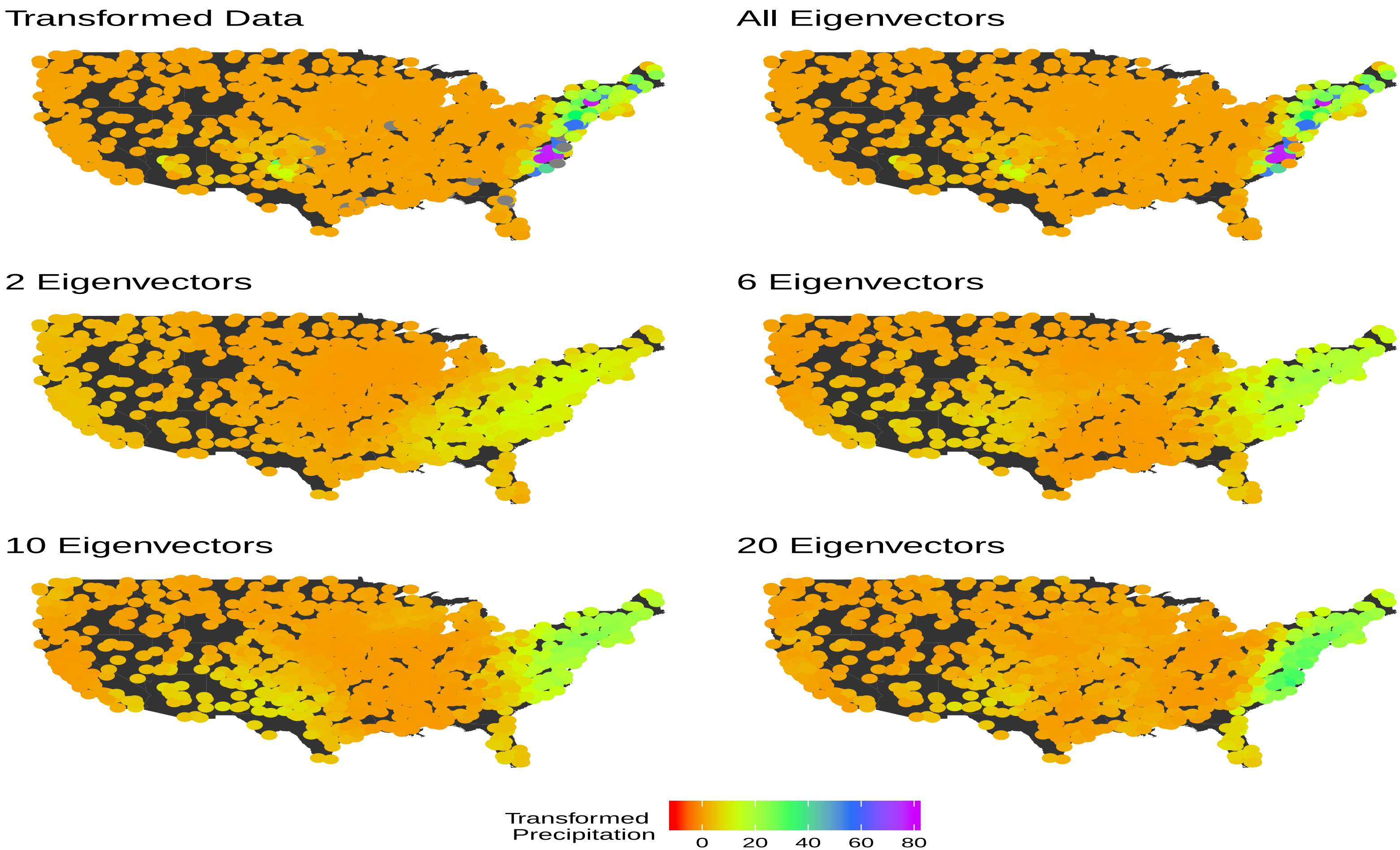


e_4: East Coast — e_5: SC vs. MW — e_6: NE vs SE



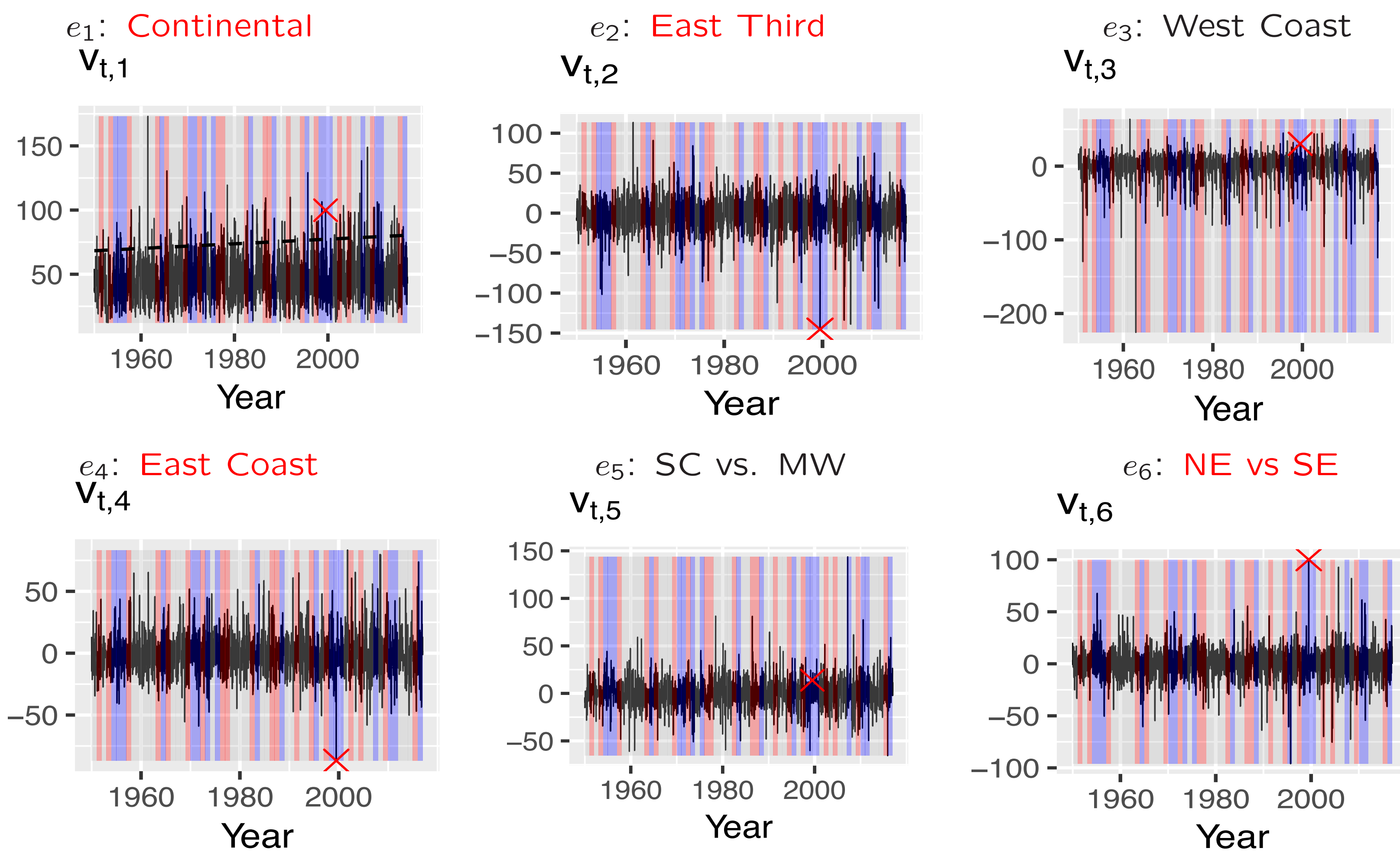
Reconstruction using Eigenvectors

September 16, 1999, Hurricane Floyd

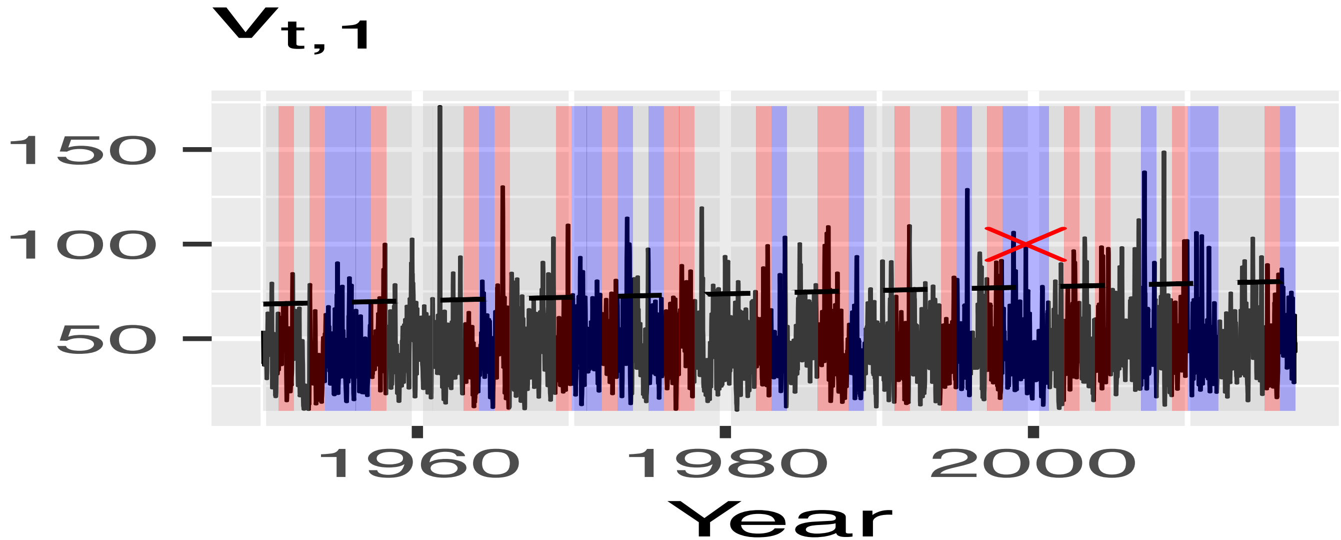


Original data, full (1140) reconstruction, 2, 6, 10, 20 eigenvectors

Time Series of Eigenvalues



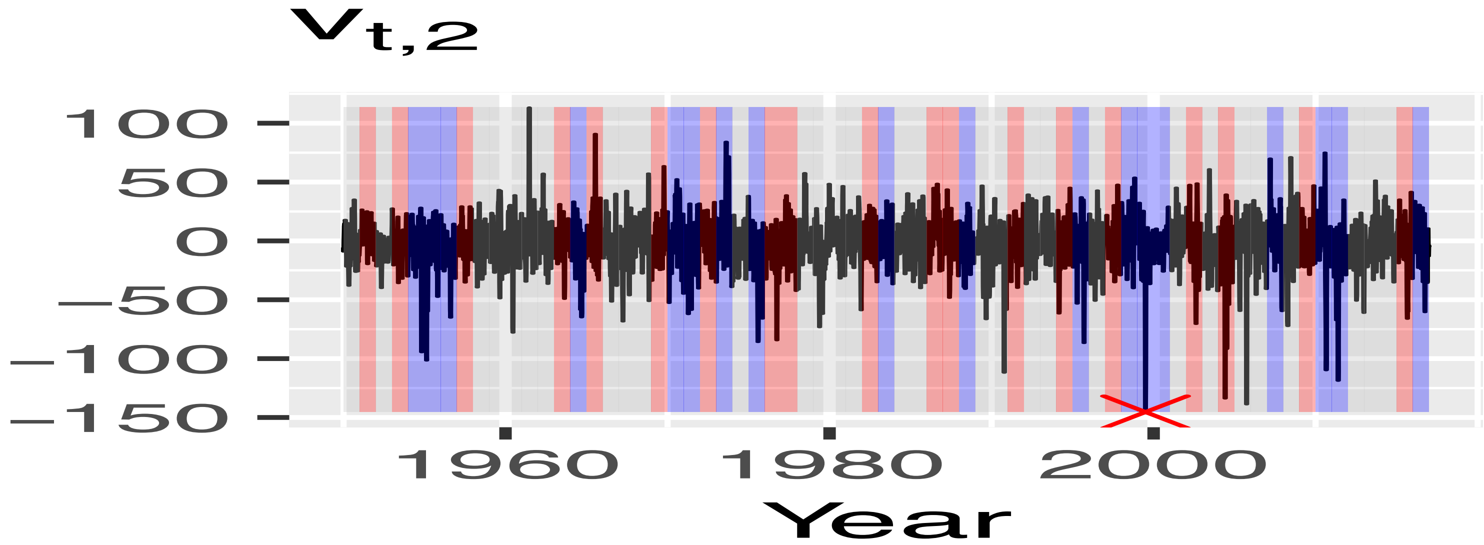
Time Series for PC1 (Continental signal)



Quantile regression at $q_{0.95}$:

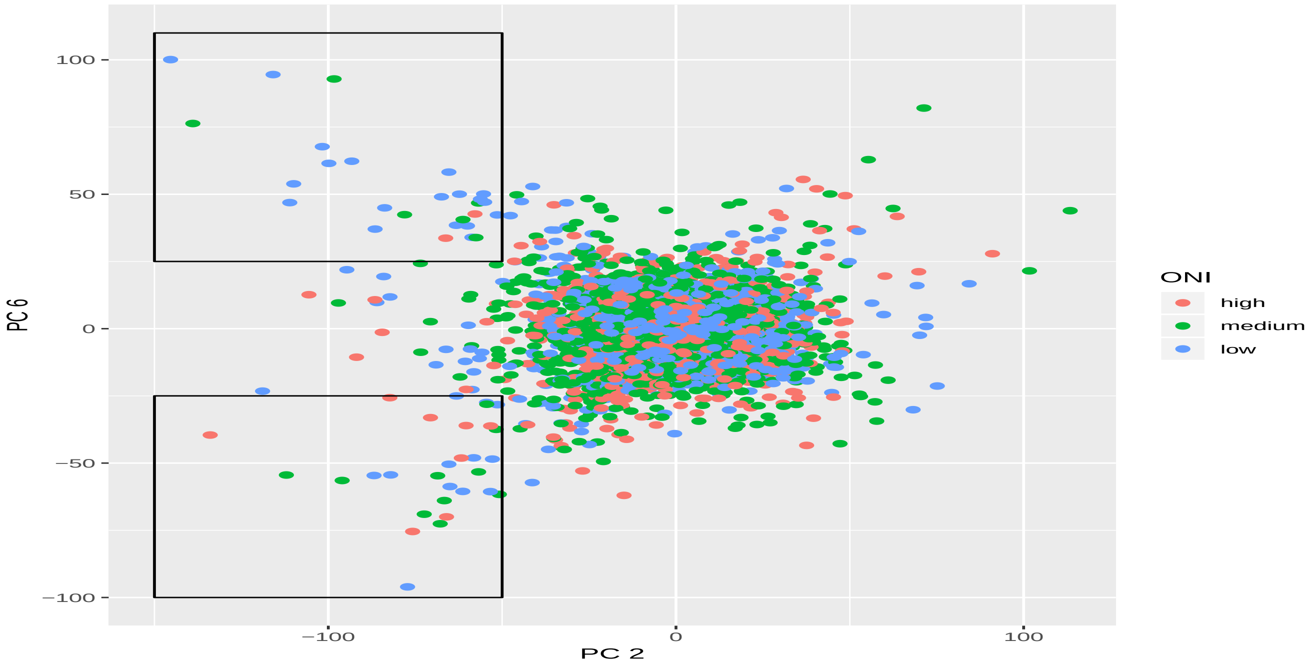
- $\widehat{\beta}_1 = 0.002$ units/year.
- p-value: 0.
- Conclusion: There is a significant positive trend in the continental signal for extreme precipitation.

Time Series for PC2 (East Coast signal)



- Large *negative* scores correspond to east coast precip.
- Low ENSO effect on exceedances of 'negative' .95 q-tile?
  - $\widehat{p}_{low} = .067$; $\widehat{p}_{notlow} = .043$; p-value = 0.0003.
- H_0: No difference in dist of large negative scores?
  - LRT on fitted GPD's, p-value = .0475.
- Low ENSO affects both freq and distn of negative exc.

Loadings 2 and 6: NE vs SE



- Blue box: Extreme precip in NE.
- Red box: Extreme precip in SE.
- Are proportions of ENSO phases the same?
  - p-value 0.022.

Wrap-up

Via the TPDM and transformed linear operations, we are able to construct useful models and methods to answer interesting questions about high-dimensional extreme value problems.

- Models/methods are relatively simple and familiar.
- Allow for new types of extremal investigations.
- Methods most useful when extremal dependence differs from that in center of distribution.
- Additional work:
  - SAR model and applied to gridded spatial data.
  - Linear prediction for extremes.
  - Partial tail correlation.