Conditional Decomposition for Modeling Concurrent Extremes

Whitney Huang
School of Mathematical and Statistical Sciences
Clemson University

School of MATHEMATICAL AND STATISTICAL SCIENCES Clemson University

✉: wkhuang@clemson.edu
🖱: https://whitneyhuang83.github.io/

Celebration of
Statistics and Data Science

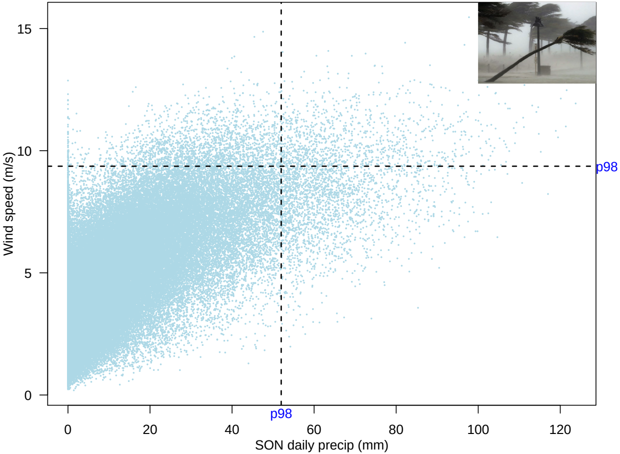
Cornell Bowers CIS Statistics and Data Science Ithaca, NY, September 6, 2024

Concurrent Extreme Events


Credit: Shuttershock (left); www.standardmedia.co.ke (right)

- Compound drought and extreme heat e.g., [Zscheischler and Seneviratne, 2017]
- Concurrent wind and precipitation extremes e.g., [Martius et al., 2016]

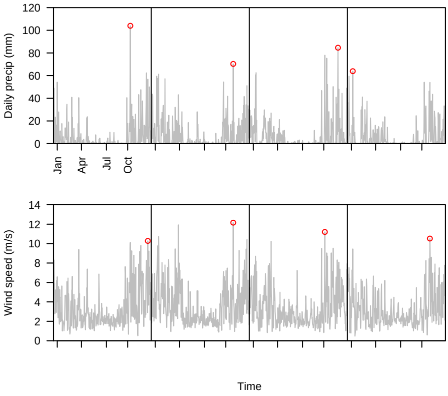
Concurrent Wind and Precipitation Extremes

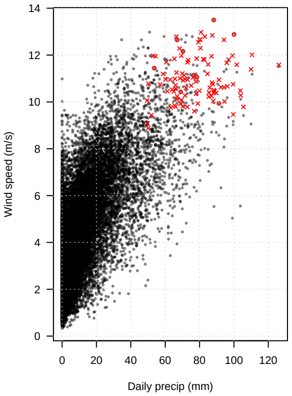


- Most climate literature focuses on *estimating the probability of concurrent extreme events*
- Aim to estimate the "**tail distribution**" via a conditional approach

Classical Multivariate Extreme Value Analysis

Modeling componentwise maxima using multivariate extreme value distribution ($\Rightarrow$ extreme-value marginals & tail copula)





Issue: Ignores event simultaneity ☹[1]

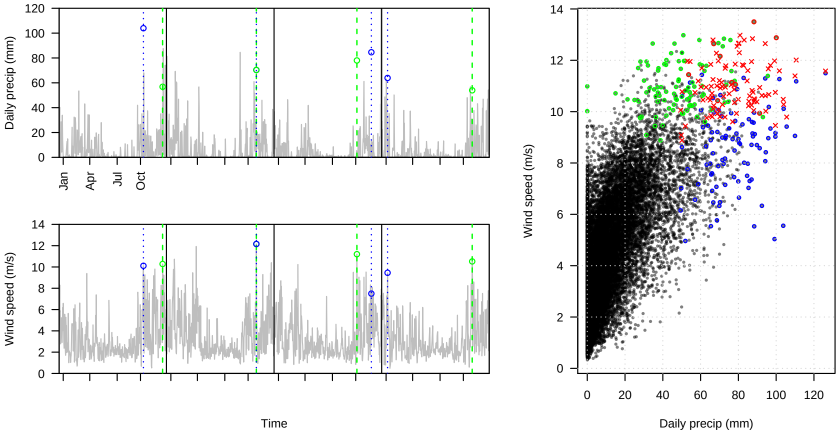
[1] *"...In practice, modeling vectors of annual maxima seems less than ideal, and it is not clear how much dependence information is lost by discarding the coincident data..."* – Cooley and Sain, Stat. Sci. 2012

Componentwise Maxima vs. Concomitants of Maxima

Red: (annual max precip, annual max wind speed)

Blue: (annual max precip, concurrent wind speed)

Green: (annual max wind speed, concurrent precip)



Conditional Approaches for Estimating Concurrent Extremes:

$$[Y, X_{\text{large}}] = \underbrace{[X\ _{\text{large}}]}_{\text{EVA}}\ \underbrace{[Y|X_{\text{large}}]}_{?}$$

i) Quantile regression

ii) Conditional extreme value models

iii) Composition of distributions & distributional regression

$$\begin{aligned}[Y, X_{\text{large}}] &= [X_{\text{large}}][Y|X_{\text{large}}] \\ &= \int [X_{\text{large}}][Y_{\text{large}}|X_{\text{large}}][Y|X_{\text{large}}, Y_{\text{large}}]\, dY_{\text{large}}\end{aligned}$$

Approximating $[Y|X_{\text{large}}]$ via Quantile Regression [Koenker and Bassett, 1978]

- **Goal**: To estimate the conditional upper quantiles, i.e., estimating $Q_Y(\tau|x) = \inf\{y : F(y|x) \geq \tau\}$, $\tau \in (0, 1)$ at a finite number of quantile levels $\tau_1, \tau_2, \cdots, \tau_J$

- Estimating each quantile separately can lead to the issue of quantile curves crossing i.e.,

$$Q_Y(\tau_i|x) > Q_Y(\tau_j|x)$$

for some $x \in \mathbb{R}$ when $0 < \tau_i < \tau_j < 1$ ☹

- We use the monotone composite quantile regression neural network [MCQRNN, Cannon, 2018] to estimate multiple **non-crossing**, **nonlinear** conditional quantile functions **simultaneously**[2]

[2] Xu and Reich (2023) developed another Bayesian approach to deal with this task

Estimating $[Y|X_{\text{large}}]$ via Extreme Value Approach

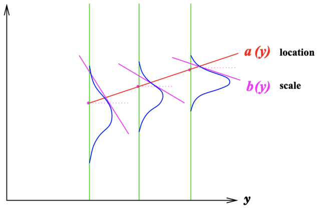
Conditional extreme value ($\texttt{CEV}$) models [Heffernan & Tawn, 04]: models $[Y|X_{\text{large}}]$ (**nonparametrically**) by assuming a **parametric** location-scale form **after marginal transformation**

Marginal modeling:

- Estimate marginal distributions of Y and X
- Transform $(Y, X)^T$ to Laplace marginals

Dependence modeling:

- $\left[\frac{\tilde{Y}-a(\tilde{X})}{b(\tilde{X})} \leq z | \tilde{X} > u\right] \overset{u \text{ large}}{\sim} G(z).$
- Assumes $a(x) = \alpha x$ and $b(x) = x^{\beta}$, $\alpha \in [-1, 1]$, $\beta \in (-\infty, 1)$

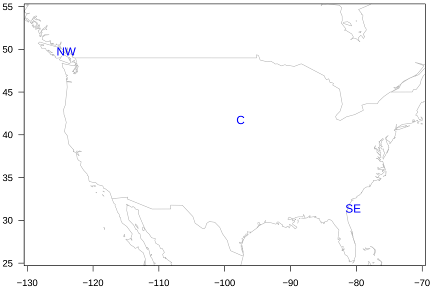


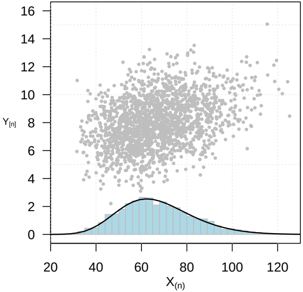
Source: Heffernan's slides give that Interface 2008 Symposium

"Data": Large Ensemble Climate Simulations

Output from Canadian Regional Climate Model 4 [CanRCM4, Scinocca et al., 2016]

- 35-member initial-condition ensemble
- Using output from 1950-1999 with CMIP5 historical forcing
- 0.44° horizontal grid ($\sim$ 50 km). We will show the results from a "Vancouver" (NW) grid cell

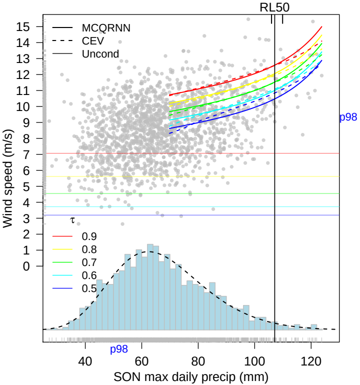




Estimating Concurrent Extremes using Large Ensemble[3]

Each run in the ensemble produces (nearly) statistically independent realizations of climate system, enabling us to:

- provide more accurate estimates in climate extremes
- assess how well statistical procedures work



- SON max precipitation ↑ concurrent wind speed ↑
- Conditional quantiles are significantly larger than unconditional ones
- `MCQRNN` and `CEV` provide similar upper quantile wind speed estimates

[3] H., Monahan, & Zwiers, *Weather and Climate Extremes*, 2021

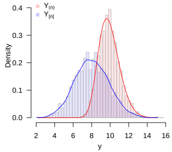
New Approach

I) Introduce an Auxiliary Variable $Y_{(n)}$ (Y_{large})

$$
\begin{aligned}
[Y_{[n]}, X_{(n)}] &= [X_{(n)}][Y_{[n]}|X_{(n)}] \\
&= \int [X_{(n)}][Y_{(n)}|X_{(n)}][Y_{[n]}|X_{(n)}, Y_{(n)}]\, dY_{(n)}
\end{aligned}
$$

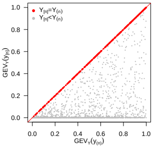
II) Composition of Distributions:[4]

Assume $Y_{[n]} \sim \mathrm{GEV}_Y^{-1}(H^{-1}(U))$, where GEV_Y is the CDF of $Y_{(n)}$, and H is a CDF on $[0, 1]$



III) Beta Distributional Regression:

H is modeled as a smooth function of $x_{(n)}$ and $y_{(n)}$, using beta regression



[4] Motivated by Naveau et al. (2016)

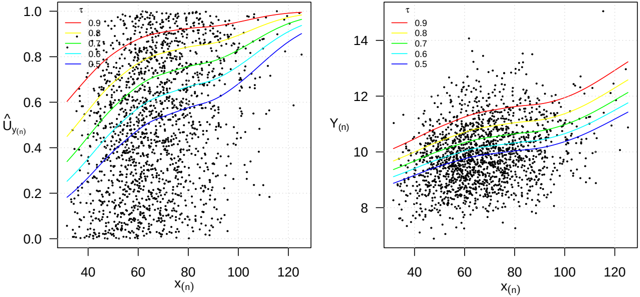
Step I: Modeling $[Y_{(n)}, X_{(n)}]$ as $[X_{(n)}][Y_{(n)}, X_{(n)}]$

- Apply probability integral transform:
  $U_{(n)} = \mathrm{GEV}_Y(Y_{(n)}) \approx U[0,1]$

- Model how $U_{(n)}$ depends on $X_{(n)} = x_{(n)}$ via beta regression

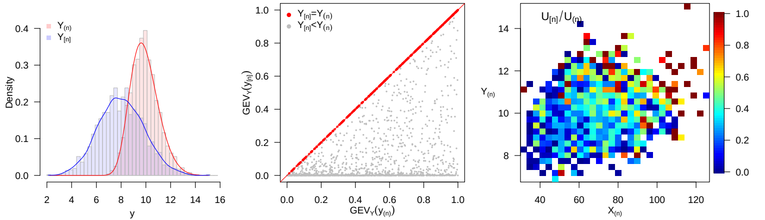
$$[U_{(n)}|X_{(n)} = x_n] \sim \mathrm{Beta}\left(\alpha\left(x_n\right), \beta\left(x_n\right)\right),$$

where $\alpha\left(x_n\right)$ and $\beta\left(x_n\right)$ are estimated using vector generalized additive models [Yee, 2015]



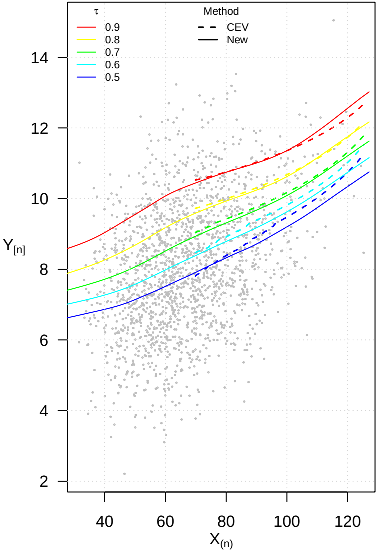
Step II: Modeling $[Y_{[n]}|Y_{(n)}, X_{(n)}]$

- Apply probability integral transform: $U_{[n]} = \mathrm{GEV}_Y(Y_{[n]})$
- Fit a non-parametric logistic regression (e.g., via `gam`, [Hastie & Tibshirani, 1996], [Woods, 2017]) to model how $\mathbb{1}_{\{Y_{[n]}=Y_{(n)}\}}$ depends on $y_{(n)}$ and $x_{(n)}$
- Model $\frac{U_{[n]}}{U_{(n)}} \mathbb{1}_{\{Y_{[n]}<Y_{(n)}\}}$ using beta regression (with parameters in generalized additive forms)



Step III: Marginalized out $Y_{(n)}$ to Estimate $[Y_{[n]}|X_{(n)}]$

- Simulate $X_{(n)}$ and then $Y_{(n)}$ given $X_{(n)} = x_{(n)}$
- Simulate $\mathbb{1}_{\{Y_{[n]}=Y_{(n)}\}}$ and $\frac{U_{[n]}}{U_{(n)}}\mathbb{1}_{\{Y_{[n]}<Y_{(n)}\}}$ given $X_{(n)} = x_{(n)}$ and $Y_{(n)} = y_{(n)}$
- Integrate out $y_{(n)}$ and apply probability integral transforms to estimate $[Y_{[n]}|X_{(n)}]$



Summary and Discussion

Summary

- We explore various conditional approaches to estimate concurrent extremes
- Large climate model ensemble is a powerful tool for studying climate extremes

Discussion

- On which variable should we condition? Comparison with joint modeling?
- Conditional thresholds exceedance version? Conditional bulk-and-tails version?
- How to handle the higher-dimensional case?
- Theoretical properties? Uncertainty estimates?

Future Directions

- Nonstationary extension account for both seasonality and long term trend for marginal and dependence structures
- Spatial extension to borrow strength across space to improve estimation of concurrent extremes
- Spatial concurrent extremes