Conditional Decomposition Approach for Modeling Concurrent Extreme Events

Whitney Huang
School of Mathematical and Statistical Sciences
Clemson University

School of MATHEMATICAL AND STATISTICAL SCIENCES Clemson University

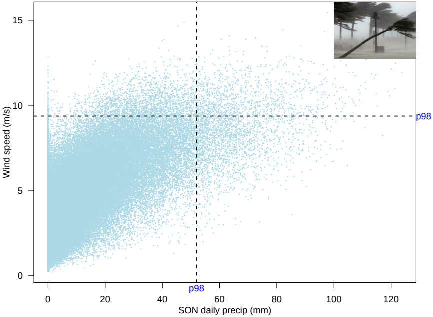
✉: wkhuang@clemson.edu
🖱: https://whitneyhuang83.github.io/

15th International Meeting on Statistical Climatology
Session on Extreme Value Analysis Methods & Theory for Climate Applications

Météo-France, Toulouse, , June 24, 2024

Concurrent Extremes: Examples & Research Question

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

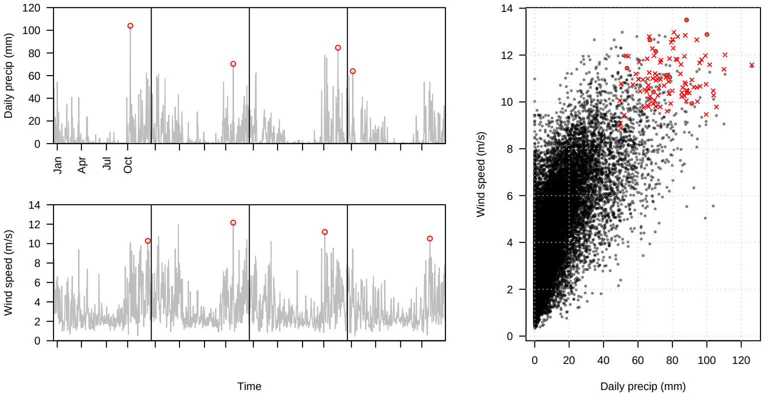


- Compound drought and extreme heat e.g., [Zscheischler and Seneviratne, 2017]
- Concurrent wind and precipitation extremes e.g., [Martius et al., 2016]

- Most (climate) literature focus on *estimating the occurrence probability of an concurrent extreme event*
- Want to estimate the "**tail distribution**" via a conditional approach

Classical Multivariate Extreme Value Analysis

Modeling componentwise maxima using multivariate extreme value distribution (⇒ extreme-value marginals & tail copula)



Issue: Ignore the event simultaneity ☹[1]

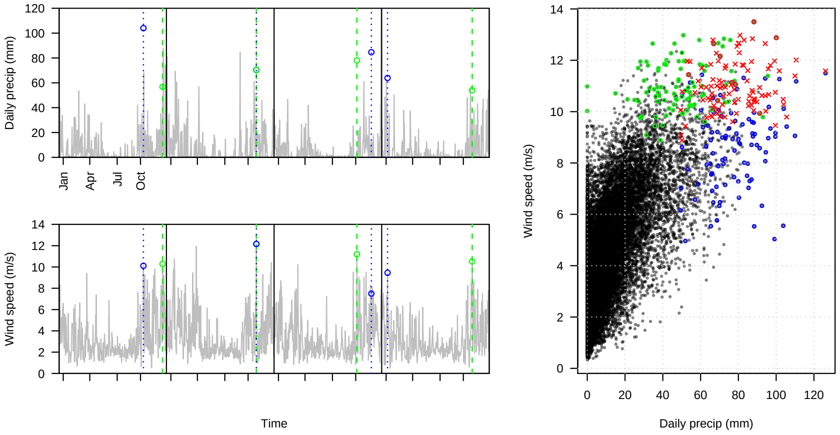
[1] "...*In practice, modeling vectors of annual maxima seems less than ideal, and it is not clear how much dependence information is lost by discarding the coincident data*..." – Cooley and Sain, Stat. Sci. 2012

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

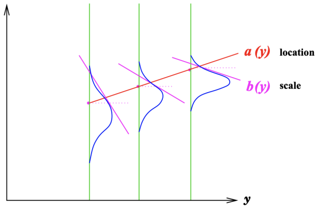
Conditional extreme value (CEV) models [Heffernan & Tawn, 04]: models $[Y|X_{\text{large}}]$ (**nonparametrically**) by assuming a **parametric** location-scale form **after marginal transformation**

Marginal modeling:

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

Dependence modeling:

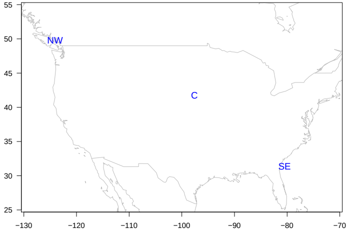
- $\left[\frac{\tilde{Y}-a(\tilde{X})}{b(\tilde{X})} \leq z|\tilde{X} > u\right] \overset{u \text{ large}}{\sim} G(z).$
- Assumes $a(x) = \alpha x$ and $b(x) = x^{\beta}$, $\alpha \in [-1, 1]$, $\beta \in (-\infty, 1)$

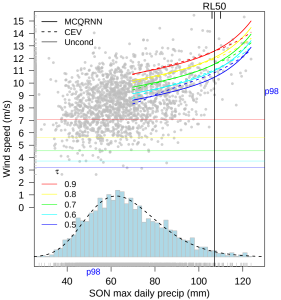


Source: Heffernan's slides give that Interface 2008 Symposium

Estimating Concurrent Extremes using Large Ensemble Climate Simulations[3] [H., Monahan, & Zwiers, WCE, 2021]

- 35-member initial-condition ensemble
- Output from 1950-1999
- 0.44° horizontal grid ($\sim$ 50 km).





- SON max precipitation ↑ concurrent wind speed ↑
- `MCQRNN` and `CEV` yield reasonably close wind speed upper quantile estimates
- Conditional quantiles are substantially larger than their unconditional counterparts

[3] We used output from Canadian Regional Climate Model 4 [CanRCM4, Scinocca et al., 2016]

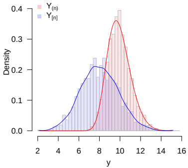
New Approach

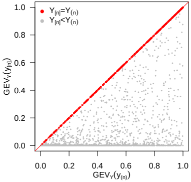
i) Introduce an Auxiliary Variable $Y_{(n)}$ (Y_{large})

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

ii) Composition of Distributions:[4]

$$Y_{(n)} \sim \text{GEV}_Y^{-1}(U)$$
$$Y_{[n]} \sim \text{GEV}_Y^{-1}(H^{-1}(U))$$





iii) Beta Distributional Regression:
H is modeled as a smooth function of $x_{(n)}$ and $y_{(n)}$, using a Beta Regression

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

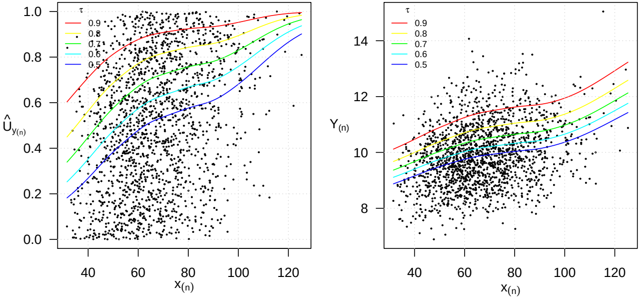
Step I: Modeling $[Y_{(n)}|X_{(n)}]$

- Given that $X_{(n)}$ and $Y_{(n)}$ can both be approximated by GEV, $[Y_{(n)}|X_{(n)}]$ can be obtained by taking the derivative of the copula function of $[Y_{(n)}, X_{(n)}]$

- However, the following could enrich this dependence structure.

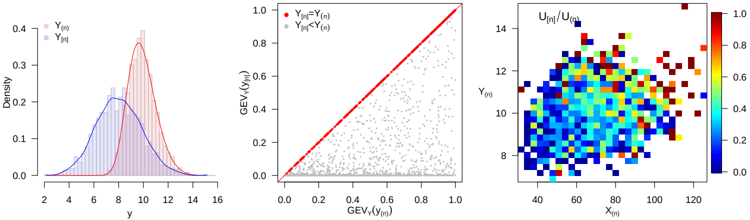
$$[\hat{U}_{Y_{(n)}}|X_{(n)} = x_n] \sim \text{Beta}\left(\alpha\left(x_n\right), \beta\left(x_n\right)\right),$$

where $\alpha\left(x_n\right)$ and $\beta\left(x_n\right)$ are estimated using `VGAM`, [Yee, 2015]



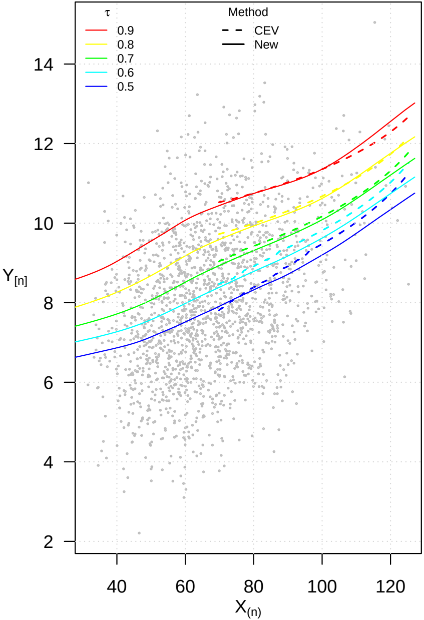
Step II: Modeling $[Y_{[n]} | Y_{(n)}, X_{(n)}]$

- Convert $Y_{[n]}$ to $U_{[n]} = \hat{\mathrm{GEV}}_Y(Y_{[n]})$
- 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 $\frac{U_{[n]}}{U_{(n)}} \mathbb{1}_{\{Y_{[n]} < Y_{(n)}\}}$ via Beta regression (with parameters as 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 estaimte $[Y_{[n]}|X_{(n)}]$



Summary and Discussion

Summary

- We explore 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?
- How to compare with Multivariate GPD?
- How to handle the higher-dimensional case?
- Comparison with joint modeling
- Theoretical properties?

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