Estimating concurrent climate extremes: A conditional approach*

Whitney Huang
CLEMSON UNIVERSITY

Joint work with Adam Monahan and Francis Zwiers

BIRS-UBCO, Climate Change Scenarios and Financial Risk
July 4, 2022

PACIFIC CLIMATE IMPACTS CONSORTIUM
University of Victoria
CANSSI INCASS
samsi NSF Duke NCSU UNC

*Huang, Whitney K., Adam H. Monahan, and Francis W. Zwiers. "Estimating concurrent climate extremes: A conditional approach." Weather and Climate Extremes (2021): 100332.

Outline of the talk

- Concurrent extremes: **simultaneous** occurrence of extreme values for **multiple** climate variables [Zscheischler et al., 2018]

- Conditional approaches for estimating concurrent extremes:

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

  - ***Quantile regression***
  - ***Conditional extreme value models***

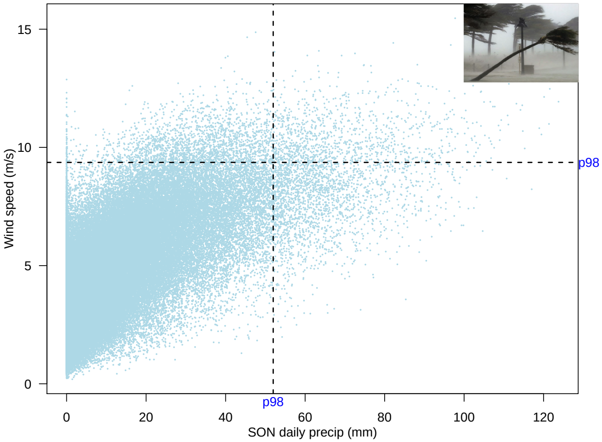
- Estimating concurrent extremes using a large ensemble climate simulations
  - Estimating concurrent wind and precipitation extremes
  - Illustrating the use of **large ensemble climate model simulations** to study **extremes**

Some examples of concurrent extreme events

Credit: Shuttershock

Source: www.standardmedia.co.ke

Concurrent wind and precipitation extremes



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

Conditional approaches for estimating concurrent extremes:

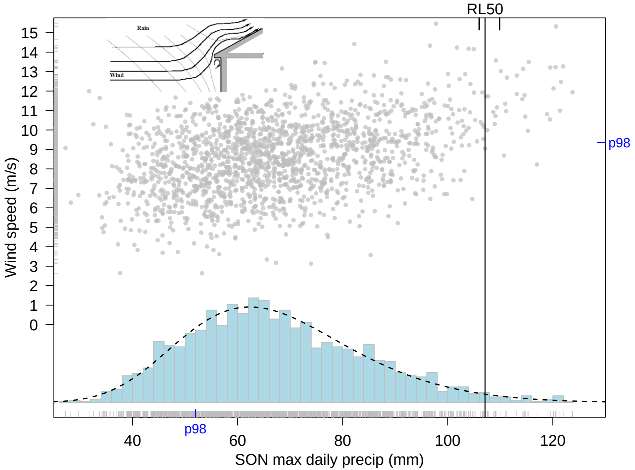
$$[Y, X \text{ large}] = \underbrace{[X \text{ large}]}_{\text{EVA}} \underbrace{[Y|X \text{ large}]}_{?}$$

- *Quantile regression*
- *Conditional extreme value models*

An illustration of conditional approach

Let X and Y be daily precipitation and wind speed

1. Condition on X being "large" e.g., annual maximum



Question: Which distribution to use to model $[X \text{ large}]$?

Extremal Types Theorem (Fisher–Tippett 1928, Gnedenko 1943)

Define $M_n = \max\{X_1, \cdots, X_n\}$ where $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} F$. If $\exists\, a_n > 0$ and $b_n \in \mathbb{R}$ such that, as $n \to \infty$, if

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \le x\right) \xrightarrow{d} \mathsf{G}(x)$$

then G must be the same type of the following form:

$$\mathsf{G}(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]_+^{\frac{-1}{\xi}}\right\}$$

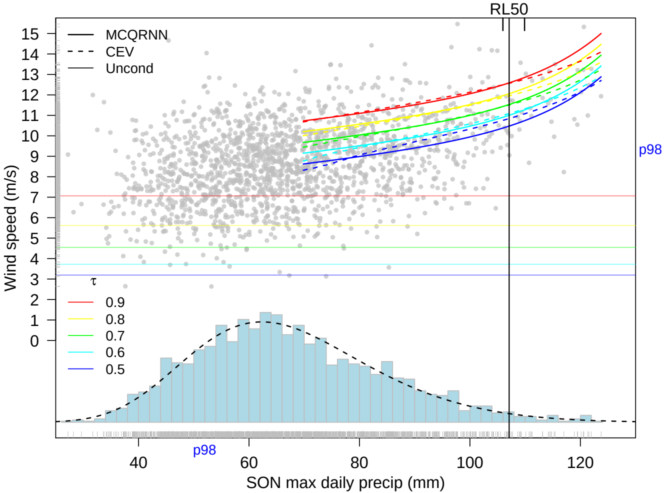
where $x_+ = \max(x, 0)$ and $G(x)$ is the distribution function of the generalized extreme value distribution (GEV(μ, σ, ξ))

- μ and σ are location and scale parameters
- ξ is a shape parameter determining the rate of tail decay, with
  - $\xi > 0$ giving the heavy-tailed case (Fréchet)
  - $\xi = 0$ giving the light-tailed case (Gumbel)
  - $\xi < 0$ giving the bounded-tailed case (reversed Weibull)

An Animation

An illustration of conditional approach

1. Condition on X being "large" e.g., annual maximum
2. Model $[Y|X\text{"large"}]$



Next, we will talk about the approaches we use for 2

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**

Estimating $[Y|X\text{“large”}]$ via Extreme Value Approach

Conditional extreme value (CEV) models [Heffernan & Tawn, 04]: models the conditional distribution by assuming a **parametric** location-scale form after marginal transformation

- **Marginal modeling**:
  1. Estimate marginal distributions of Y and X
  2. Transform $(Y, X)^T$ to Laplace marginals $(\tilde{Y}, \tilde{X})^T$
- **Dependence modeling**:

  Assume for large u,

$$\left[\frac{\tilde{Y} - a\left(\tilde{X}\right)}{b\left(\tilde{X}\right)} \leq z | \tilde{X} > u\right] \sim G(z),$$

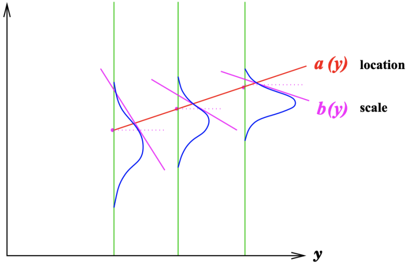
where $a(x) = \alpha x$ and $b(x) = x^{\beta}$, $\alpha \in [-1, 1]$, $\beta \in (-\infty, 1)$

A cartoon illustration of the CEV dependence modeling

Assume for large u,

$$\left[\frac{\tilde{Y} - a\left(\tilde{X}\right)}{b\left(\tilde{X}\right)} \le z | \tilde{X} > u\right] \sim G(z),$$

where $a(x) = \alpha x$ and $b(x) = x^{\beta}$, $\alpha \in [-1, 1]$, $\beta \in (-\infty, 1)$



Source: Heffernan's slides given at the Interface 2008 Symposium

- $\tilde{Y} = \alpha\tilde{X} + \tilde{X}^{\beta} Z$,
  $\Rightarrow Z = \frac{\tilde{Y} - \alpha\tilde{X}}{\tilde{X}^{\beta}} \sim G$
- α and β are estimated by making a parametric assumption of $\tilde{Y}$
- G estimated nonparametrically

"Data"

Output from CanRCM4, Canadian Regional Climate Model 4

- 35-member initial-condition ensemble
- Using output from 1950-1999 with CMIP5 historical forcings
- North American region, 0.44° horizontal grid (~ 50 km). We will show the results from a "Vancouver" (NW) grid cell

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

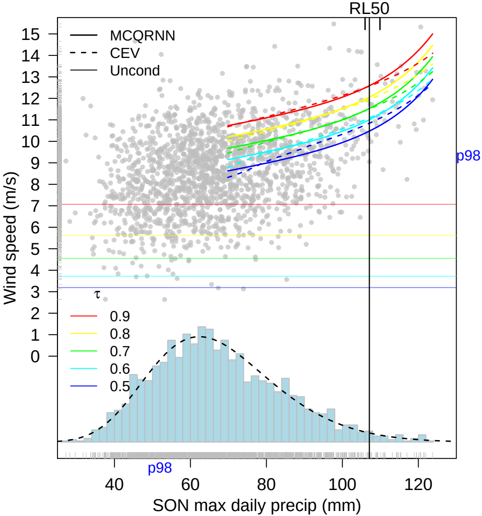
+ provide more accurate estimates in climate extremes

+ assess how well statistical procedures work

Estimating concurrent extremes using large ensemble climate simulations

- Estimating concurrent wind and precipitation extremes
- Illustrating the use of large ensemble climate model simulations to evaluate statistical methods

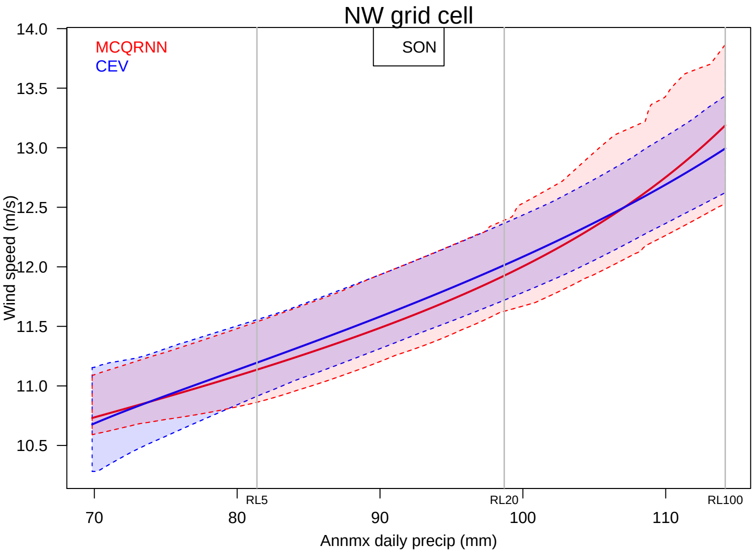
Estimating conditional quantiles using MCQRNN and CEV



- 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

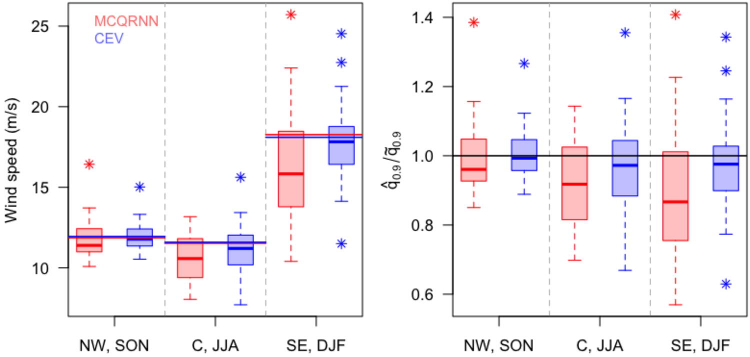
Bootstrap ensemble runs to obtain uncertainty estimates

Here we show the bootstrap confidence interval for 0.9 quantile function estimates



Assessing statistical model performance via large ensemble**

- We treat the fitted conditional quantile function at $\tau = 0.9$ using all 35 ensemble members as the "truth"
- We assess the model performance by fitting CEV and MCQRNN for each individual ensemble member



** See Sec. 2.3 of "Some Statistical Issues in Climate Science", 2019 Stat. Sci. by Michael Stein

Summary & discussion

- We explore conditional approaches to estimate the concurrent wind and precipitation extremes
- Large climate model ensemble is a powerful tool for studying climate extremes
- **Ongoing work**
  - 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

Summary & discussion

- We explore conditional approaches to estimate the concurrent wind and precipitation extremes
- Large climate model ensemble is a powerful tool for studying climate extremes
- **Ongoing work**
  - 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

Thank you for your attention!
Paper: www.sciencedirect.com/science/article/pii/S221209472100030X.
Code:
https://github.com/whitneyhuang83/ConcurrentExtremes

Some thoughts on financial risk

Nice review paper: Nolde, N., & Zhou, C. (2021). Extreme value analysis for financial risk management. *Annual Review of Statistics and Its Application*, 8, 217-240.

- Estimation of
  - marginal expected shortfall (MES)

$$\mathrm{MES}_p = \mathbb{E}[Y|X > \mathrm{VaR}_p(X)]$$

  - Conditional value-at-risk (CoVaR)

$$\mathbb{P}(X > \mathrm{CoVaR_p}|Y > \mathrm{VaR}_p(Y)) = 1 - p$$

- Hedging against climate risks using weather derivatives
- Large ensemble in finance (GARCH-like stochastic differential equations)?

Backup Slides

Estimating the magnitude of concurrent extremes

Consider the bivariate case, i.e., $\boldsymbol{X} = (X_1, X_2)^T$

- There is no natural ordering to define an extreme value for multivariate data

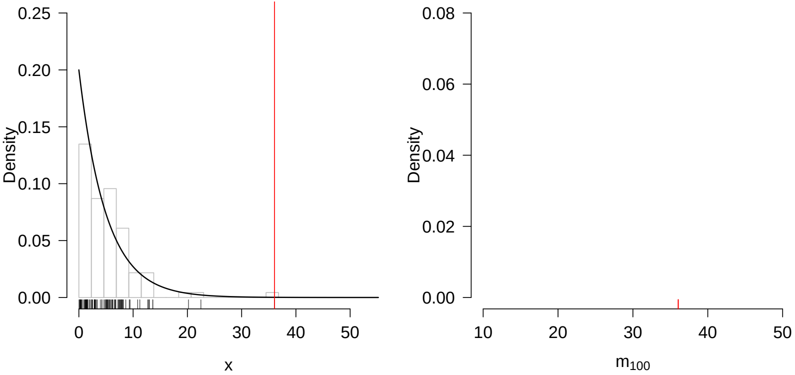
  "order properties ... exist only in one dimension" – Kendall (1966)
  "there is no natural concept of rank for bivariate data"–Bell and Haller (1969)

- Traditional multivariate extreme value analysis mainly focus on modeling **component-wise maximum** ⇒ may lead to "wrong" events ☹

- It is important to account for "**event simultaneity**" for modeling concurrent extremes ⇒ we do this by conditioning on one variable being extremes

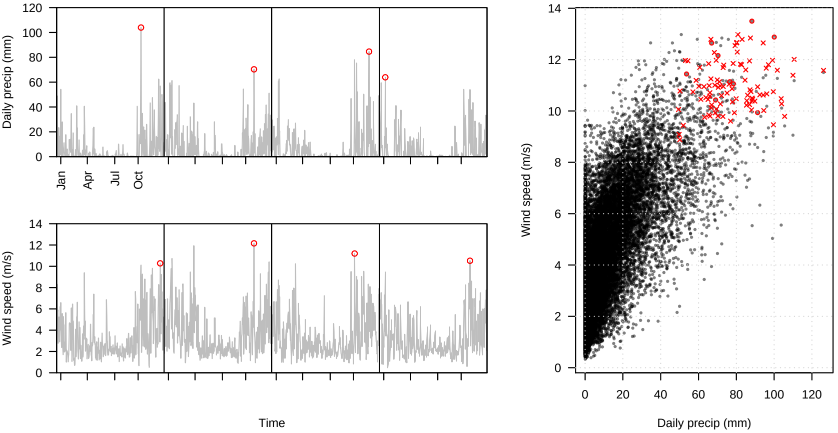
Extremal Types Theorem in Action

1. Generate 100 (n) random numbers from an Exponential distribution (population distribution)
2. Compute the sample maximum of these 100 random numbers
3. Repeat this process 120 times



Classical multivariate extreme value analysis

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



Issue: Ignore the event simultaneity

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)