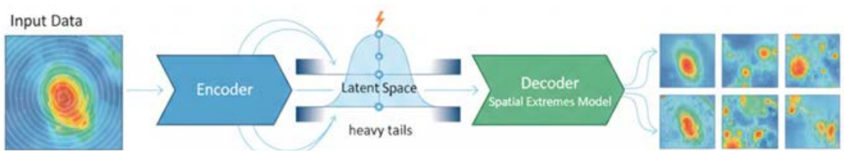
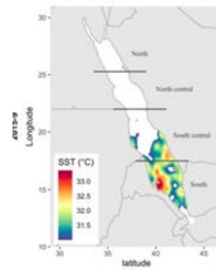
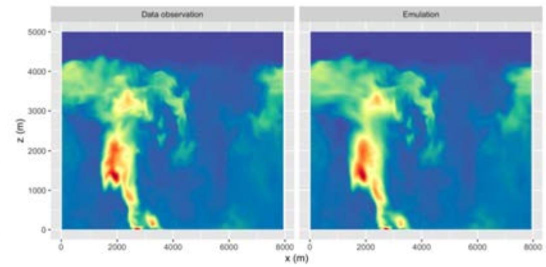






Flexible and Efficient Spatial Extremes Estimation and Emulation via Variational Autoencoders

Christopher K. Wikle
Department of Statistics and Data Science
Center for Spatio-Temporal Statistics and AI
University of Missouri

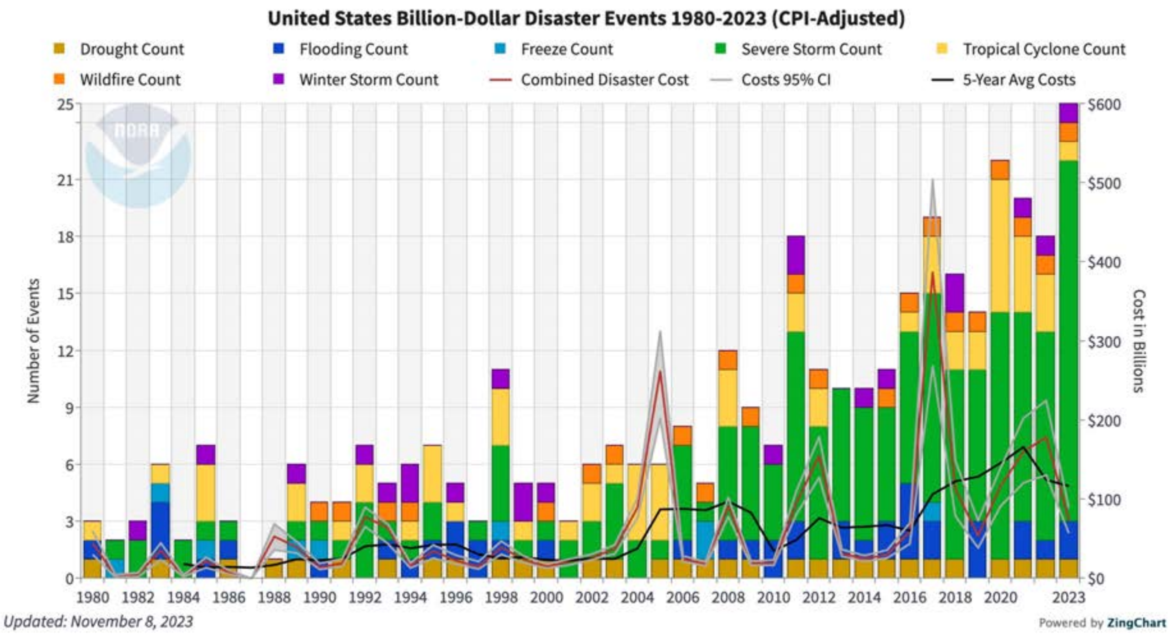
Collaborators: Likun Zhang (Statistics), Xiaoyu Ma (PhD Student, Statistics), Raphael Huser (KAUST), Kiran Bhaganagar (University of Texas-San Antonio)

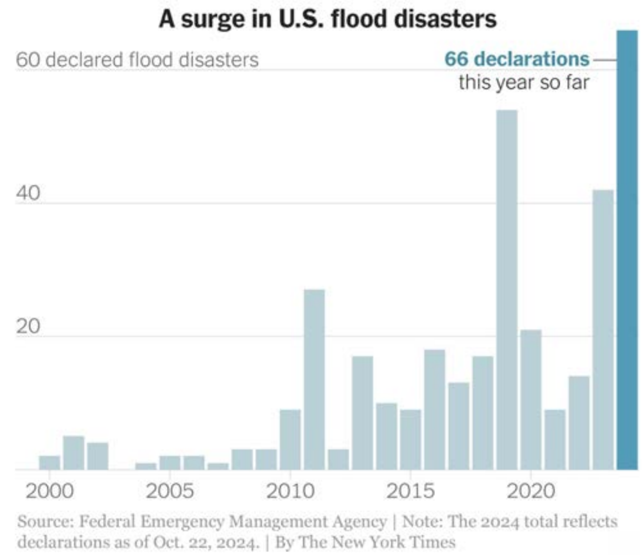
STAMPS: April 17, 2026

Extremes!

https://www.paddleyourownkanoo.com/2023/07/11/passenge
rs-and-cabin-crew-hit-ceiling-after-severe-clear-air-turbulence-
shakes-air-china-flight-to-beijing/

The number and intensity, and cost of extreme weather events is increasing!

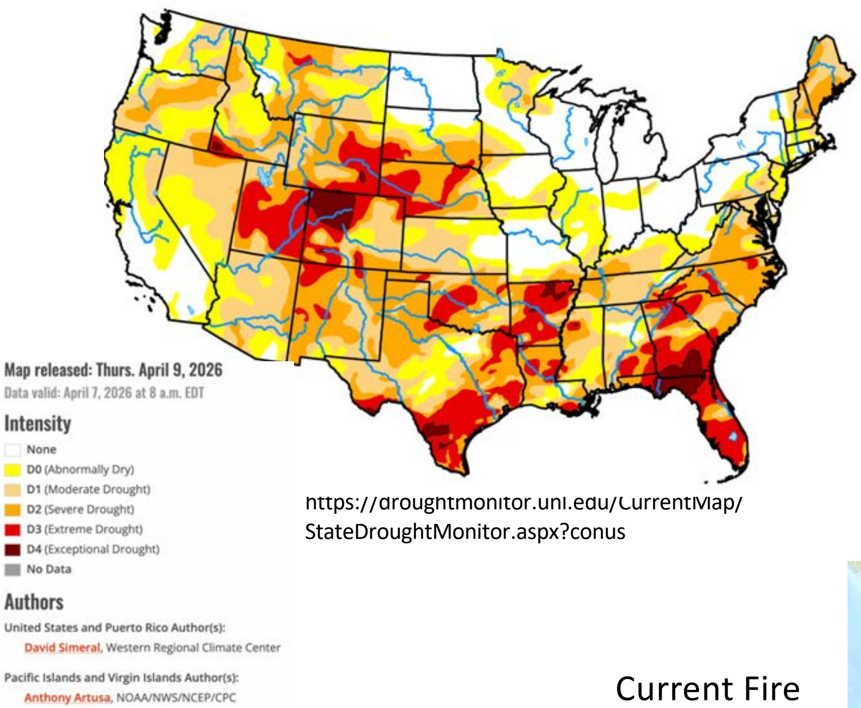




NOAA Centers for Environmental Information
https://www.ncei.noaa.gov/access/billions/time-series

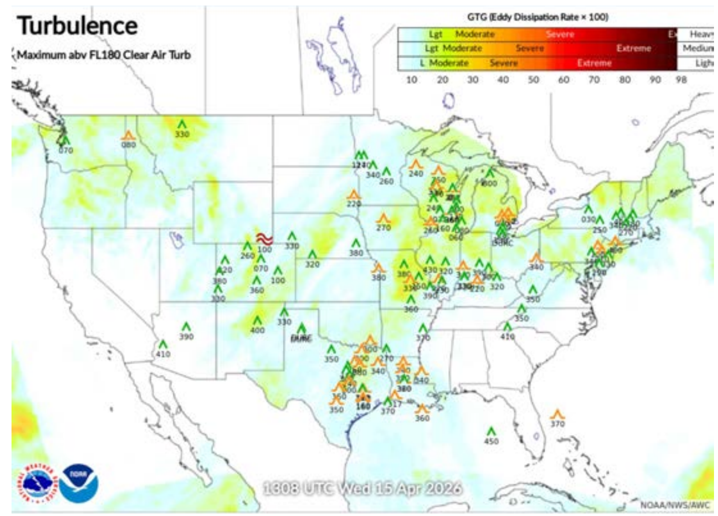
Extreme Events in the Environment Occur in Clusters

US Drought Monitor (04/09/2026)



https://droughtmonitor.unl.edu/CurrentMap/StateDroughtMonitor.aspx?conus

Clear Air Turbulence (04/15/2026)



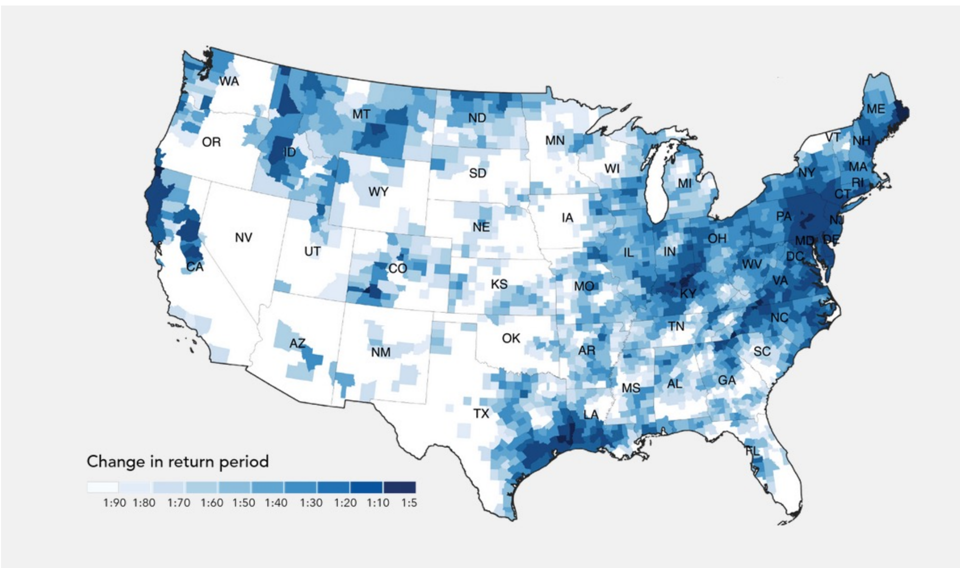
https://www.turbulenceforecast.com/clear-air-turbulence

Current Fire Incidents (04/15/2026)

https://fire.airnow.gov/#3.54/38.02/-103.08

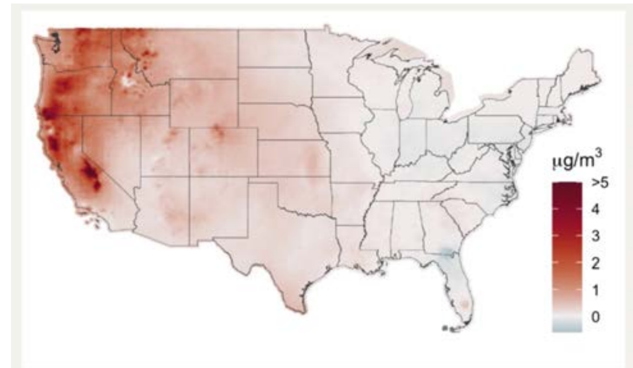
“Times they are a-changin’”!

1-in-100-year Flood Events
Now Happen Every 8 years!

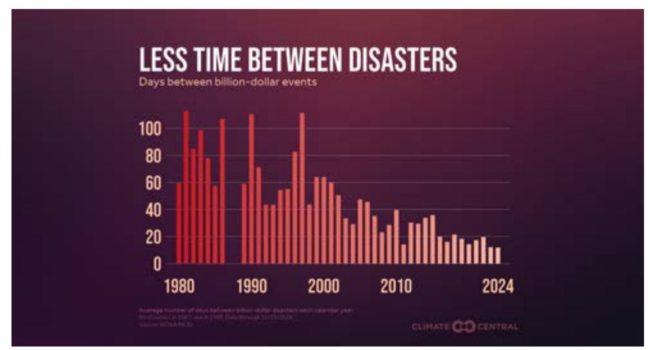


https://firststreet.org/press/noaas-1-in-100-year-flooding-can-now-be-expected-every-8-years/

Increasing Fine Particulate from Wildfires
(Stanford News, 9/22/2022)



Over the last decade, PM2.5 from wildfire smoke has increased in much of the U.S., particularly in Western states, but some areas in the South and East have seen modest declines. This map shows the decadal change in smoke PM2.5, meaning the difference in daily average smoke PM2.5 during 2006–2010 compared to 2016–2020. *(Image credit: Childs et al. 2022, Environmental Science & Technology)*



https://www.climatecentral.org/climate-matters/billion-dollar-disasters-2025

Example: Change in ocean extremes

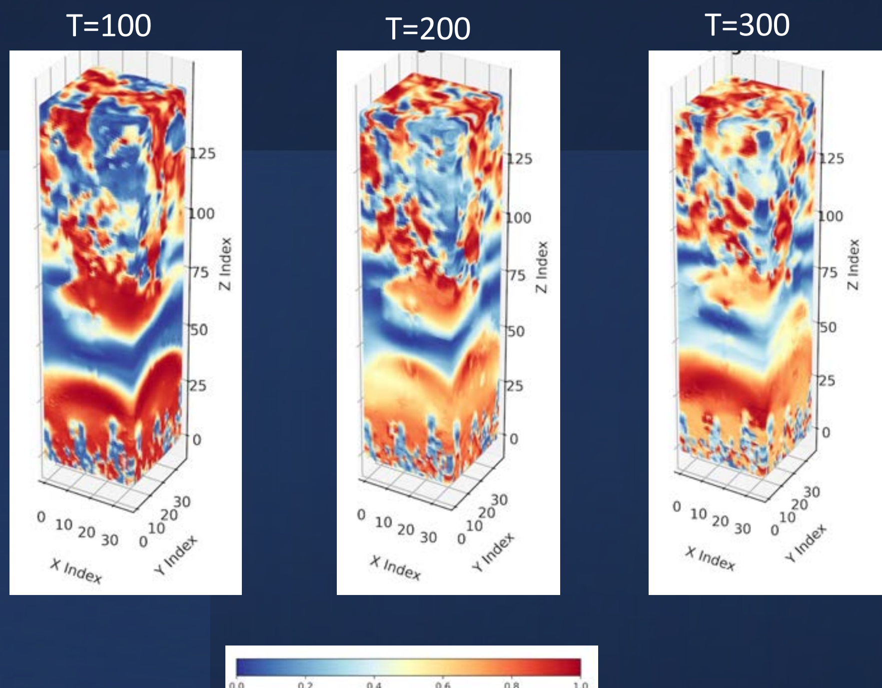
Coral Bleaching

I'll come back to this problem in my applications!

A reef in American Samoa is seen before, during, and after a coral bleaching event.
Credit: XL Catlin Seaview Survey

https://insideclimatenews.org/news/08062016/coral-bleaching-alarms-scientists-climate-change-global-warming-great-barrier-reef/

Emulating Complex Fluid Dynamic Fire Plume Model



3-D Turbulent Plume Numerical Model (modeling a wildfire plume)

I'll come back to this problem in my applications!

Why Statistics?

- Inference with uncertainty quantification on high-dimensional data that have spatial clusters of extremes
- Assessment of extremes via "one shot" data sets and complex, high-dimensional simulation models
  - Simulation models can be difficult to calibrate and account for uncertainty due to parameterizations, inputs and initial conditions
  - Emulation of simulation model output or data sets with surrogate models can provide an inexpensive way to calibrate, evaluate the model under different inputs, and/or provide Monte Carlo uncertainty quantification

Why Extremes?

Traditional environmental models used for one shot modeling and emulation (e.g., Gaussian processes, Markov random fields, polynomial chaos, proper orthogonal decomposition, neural networks, etc.) **do not natively account for extreme events**, which are often the most important outcomes of a complex simulation model

- e.g., drought in a climate model, hot spots in sea surface temps, fire plumes

Why not? Historically, it has been very difficult to build extreme-value-based statistical models for dependent extremes that are flexible (account for different dependence types over large domains) and can be applied to high-dimensional data.

This Talk

- Some background on spatial extremes
- A new flexible non-stationary extremes model
- Brief introduction to variational autoencoders (VAEs)
- XVAE: Implementing our flexible extremes model in a VAE
- Examples
  - Simulation
  - Red Sea Temperature
  - Complex Fire Plume Model Emulation
  - Extensions
- Brief Discussion

Note: this is a fairly non-technical talk, but the technical details are available in the papers.

Spatial Extremes

Many spatial extremes models have been developed to characterize the simultaneous occurrences of extremal events, with a specific focus on characterizing the rates of joint tail decay in different parts of the spatial domain.

Consider the bivariate coefficient:

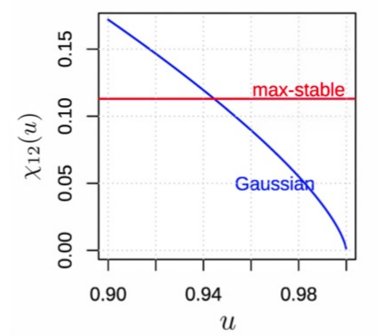
$$\chi_{12}(u) = P\{F_2(X_2) > u \mid F_1(X_1) > u\} = \frac{P\{F_2(X_2) > u, F_1(X_1) > u\}}{P\{F_1(X_1) > u\}},$$

in which $u \in (0, 1)$ and X_1 and X_2 are two spatial variables with marginal distribution functions F_1 and F_2.

When the quantile u is close to one, $\chi(u)$ quantifies the probability of simultaneous occurrence of events exceeding the high extremity level u given one variable is extreme.

Spatial Extremes

$$\chi_{12}(u) = P\{F_2(X_2) > u \mid F_1(X_1) > u\}$$



- If $lim_{u \to 1}\ \chi(u) = 0$, X_1 and X_2 are asymptotically independent (AI)

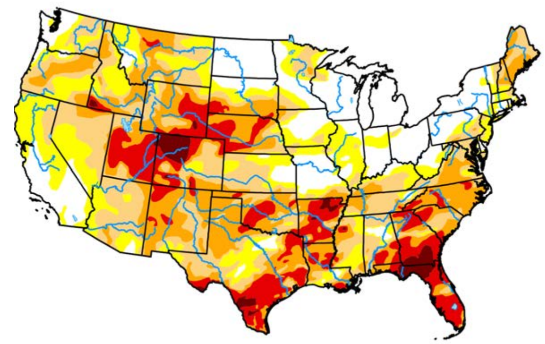
E.g., Gaussian processes always AI regardless of proximity

- If $lim_{u \to 1}\ \chi(u) > 0$, X_1 and X_2 are asymptotically dependent (AD)

E.g., max-stable (e.g., Padoan et al., 2010) , Pareto (Thibaud and Optiz, 2015) always AD regardless of proximity

But, in practice χ(u) typically decays as $u \to 1$ in environmental processes (extreme events become spatially more localized as their intensity increases).

Spatial Extremes



There are increasingly more options for spatial extremes models that transcend traditional extreme value theory (see overview in Huster and Wadsworth, 2022). However, these typically assume stationary dependence across space and time and are too computationally demanding for large spatial data sets.

A Good Model Should Allow:

- Close locations to have stronger dependence than locations far apart
- Dependence strength to decay as "extremeness" intensifies
- Dependence structure to be heterogeneous (non-stationary) across large domains (e.g., continental US)
- Allow both long-range AI and short-range AI or AD to be present
- Implementation on large data sets

Max-ID Spatial Extremes Model

Reich and Shaby (2012) and Bopp et al. (2021) developed a max-id (max-infinitely divisible) spatial model that uses a low rank basis representation:

obs

$$X(\mathbf{s}) = \epsilon(\mathbf{s})Y(\mathbf{s})$$

$$Y(\mathbf{s}) = \left\{\sum_{k=1}^{K} \omega_k(\mathbf{s})^{1/\alpha} Z_k\right\}^{\alpha},$$

Low rank basis expansion;
"global" spatial basis functions

where $\epsilon(\mathbf{s})$ is a white noise process with $\text{Fréchet}(0, 1, 1/\alpha)$ marginals, $\omega_k(\mathbf{s})$ are spatial basis functions, and $\{Z_k : k = 1, \ldots, K\}$ are exponentially tilted, positive-stable (PS) random variables:

$$H(x;\alpha,\theta) = \frac{f_\alpha(x)\exp(-\theta x)}{\exp(-\theta^\alpha)}, x > 0$$

$$Z_1, \ldots, Z_k \sim H(\alpha, \theta),$$

where $\alpha \in (0, 1)$ controls the rate at which the power-law tail drops off, and $\theta \geq 0$ is the "tilting" parameter, with larger values of θ leading to lighter tails.

Max-ID Spatial Extremes Model

This model is quite nice because it handles AI when $\theta > 0$ or AD when $\theta = 0$ and $\alpha < 1$. But,

- Only one dependence class for a given θ value;
- Long-range AI and short-range AD cannot be satisfied at the same time;
- $\epsilon(s) \sim \text{Fréchet}(0, 1, 1/\alpha)$ and α controls both the white noise distribution and the basis coefficients dependence properties.

This is not flexible enough for us; we need the potential for long- and short-range AI and short-range AD in the same domain.

We modify this model to allow for both AI and AD in the same domain.

Flexible Non-Stationary Max-ID Model

Zhang et al. (2026)

(Our enhancements are in red)

$$X(\mathbf{s}) = \epsilon(\mathbf{s})Y(\mathbf{s}),$$

$$Y(\mathbf{s}) = \left\{\sum_{k=1}^{K} \omega_k(\mathbf{s}, r_k)^{\frac{1}{\alpha}} Z_k\right\}^{\alpha_0},$$

in which $\epsilon(\mathbf{s})$ is a white noise process with Fréchet$(0, \tau, 1/\alpha_0)$ marginals, $\omega_k(\mathbf{s}, r_k)$ are compact-supported (Wendland) spatial basis functions centered at K knots (r_k), $\alpha \in (0, 1)$ is a concentration parameter, and $\{Z_k : k = 1, \ldots, K\}$ are exponentially tilted, positive-stable variables:

$$Z_1, \ldots, Z_k \sim H(\alpha, \theta_k),$$

where $\theta_k \geq 0, k = 1, \ldots, K.$

We further allow these parameters to vary in time w/ space-time data.

- **Compactly supported basis functions:** introduces long-range AI for two far-apart stations that are impacted by disjoint sets of knots
- **Tilting indices (θ_k) vary across domain and can be 0 or positive:** allows non-stationary tail dependence (local AD if $\theta_k = 0$; AI if $\theta_k > 0$)
- **Including both α and α_0:** decouples dependence structure from rate of marginal tail decay
- **Fréchet$(0, \tau, 1/\alpha_0)$ marginals:** concentrates around 1 when shape $1/\alpha_0 > 1$ and scale $\tau > 1$, acting like measurement error

Fréchet $\Pr\{\epsilon(\boldsymbol{s}) \leq x\} = \exp\{-(x/\tau)^{1/\alpha_0}\}$

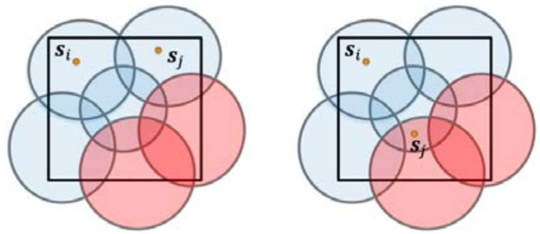
Properties of Flexible Max-ID Extremes Model

We investigate rigorously the marginal and joint dependence properties of our model; **see Zhang et al. (2026) for formal statements and proofs.**

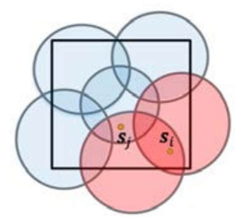
We show: the modeled spatial process has non-stationary tail-dependence structure (local AI or AD, long-range AI)

- *The local dependence strength is proportional to the tail-heaviness of the latent variable at the closest knot.*
- *The compactness of the basis function support automatically introduces long-range exact independence (thus, also AI) for two far-apart stations that are impacted by disjoint sets of basis functions*

AI $\chi(\boldsymbol{s}_i, \boldsymbol{s}_j) = 0$



AD $\chi(\boldsymbol{s}_i, \boldsymbol{s}_j) > 0$



Implementation

- Ideally, we would implement this hierarchical model in a Bayesian inferential framework
- This would be very computationally intensive and impractical for large domains and data volumes
- Can we utilize deep neural models?

Dependent Extremes and Neural Computing

There has recently been a fast-growing body of work to combine extremes modeling and deep neural networks (NNs)

- Richards and Huser (2022) and Pasche and Engelke (2024) use deep extremal quantile regression models to improve the modeling of marginal extremes in spatial and temporal settings, respectively.
- Boulaguiem et al. (2022) use a deep convolutional GAN (called extGAN) to learn the dependence structure of spatial extremes; their approach does not impose any parametric constraint on the extremal dependence structure, which leads to asymptotic independence.
- Lafon et al. (2023) develop a VAE tailored to multivariate regularly varying (i.e., jointly heavy-tailed) data; their approach thus only applies to asymptotic dependent data, and it has so far only been validated in small dimensions (specifically, 5 sites in their application)
- Lenzi et al. (2023), Sainsbury-Dale et al. (2024, 2023), Richards et al. (2024), Walchessen et al. (2024) use amortized deep learning methods for fast likelihood-free inference with parametric spatial extreme-value models. Yet cannot easily handle highly-parameterized models (such as nonstationary processes) and are designed more to provide parameter estimates rather than generative emulation

What we need from an extremes-aware neural estimator

We require methods that simultaneously:

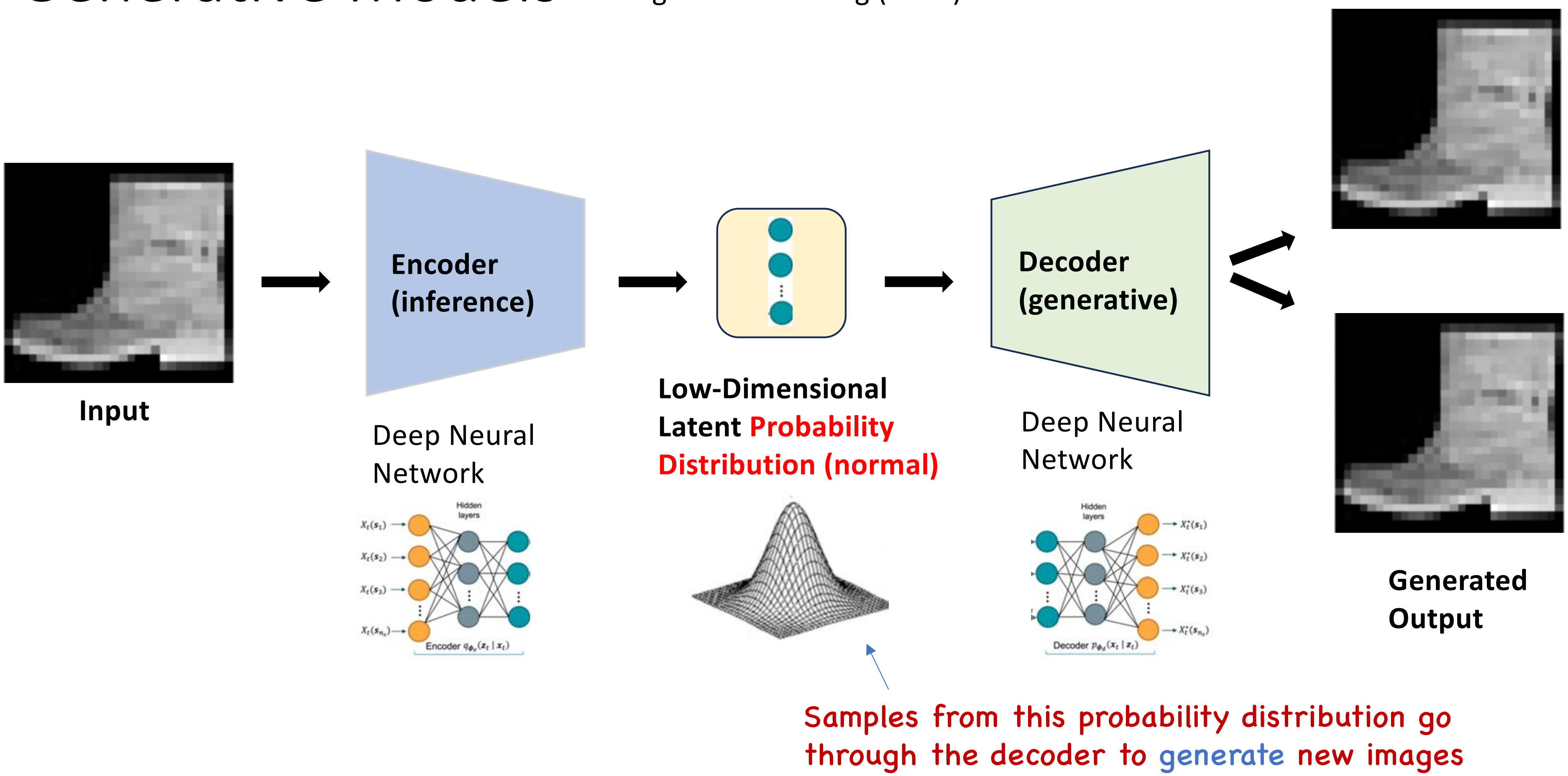
- provide flexible modeling options (e.g., non-stationary asymptotic dependence and independence) over large domains
- allow efficient inference
- allow fast accurate simulation

A variational autoencoder (VAE) is great for this, especially if we make it "extremes aware" by using our flexible model.

First, what's an autoencoder?

Variational Autoencoder (VAE) – Probabilistic Generative Models

Kingma and Welling (2013)



VAE For Spatio-Temporal Data (Bayes Rule) - Details

Consider a replicate of a process $\mathbf{X}_t = \{X_t(\mathbf{s}_j) : j = 1, \ldots, n_s\}$, and its encoded K-dimensional latent random vector $\mathbf{Z}_t = \{Z_{kt} : k = 1, \ldots, K\}$, $t = 1, \ldots, n_t$.

- We denote the joint model as $p_{\phi_d}(\mathbf{x}_t, \mathbf{z}_t)$ (with parameters ϕ_d) and
  - Encoder: $p_{\phi_d}(\mathbf{z}_t \mid \mathbf{x}_t)$
  - Decoder: $p_{\phi_d}(\mathbf{x}_t \mid \mathbf{z}_t)$.
- In principle, we can obtain the encoder via Bayes' rule: $p_{\phi_d}(\mathbf{z}_t \mid \mathbf{x}_t) \propto p_{\phi_d}(\mathbf{x}_t \mid \mathbf{z}_t) p_{\phi_d}(\mathbf{z}_t)$, but it is often intractable.
- A VAE uses a variational-Bayes-based approximation $q_{\phi_e}(\mathbf{z}_t \mid \mathbf{x}_t)$, in which $\phi_e$is shared across all replicates.

VAE For Spatial Data: Variational Inference

The marginal distribution of $\mathbf{x}_t$ can then be rewritten as follows:

$$\log p_{\phi_d}(\mathbf{x}_t) = \quad E_{q_{\phi_e}(\mathbf{z}_t|\mathbf{x}_t)}\left\{\log \frac{p_{\phi_d}(\mathbf{x}_t, \mathbf{z}_t)}{q_{\phi_e}(\mathbf{z}_t \mid \mathbf{x}_t)}\right\} + D_{KL}\left\{q_{\phi_e}(\mathbf{z}_t \mid \mathbf{x}_t) \,||\, p_{\phi_d}(\mathbf{z}_t \mid \mathbf{x}_t)\right\}$$

- the second term on the right is the Kullback-Leibler (KL) divergence between $q_{\phi_e}(\mathbf{z}_t \mid \mathbf{x}_t)$ and $p_{\phi_d}(\mathbf{z}_t \mid \mathbf{x}_t)$
- the first term on the right is the variational evidence lower bound (ELBO), and it is sufficient to maximize this (numerically via MC):

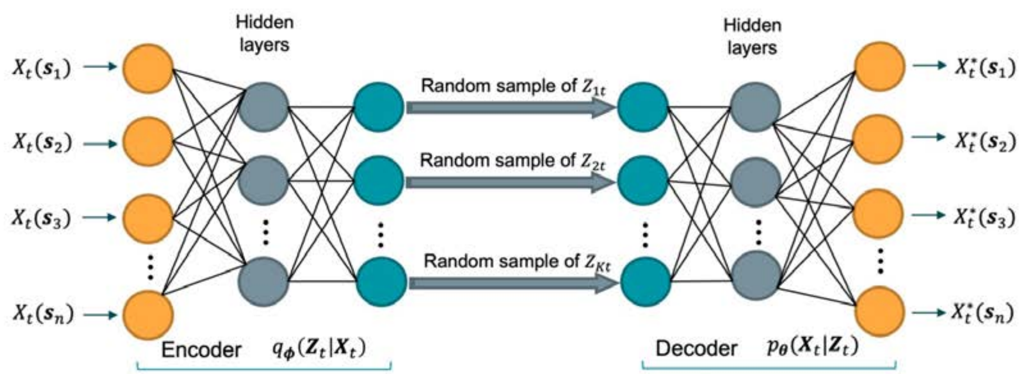
$$\mathcal{L}_{\phi_e,\phi_d}(\mathbf{x}_t) = E_{q_{\phi_e}(\mathbf{z}_t|\mathbf{x}_t)}\left\{\log \frac{p_{\phi_d}(\mathbf{x}_t, \mathbf{z}_t)}{q_{\phi_e}(\mathbf{z}_t \mid \mathbf{x}_t)}\right\} \approx \frac{1}{L}\sum_{\ell=1}^{L} \log \frac{p_{\phi_d}(\mathbf{x}_t, \mathbf{z}^{\ell})}{q_{\phi_e}(\mathbf{z}^{\ell} \mid \mathbf{x}_t)}.$$

balances the log-likelihood and the KL divergence

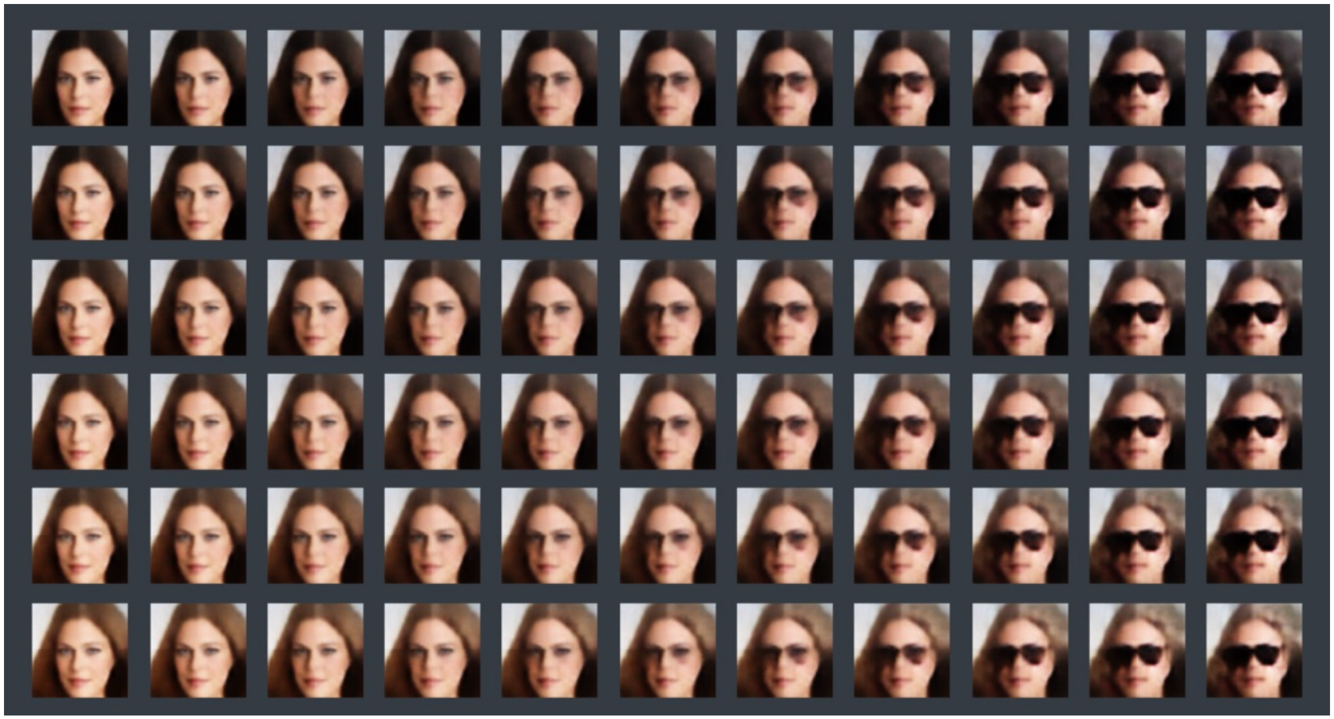
See Kingma and Welling (2013,2019) for details

VAE For Spatial Data: Variational Inference

- Typically, $q_{\phi_e}(\mathbf{z}_t \mid \mathbf{x}_t)$ are normal distributions and the means and variances are learned via a deep neural network
- The prior distribution $p_{\phi_d}(\mathbf{z}_t)$ is typically $N(\mathbf{0}, \mathbf{I}_K)$
- The data model $p_{\phi_d}(\mathbf{x}_t \mid \mathbf{z}_t)$ is usually Normal or does not account for extremal dependence; again, parameter dependence on $\mathbf{z}_t$ is specified with a deep neural net
- Update ϕ_d, ϕ_e with stochastic gradient descent



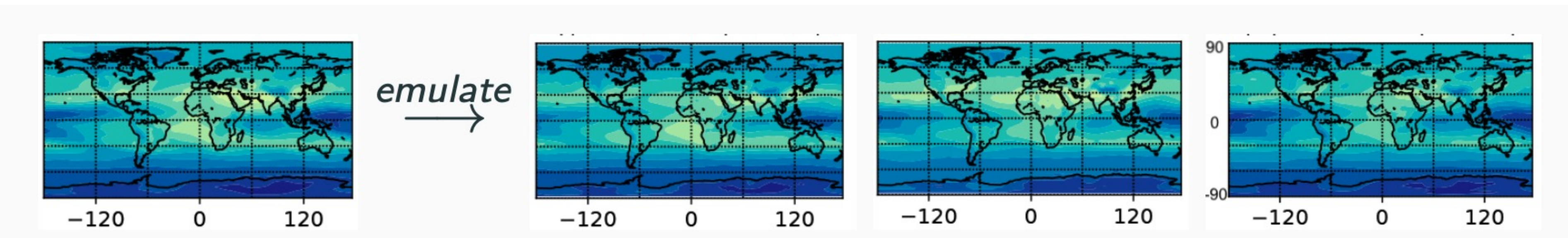
VAEs work well for many emulation tasks. E.g., Deep Fake Generation



Variational Autoencoders are Beautiful, Steven Flores
(Monday, Apr 15, 2019) https://www.compthree.com/blog/autoencoder/

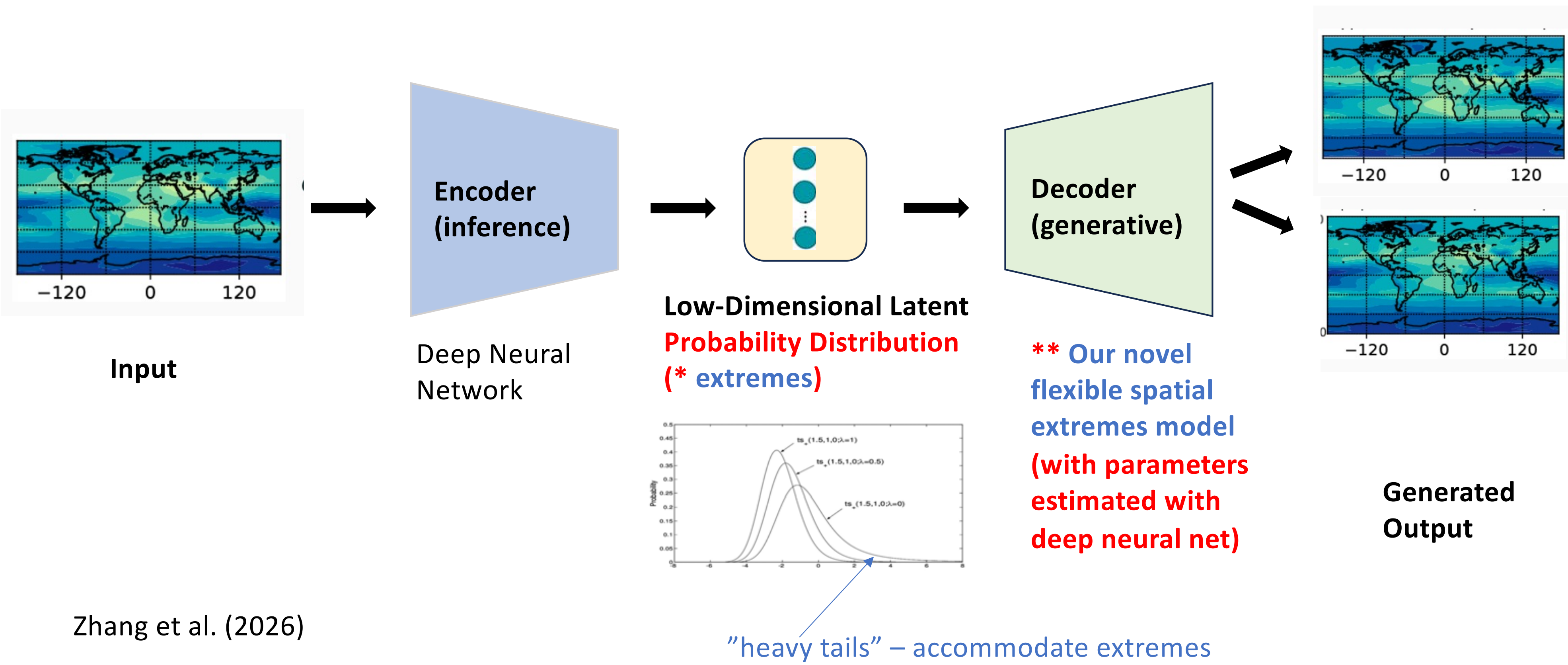
However, we have extremes!

We are interested in emulating processes that can exhibit extremes (e.g., from climate models, complex simulators, or large data sets).



Traditional probabilistic emulator approaches (including VAEs) don't naturally handle extremes, especially those that that are clustered.

Extreme-aware Variational Autoencoder (XVAE)



Zhang et al. (2026)

XVAE: Details

Encoder:

Construct the latent dimensions for $q_{\phi_e}(\mathbf{z}_t \mid \mathbf{x}_t)$ using the *reparameterization trick* (Kingma and Welling, 2013):

$$\mathbf{z}_t = \boldsymbol{\mu}_t + \boldsymbol{\sigma}_t \odot \boldsymbol{\eta}_t,$$
$$\eta_{kt} \overset{\text{i.i.d.}}{\sim} \text{Normal}(0, 1),$$
$$(\boldsymbol{\mu}_t, \log \boldsymbol{\sigma}_t) = \text{EncoderNeuralNet}_{\phi_e}(\mathbf{x}_t),$$

See paper for details about the neural net architecture

in which $\boldsymbol{\eta}_t = \{\eta_{kt} : k = 1, \ldots, K\}$ and $\odot$ is the elementwise product.

Latent Process Prior:

$$p_{\phi_d}(\mathbf{z}_t) = \prod_{k=1}^{K} H(z_{kt}; \alpha_t, \theta_k)$$

Exponentially tilted positive-stable random variables

XVAE: Details

Decoder: (this is our flexible extremes model!)

The parameters in the decoder are given by $\boldsymbol{\phi}_d = (\alpha_0, \tau, \alpha_t, \boldsymbol{\theta}_t, \mathbf{W})$, and

$$p_{\phi_d}(\mathbf{x}_t \mid \mathbf{z}_t) = \left(\frac{1}{\alpha_0}\right)^n \left\{\prod_{j=1}^{n} \frac{1}{x_{jt}} \left(\frac{x_{jt}}{\tau y_{jt}}\right)^{-1/\alpha_0}\right\} \exp\left\{-\sum_{j=1}^{n} \left(\frac{x_{jt}}{\tau y_{jt}}\right)^{-1/\alpha_0}\right\},$$

in which $y_{jt} = \sum_{k=1}^{K} \omega_{kj}^{1/\alpha_t} z_{kt}$.

This is where we use a neural network: (neural estimation)

$$(\hat{\alpha}_t, \hat{\boldsymbol{\theta}}_t) = \text{DecoderNeuralNet}_{\phi_d}(\mathbf{z}_t)$$

See paper for details about the neural net architecture

XVAE: Implementation

We estimate the parameters of the decoder (our model) ϕ_d and encoder (deep neural network) ϕ_e using stochastic gradient descent on the variational evidence lower bound (ELBO), given by: $\mathcal{L}_{\phi_e,\phi_d}(\boldsymbol{x}_t)$

We use automatic differentiation to calculate the gradient using `R torch`.

See paper for implementation details

Our algorithm is very efficient – many times faster than if we tried to implement our extremes model in a hierarchical Bayesian context. But, we do amortize the computational effort up front so the generation of emulated fields is very fast.

Semi-amortized inference is crucial

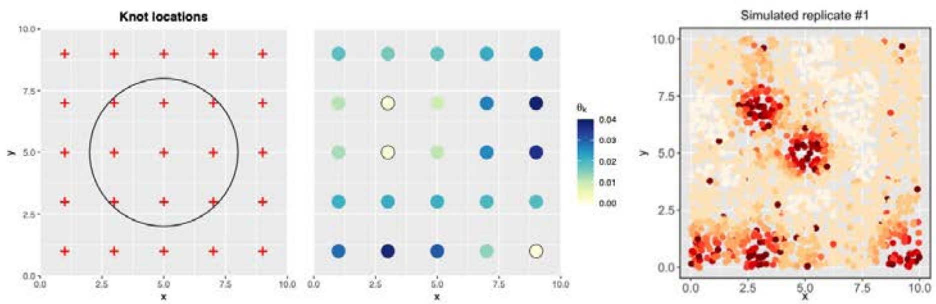
XVAE Application

Consider some examples:

- Simulation (note, more comprehensive simulation study in our paper, Zhang et al. 2026)
- Application to Red Sea temperatures
- Application to turbulent geophysical fire plume simulation models

Simulation Settings

Using $n = 2,000$ sites on $[0, 10] \times [0, 10]$ and $T = 100$ independent replicates, we generated data from the enhanced max-id model:



More details on NN architecture and additional simulations in the paper.

- $\alpha_0 = 0.25$, $\alpha = 0.5$, $\tau = 1$
- $K = 25$ (knots)
- All NNs in the extreme VAE are implemented in `R` using `torch`.
- Converges in ≈ 7000 epochs (about 10 mins) on a Mac Studio (M1 Ultra) (compared to days with competing models)

Simulation Results

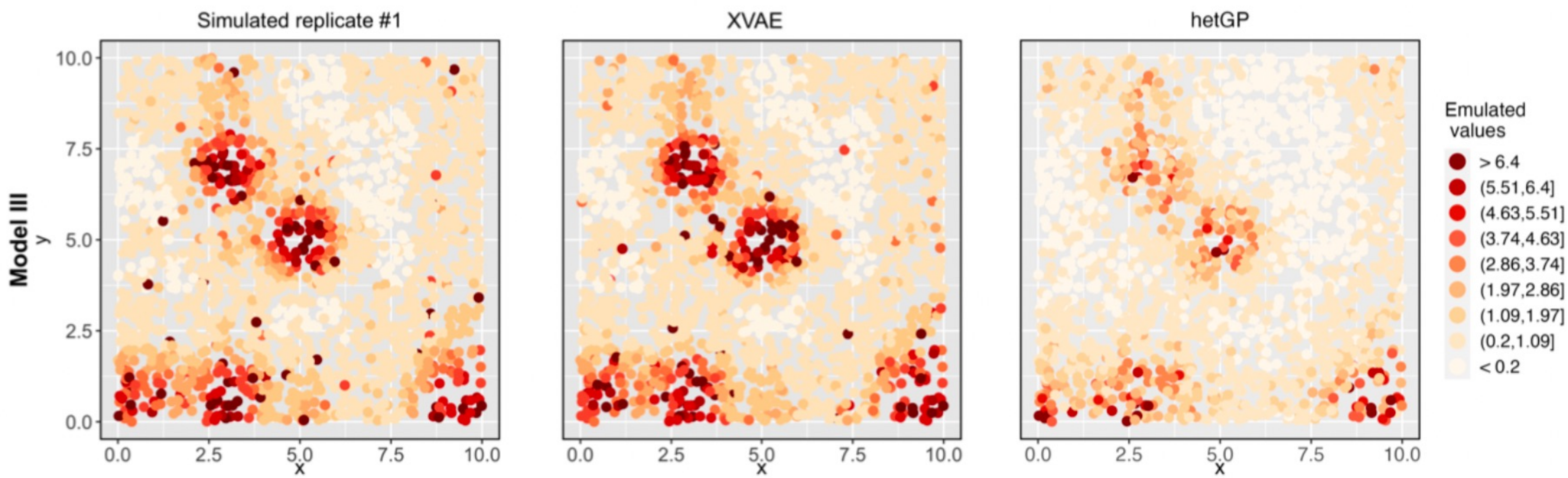
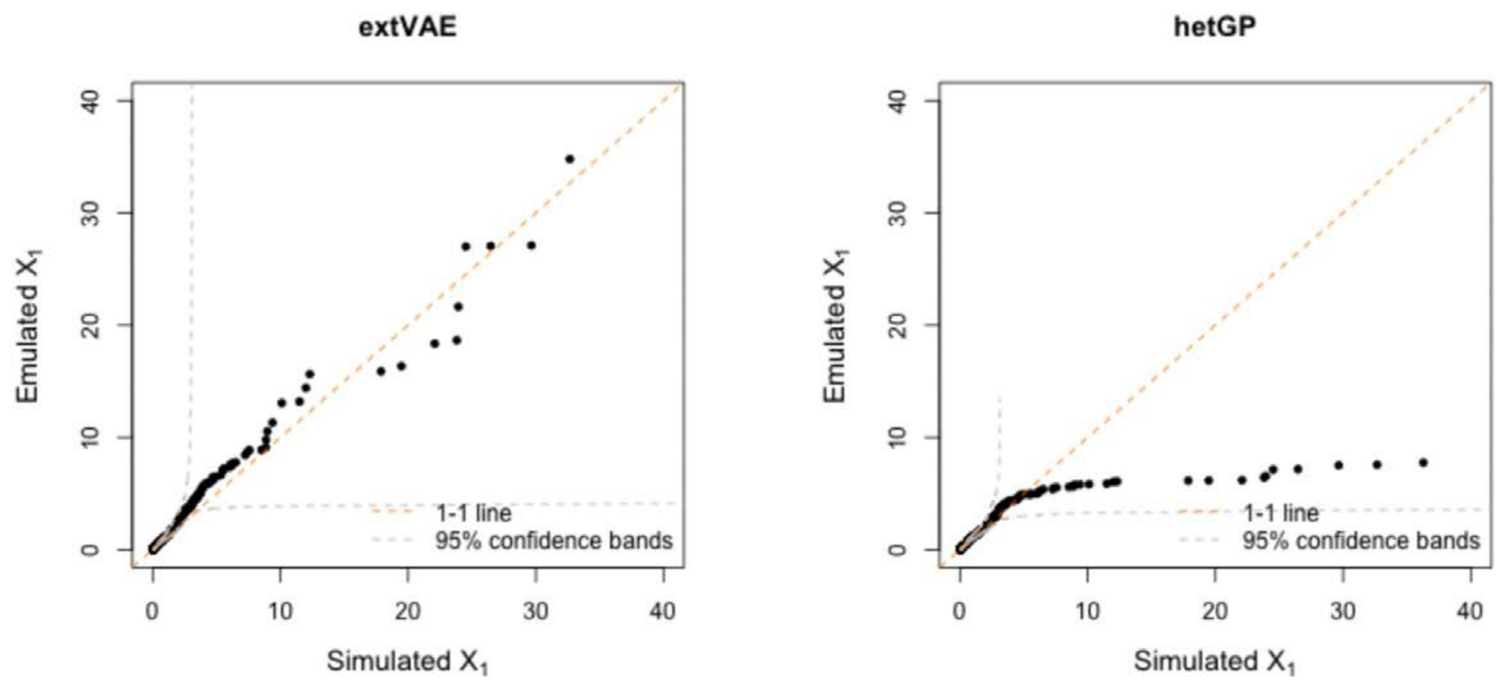


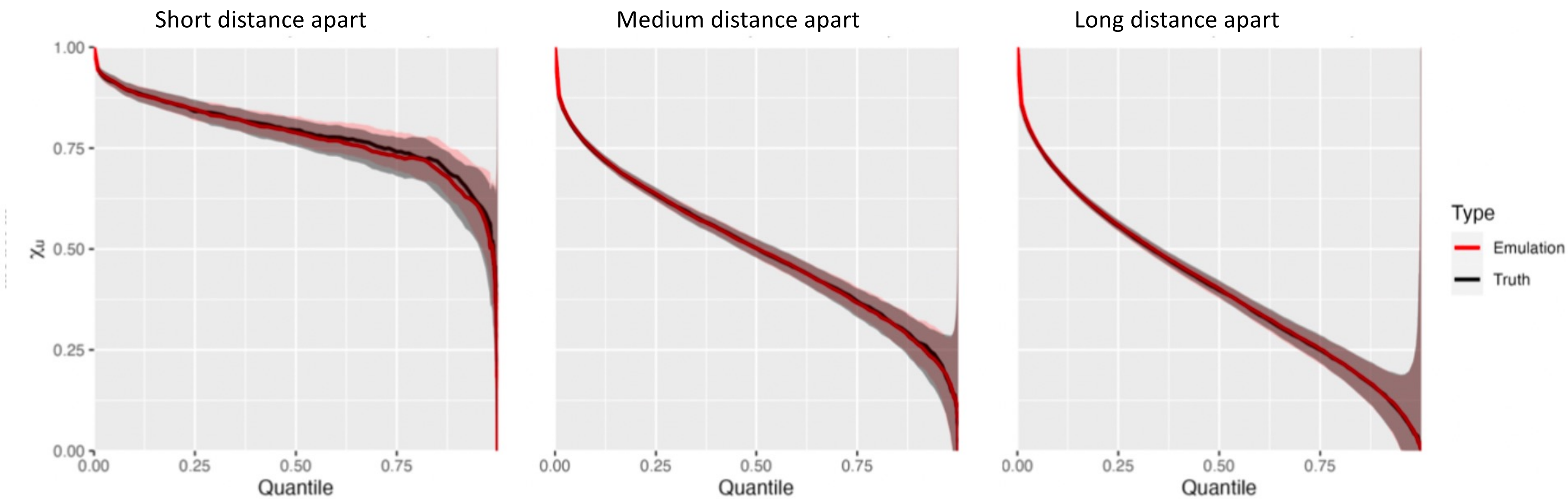
Illustration of a simulated spatial response (left panel), the emulated field from our extreme VAE model (XVAE, middle panel), and comparison to a heterogenous Gaussian Process (GP) emulation from the `hetGP` package (right panel).

Simulation Results



qq-plot for the emulated field from our extreme VAE model (XVAE, left panel) and the heterogenous Gaussian Process emulated field (right panel).

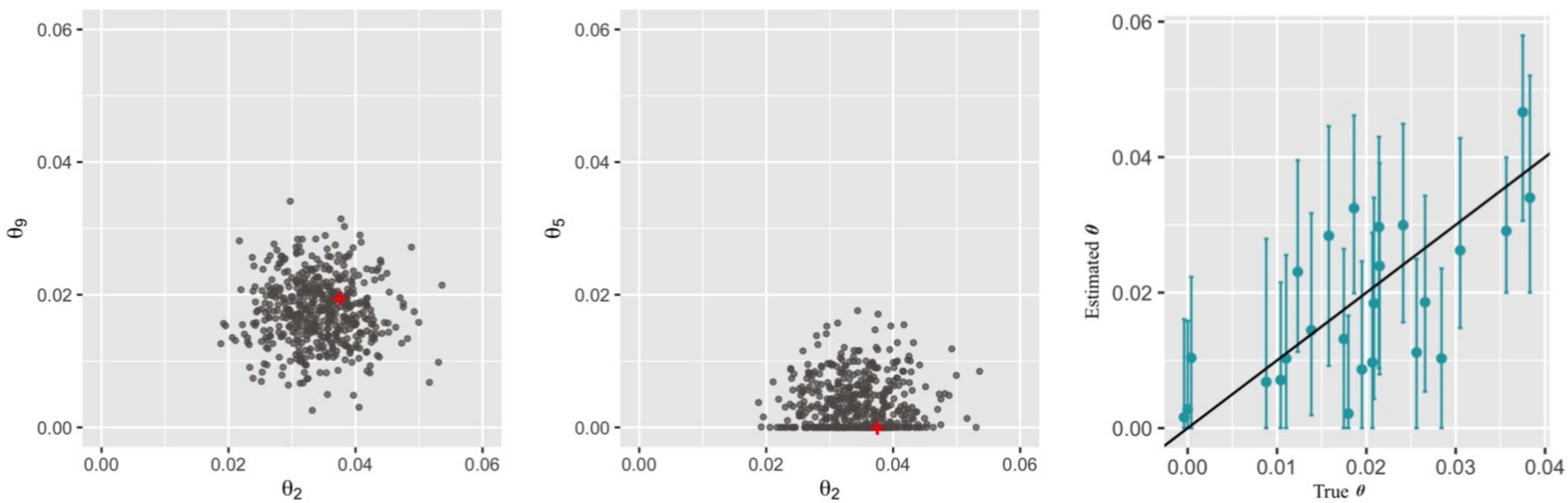
Simulation Results



Empirical estimate of $\chi_h(u)$ for $h = 0.5, 2, 5$ (left, center, right); simulated (black) and XVAE emulated (red) (also, 95% UQ intervals)

Simulation Results: **Parameter Uncertainty Quantification**

e.g., tilting parameters



Modeling Red Sea Surface Temperature Extremes

Goal: to accurately assess marine heatwave risks and identify regions susceptible to coral bleaching in the Red Sea (a biodiversity hotspot).

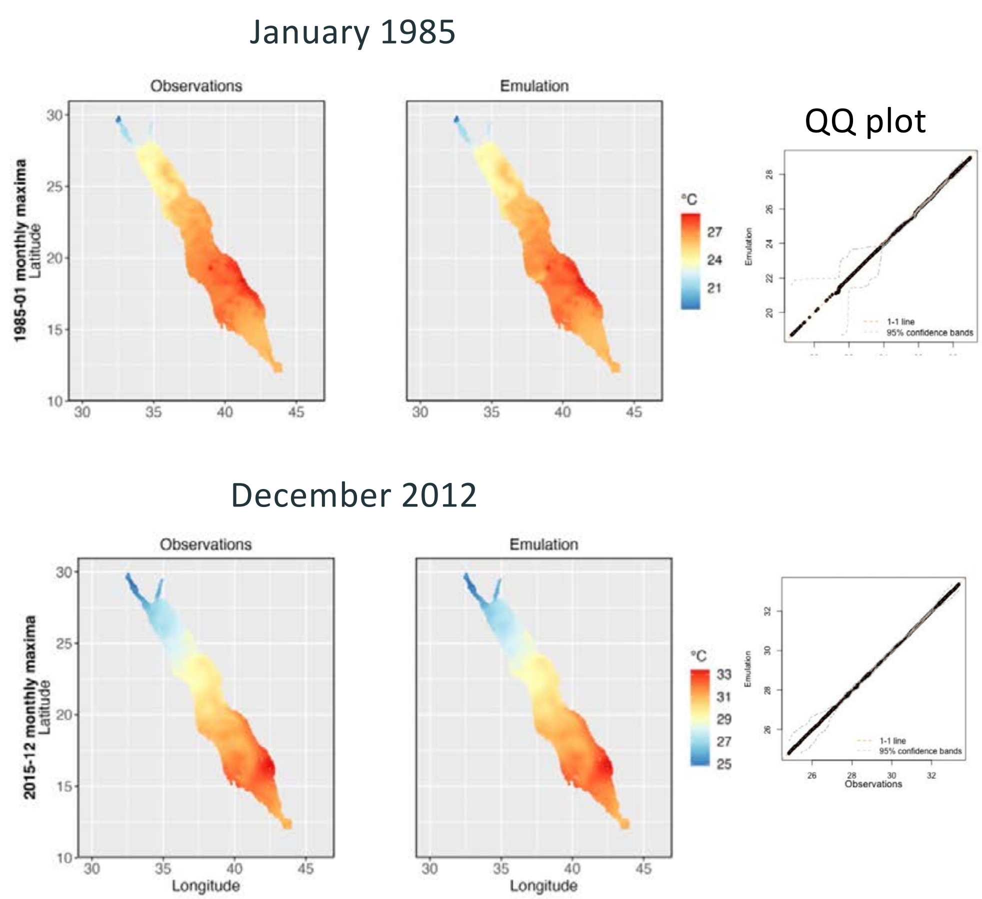
- Need to consider the distribution of area of extreme sea surface temperature (SST) exceedances.
- Satellite/model data fusion are state-of-the-art, but the models are too expensive to generate the number of ensembles required to understand the distribution of exceedances.
- We use our XVAE to emulate realizations of Red Sea SST fields across time.

Modeling Red Sea Surface Temperature Extremes

The dataset contains satellite-derived daily SST estimates from 16,703 locations on 0.05° grid and time stamps from 1985-01-01 to 2015-12-31 (11,315 days).

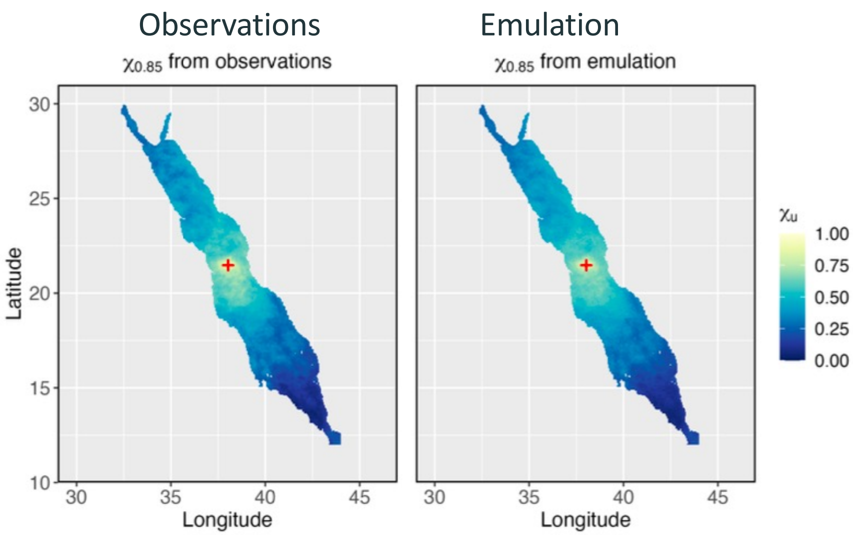
First, we remove seasonality, extract monthly maxima, and perform GEV marginal transformations (372 months; combined **6.2 million observations**).

Our model trains in under one hour – arguably, no other flexible spatial extremes procedure can handle this large of a data set (at least not in 2023).



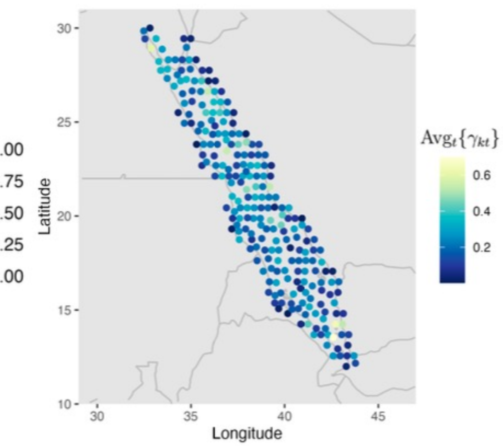
See paper for other verification metrics

Modeling Red Sea Surface Temperature Extremes

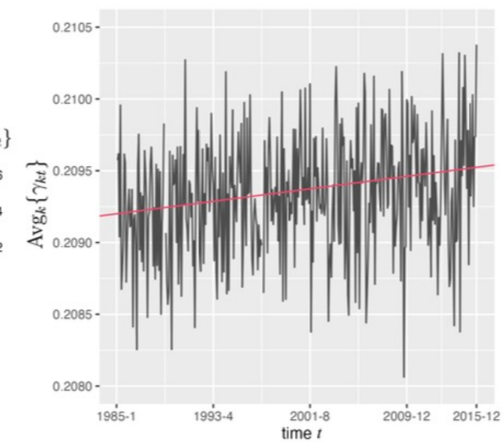


Empirical $\chi_{.85}$ tail dependence between the location at the red + and all other locations.

(Neural) Estimates of Model Tilting Parameters



Estimated θ_k tilting parameters corresponding to K=243 knots, averaged over time showing spatial inhomogeneity.



Estimated θ_k tilting parameters corresponding to K=243 knots, averaged over space showing upward time trend (extreme events are becoming more localized over time).

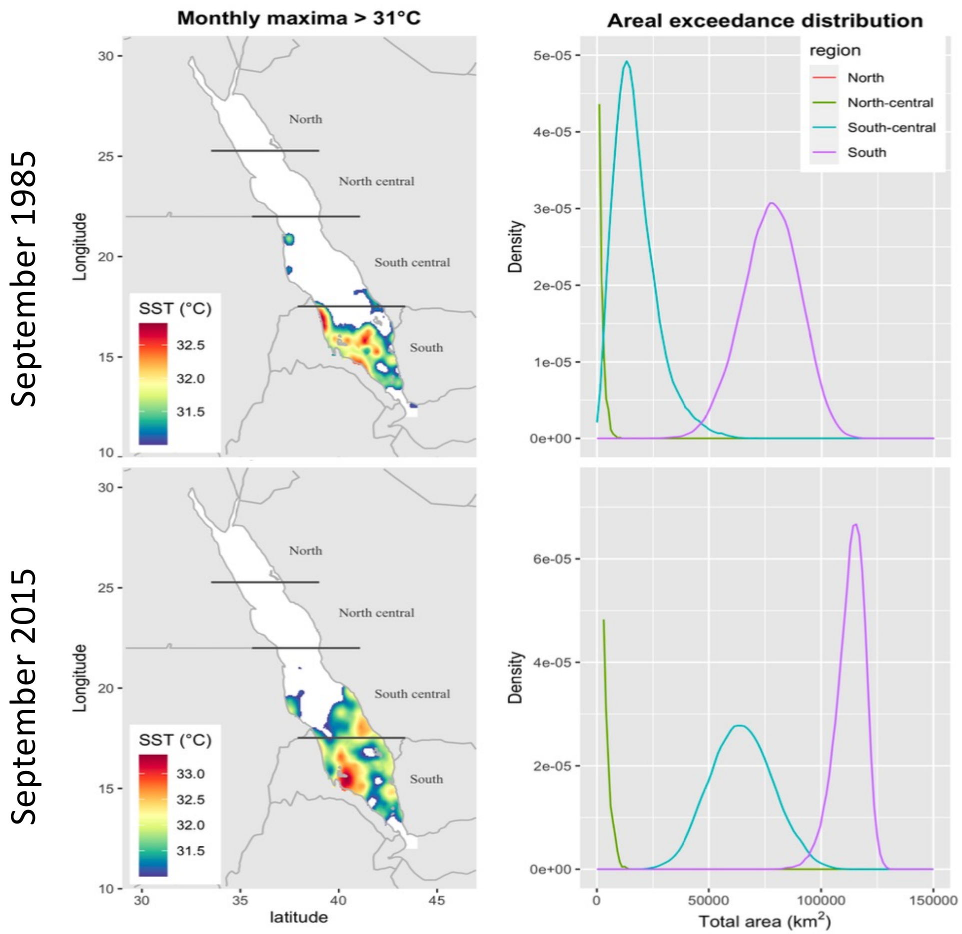
So, what does this tell us about potential coral bleaching extreme events?

Red Sea SST Exceedance Areas from XVAE

Monthly maxima areas exceeding 31 deg C. in the 4 regions defined in Raitsos et al. (2013).

Based on 30,000 emulated realizations from the XVAE.

Note:
Area of exceedances are increasing over time in the south and south-central regions.



Emulating Complex Fire Plume Simulator

Zhang et al. (2025)

Joint work with Kiran Bhaganagar, Mechanical Engineering, U. Texas San Antonio

- Processes such as wildfires, volcano emissions, air-pollution, and chemical releases are examples of "buoyant plume transport"
- Such turbulent events include extremes as energy is focused in small regions
- Modeling such phenomena requires very high-resolution fluid simulations that can account for turbulent effects
  - These models are extremely computationally intensive, particularly at high Reynolds numbers

Emulating Complex Fire Plume Simulator

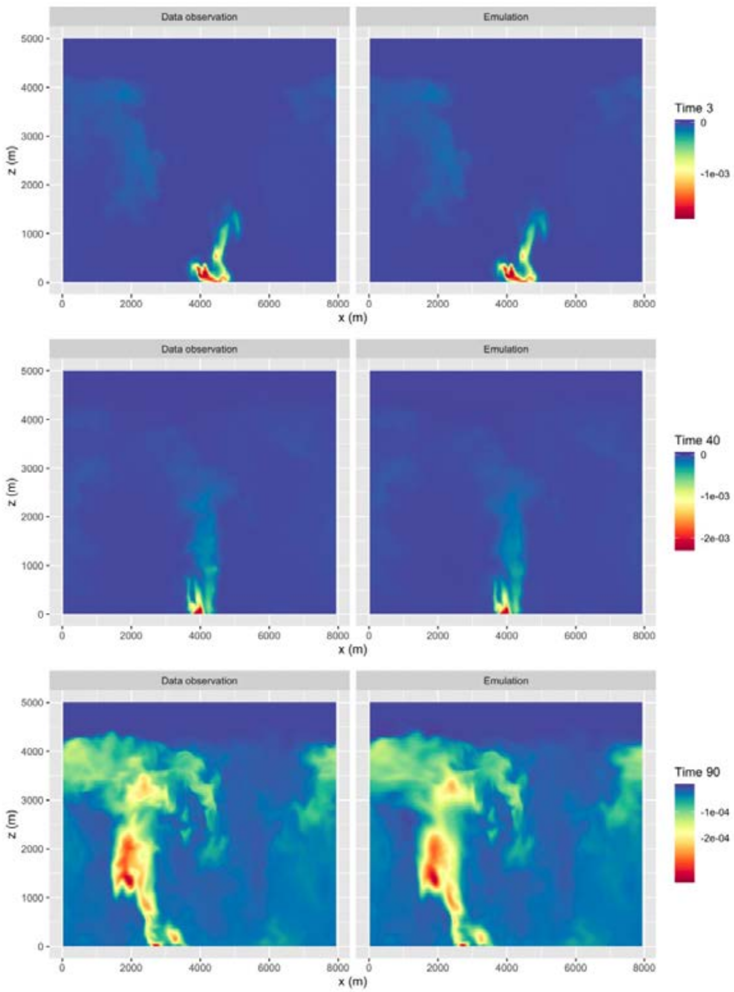
Zhang, L., Bhaganagar, K., & Wikle, C. K. (2025; *Neural Computing and Applications,* in press)

- Turbulent plume simulated generated using high-resolution LES methodology—bplume (Bhaganagar and Bhimireddy, 2020, 2021)
  - WRF-LES model is coupled with an active scalar transport equation (3-d)
- Simulated the September 2022 Mosquito wildland fire (California's largest wildland fire in 2022).
  - Realistic boundary conditions for the simulation were obtained from NOAA's High-Resolution Rapid Refresh (HRRR) model.
- We consider a 2-d slice and emulate it with our XVAE model
  - **70,000 spatial locations, 100 time periods**
  - Main difference between this and Red Sea example: Here, we use non-negative matrix factorization to find empirical spatial basis functions and compare to principle orthogonal decomposition (POD) and vanilla VAE.

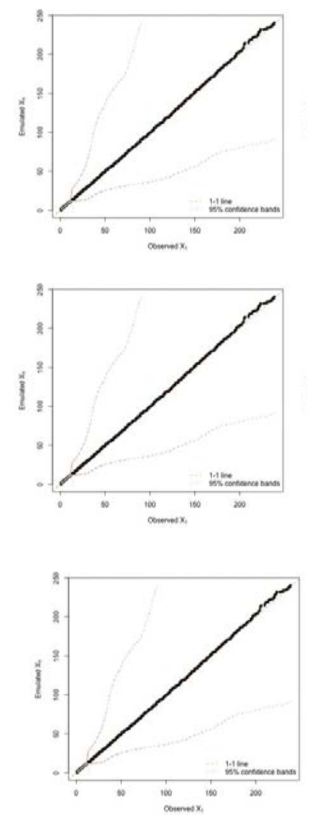
Fire Plume Simulation and Emulation: Time 3, 30, 90

data **emulation**

Fluid Density



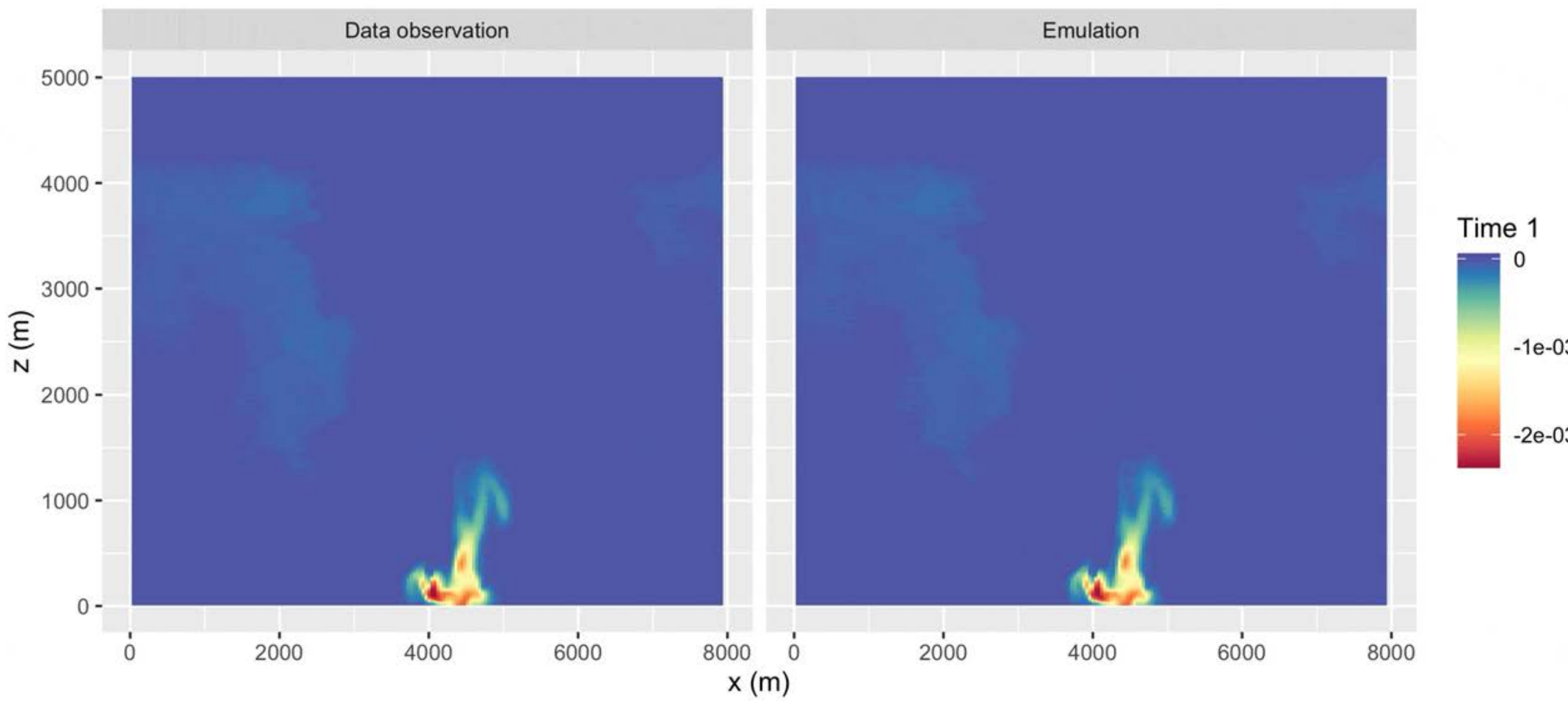
QQ plots



Note: Also, qq-plots and χ-plots show this captures extremes much better than proper orthogonal decomposition (POD) emulation and vanilla VAE.

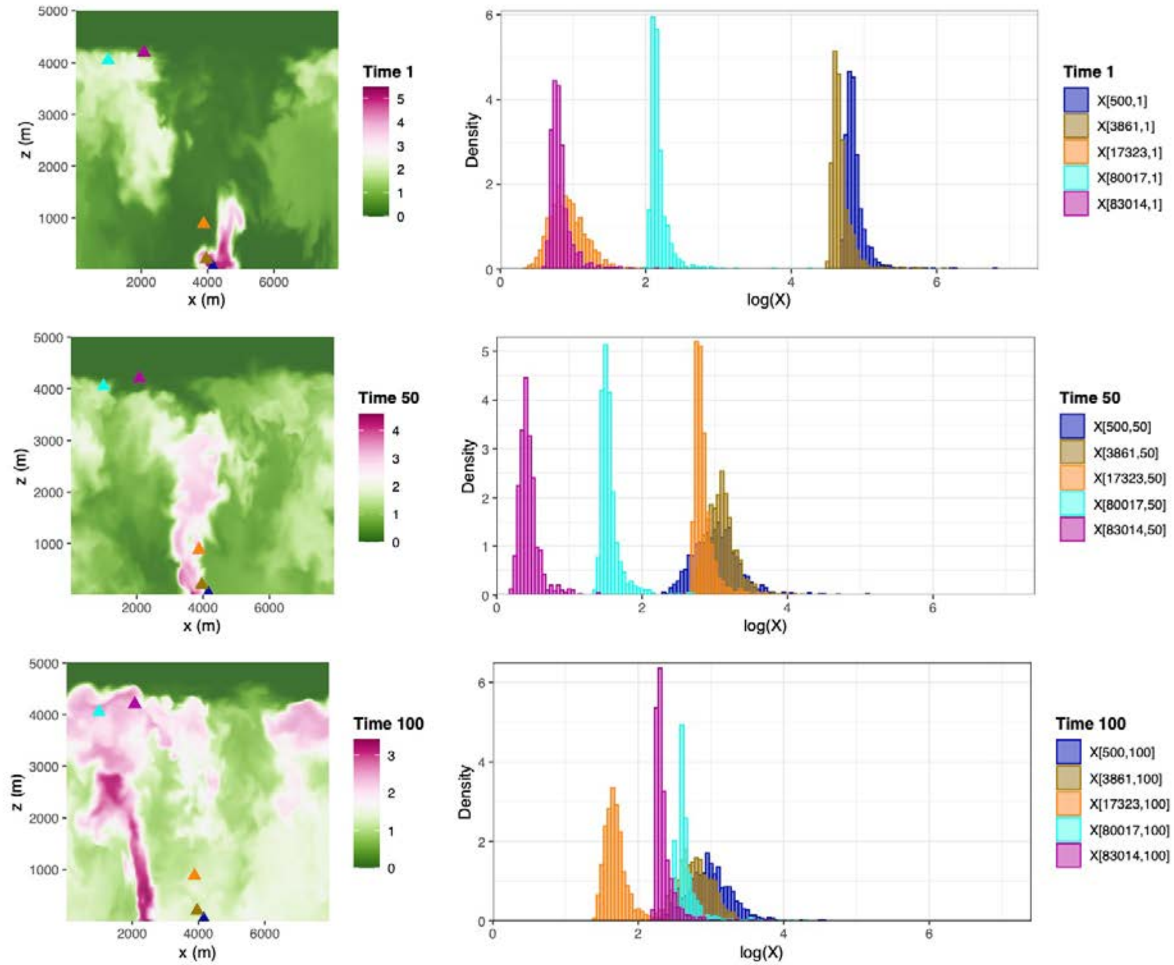
Fire Plume Animation

Original model simulation | Our model emulation

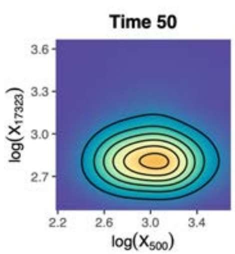


Uncertainty: Posterior Distribution

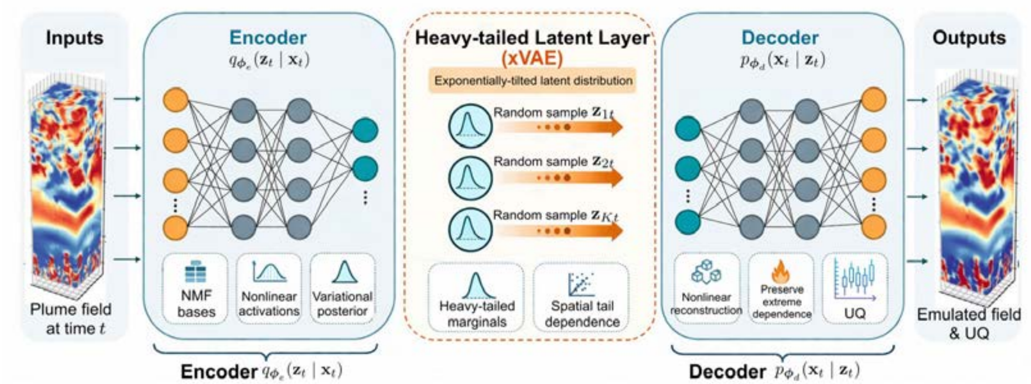
Posterior density at 5 locations for time 1, 50, 100 based on 1000 emulated samples.

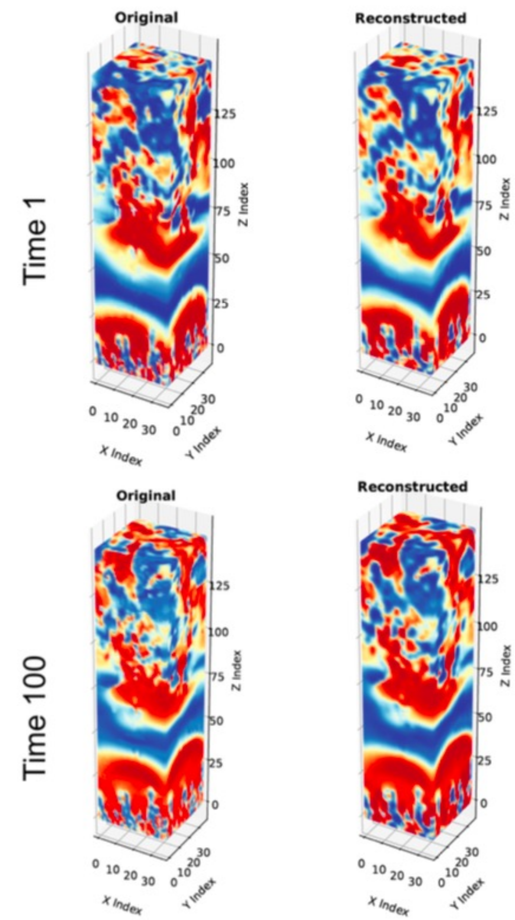


Can also get joint distributions: e.g.,



Extension: 3-D Turbulent Plume Emulation
(very preliminary results)





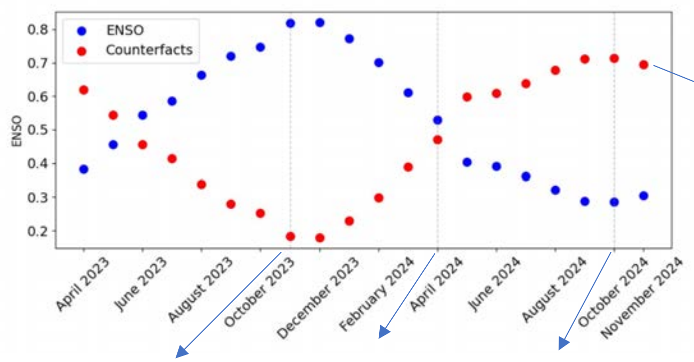
Normalized vertical velocity

(This is part of Michael Hatley's dissertation research; work in progress)

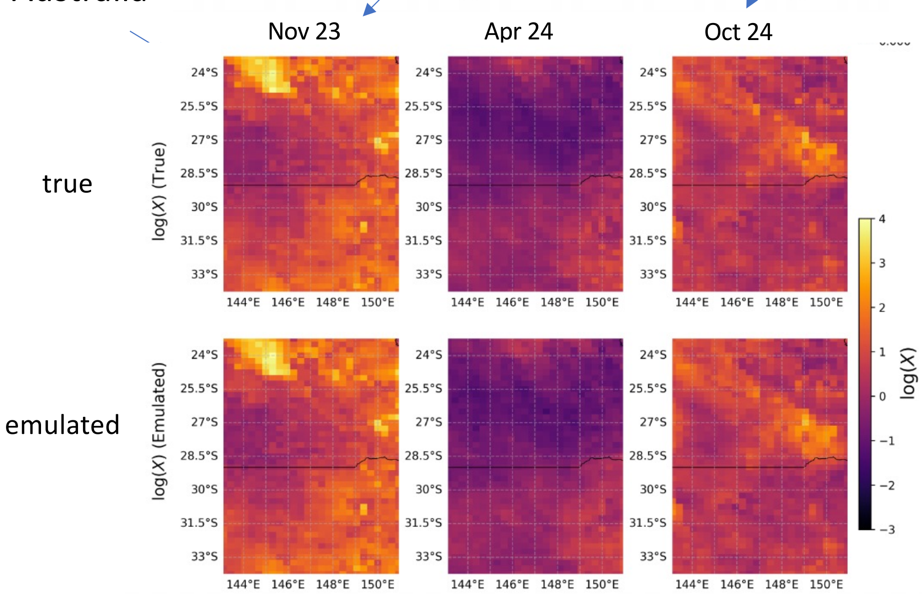
Extension: Conditional XVAE

(preliminary work in progress; from Xiaoyu Ma's dissertation)

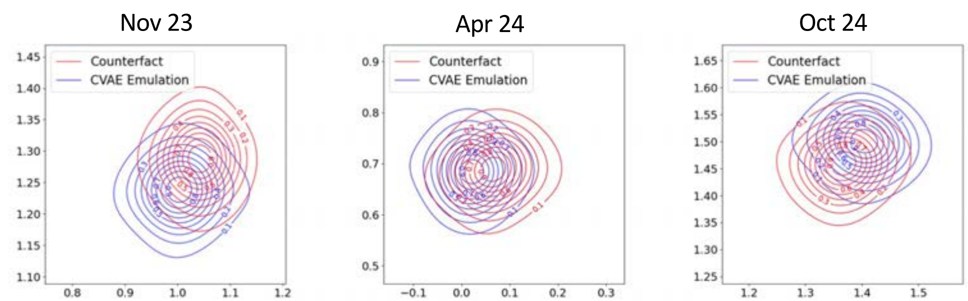
ENSO Index

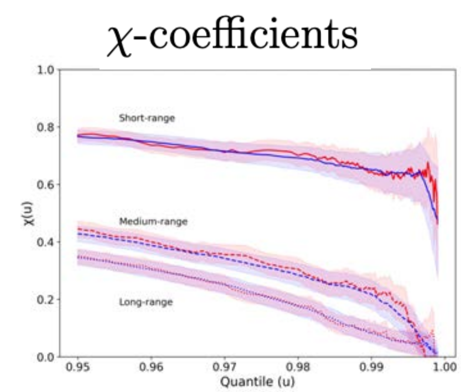


Fire Weather Index, Australia



Counterfactual Comparison





Blue – emulated
Red - observations

Summary

- Extreme weather-related events are increasing in number and have serious economic, ecosystem and health consequences
- Extreme events tend to cluster in space, and this is difficult to model in high dimensions
- In the last few years, many new statistical models for spatial extremes have been developed that can handle either AI or AD, but only a single dependence class over the entire domain
- We extend this model to make it flexible to handle non-stationary extremal dependence (AI or AD within the domain and over time)
- By embedding this in a variational autoencoder, we can implement the model on data sets of unprecedented size (for dependent extremes); amortized inference facilitates uncertainty quantification
- This can be used for inference, prediction, model emulation
- **Extensions:** realistic time dependence/dynamics!

Thank You!

Main Papers (with references cited here)

Zhang, L., Ma, X., Wikle, C.K., & R. Huser (2026). Fast and flexible emulation of spatial extremes processes via variational autoencoders. *Journal of the American Statistical Association* (in press).

Zhang, L., Bhaganagar, K., & Wikle, C. K. (2025). Capturing Extreme Events in Turbulence using an Extreme Variational Autoencoder (xVAE). *Neural Computing and Applications*. (in press) *arXiv preprint arXiv:2502.04685*.

wiklec@missouri.edu
(website)