Cascading Extremes

Prof. Miguel de Carvalho

Chair of Statistical Data Science

Joint with: Clemente Ferrer & Ronny Vallejos

THE UNIVERSITY of EDINBURGH
School of Mathematics

universidade de aveiro
theoria poiesis praxis

The Royal Society of Edinburgh
KNOWLEDGE MADE USEFUL

LEVERHULME TRUST

amazon
(submitted)

A Kolmogorov–Arnold Neural Model for Cascading Extremes

Miguel de Carvalho
School of Mathematics
University of Edinburgh
EH9 3FD, Edinburgh, UK
Miguel.deCarvalho@ed.ac.uk

Clemente Ferrer
Department of Mathematics
Universidad Técnica Federico Santa Maria
Valparaíso, Chile
Clemente.Ferrer@usm.cl

Ronny Vallejos
Department of Mathematics
Universidad Técnica Federico Santa Maria
Valparaíso, Chile
Ronny.Vallejos@usm.cl

Abstract

This paper addresses the growing concern of cascading extreme events, such as an extreme earthquake followed by a tsunami, by presenting a novel method for risk assessment focused on these domino effects. The proposed approach develops an extreme value theory framework within a Kolmogorov–Arnold network (KAN) to estimate the probability of one extreme event triggering another, conditionally on a feature vector. An extra layer is added to the KAN's architecture to enforce the definition of the parameter of interest within the unit interval, and we refer to the resulting neural model as KANE (KAN with Natural Enforcement). The proposed method is backed by exhaustive numerical studies and further illustrated with real-world applications to seismology and climatology.

Keywords Bernoulli process · Chained extreme events · KAN · Kolmogorov superposition theorem · Neural network · Multivariate extremes · Regression models for extremes

1 Introduction

Record-breaking extreme events—such as catastrophic wildfires, unprecedented flooding, intense hurricanes, and unparalleled heatwaves—underscore the urgent need to strengthen our quantitative understanding of these occurrences. Extreme Value Theory (EVT) offers a solid mathematical framework, leveraging regular variation and asymptotic principles to estimate risks of such events by extrapolating beyond the limits of available data, into the tails of a distribution (Coles, 2001; Beirlant et al., 2004; de Haan and Ferreira, 2006; Resnick, 2007).

While it is widely recognized by practitioners that extreme events tend to occur in complex sequential forms (Cutter, 2018; Raymond et al., 2020), statistical modelling of this context from an EVT viewpoint is still underdeveloped. Multivariate EVT, though a natural starting point, falls short by: i) disregarding the triggering role of certain events; ii) overlooking the order and sequential nature of extreme event cascades; iii) lacking the ability to model feedback loops between events.

Inspired by the multivariate EVT framework, this paper introduces a novel concept—the POC (Probability of Cascade) surface—which assesses the probability of domino effects between extreme events conditionally on a covariate or feature vector $\mathbf{x} = (x_1, \ldots, x_p)^\top$. As it will be shown below, the POC surface can be interpreted as the probability of a cascading extremal event, as it quantifies the probability that a trigger event (such as an earthquake exceeding magnitude u) results in a follow-up event (like a subsequent tsunami) as a function of a covariate. The proposed POC-based

Part I

Neural Statistical Modeling of Cascading Extremes

- While it is widely recognized by practitioners that extreme events tend to occur in complex sequential forms (Cutter, 2018; Raymond et al., 2020), statistical modelling of such context from an EVT viewpoint is still underdeveloped.
- Multivariate EVT, though a natural starting point, falls short by:
  - disregarding the triggering role of certain events;
  - overlooking the order and sequential nature of extreme event cascades;
  - lacking the ability to model feedback loops between events.

- Inspired by the multivariate EVT framework, in this talk I introduce a novel concept, the POC (Probability of Cascade) surface, to assess the probability of domino effects between extreme events conditionally on a covariate or feature vector $\mathbf{x} \in \mathbb{R}^p$.

- The proposed POC-based approach is fully general in the sense that the focus can be placed beyond the case where follow-up event is binary.

- In particular, we extend the framework to a multi-class setting, allowing for different types of follow-up extreme events.

- The case where the follow-up event is continuous includes as a particular case the conditional coefficient of extremal dependence introduced by Lee et al. (2024).

Background

- To learn about the POC surface from the data, we develop a neural model grounded on Kolmogorov's superposition theorem.
- Superpositions are functions of functions.

Example (Superposition of univariate and bivariate functions)

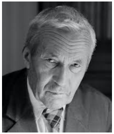
$$f(x_1, x_2, x_3) = g\big(a(\alpha(x_1), \beta(x_2, x_3)), b(x_1, x_2)\big).$$

Theorem (Kolmogorov's superposition theorem)

Let $f : [0,1]^d \to \mathbb{R}$ be a continuous function. Then,

$$f(\mathbf{x}) = \sum_{i=1}^{2d+1} \Phi_i^{(2)} \left(\sum_{j=1}^{d} \Phi_{i,j}^{(1)}(x_j) \right), \qquad \mathbf{x} = (x_1, \ldots, x_d)^{\mathrm{T}},$$

for some continuous one-dimensional functions $\Phi_{i,j}^{(1)}$ and $\Phi_j^{(2)}$.



A. Kolmogorov

Background

- From an AI perspective, the theorem reveals a two-layer neural network architecture recently popularized by Liu et al. (2024) following their extension to deeper settings.
- While multi-layer perceptrons are inspired by the universal approximation theorem (e.g., Berlyand and Jabin, 2023), Kolmogorov–Arnold Networks (KAN) are a novel and fast-evolving addition to the AI toolbox, and are rooted on Kolmogorov's superposition theorem.
- A particularly impressive aspect of KAN is that they are based on the principle any multivariate continuous function can be expressed exactly using only $2d + 1$ outer functions and d inner functions.
- In addition to the many developments following Liu et al., it should be noted that other neural approaches based on this theorem had already appeared in the literature (Lin and Unbehauen, 1993; Sprecher and Draghici, 2002; Montanelli and Yang, 2020; Fakhoury et al., 2022).

- Let $I = \{I_u : u \in \mathbb{R}\}$ be a Bernoulli process and $Y \sim F_Y$ be a continuous rv.
- We start by introducing the following functional, referred to as alpha, which plays a central role in our developments:

$$\alpha \equiv \alpha_I = \lim_{u \to y^*} P(I_u = 1 \mid Y > u).$$

- Notation: $y^* = \sup\{y : F_Y(y) < 1\}$ is the right endpoint of F_Y.
- Loosely, α is the probability of a follow-up event (like a tsunami $I_u = 1$), given a trigger event (such as an earthquake exceeding magnitude u).
- The nature of the Bernoulli process I defining the follow-up event opens up a variety of modeling possibilities as illustrated below.

Examples of Alpha

Example (Tail dependence coefficient)

If $I_u = I(Z > u)$, where Y and Z have common distribution, then

$$\alpha^{\mathrm{TDC}} \equiv \alpha_I = \lim_{u \to y^*} P(Z > u \mid Y > u).$$

Thus, α in (8) includes the well-known tail dependence coefficient as a special case when the follow-up event involves Z being extreme and observable. ■

Example (Extremal probabilistic index)

If $I_u = I(Z > Y)$, then

$$\alpha^{\mathrm{PI}} \equiv \alpha_I = \lim_{u \to y^*} P(Z > Y \mid Y > u),$$

which can be regarded as extremal version of the probabilistic index (Thas et al., 2012). ■

POC Surface

Setup and Definition

- Our setup keeps in mind that for some applications the variable Z in the above examples might be latent, but it assumes that the Bernoulli process I is always observable.
- In practice it is desirable to assess how the α functional may be impacted by a covariate or feature.

Definition (POC Surface)

Let $\mathbf{x} = (x_1, \dots, x_d)^{\mathrm{T}} \in \mathcal{X} \subseteq \mathbb{R}^d$. The probability of cascade surface is defined as

$$\mathrm{POC} = \{(\mathbf{x}, \alpha_I(\mathbf{x})) : \mathbf{x} \in \mathcal{X}\}, \quad \alpha_I(\mathbf{x}) = \lim_{u \to y^*} P(I_{u,\mathrm{x}} = 1 \mid Y_{\mathrm{x}} > u),$$

where $I = \{I_{u,\mathrm{x}} : (u, \mathbf{x}) \in \mathbb{R} \times \mathcal{X}\}$ is a random field with Bernoulli marginal distributions and $\{Y_{\mathrm{x}} : \mathbf{x} \in \mathcal{X}\}$ is a random field.

Starting Point (KAN):

$$\alpha_l(\mathbf{x}) = \sum_{i=1}^{2d+1} \Phi_i^{(2)} \left(\sum_{j=1}^{d} \Phi_{i,j}^{(1)}(x_j) \right).$$

Or in function matrix notation

$$\alpha_l(\mathbf{x}) = (\mathbf{\Phi}^{(2)} \circ \mathbf{\Phi}^{(1)})\,\mathbf{x},$$

where

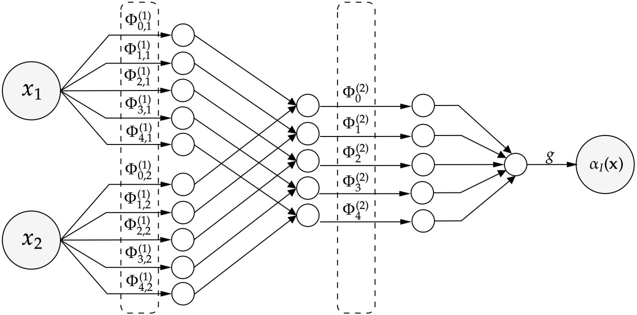
$$\mathbf{\Phi}^{(1)} = \begin{pmatrix} \Phi_{1,1}^{(1)} & \cdots & \Phi_{1,d}^{(1)} \\ \vdots & \ddots & \vdots \\ \Phi_{2d+1,1}^{(1)} & \cdots & \Phi_{2d+1,d}^{(1)} \end{pmatrix}, \qquad \mathbf{\Phi}^{(2)} = \begin{pmatrix} \Phi_1^{(2)} \\ \vdots \\ \Phi_{2d+1}^{(2)} \end{pmatrix}^{\mathrm{T}}.$$

Issue: $\alpha_l(\mathbf{x})$ may not be in $[0, 1]$!

Refined Version (KANE): Let $g : \mathbb{R} \to [0, 1]$, and set

$$\alpha_I(\mathbf{x}) = g\left(\sum_{i=1}^{2d+1} \Phi_i^{(2)}\left(\sum_{j=1}^{d} \Phi_{i,j}^{(1)}(x_j)\right)\right).$$

Or in function matrix notation $\alpha_I(\mathbf{x}) = g(\mathbf{\Phi}^{(2)} \circ \mathbf{\Phi}^{(1)})\,\mathbf{x}$. **Now**: $\alpha_I(\mathbf{x}) \in [0, 1]$.



Deep POC Surface

A Deep Version of the Model *a la* Liu et al.

Deep Version (L Layer Model): Let $g : \mathbb{R} \to [0, 1]$, and set

$$\alpha_l(\mathbf{x}) = g(\boldsymbol{\Phi}^{(L-1)} \circ \cdots \circ \boldsymbol{\Phi}^{(1)}),$$

where

$$\boldsymbol{\Phi}^{(l)} = \begin{pmatrix} \Phi^{(l)}_{1,1} & \cdots & \Phi^{(l)}_{1,n_l} \\ \vdots & \ddots & \vdots \\ \Phi^{(l)}_{n_l+1,1} & \cdots & \Phi^{(l)}_{n_{l+1},n_l} \end{pmatrix},$$

where n_l is the number of nodes in the lth layer.

- Consider $m+1$ equally-spaced knots, $t_0 < \cdots < t_m$. We model the inner and outer functions as a linear combination of B-spline basis functions, that is,

$$\Phi_{i,j}^{(l)}(x) = \sum_{k=1}^{K} \beta_{i,j,k}^{(l)} B_k^p(x),$$

for $i = 1, \ldots, d$, where $B_k^p(x)$ is a B-spline basis function of degree p evaluated at x and $K = p + m$.

- The parameter of interest is given by the following collection of matrices

$$\beta_k^{(l)} = \begin{pmatrix} \beta_{n_l,1,k}^{(l)} & \cdots & \beta_{n_l,d,k}^{(l)} \\ \vdots & \ddots & \vdots \\ \beta_{n_{l+1},1,k}^{(l)} & \cdots & \beta_{n_{l+1},d,k}^{(l)} \end{pmatrix},$$

where $k = 1, \ldots, K$ and $l = 1, \ldots, L$.

Consequences & Extensions

Multi-Trigger Systems

- In practice, multiple competing incidents can contribute to trigger in the follow-up event.
- To address this we define a multi-trigger system where we consider $\mathcal{Y}_1, \dots, \mathcal{Y}_K$ as a sequence of identically distributed random fields with

$$\mathcal{Y}_k = \{Y_{k,\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}, \qquad k = 1, \dots, K.$$

- Our framework extends to the framework of K trigger events by considering

$$\alpha_I(\mathbf{x}) = \lim_{u \to y^*} P(I_{u,\mathbf{x}} = 1 \mid Y_{1,\mathbf{x}} > u \vee \cdots \vee Y_{K,\mathbf{x}} > u).$$

- This formula simplifies to the original single-trigger case by defining $Y_\mathbf{x} = \min\{Y_{1,\mathbf{x}}, \dots, Y_{K,\mathbf{x}}\}$, and hence the theory and methods discussed earlier readily apply to this context as well.

- In real-world applications, the follow-up extreme event may come in different flavors or categories.

Example

For example, $j = 0$ may represent *no tornado*, $j = 1$, *supercell tornado*, and $j = 2$ a *non-supercell tornado*.

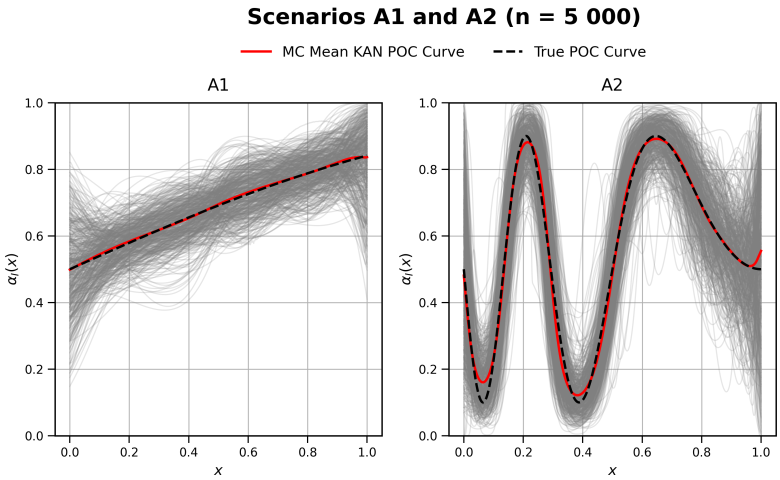
- The proposed cascading probability surfaces naturally extend to this context as follows

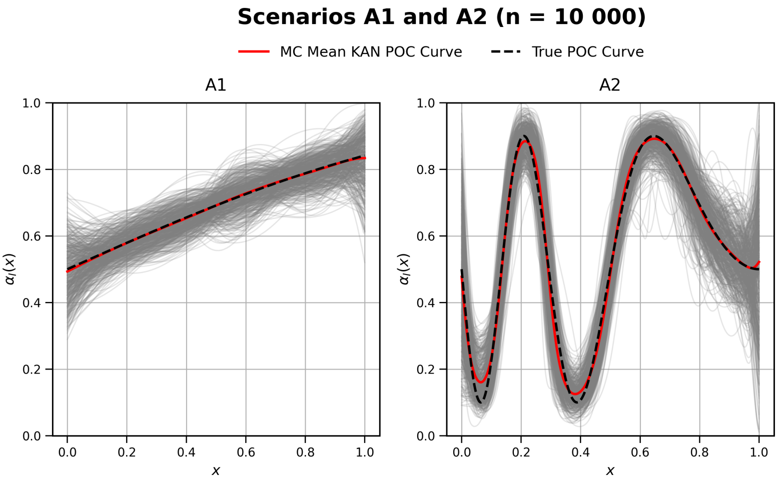
$$\alpha_I(\mathbf{x})^{(j)} = \lim_{u \to y^*} P(I_{u,\mathrm{x}} = j \mid Y_{\mathrm{x}} > u),$$

where $j = 0, \dots, J$.

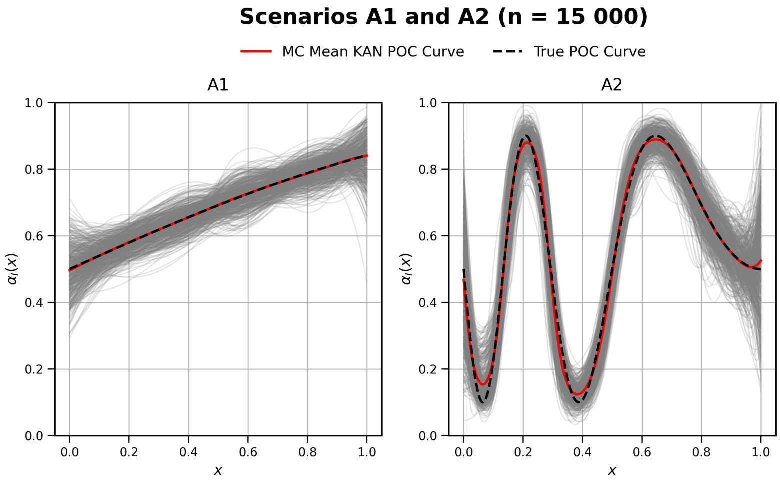
Monte Carlo Simulation Study

Scenarios A





Scenarios A1 and A2 (n = 10 000)
MC Mean KAN POC Curve
True POC Curve
A1
A2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
$\alpha_l(x)$
x


Scenarios A1 and A2 (n = 15 000)
MC Mean KAN POC Curve
True POC Curve
A1
A2
$\alpha_l(x)$
x
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Real Data Illustration

Applied Rationale and Data Description

- Coastal regions are highly vulnerable to tsunamis, with their impacts often compounded by extreme earthquakes that act as primary triggers.
- We now apply the proposed method to quantify the probability of cascade for tsunami occurrence triggered by extreme earthquakes.
- Our analysis uses data from the NCEI/WDS Global Significant Earthquake Database, provided by the NOAA National Centers for Environmental Information.
- The dataset contains over 5 700 significant earthquakes from 2150 B.C. to present, defined by criteria such as fatalities, damages over $1 million, Modified Mercalli Intensity (MMI) X or greater, or the earthquake generated a tsunami.
- Each observation includes event date, location, depth, magnitude, MMI, and socio-economic impacts (casualties, injuries, property damage), with references and notes on related events like tsunamis and eruptions.

- We threshold the data at their 95% threshold for fitting the POC surface and consider the features,

$$(\text{latitude}, \text{longitude}, \text{depth}).$$

- We transform longitude and latitude coordinates to the unit square for modeling purposes.
- Finally, we fit a KAN model with three layers, using a sigmoid activation function at the output.

Real Data Illustration

Visualization

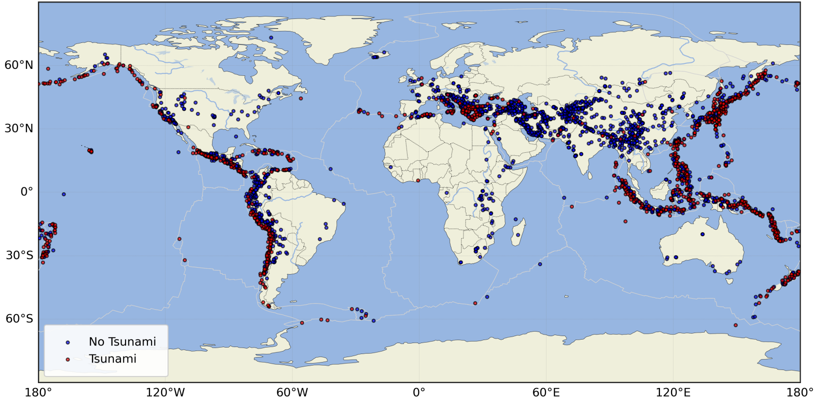
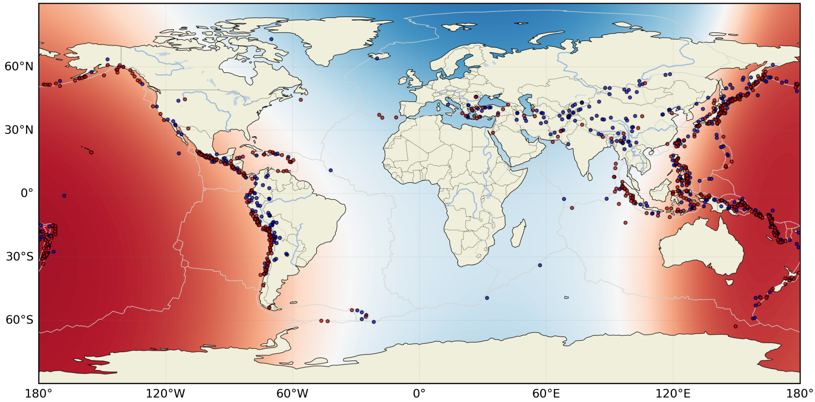


Figure: Point pattern of earthquakes (red) and associated tsunami occurrences (blue).

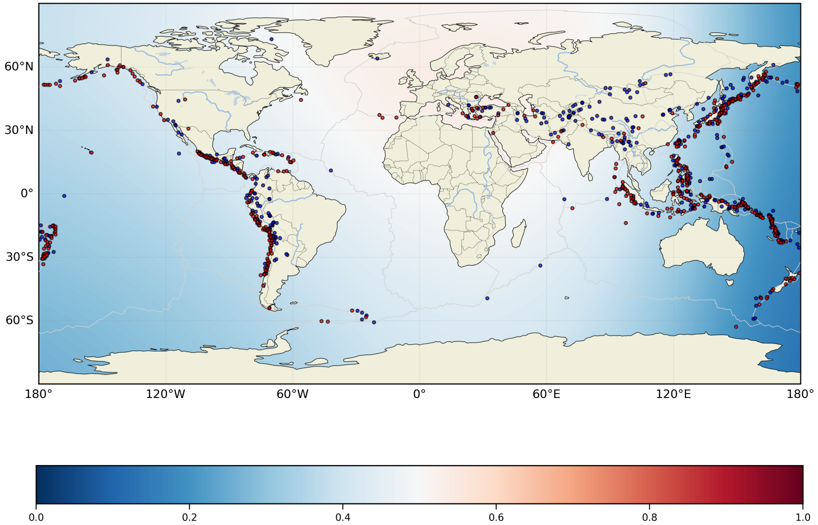
Real Data Illustration

POC Surface—Depth: Percentile 5



Real Data Illustration

POC Surface—Depth: Percentile 85



Part II

Generative Transformer-Based Approaches for Cascading Extremes

Rationale

Cascades are Natural Models for Generative Extremes

- Chains of extreme events and cascades offer a natural starting point for thinking about generative extremes.

NVIDIA $\downarrow 3.6\%$ $\rightarrow$ TSMC $\downarrow 3.1\%$ $\rightarrow$ AMD $\downarrow 4.2\%$ $\rightarrow$ $\cdots$

Earthquake (M 7.8) $\rightarrow$ Tsunami (Wave height 8m) $\rightarrow$ Power outage (Region-wide) $\rightarrow$ $\cdots$

- The starting point for the construction of our first generative AI approach for extreme events is based on notions of multi-headed attention as well as of transformers, as introduced in Vaswani et al (2017); for an introduction see Bishop & Bishop (2023; §12).

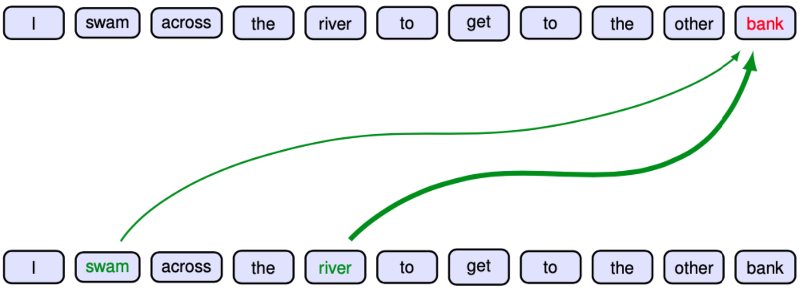


Figure 12.1 Schematic illustration of attention in which the interpretation of the word 'bank' is influenced by the words 'river' and 'swam', with the thickness of each line being indicative of the strength of its influence.

Context

- Whereas the unit of analysis in standard transformers is a sequence of tokens, in our EVT-based transformer is a sequence of extreme events.

NVIDIA $\downarrow 3.6\%$ $\longrightarrow$ TSMC $\downarrow 3.1\%$ $\longrightarrow$ AMD $\downarrow 4.2\%$ $\longrightarrow$ $\cdots$

- By analogy with language models, where the meaning of a token depends on its context, the interpretation of an extreme event similarly depends on the surrounding events.

Example

The sequence of extreme stock losses shown in the above figure is more indicative of a tech-specific cascade than of a broader macro-financial shock.

The proposed model is conceived to learn from data both:

- the most probable continuations of extreme event sequences;
- the probabilities of subsequent extremes.

Closing Remarks

Summary

- This talk presented a novel statistical framework to tackle the rising concern of cascading extreme events—like tsunamis followed by earthquakes or heatwaves sparking wildfires, which in turn lead to further losses.
- The proposed approach aims to offer a novel outlook into triggering extremal events and their domino effects.
- KANE, a neural model based on Kolmogorov's superposition theorem, was developed to learn about the proposed POC surface.
- In addition, I offered some remarks on how cascades offer a natural starting point for thinking about generative extremes.
- Whereas the unit of analysis in standard transformers is a sequence of tokens, in our EVT-based transformer is a sequence of extreme events.

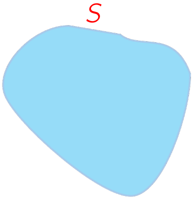
Thank you!
Questions?

Miguel.deCarvalho@ed.ac.uk

Other Directions

Future Work, I

- In practice, either the follow-up or the trigger event may be functional in the sense of FDA (Functional Data Analysis) (Horváth and Kokoszka, 2012; Kokoszka, 2017).
- For example, we may observe $I_{u,\mathbf{x}} = 1$ for $\mathbf{x}$ over a *continuum* rather than over a point $\mathbf{x} \in \mathbb{R}^d$.
- As a concrete instance of this, in the earthquake data application, the follow-up event could represent the full *region* $S \subset \mathbb{R}^2$ affected by the tsunami, in which case $I_{u,\mathbf{x}} = 1$ for all $\mathbf{x} \in S$.
- While our theory also accommodates this framework, further investigation is needed to incorporate such functional events into the inferences in a fully FDA-aligned fashion.



Other Directions

Future Work, II

- The proposed approach may set the stage for extremal chains of random length ($k \leq d$), and also factoring in the role of the chain's pathways.
- TDMs (Tail Dependence Matrices) (Embrechts et al., 2016) can perhaps be used govern a Markovian pathway of the extremal cascade, with the order of events dictated by a random permutation of $\{1, \ldots, k\}$ based on TDMs' transition probabilities between extremal events.
- This will allow for modelling and learning from the data the length or size of the extremal cascade (k) as well as the distribution of the pathway taken by it (say, $4 \rightarrow 1 \rightarrow 2 \rightarrow 3$ or $3 \rightarrow 2 \rightarrow 1 \rightarrow 4$ as two examples of realizations of the chain of extremes), where trigger can vary (e.g., Y_4 or Y_3).

Supporting Information

Model Checking

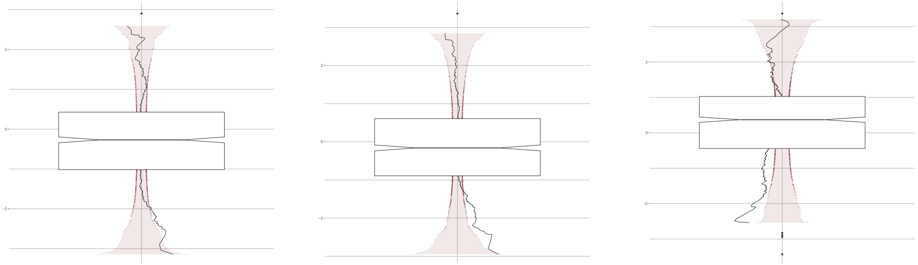
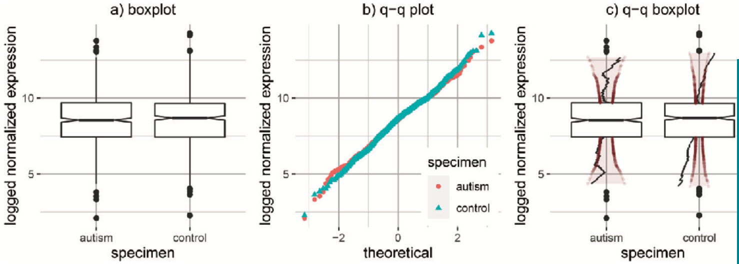


Figure: QQ boxplot of Dunn–Smyth residuals.

QQ Boxplot

Rodu and Kafadar (2022; *Journal of Computational and Graphical Statistics*)



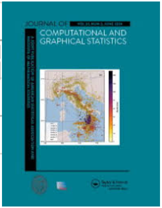


Figure 1. A comparison of the boxplot, q-q plot, and q-q boxplot highlights advantages of the q-q boxplot. Logged, normalized gene expression data for a patient with autism (left) and a "control" patient (right) as displayed by (a) boxplot; (b) q-q plot referenced to the normal distribution[3]; and (c) q-q boxplot referenced to the normal distribution. Data come from a random sample of the observations obtained from the Expression Atlas (Papatheodorou et al. 2019) (https://www.ebi.ac.uk/gxa/experiments/E-GEOD-30573/Results).

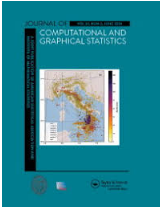
Randomized Quantile Residuals

Dun and Smyth (1996; *Journal of Computational and Graphical Statistics*)

3. RANDOMIZED QUANTILE RESIDUALS

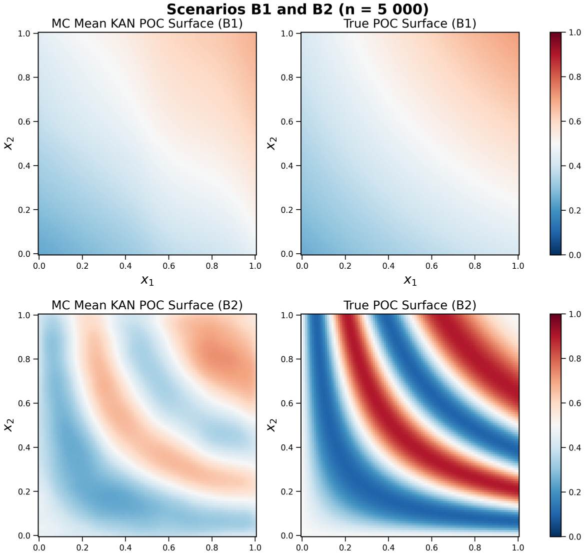
Let $F(y; \mu, \phi)$ be the cumulative distribution function of $\mathcal{P}(\mu, \phi)$. If F is continuous, then the $F(y_i; \mu_i, \phi)$ are uniformly distributed on the unit interval. In this case, the quantile residuals are defined by

$$r_{q,i} = \Phi^{-1}\{F(y_i; \hat{\mu}_i, \hat{\phi})\},$$



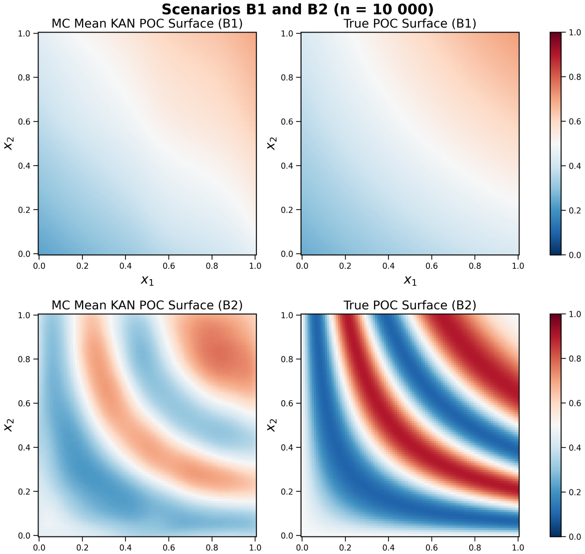
Randomized Quantile Residuals

Scenarios B



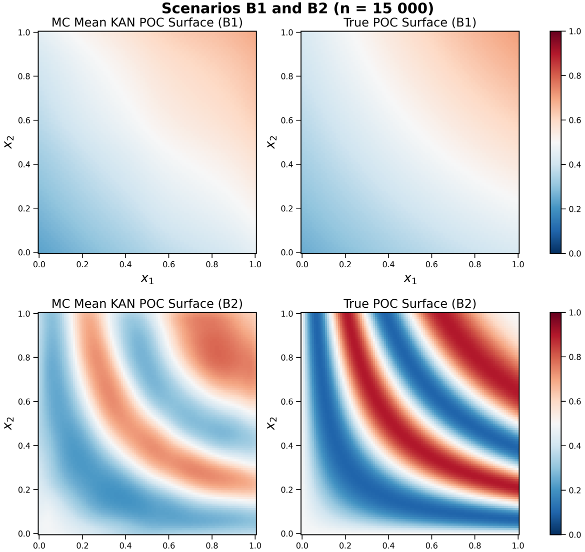
Randomized Quantile Residuals

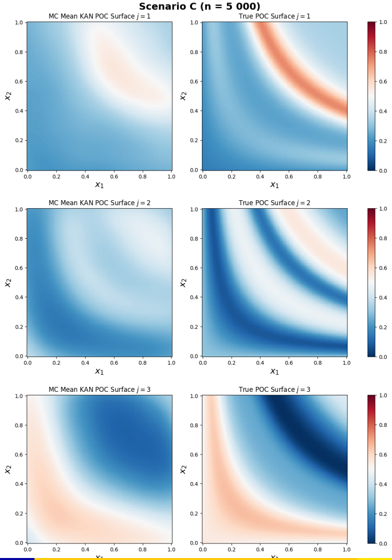
Scenarios B



Randomized Quantile Residuals

Scenarios B

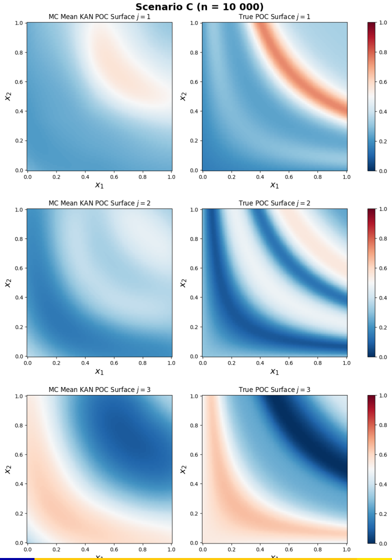




Scenario C (n = 5 000)
MC Mean KAN POC Surface $j = 1$
True POC Surface $j = 1$
MC Mean KAN POC Surface $j = 2$
True POC Surface $j = 2$
MC Mean KAN POC Surface $j = 3$
True POC Surface $j = 3$
x_1
x_2
0.0
0.2
0.4
0.6
0.8
1.0

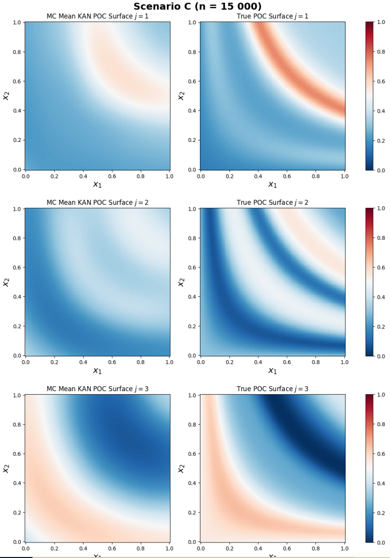
Randomized Quantile Residuals

Scenarios C



Randomized Quantile Residuals

Scenarios C



Universal Approximation Theorem

Theorem (Universal Approximation Theorem)

Suppose f is a continuous function on a compact space $\mathcal{X} \subset \mathbb{R}^d$ and σ is not a polynomial. Then, for any $\varepsilon > 0$, there exists a one-hidden layer neural network $f_{\boldsymbol{\theta}}$ such that

$$\sup_{\mathbf{x} \in \mathcal{X}} |f(\mathbf{x}) - f_{\boldsymbol{\theta}}(\mathbf{x})| < \varepsilon.$$

Notation: Here $f_{\boldsymbol{\theta}} : \mathbb{R}^d \to \mathbb{R}$ is a one-hidden layer feedforward neural network composed of K neurons

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = b^{(2)} + \sum_{k=1}^{K} w_k^{(2)} \sigma(\langle \mathbf{w}_k^{(1)}, \mathbf{x} \rangle + b_k^{(1)}),$$

where

$$\boldsymbol{\theta} := \{b^{(2)}\} \cup \{\mathbf{w}_k^{(1)}, w_k^{(2)}, b_k^{(1)}\}_{k=1}^{K} \in \Theta := \mathbb{R} \times (\mathbb{R}^d \times \mathbb{R} \times \mathbb{R})^K,$$

and where $\langle \cdot, \cdot \rangle$ is a scalar product on $\mathbb{R}^d$, and $\sigma : \mathbb{R} \to \mathbb{R}$ is a non-linear activation function.

Allouche, M., Girard, S., and Gobet, E. (2022), "EV-GAN: Simulation of extreme events with ReLU neural networks," *Journal of Machine Learning Research*, 23, 1–39.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*, Wiley, Hoboken, NJ: Wiley.

Berlyand, L. and Jabin, P.-E. (2023), *Mathematics of deep learning: An introduction*, Boston: de Gruyter.

Bhatia, S., Jain, A., and Hooi, B. (2021), "Exgan: Adversarial generation of extreme samples," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 6750–6758.

Braun, J. and Griebel, M. (2009), "On a constructive proof of Kolmogorov's superposition theorem," *Constructive Approximation*, 30, 653–675.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, London: Springer.

Cutter, S. L. (2018), "Compound, cascading, or complex disasters: What's in a name?" *Environment: Science and Policy for Sustainable Development*, 60, 16–25.

de Haan, L. and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, New York: Springer.

Embrechts, P., Hofert, M., and Wang, R. (2016), "Bernoulli and tail-dependence compatibility," *The Annals of Applied Probability*, 26, 1636–1658.

Fakhoury, D., Fakhoury, E., and Speleers, H. (2022), "ExSpliNet: An interpretable and expressive spline-based neural network," *Neural Networks*, 152, 332–346.

Horváth, L. and Kokoszka, P. (2012), *Inference for Functional Data with Applications*, vol. 200, New York: Springer.

Karpov, O. E., Grubov, V. V., Maksimenko, V. A., Kurkin, S. A., Smirnov, N. M., Utyashev, N. P., Andrikov, D. A., Shusharina, N. N., and Hramov, A. E. (2022), "Extreme value theory inspires explainable machine learning approach for seizure detection," *Scientific Reports*, 12, 11474.

Kokoszka, P. (2017), *Introduction to Functional Data Analysis*, Boca Raton, FL: Chapman & Hall/CRC.

Lee, J., de Carvalho, M., Rua, A., and Avila, J. (2024), "Bayesian smoothing for time-varying extremal dependences," *Journal of the Royal Statistical Society, Ser. C*, 73, 581–597.

Lin, J.-N. and Unbehauen, R. (1993), "On the realization of a Kolmogorov network," *Neural Computation*, 5, 18–20.

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., and Tegmark, M. (2024), "KAN: Kolmogorov–Arnold networks," *arXiv:2404.19756*.

Montanelli, H. and Yang, H. (2020), "Error bounds for deep ReLU networks using the Kolmogorov–Arnold superposition theorem," *Neural Networks*, 129, 1–6.

NS, K. S., Wang, Y., Schram, M., Drgona, J., Halappanavar, M., Liu, F., and Li, P. (2024), "Extreme risk mitigation in reinforcement learning using extreme value theory," *Transactions on Machine Learning Research*, 05.

Raymond, C., Horton, R. M., Zscheischler, J., Martius, O., AghaKouchak, A., Balch, J., Bowen, S. G., Camargo, S. J., Hess, J., Kornhuber, K., et al. (2020), "Understanding and managing connected extreme events," *Nature Climate Change*, 10, 611–621.

Resnick, S. I. (2007), *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, New York: Springer.

Sprecher, D. A. and Draghici, S. (2002), "Space-filling curves and Kolmogorov superposition-based neural networks," *Neural Networks*, 15, 57–67.

Thas, O., Neve, J. D., Clement, L., and Ottoy, J.-P. (2012), "Probabilistic index models," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 74, 623–671.

Temporary page!

LATEX was unable to guess the total number of pages correctly. As th
some unprocessed data that should have been added to the final pag
page has been added to receive it.
If you rerun the document (without altering it) this surplus page will
because LATEX now knows how many pages to expect for this docum