Deep Compositional Models for Non-Stationary Extremal Dependence

Xuanjie Shao Jordan Richards* Raphaël Huser

Statistics Program, CEMSE Division, KAUST

جامعة الملك عبدالله للعلوم والتقنية

King Abdullah University of Science and Technology

Introduction

- Modeling **non-stationarity** in **extremal dependence** is challenging, while inference for stationary and isotropic models is considerably easier.
- **Deformation approach** (Sampson and Guttorp, 1992): warp the original domain $\mathcal{G}$ to a latent space $\mathcal{W}$ where stationarity/isotropy can be reasonably assumed.
- Consider a spatial process $Z(\cdot)$ on $\mathcal{G}$ with finite variance for each $\boldsymbol{s} \in \boldsymbol{S}_0 \subset \mathcal{G}$. Estimation of the warping function $\boldsymbol{f} : \mathcal{G} \to \mathcal{W}$ is problematic: computationally expensive, lacking in flexibility, and not injective.
- **Deep composition modeling approach** (Zammit-Mangion et al., 2022): transformation is constructed using a series of n injective warping functions,

$$\boldsymbol{f}(\cdot) \equiv \boldsymbol{f}_n \circ \boldsymbol{f}_{n-1} \circ \cdots \circ \boldsymbol{f}_1(\cdot).$$

- We present an extension of this methodology to spatial extremal processes, using the $\mathbb{R}$ interface to **TensorFlow 2**.

Deep Compositional Model (DCM)

The DCM comprises n layers in a neural network, with

$$\boldsymbol{f}_i(\boldsymbol{s}) = \boldsymbol{W}_i \boldsymbol{\phi}_i(\boldsymbol{s}; \boldsymbol{\Theta}_i) : D_{i-1} \to D_i, \quad i = 1, \dots, n,$$

mapping the input to a d_i-dimensional output for $D_i \subset \mathbb{R}^{d_i}$, $D_0 \equiv \mathcal{G}$, and $D_n \equiv \mathcal{W}$.

- $\boldsymbol{\phi}_i(\cdot; \boldsymbol{\Theta}_i)$: basis functions at the i-th layer with parameters $\boldsymbol{\Theta}_i$.
- $\boldsymbol{W}_i$: weights for the basis-functions at the i-th layer.

Different choices for $\boldsymbol{\phi}_i$ lead to different types of warping units.

- **Axial Warping Unit (AWU)**: $\boldsymbol{f}_i$ models nonlinear scaling of one input dimension.
- **Radial Basis Function (RBF)**: $\boldsymbol{f}_i$ describes local expansions/contractions at various resolutions.
- **Möbius Transformation**: $\boldsymbol{f}_i$ performs a rotation in the complex plane.

Brown–Resnick Process

Brown-Resnick (BR) Process

- The Brown–Resnick process is a max-stable process that can be expressed as

$$Z(\boldsymbol{s}) = \sup_{i \geq 1} U_i(\boldsymbol{s})/P_i,$$

where P_i's are Poisson point processes on $\mathbb{R}_+$ with unit rate intensity, and $U_i(\boldsymbol{s})$ are i.i.d. copies of a non-negative stochastic process

$$U(\boldsymbol{s}) = \exp\{\epsilon(\boldsymbol{s}) - \sigma^2(\boldsymbol{s})/2\}, \quad \boldsymbol{s} \in \mathcal{S},$$

for an intrinsically-stationary Gaussian process $\epsilon(\boldsymbol{s})$ with variance $\sigma^2(\boldsymbol{s})$, $\epsilon(\boldsymbol{0}) \overset{a.s.}{=} 0$, and semivariogram $\gamma(h) = (h/\varphi)^\kappa$ with range $\varphi > 0$ and smoothness $\kappa \in (0, 2]$.
- The bivariate joint distribution function of Z is $\Pr\{Z(\boldsymbol{s}_1) \leq z_1, Z(\boldsymbol{s}_2) \leq z_2\} = \exp\{-V(z_1, z_2)\}$, where V is the **bivariate exponent function**

$$V(z_1, z_2) = z_1^{-1}\Phi\left\{a/2 - a^{-1}\log(z_1/z_2)\right\} + z_2^{-1}\Phi\left\{a/2 - a^{-1}\log(z_2/z_1)\right\},$$

with $a = \{2\gamma(h)\}^{1/2}$ and $h = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$.
- **Extremal coefficient (EC)** measures extremal dependence between sites $\boldsymbol{s}_1, \boldsymbol{s}_2$, denoted $\theta(\boldsymbol{s}_1, \boldsymbol{s}_2) = V(1, 1) = 2\Phi\left\{\sqrt{2\gamma(h)}/2\right\} \in [1, 2]$.

Inference

- **Pairwise likelihood (PL)**: ℓ_{PL} with dependence parameters $\boldsymbol{\psi}$ is

$$\ell_{PL}(\boldsymbol{\psi}) = \sum_{t=1}^{T} \sum_{\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}} \ell_{PL}(z_{t,1}, z_{t,2}; \boldsymbol{\psi})$$

where $z_{t,i}$ is t-th block maxima at $\boldsymbol{s}_i \in \mathcal{W}$, and ℓ_{PL} represents the corresponding pairwise log-density.
To save time $\Rightarrow$ Careful choice of a suitable smaller number of pairs.
- **Least squares inference through extremal coefficients**: using empirical bivariate extremal coefficients $\hat{\theta}_2(\boldsymbol{s}_1, \boldsymbol{s}_2)$,

$$\ell_{EC}(\boldsymbol{\psi}) = \sum_{\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}} \left\{\hat{\theta}_2(\boldsymbol{s}_1, \boldsymbol{s}_2) - \theta_2(\boldsymbol{s}_1, \boldsymbol{s}_2 | \boldsymbol{\psi})\right\}^2.$$

- **The joint loss function**: with the original spatial locations $\boldsymbol{S}_0 \subset \mathcal{G}$,

$$L(\boldsymbol{\psi}, \boldsymbol{\Theta}, \boldsymbol{W}; \boldsymbol{S}_0) = \ell(\boldsymbol{\psi}; \boldsymbol{S}_n),$$

where the warped locations $\boldsymbol{S}_n = \boldsymbol{f}(\boldsymbol{S}_0; \boldsymbol{\Theta}, \boldsymbol{W}) \subset \mathcal{W}$. Here $\boldsymbol{W} = \{\boldsymbol{W}_1, \dots, \boldsymbol{W}_n\}$, $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_n\}$ are weights and parameters in the DCM.
We can take $\ell(\boldsymbol{\psi}; \boldsymbol{S}_n)$ to be either $-\ell_{PL}(\boldsymbol{\psi})$ or $\ell_{EC}(\boldsymbol{\psi})$.

References

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119.

Shao, X., Hazra, A., Richards, J., and Huser, R. (2022). Flexible modeling of nonstationary extremal dependence using spatially-fused LASSO and ridge penalties. *arXiv preprint arXiv:2210.05792*.

Zammit-Mangion, A., Ng, T. L. J., Vu, Q., and Filippone, M. (2022). Deep compositional spatial models. *Journal of the American Statistical Association*, 117(540):1787–1808.

Simulation Study

- Simulated data over a warped domain (generated with 2 AWUs and an RBF unit) on a 101×101 grid are generated from a **stationary BR process** with $\varphi = 0.1$ and $\kappa = 1$. $D = 2000$ locations are chosen randomly for inference, with 100 independent replications held out for training.
- The deep compositional model consists of AWUs for each input axis, one RBF, and a Möbius transformation. Both PL and EC inference methods are used to reconstruct the warped space.
- Computation time for PCL inference using **1% (100%) pairs** is **1h (7h)**; computation time for EC inference is generally **less than 1h**.

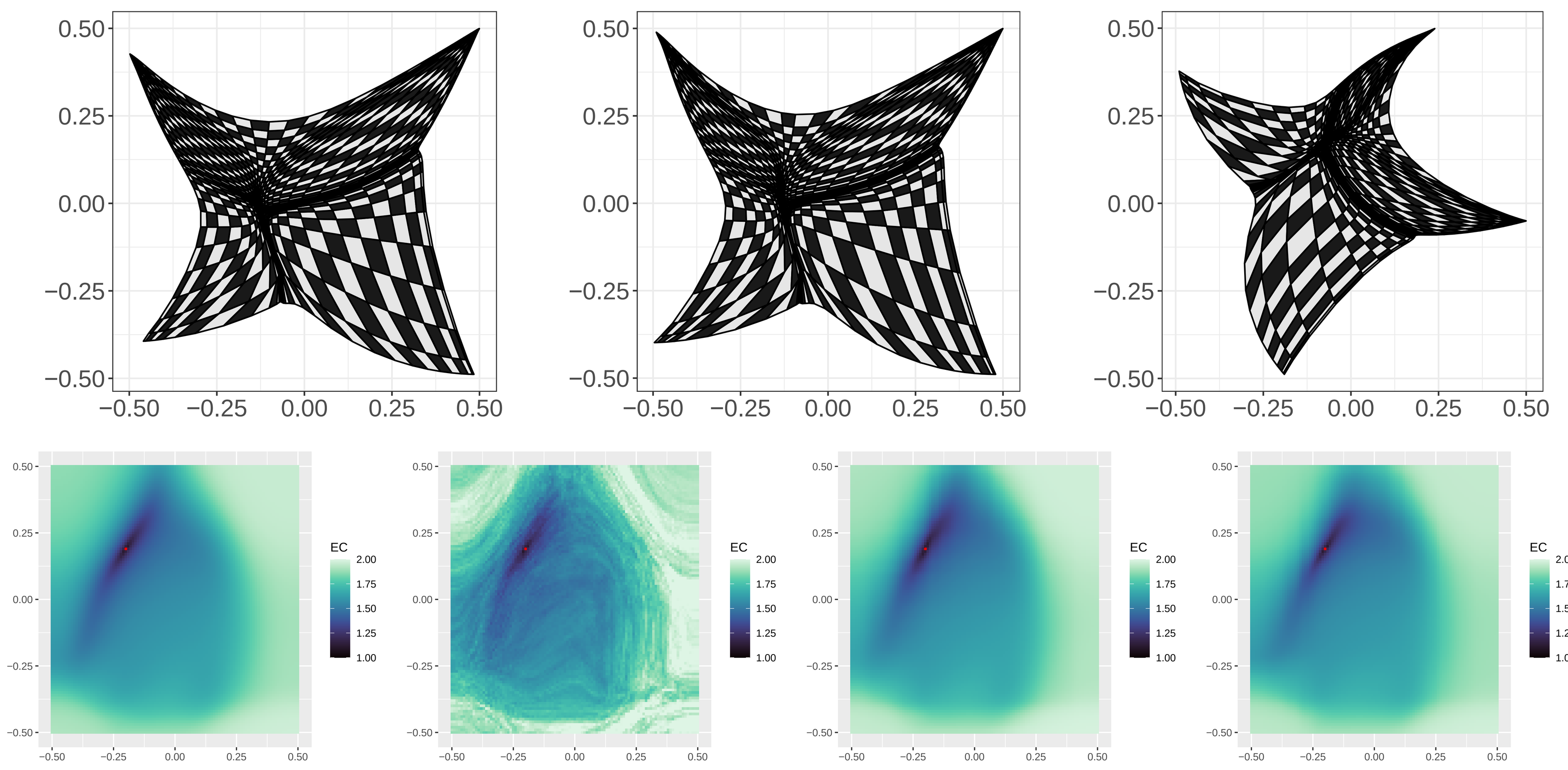

Figure 1. **Top**: **true** (left), and **fitted** (PCL: middle; EC: right) space;
Bottom: **theoretical and empirical pairwise ECs** (1st, 2nd), **fitted pairwise ECs** (PCL: 3rd; EC: 4th) relative to the reference site in red.

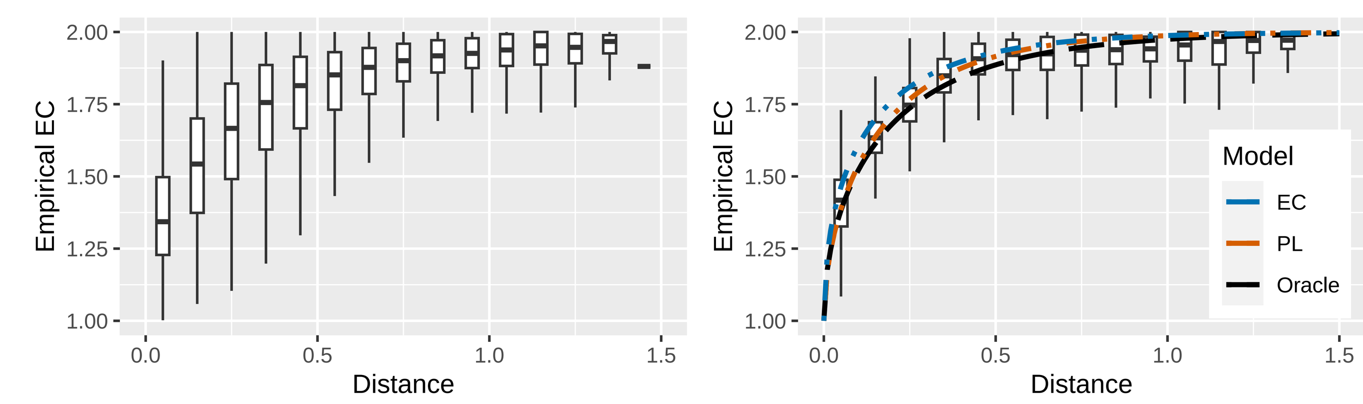

Figure 2. Boxplot of empirical ECs against distance (left: original space; right: warped space), and model EC curves.

Application: Monthly Temperature Maxima over Nepal

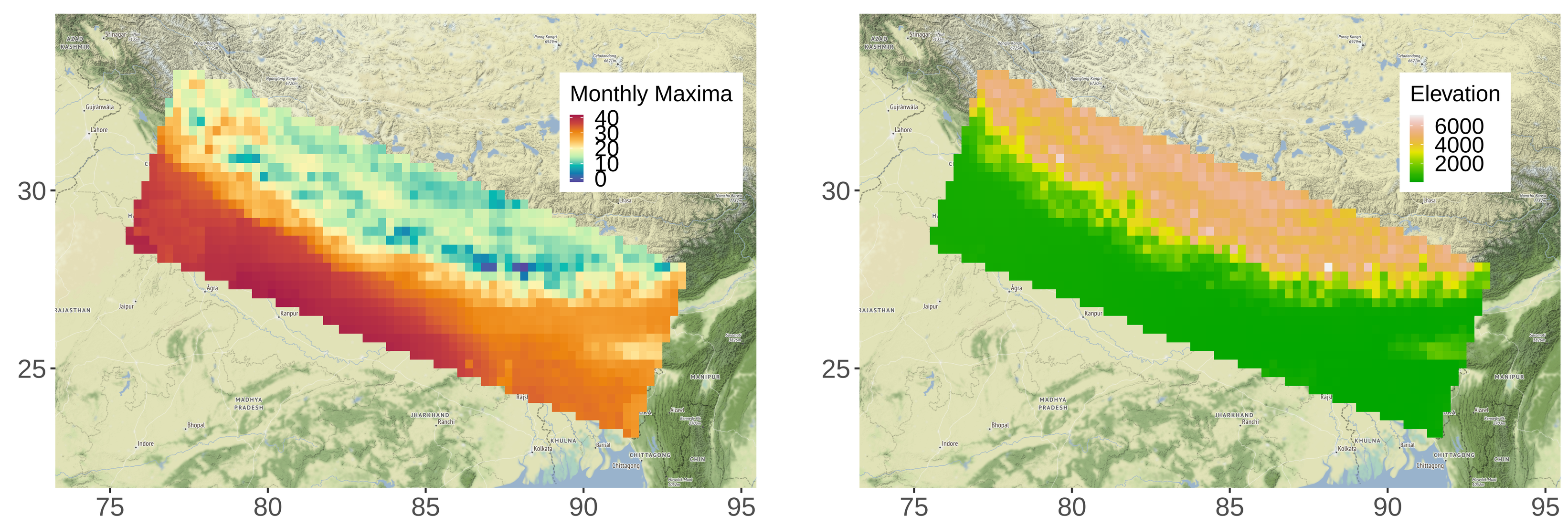

Figure 3. Left: one observed monthly maxima (°C); right: elevation of the spatial domain.

- **Monthly maxima** of daily average temperatures with $D = 1417$ sites and $T = 192$ replications, with **unit Fréchet margins**.
- Previously analysed by Shao et al. (2022) with a locally-stationary model.
- Elevation information is further introduced, which improve the modeling of some spatial areas, e.g., around Everest.

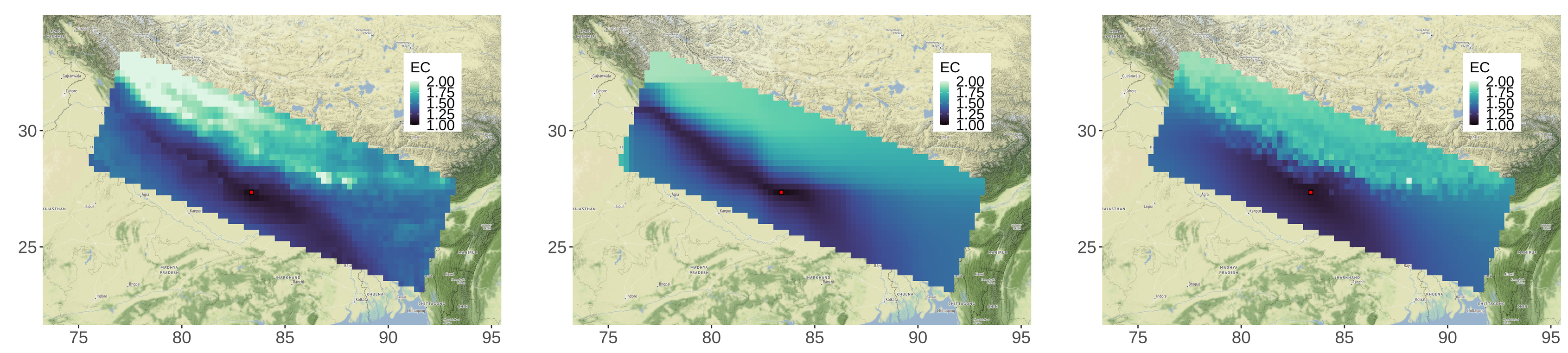

Figure 4. Left: empirical pairwise ECs; middle: fitted pairwise ECs (relative to a reference site; red) based on the DCM with a Brown-Resnick process (using least squares inference); right: fitted pairwise ECs incorporating elevation in $\mathcal{G}$.

Future Work

- Extension to **other extremal processes**, e.g., r-Pareto processes, inverted BR.
- Modeling temporal changes, uncertainty assessment, new warping units.