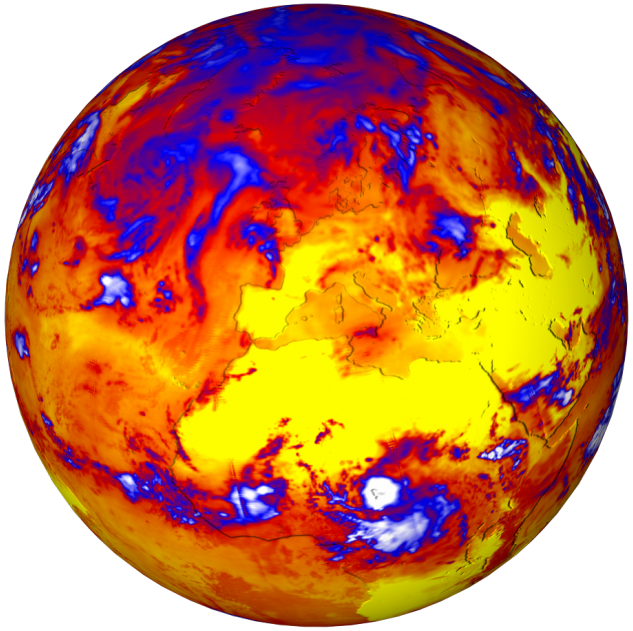
Fully Bayesian Inference for Spatial Extremes Using Hierarchical Extreme Value Processes

Ben Shaby

Joint work with Brian Reich



(Photo: NASA)

Outline

1. Introduction
2. Max-stable Processes
3. A Hierarchical Representation
4. Back to the Motivating Example
5. Discussion

Outline

1. Introduction
2. Max-stable Processes
3. A Hierarchical Representation
4. Back to the Motivating Example
5. Discussion

Motivating Example: Crops, heat shocks, and food security

From Shaby and Reich (2013):

Very high temperatures are detrimental to agriculture.

- Key concern: If one farm experiences a heat shock, it's a problem. If many farms do, it's a disaster.

Which leads to ...

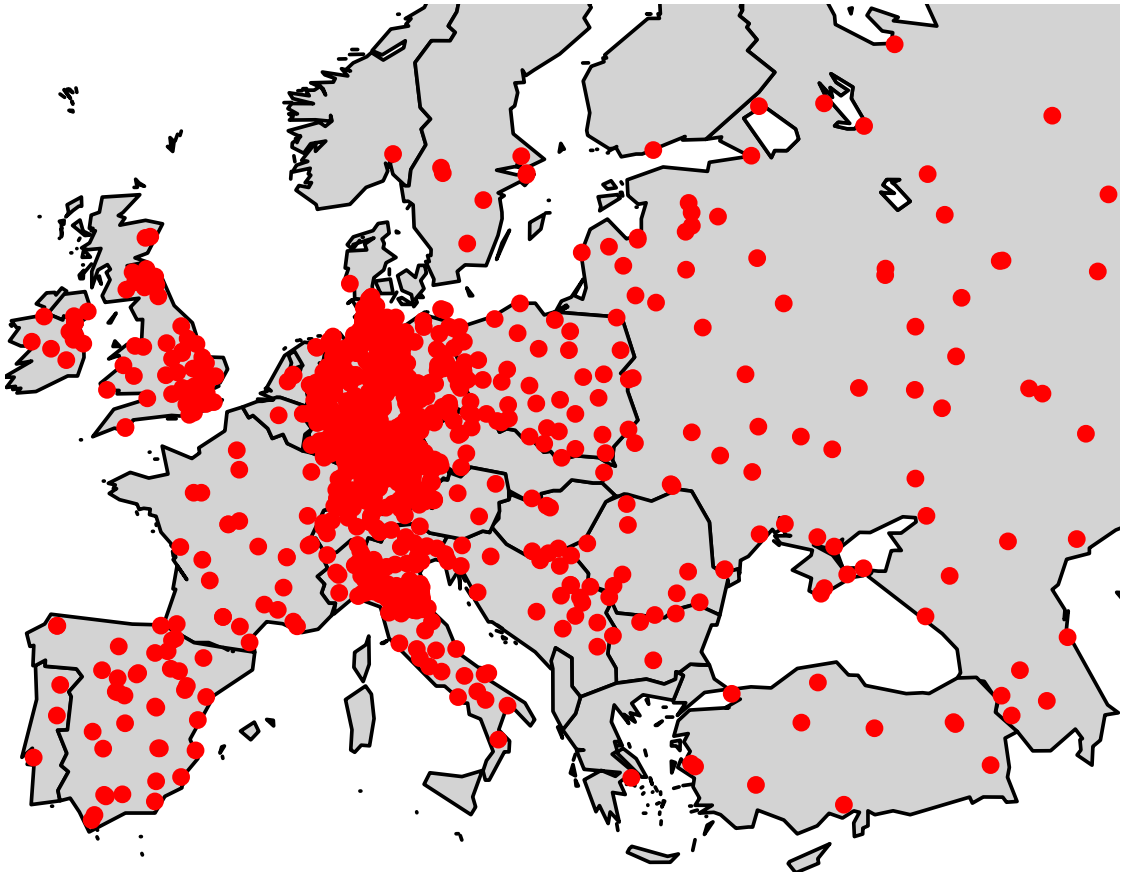
- Big question: For given values p and T, what is the probability that p percent of the agricultural land will experience a temperature higher than T in a given year?

And the follow-up:

- Is that probability changing?

What is the Data?

- Annual max temperatures at 985 monitoring locations for 1900–2011.



Goals and Strategy

We want to define a class of stochastic processes that

1. possess properties required by extreme value theory.
2. capture spatial dependence.
3. can be fit using Bayesian methods.

Goals and Strategy

We want to define a class of stochastic processes that

1. possess properties required by extreme value theory.
2. capture spatial dependence.
3. can be fit using Bayesian methods.

We will get there by

1. truncating a spectral representation to give a finite collection of parameters.
2. building a spatial process using the truncated spectrum and latent variables.
3. conditioning on the spatial process to form a data likelihood.

* Bonus: this also allows us to do spatial prediction.

Advantages to a Bayesian Approach

In addition to the usual advantages,

- Models that assume unit Fréchet margins will require many marginal transformations.
- by definition we are data poor, and parameters are hard to estimate (especially ξ)!
  $\rightarrow$ Use priors to trade space for time.
- answers to the interesting questions are often probability statements about complicated transformations of unknown parameters.
- likelihoods of existing models are intractable.
  $\rightarrow$ Build process models conditionally.

Outline

1. Introduction
2. Max-stable Processes
3. A Hierarchical Representation
4. Back to the Motivating Example
5. Discussion

Max-stable processes

Max stability characterizes extreme value processes.

- Pointwise maxima of stochastic processes can only converge to max-stable processes.
- In this sense the only asymptotically justified class of maximal processes are max-stable.
- Can we build spatial processes that are max-stable?

A Spectral Representation

There is a more suggestive way to characterize extreme value processes (Schlather, 2002).

Let

- $\{H_j\}_{j=1}^{\infty}$ be a Poisson process on $\mathbb{R}_+$ with intensity measure dh/h^2.
- $\{W_j'(\mathbf{s})\}_{j=1}^{\infty}$ be independent copies of a stationary process on $\mathcal{D} \subset \mathbb{R}^d$, and $W_j(\mathbf{s}) = \max\{0, W_j'(\mathbf{s})\}$ for $j = 1, \ldots, \infty$ such that $\mathsf{E}[W_j(\mathbf{s})] = 1$ for all $\mathbf{s}$.

Then

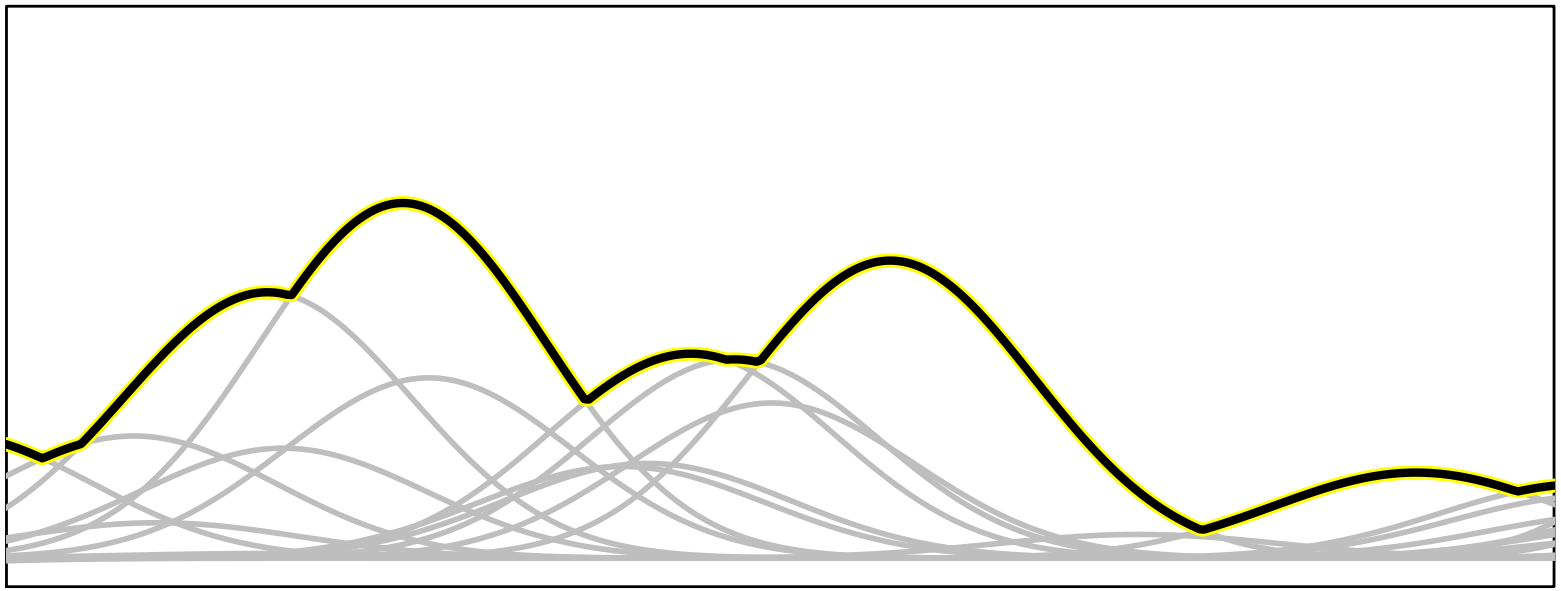
$$X(\mathbf{s}) = \max_j \{H_j W_j(\mathbf{s})\}$$

is a stationary max-stable process on $\mathcal{D}$ with unit Fréchet margins.

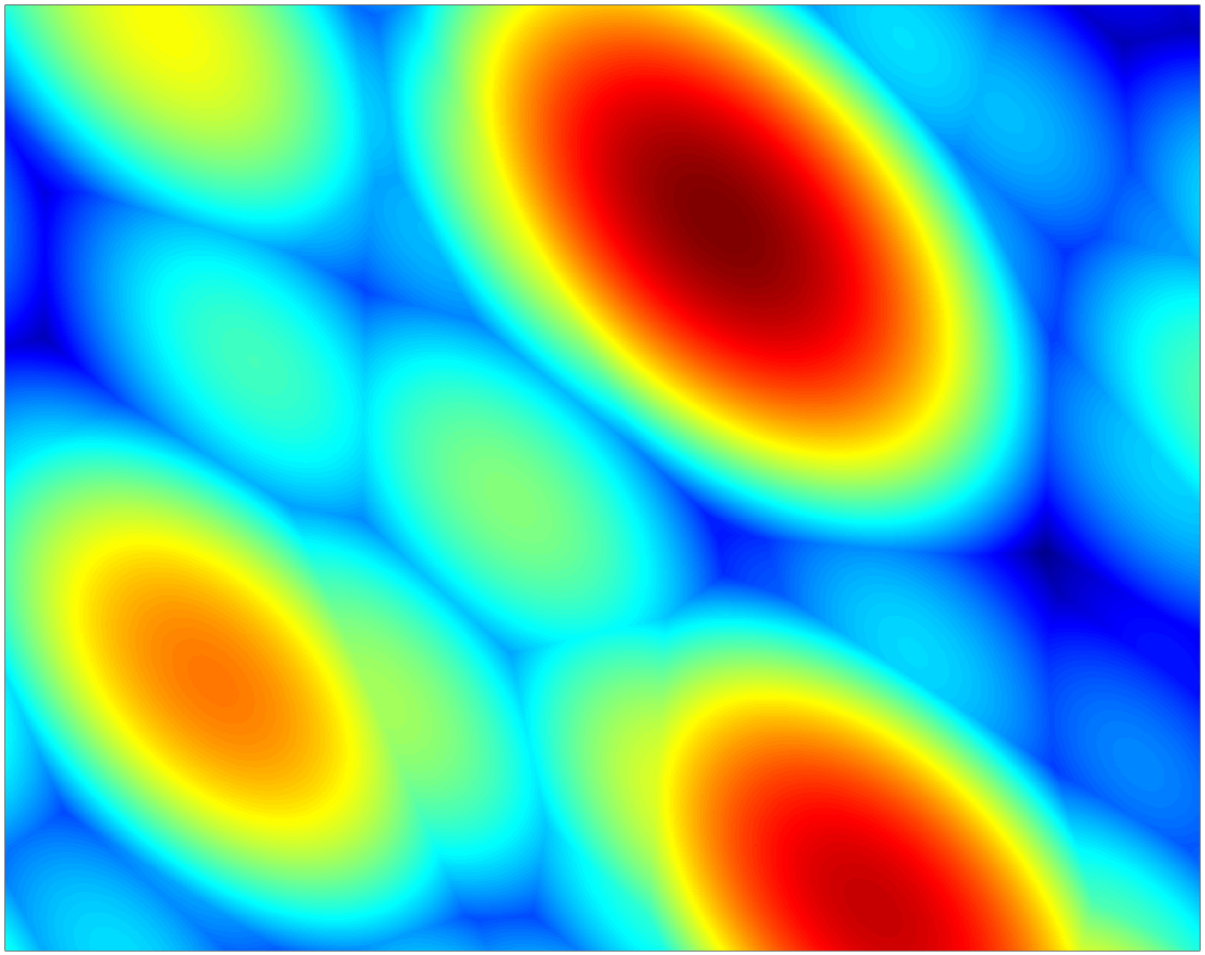
Example: Gaussian Extreme Value Process (GEVP)

To construct the "Smith process" (Smith, 1990), let

- $W_j(\mathbf{s}) = \varphi(\mathbf{s}; \mu_j, \mathbf{\Sigma})$ are normal kernels with mean μ_j and covariance matrix Σ, $j = 1, \ldots, \infty$,
- $\{\mu_j\}_{j=1}^{\infty}$ is a homogeneous Poisson process on $\mathcal{D}$.



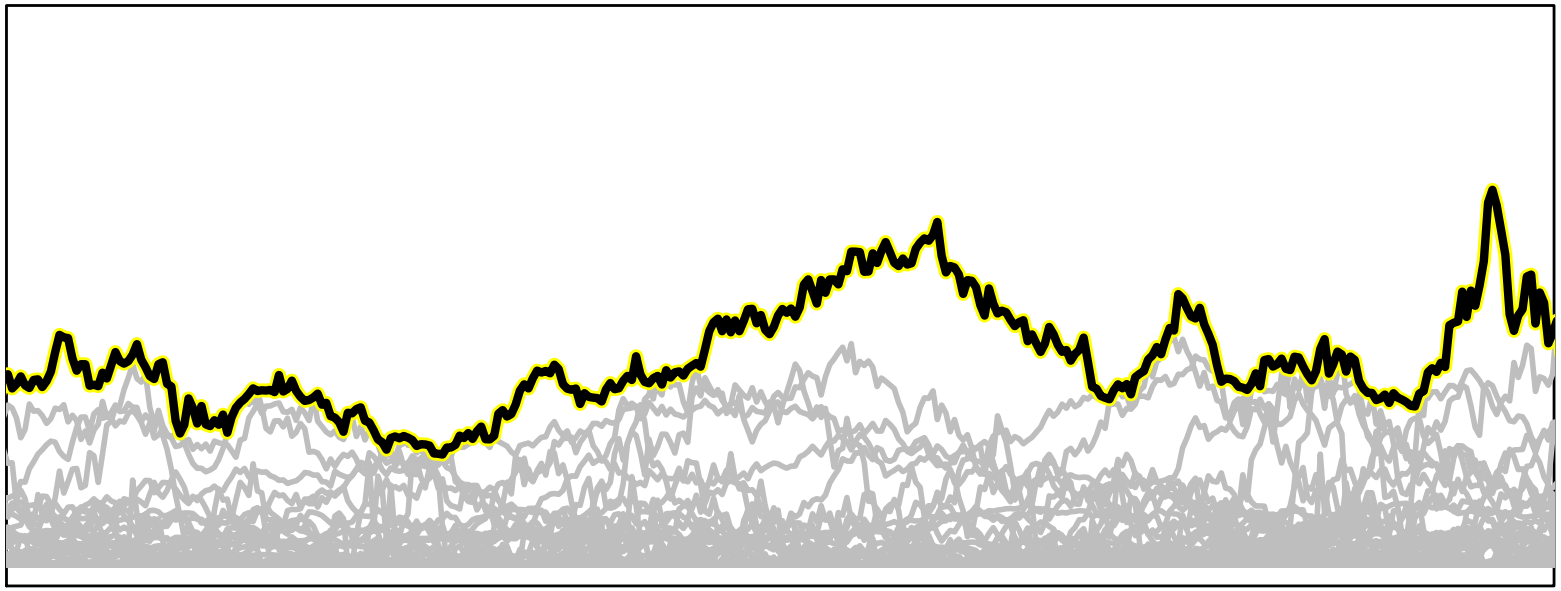
Example: Gaussian Extreme Value Process



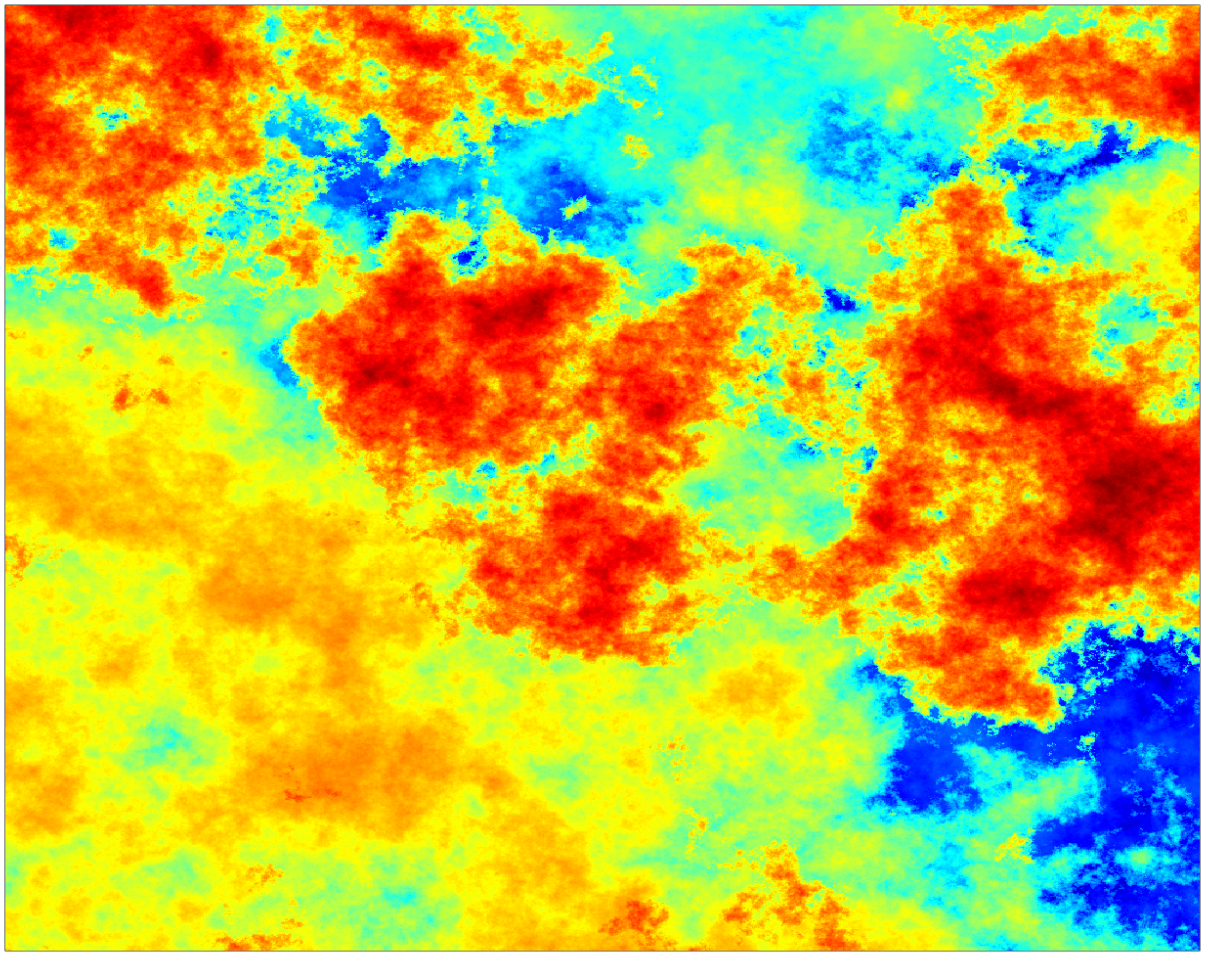
Example: Extremal Gaussian Process

To construct the "Schlather process" (Schlather, 2002), let

- $\{W_j'(\mathbf{s})\}_{j=1}^{\infty}$, be independent copies of a stationary Gaussian process with correlation function ρ, scaled such that $\mathrm{E}[\max\{0, W_j'(\mathbf{s})\}] = 1$.



Example: Extremal Gaussian Process



Outline

1. Introduction
2. Max-stable Processes
3. A Hierarchical Representation
4. Back to the Motivating Example
5. Discussion

Truncating the Spectral Representation

$$X(\mathbf{s}) = \max\{H_1 W_1(\mathbf{s}), H_2 W_2(\mathbf{s}), \ldots\}$$

- The first step is deciding that maybe we don't need an *infinite* number of W_js.
- Simply truncating this spectrum does not yield a likelihood.
- We seek a model with a finite number of components that
  1. is max-stable.
  2. gives a tractable likelihood.
- One key to producing a likelihood will be using latent positive stable variables (Fougères et al., 2009; Stephenson, 2009).

Positive Stable Random Variables

Let $A \sim \mathsf{PS}(\alpha)$ denote a positive stable random variable with characteristic exponent $\alpha \in (0, 1)$.

Fact to keep in mind about $\mathsf{PS}(\alpha)$:

- As $\alpha \to 1$, $f_A(a \,|\, \alpha) \to \delta(1)$

Truncated Spectrum with Latent PS(α) Random Variables

1. First, choose a large but finite number L "basis functions" $W_1(\mathbf{s}), \ldots, W_L(\mathbf{s})$.
2. Normalize each such that $W_{w,l}(\mathbf{s}) = W_l(\mathbf{s}) / \sum_{j=1}^{L} W_j(\mathbf{s})$. (replaces the moment condition)
3. Now define the spatial effect

$$X^*(\mathbf{s}) = \Big(\sum_{l=1}^{L} A_l W_{w,l}(\mathbf{s})^{1/\alpha} \Big)^{\alpha},$$

where $A_l \sim$ PS(α) independently for $l = 1, \ldots, L$.

The Role of α in $X^*(\mathbf{s})$

$$X^*(\mathbf{s}) = \Big(\sum_{l=1}^{L} A_l W_{w,l}(\mathbf{s})^{1/\alpha} \Big)^{\alpha},$$

- As $\alpha \to 0$,
  - $\left(\sum_{l=1}^{L} x^{1/\alpha} \right)^{\alpha} \to \max_l \{x_l\}$
  - so $X^*(\mathbf{s})$ becomes the max of randomly scaled spatial processes.

- As $\alpha \to 1$,
  - $A_l \xrightarrow{\mathcal{D}} 1$
  - so $X^*(\mathbf{s})$ becomes the constant 1.

Properties of $X^*(\mathbf{s})$

Let $U(\mathbf{s}) \overset{\text{iid}}{\sim} \text{GEV}(1, 1, 1)$.

Fact 1

$U(\mathbf{s})^\alpha X^*(\mathbf{s})$ *is max-stable.*

Fact 2

$U(\mathbf{s})^\alpha X^*(\mathbf{s})$ *has unit Fréchet margins for all* $\mathbf{s} \in \mathcal{D}$.

The Object $U(\mathbf{s})^{\alpha} X^{*}(\mathbf{s})$

$U(\mathbf{s})^{\alpha} X^{*}(\mathbf{s})$ is max-stable with unit Fréchet margins.

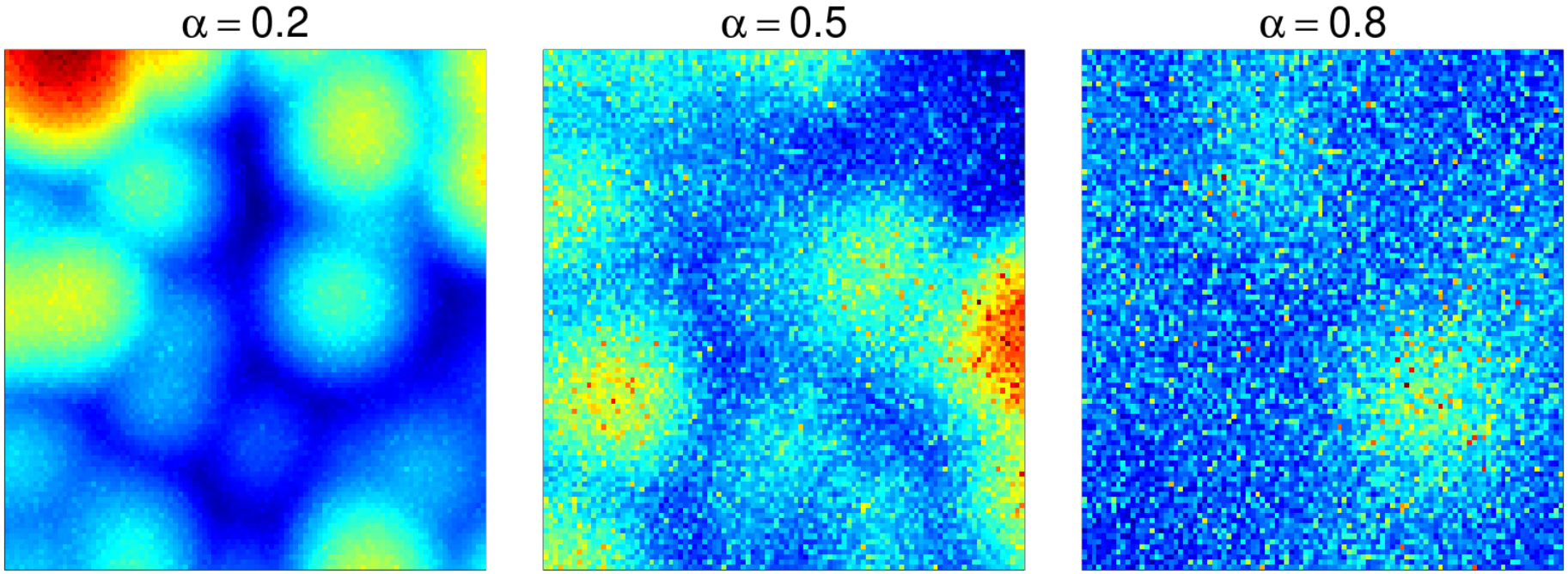
- $U(\mathbf{s}) \overset{\text{iid}}{\sim} \text{GEV}(1, 1, 1)$ is the extremal "nugget effect."
- This expression makes explicit that α trades off spatial with non-spatial variability.
- We know how to transform from $\text{GEV}(1, 1, 1)$ to $\text{GEV}(\mu, \sigma, \xi)$.
  - $\rightarrow$ Gives us a likelihood!

Example: The Hierarchical Kernel Extreme Value Process (HKEVP)

From Reich and Shaby (2013):

1. Choose a large but finite number L of fixed knots, $\mathbf{v}_1, ..., \mathbf{v}_L$.
2. Choose $W_1(\mathbf{s}), \ldots, W_L(\mathbf{s})$, Gaussian kernels centered at the knots to construct $X^*(\mathbf{s}) = \left(\sum_{l=1}^{L} A_l W_{w,l}(\mathbf{s}|\mathbf{v}_l)^{1/\alpha} \right)^{\alpha}$.

The effect of α



Fact 3

For equally spaced knots, the HKEVP converges in distribution to the GEVP as $\alpha \to 0$ and $L \to \infty$.

Example: The Hierarchical Log-Gaussian Extreme Value Process

Cribbed from from Davison et al. (2012):

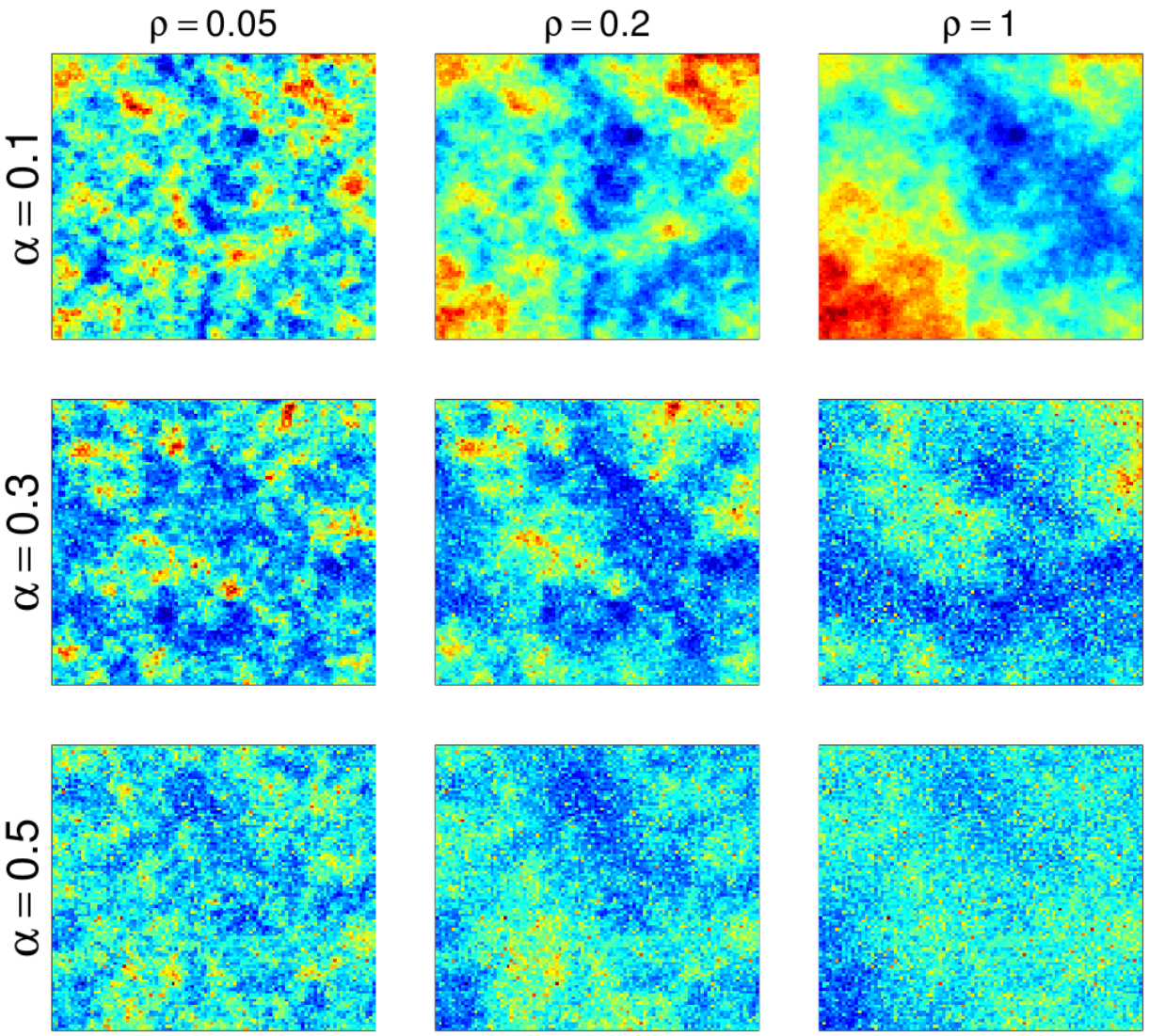
1. Choose $W_1(\mathbf{s}), \ldots, W_L(\mathbf{s})$ as independent copies of log Gaussian processes, with

$$\{W_l(\mathbf{s} \,|\, \sigma_W, \boldsymbol{\theta}_W)\}_{l=1}^{L} = \exp\{\sigma_W \varepsilon_l(\mathbf{s} \,|\, \boldsymbol{\theta}_W) - \sigma_W^2/2\}.$$

   - $\varepsilon_l(\mathbf{s} \,|\, \boldsymbol{\theta}_W)$ are standard Gaussian processes with covariance parameters $\boldsymbol{\theta}_W$.

2. Scale appropriately to meet sum-to-one constraint, and construct $X^*(\mathbf{s}) = \left(\sum_{l=1}^{L} A_l W_{w,l}(\mathbf{s})^{1/\alpha} \right)^{\alpha}$ as before.

The effect of α and the Range Rarameter

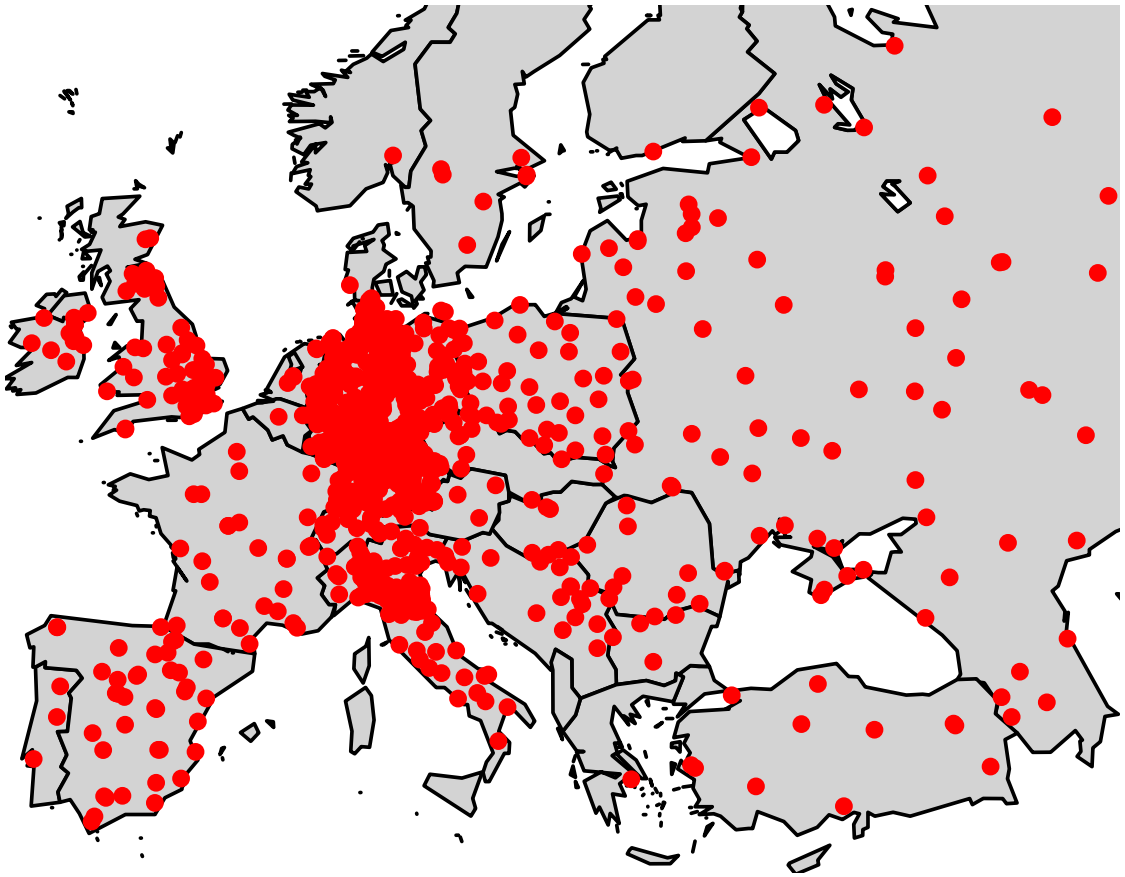


Outline

1. Introduction
2. Max-stable Processes
3. A Hierarchical Representation
4. Back to the Motivating Example
5. Discussion

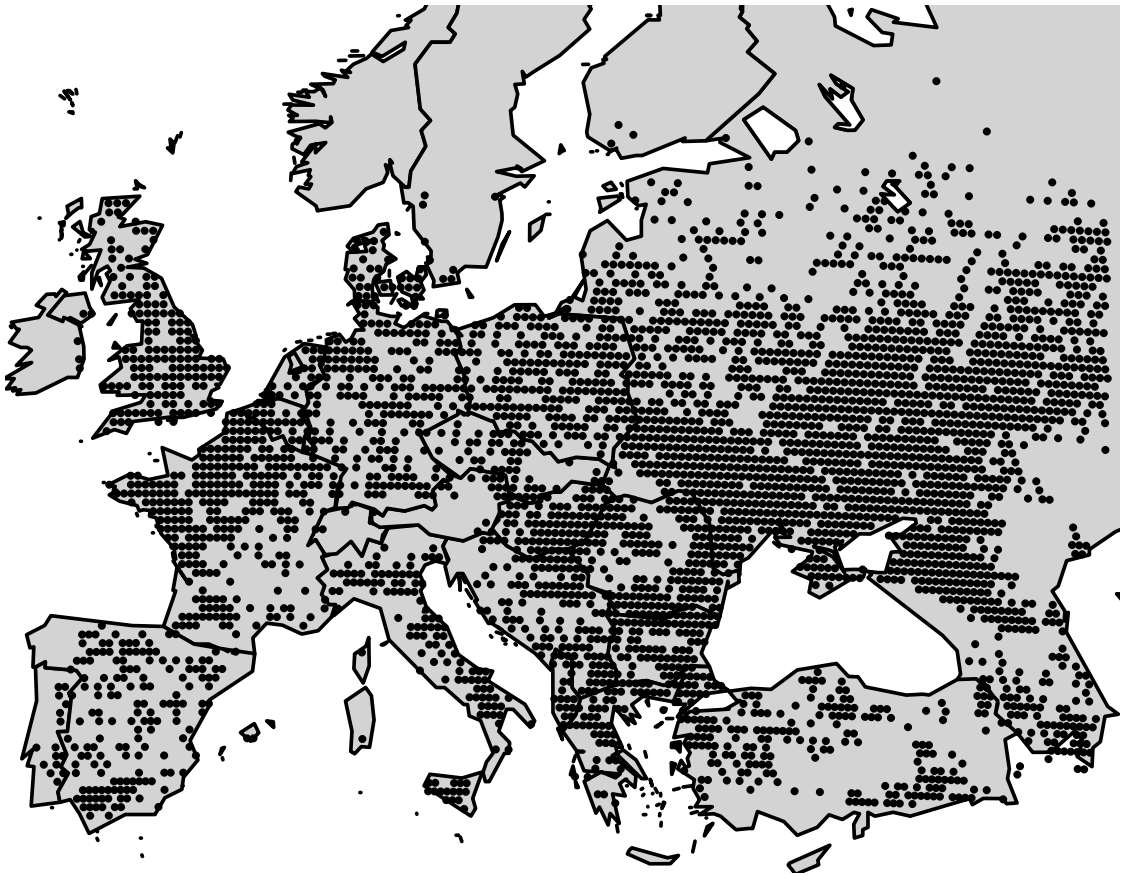
Data Locations

- Annual max temperature data from the European Climate Assessment (`http://eca.knmi.nl/`).
- Extracted years 1900–2011 from 985 monitoring locations.

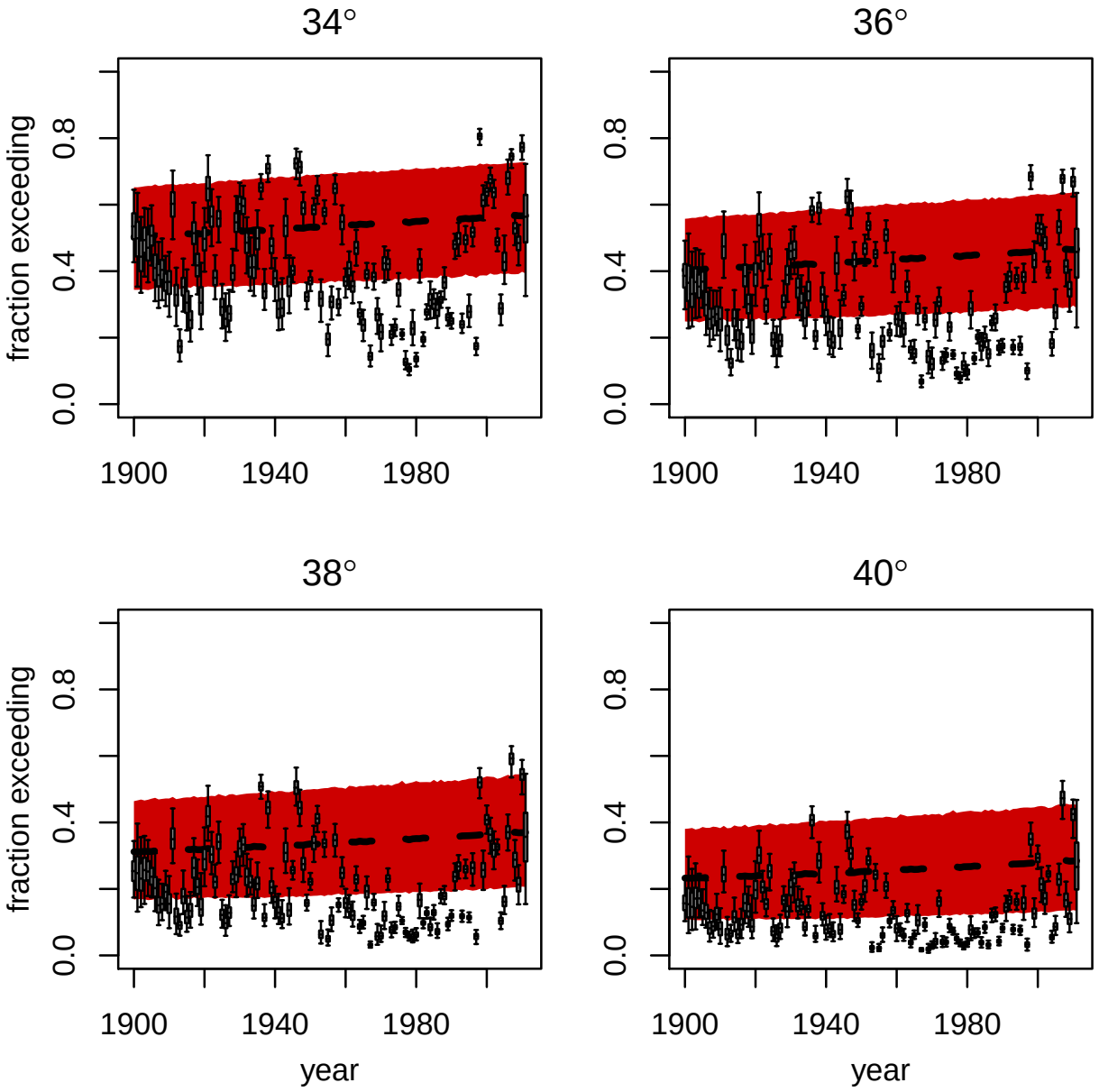


Prediction Locations

- Landcover data from MODIS MOD12 data product (`http://www-modis.bu.edu/landcover/`).
- Regular prediction grid on an equal area projection.



Fraction of Cropland Above Threshold



Outline

1. Introduction
2. Max-stable Processes
3. A Hierarchical Representation
4. Back to the Motivating Example
5. Discussion

Thoughts on the Future

1. Considering only annual maxes ignores valuable information.
   - Can use max-stable processes for threshold data (e.g. Jeon and Smith, 2013; Huser and Davison, 2013).
     - ? But how to use censored likelihoods?
   - Possible approach is Bortot and Gaetan (2013).
     - Won't work for temperatures.
     - Not as straightforward as I had hoped in the spatial context.
   - Generalized Pareto processes?

Thoughts on the Future

1. Considering only annual maxes ignores valuable information.
   - Can use max-stable processes for threshold data (e.g. Jeon and Smith, 2013; Huser and Davison, 2013).
     - ? But how to use censored likelihoods?
   - Possible approach is Bortot and Gaetan (2013).
     - Won't work for temperatures.
     - Not as straightforward as I had hoped in the spatial context.
   - Generalized Pareto processes?

2. Maybe asymptotic dependence is too strong.
   - Inverted max-stable processes (Wadsworth and Tawn, 2012).
     - Seems doable.

Summary

- The hierarchical framework produces spatial max-stable processes that have tractable likelihoods.

- By truncating and adding a PS(α) random effect, we managed to
  1. be Bayesian
  2. remain max-stable.

- This allows us
  1. modeling flexibility
  2. theoretical justification
  3. a platform for building more exotic max-stable models.

- We can do better!

The End

Thanks!

References

Bortot, P. and Gaetan, C. (2013), "A latent process model for temporal extremes," *Scand. J. Statist.*

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012), "Statistical modeling of spatial extremes," *Statist. Sci.*, 27, 161–186.

Fougères, A.-L., Nolan, J. P., and Rootzén, H. (2009), "Models for dependent extremes using stable mixtures," *Scand. J. Stat.*, 36, 42–59.

Huser, R. and Davison, A. C. (2013), "Space-time modelling of extreme events," *J. R. Stat. Soc. Ser. B Stat. Methodol.*

Jeon, S. and Smith, R. L. (2013), "Dependence structure of spatial extremes using threshold approach," .

Reich, B. J. and Shaby, B. A. (2013), "A hierarchical max-stable spatial model for extreme precipitation," *Ann. of Appl. Stat.*, in press.

Schlather, M. (2002), "Models for stationary max-stable random fields," *Extremes*, 5, 33–44.

Shaby, B. A. and Reich, B. J. (2013), "Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland," *Environmetrics*, in Press.

Smith, R. (1990), "Max-stable processes and spatial extremes," *Unpublished manuscript*.

Stephenson, A. G. (2009), "High-dimensional parametric modelling of multivariate extreme events," *Aust. N. Z. J. Stat.*, 51, 77–88.

Wadsworth, J. L. and Tawn, J. A. (2012), "Dependence modelling for spatial extremes," *Biometrika*, 99, 253–272.