Applying Max-and-Smooth to the UKCP data

Brynjólfur Gauti Guðrúnar Jónsson

2024-12-04

Introduction

This document describes the implementation of the Max-and-Smooth algorithm for fast approximate Bayesian inference in spatial extreme value analysis of climate projections provided by the UKCP. The algorithm is specifically applied to Generalized Extreme Value (GEV) distributions and is implemented in C++ with R interfaces using Rcpp and RcppEigen as well as Stan.

Package Overview

The `maxandsmooth` R package provides tools for fast approximate Bayesian inference for spatial GEV models. The core of the package is implemented in C++ for efficiency, with R wrappers for ease of use.

Key features of the package include:

- Implementation of the Max-and-Smooth algorithm
- Efficient C++ code using automatic differentiation and Eigen
- Spatial modeling of GEV parameters using Stan's efficient HMC sampler
- R interface for easy integration into existing extreme value analysis workflows

Algorithm Description

The Max-and-Smooth algorithm, as applied to spatial GEV models, consists of two main steps:

1. **Max Step**: Maximum likelihood estimation of GEV parameters at each spatial location using C++
2. **Smooth Step**: Spatial smoothing of the maximum likelihood estimates using a BYM2 model implemented in Stan

The algorithm treats the ML estimates as sufficient statistics for a latent Gaussian field, providing a fast approximation to full Bayesian inference for spatial extreme value models.

Code Structure

The package is organized into several key files:

1. `src/max.cpp`: Implements the Max step (maximum likelihood estimation for GEV)
2. `src/gev.cpp`: Implements GEV-specific functions for likelihood and gradient calculations
3. `Stan/stan_smooth_bym2.stan` Implements the Smooth step using Stan

Max Step

The Max step involves computing location-wise maximum likelihood estimates (MLEs) for the GEV model parameters. This step is performed independently for each location, treating the data as if it were independent across locations.

Data Structure

The input data Y, calculated from the UKCP, is structured as a matrix, where:

- Rows represent observations of hourly maximum rainfall in yearly blocks
- Columns represent different spatial locations over Great Britain.

Maximum Likelihood Estimation

For each location (column in Y), we compute the MLEs for the three GEV parameters: location (μ), scale (σ), and shape (ξ). The log-likelihood function for the GEV distribution at a single location is:

$$\ell(\mu, \sigma, \xi | y) = -n \log \sigma - (1 + \frac{1}{\xi}) \sum_{i=1}^{n} \log \left(1 + \xi \frac{y_i - \mu}{\sigma} \right) - \sum_{i=1}^{n} \left(1 + \xi \frac{y_i - \mu}{\sigma} \right)^{-1/\xi}$$

where n is the number of observations at the location.

The MLEs are obtained by maximizing this likelihood function with respect to the parameters:

$$(\hat{\mu}_i, \hat{\sigma}_i, \hat{\xi}_i) = \arg\max_{(\mu,\sigma,\xi)} \ell(\mu, \sigma, \xi | Y_i)$$

where Y_i is the data for location i.

Link Functions

Instead of directly maximizing the likelihood with the original parameters, we transform the parameters using a link function

$$(\psi, \tau, \phi) = h(\mu, \sigma, \xi) = (\log(\mu), \log(\sigma) - \log(\mu), \text{logit}(\xi))$$

Implementation

The maximization is performed using numerical optimization techniques. In our C++ implementation, we use the following approach:

1. Parameters are transformed with a link function
2. The negative log-likelihood and its gradient are computed on the unconstrained scale
3. A numerical optimizer is used to find the MLEs as well as the Hessians at each location's optimum.

The output of this step includes:

1. A vector of parameter estimates, $\hat{\eta}$, ordered such that each station's location parameter appears first, followed by scale parameters, then shape parameters:

$$\hat{\eta} = (\hat{\psi}_1, ..., \hat{\psi}_n, \hat{\tau}_1, ..., \hat{\tau}_n, \hat{\phi}_1, ..., \hat{\phi}_n)^T$$

   where n is the number of stations.

2. A precision matrix, $Q_{\eta y}$, constructed from the negative Hessians of the log-likelihood at the MLE estimates. Due to the parameter ordering in $\hat{\eta}$, $Q_{\eta y}$ can be described as a 3×3 block matrix:

$$Q_{\eta y} = \begin{bmatrix} Q_{\psi\psi} & Q_{\psi\tau} & Q_{\psi\phi} \\ Q_{\tau\psi} & Q_{\tau\tau} & Q_{\tau\phi} \\ Q_{\phi\psi} & Q_{\psi\tau} & Q_{\phi\phi} \end{bmatrix}$$

   where each block Q_{ij} is an $n \times n$ diagonal matrix. The diagonal elements of Q_{ii} correspond to the negative second derivatives of the log-likelihood with respect to the i-th parameter

at each station. The off-diagonal blocks Q_{ij} (where $i \neq j$) contain the negative mixed partial derivatives of the log-likelihood with respect to the i-th and j-th parameters.

For example, the elements of $Q_{\psi\psi}$ are the conditional precisions of the location parameters:

$$
\begin{aligned}
Q_{\psi\psi} &= \mathrm{diag}\left(\tau_1^{\psi\psi}, \dots, \tau_n^{\psi\psi}\right) \\
&= \mathrm{diag}\left(-\frac{\partial^2 \ell(Y_1|\psi_1, \tau_1, \phi_1)}{\partial \psi_1^2}, \dots, -\frac{\partial^2 \ell(Y_n|\psi_n, \tau_n, \phi_n)}{\partial \psi_n^2}\right),
\end{aligned}
$$

and the elements of $Q_{\mu\xi}$ are the conditional dependencies between the location and shape parameters:

$$
\begin{aligned}
Q_{\psi\phi} &= \mathrm{diag}\left(\tau_1^{\psi\phi}, \dots, \tau_n^{\psi\phi}\right) \\
&= \mathrm{diag}\left(-\frac{\partial^2 \ell(Y_1|\psi_1, \tau_1, \phi_1)}{\partial \psi_1 \partial \phi_1}, \dots, -\frac{\partial^2 \ell(Y_n|\psi_n, \tau_n, \phi_n)}{\partial \psi_n \partial \phi_n}\right)
\end{aligned}
$$

This structure reflects the independence assumption between stations in the Max step, while capturing the parameter dependencies within each station. The outputs, $\hat{\eta}$ and $Q_{\eta y}$, serve as inputs into the Smooth step.

Smooth Step

The Smooth step involves Bayesian inference on a latent Gaussian field, using the maximum likelihood estimates from the Max step as data. We implement this step using Stan, which provides efficient Hamiltonian Monte Carlo sampling.

Model Structure

Let η be the latent field of parameters, and $\hat{\eta}$ be the maximum likelihood estimates from the Max step. The model follows a BYM2 (Besag-York-Mollié) structure, which decomposes the spatial effect into structured and unstructured components:

For each parameter $p \in \{\psi, \tau, \phi\}$:

$$
\eta_p = \mu_p + \sigma_p \left(\eta_p^{\mathrm{spatial}} \sqrt{\frac{\rho_p}{c}} + \eta_p^{\mathrm{random}} \sqrt{1 - \rho_p}\right)
$$

where:

- μ_p is the overall mean for parameter p
- σ_p is the marginal standard deviation
- ρ_p is the mixing parameter determining the balance between spatial and random effects
- c is a scaling factor that ensures the marginal variance of the spatial component is approximately 1
- η_p^{spatial} follows an intrinsic conditional autoregressive (ICAR) prior
- η_p^{random} follows a standard normal distribution

Data Level

The data level models the relationship between the observed maximum likelihood estimates $\hat{\eta}$ and the true latent field η using a multivariate normal distribution with precision matrix $Q_{\eta y}$ from the Max step:

$$\hat{\eta}|\eta \sim N(\eta, Q_{\eta y}^{-1})$$

The precision matrix $Q_{\eta y}$ is passed to Stan in a sparse representation and a custom likelihood function for multivariate Gaussian distributions with know Cholesky decomposed precision matrix is used.

Prior Distributions

The model uses the following prior distributions: what

- $\sigma_p \sim \text{Exponential}(1)$ for the marginal standard deviations
- $\rho_p \sim \text{Beta}(1,1)$ for the mixing parameters
- $\mu_p \sim 1$ for the overall means