SpatialGEV: Fast Bayesian inference for spatial extreme value models in R

June 12, 2024

Summary

Extreme weather phenomena such as floods and hurricanes are of great concern due to their potential to cause extensive damage. To develop more reliable damage prevention protocols, statistical models are often used to infer the chance of observing an extreme weather event at a given location (Coles and Casson 1998; Cooley, Nychka, and Naveau 2007; Sang and Gelfand 2010). Here we present **SpatialGEV**, an R package providing a fast and convenient toolset for analyzing spatial extreme values using a hierarchical Bayesian modeling framework. In this framework, the marginal behavior of the extremes is given by a generalized extreme value (GEV) distribution, whereas the spatial dependence between locations is captured by modeling the GEV parameters as spatially varying random effects following a Gaussian process (GP). The general form of the GEV-GP model that can be fit using **SpatialGEV** is described in the package vignette. Model inference is carried out using an efficient implementation of the Laplace approximation, which produces highly accurate posterior estimates several orders of magnitude faster than Markov Chain Monte Carlo (MCMC) methods. Users are provided with a streamlined way to build and fit various GEV-GP models in R, which are compiled in C++ under the hood. For downstream analyses, the package offers methods for Bayesian parameter estimation and forecasting of extreme events.

Statement of need

In a Bayesian context, the posterior distribution $p(z_p(\boldsymbol{x}) \mid \boldsymbol{Y})$ conditional on all data $\boldsymbol{Y}$ is very useful for forecasting extreme weather events. Traditionally, MCMC methods are used to sample from the posterior distribution of the GEV model (e.g., Cooley, Nychka, and Naveau 2007; Schliep et al. 2010; Dyrrdal et al. 2015). However, this can be extremely computationally intensive when the number of locations is large. **SpatialGEV** implements an approximate Bayesian inference approach as an alternative to MCMC, making large-scale spatial analyses orders of magnitude faster while achieving roughly the same accuracy as MCMC. We construct a Normal approximation to the joint posterior distribution of both GEV parameters and GP hyperparameters $p(u(\boldsymbol{x}), \boldsymbol{\eta}_u, \boldsymbol{\beta}_u \mid \boldsymbol{Y})$, which is then used to estimate the return level posterior. This is done via an efficient Laplace approximation to the marginal hyperparameter posterior $p(\boldsymbol{\eta}_u, \boldsymbol{\beta}_u \mid \boldsymbol{Y})$ transforming a high-dimensional MCMC into a nested optimization problem that is faster to solve (Tierney and Kadane 1986; Kristensen et al. 2016; Chen, Ramezan, and Lysy 2024). The Laplace approximation is carried out using the R/C++ package **TMB** (Kristensen et al. 2016). Details of the inference method can be found in Chen, Ramezan, and Lysy (2024).

The R package **SpatialExtremes** (Ribatet, Singleton, and R Core team 2022) is a popular software for fitting spatial extreme value models including GEV-GP. Although it supports a wider range of model classes, its inference for GEV-GP models relies on a basic Gibbs sampler updating each of the hyperparameters and random effects one at a time, which tends to converge very slowly since these variables are highly correlated with each other. Furthermore, GP computation in **SpatialExtremes** scales as $\mathcal{O}(n^3)$ with the number of locations n, whereas **SpatialGEV** offers an option for approximate GP computation scaling as $O(n^{3/2})$ (Lindgren, Rue, and Lindström 2011). Coupled with the Laplace approximation, this allows

SpatialGEV to fit GEV-GP models to several hundreds spatial locations on a personal computer in minutes (Chen, Ramezan, and Lysy 2024). A more efficient MCMC algorithm for hierarchical spatial models is Hamiltonian Monte Carlo and its variants (Neal 2011; Hoffman and Gelman 2014), for which a highly efficient and self-tuning implementation is provided by the R/C++ package **RStan** (Stan Development Team 2020). Chen, Ramezan, and Lysy (2024) compares the speed and accuracy of the Laplace method implemented in **SpatialGEV** to **RStan**. It is found that, while **SpatialGEV** tends to underestimate the posterior variance of the hyperparameters, it accurately estimates the posteriors of both GEV parameters and return levels – and does this three orders of magnitude faster than **RStan**. A well-known alternative to MCMC is the integrated Laplace approximation (INLA) method, whose R implementation is provided in the **R-INLA** package (Lindgren and Rue 2015). As an extension of the Laplace approximation, INLA is typically more accurate. However, **R-INLA** is inapplicable to GEV-GP models in which two or more GEV parameters are modeled as random effects following different Gaussian processes. In contrast, **SpatialGEV** offers more flexibility as it is straightforward for the user to choose what GEV parameters are spatial random effects. Wood (2023) and Youngman (2022) provide another means for estimating spatially varying GEV parameters via a scalable basis representation reducing the number of random effects in the model. Compared to **SpatialGEV** which keeps all random effects for inference, the basis function expansion approach is less accurate for estimating spatial processes that are not smooth or exhibit short-range correlation (Wood 2020; Lindgren, Bolin, and Rue 2021).

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada, grant numbers RGPIN-2018-04376 (Ramezan), DGECR-2018-00349 (Ramezan) and RGPIN-2020-04364 (Lysy).

References

Chen, M., R. Ramezan, and M. Lysy. 2024. "Fast and Scalable Approximate Inference for Spatial Extreme Value Models." *Canadian Journal of Statistics (Accepted)*. https://doi.org/10.48550/arXiv.2110.07051.

Coles, S. G., and E. Casson. 1998. "Extreme Value Modelling of Hurricane Wind Speeds." *Structural Safety* 20: 283–96. https://doi.org/10.1111/1467-9876.00189.

Cooley, D., D. Nychka, and P. Naveau. 2007. "Bayesian Spatial Modeling of Extreme Precipitation Return Levels." *Journal of the American Statistical Association* 102: 824–40. https://doi.org/10.1198/016214506000000780.

Dyrrdal, A. V., A. Lenkoski, T. L. Thorarinsdottir, and F. Stordal. 2015. "Bayesian Hierarchical Modeling of Extreme Hourly Precipitation in Norway." *Environmetrics* 26: 89–106. https://doi.org/10.1002/env.2301.

Hoffman, M. D., and A. Gelman. 2014. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research* 15: 1593–623. https://doi.org/10.5555/2627435.2638586.

Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. M Bell. 2016. "TMB: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software* 70 (5): 1–21. https://doi.org/10.18637/jss.v070.i05.

Lindgren, F. K., D. Bolin, and H. Rue. 2021. "The SPDE Approach for Gaussian and non-Gaussian Fields: 10 Years and Still Running." *Spatial Statistics*. https://doi.org/10.1016/j.spasta.2022.100599.

Lindgren, F. K., and H. Rue. 2015. "Bayesian Spatial Modelling with R-INLA." *Journal of Statistical Software* 63: 10.18637/jss.v063.i19.

Lindgren, F. K., H. Rue, and J. Lindström. 2011. "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach." *Journal of the Royal Statistical Society, Series B* 73: 423–98. https://doi.org/10.1111/j.1467-9868.2011.00777.x.

Neal, R. M. 2011. "MCMC Using Hamiltonian Dynamics." In *The Handbook of Markov Chain Monte Carlo*. Chapman & Hall / CRC Press.

Ribatet, M., R. Singleton, and R Core team. 2022. *SpatialExtremes: Modelling Spatial Extremes*. Version 2.1-0. https://CRAN.R-project.org/package=SpatialExtremes.

Sang, H., and A. E. Gelfand. 2010. "Continuous Spatial Process Models for Spatial Extreme Values." *Journal* *of Agricultural, Biological, and Environmental Statistics* 15: 49–56. https://doi.org/10.1007/s13253-009-0010-1.

Schliep, E. M., D. Cooley, S. R. Sain, and J. A. Hoeting. 2010. "A Comparison Study of Extreme Precipitation from Six Different Regional Climate Models via Spatial Hierarchical Modeling." *Extremes* 13: 219–39. https://doi.org/10.1007/s10687-009-0098-2.

Stan Development Team. 2020. "RStan: The R Interface to Stan." http://mc-stan.org/.

Tierney, L., and J. Kadane. 1986. "Accurate Approximations for Posterior Moments and Marginal Densities." *Journal of the American Statistical Association* 81: 82–86. https://doi.org/10.1080/01621459.1986.10478240.

Wood, Simon N. 2020. "Inference and Computation with Generalized Additive Models and Their Extensions." *Test* 29: 307–39.

———. 2023. *`mgcv`: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*. Version 1.9-1. https://CRAN.R-project.org/package=mgcv.

Youngman, Benjamin D. 2022. "`evgam`: An R Package for Generalized Additive Extreme Value Models." *Journal of Statistical Software* 103: 1–26. https://doi.org/10.18637/jss.v103.i03.