ABC model selection for spatial max-stable models applied to South Australian maximum temperature data

Xing Ju Lee[†], **Markus Hainy**[†‡], **James P. McKeone**[†],
Christopher C. Drovandi[†], **Anthony N. Pettitt**[†]

† School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia
‡ Department of Applied Statistics, Johannes Kepler University, Linz, Austria

Bayes on the Beach 2017, 13 – 15 November 2017, Surfers Paradise

Introduction

One possibility to model spatial extremes is via max-stable processes. Our aim is to perform Bayesian model selection concerning the spatial dependence structure.

For moderate to large dimensions, max-stable models are intractable. We suggest a model selection procedure based on approximate Bayesian computation (ABC) using a semi-automatic summary selection scheme inspired by Fearnhead and Prangle (2012).

Max-stable processes

The spectral representation of a max-stable process $\{Z(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{X}\}$ with unit Fréchet margins is

$$Z(\boldsymbol{x}) = \max_{i \geq 1} \zeta_i Y_i(\boldsymbol{x}), \ \boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^d,$$

where the ζ_i are Poisson process points on $(0, \infty)$ with intensity $d\Lambda(\zeta) = \zeta^{-2} d\zeta$, and the $Y_i(\boldsymbol{x})$ follow a non-negative stochastic process $Y(\boldsymbol{x})$ with $E[Y(\boldsymbol{x})] = 1$.

The different max-stable models differ in the specification of $Y(\boldsymbol{x})$. We consider four models: the extremal-t model with Whittle-Matérn, Cauchy, and powered exponential correlation function, and the Brown-Resnick model. We also incorporate geometric anisotropy into all models.

All models have a range (c_2) and a smoothness (κ) parameter and two parameters controlling the geometric anisotropy (rotation angle (α) and principal axes ratio (r) of the confidence ellipse). In addition, the extremal-t models feature a degrees of freedom parameter (ν).

Semi-automatic ABC for model selection and parameter estimation

Our aim is to generate draws from the joint ABC posterior of the model indicator and the parameters:

$$\pi_\epsilon(\boldsymbol{\phi}_k, k | \, \boldsymbol{z}) \propto \Pr(\mathcal{M} = k) \, \pi(\boldsymbol{\phi}_k | \, \mathcal{M} = k) \int_{\boldsymbol{u}} f_k(\boldsymbol{u} | \, \boldsymbol{\phi}_k) \, \mathbb{1}\{d_T(\boldsymbol{z}, \boldsymbol{u}) < \epsilon\} \, \mathrm{d}\boldsymbol{u}.$$

We use the sequential Monte Carlo ABC replenishment algorithm of Drovandi and Pettitt (2011) to obtain samples from $\pi_\epsilon(\boldsymbol{\phi}_k, k | \, \boldsymbol{z})$.

To select suitable ABC summary statistics for parameter estimation and model selection, we employ the semi-automatic selection schemes of Fearnhead and Prangle (2012) and Prangle et al. (2014). They propose to use estimates of the posterior means of the parameters as ABC summary statistics for parameter estimation and to use estimates of the posterior model probabilities as ABC summary statistics for model selection.

The estimates are obtained by using predictions from regressions run on a large sample from the prior predictive distribution. Linear regressions are used to regress the individual parameter samples on a large set of potentially informative regression summary statistics computed from the simulated observations, while multinomial logistic regression is used to regress the model indicators on the same set of regression summary statistics. Superfluous regression summary statistics are eliminated via a stepwise procedure.

The discrepancy functions for model selection and parameter estimation are

$$d_M(\boldsymbol{z}, \boldsymbol{u}) = \sum_{k=1}^{K-1} \left[\widehat{\Pr}(\mathcal{M} = k | \, \boldsymbol{z}) - \widehat{\Pr}(\mathcal{M} = k | \, \boldsymbol{u}) \right]^2,$$

$$d_P(\boldsymbol{z}, \boldsymbol{u}) = \sum_{k=0}^{K-1} \sum_{j=1}^{Q_k} \left[\frac{\hat{\phi}_{k,j}(\boldsymbol{z}) - \hat{\phi}_{k,j}(\boldsymbol{u})}{\widehat{\mathrm{sd}}(\hat{\phi}_{k,j})} \right]^2,$$

and the overall discrepancy function is

$$d_T(\boldsymbol{z}, \boldsymbol{u}) = \log \{ d_M(\boldsymbol{z}, \boldsymbol{u}) \cdot d_P(\boldsymbol{z}, \boldsymbol{u}) \}.$$

Implementation details

The data set comprises annual maximum temperature measurements obtained at 25 weather stations in South Australia from 1979 to 1996.

The set of regression summary statistics includes: group means and standard deviations of grouped estimates of pairwise F-madograms, pairwise and tripletwise extremal coefficients, and pairwise extremal concurrence probabilities, as well as composite score vectors for the parameters of all models.

Prior distributions: $\log(c_2) \sim \mathcal{N}(1, 4)$, $\kappa \sim \mathcal{U}(0, 2)$, $\alpha \sim \mathcal{U}(0, \pi/2)$, $\log(r) \sim \mathcal{N}(0, 8)$, $\log(\nu) \sim \mathcal{N}(0, 1)$ truncated on $[-2.5, 2.5]$.

FP step: 2,500 simulations from each model from the prior predictive distribution.

SMC ABC step: $N = 2{,}000$ particles.

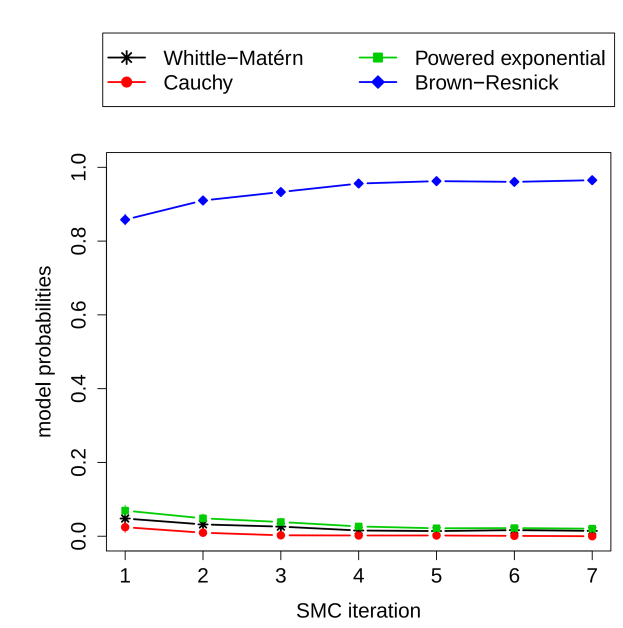
FP step

Misclassification matrix obtained by FP step:

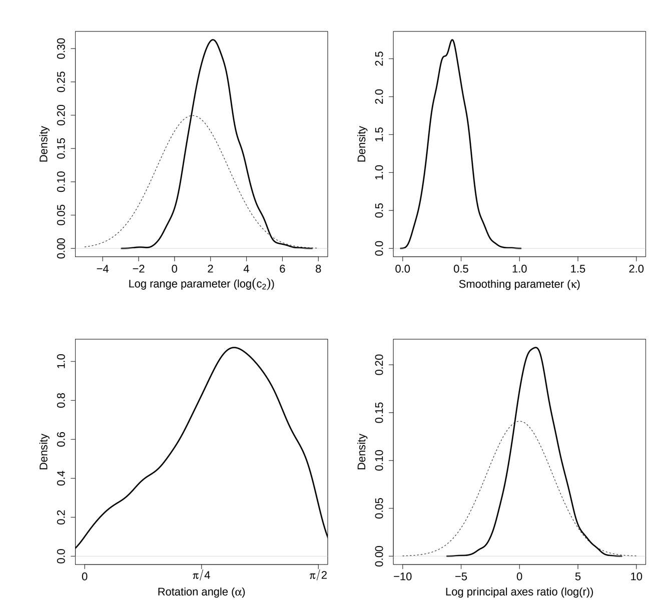
$$\begin{array}{l} \\ j=1 \text{ (WM)} \\ j=2 \text{ (Cauchy)} \\ j=3 \text{ (pow. exp.)} \\ j=4 \text{ (BR)} \end{array} \begin{array}{c} \begin{array}{cccc} \Pr(k=1|j) & \Pr(k=2|j) & \Pr(k=3|j) & \Pr(k=4|j) \end{array} \\ \left[\begin{array}{cccc} 0.26 & 0.29 & 0.29 & 0.16 \\ 0.14 & 0.55 & 0.20 & 0.11 \\ 0.14 & 0.22 & 0.46 & 0.17 \\ 0.11 & 0.08 & 0.05 & 0.76 \end{array} \right] \end{array}$$

ABC results

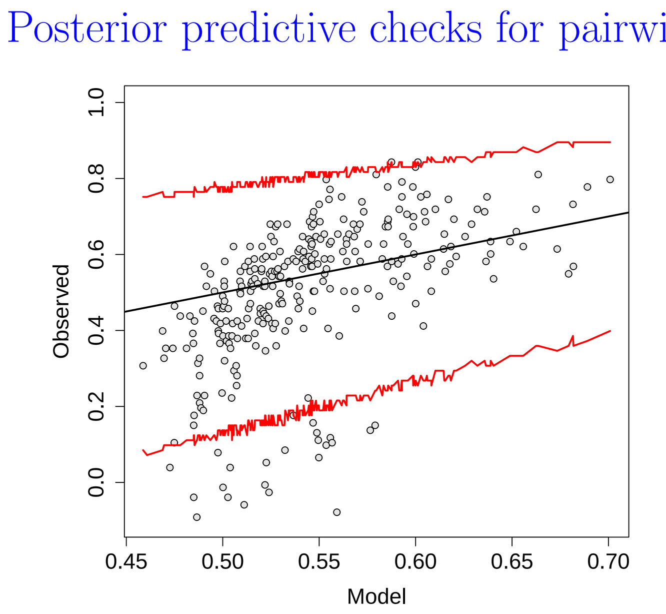
Posterior model probabilities:



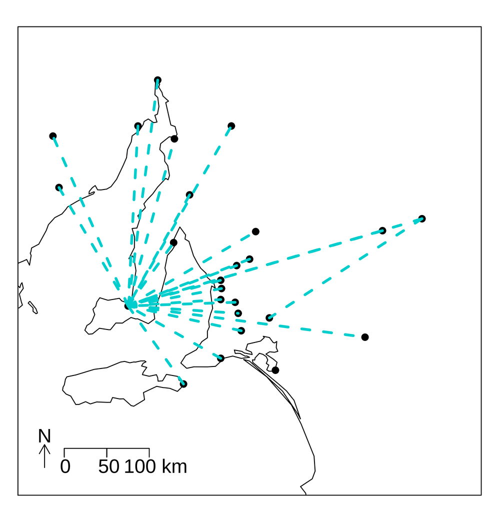
Posterior distributions (Brown-Resnick):



Posterior predictive checks for pairwise extremal concurrence probabilities:



Observed values vs.
posterior predictive distribution



Outlier map

References

Drovandi, C. C. and Pettitt, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics, 67*, 225–233.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74*, 419–474.

Lee, X. J., Hainy, M., McKeone, J. P., Drovandi, C. C., and Pettitt, A. N. (2017). ABC model selection for spatial max-stable models applied to South Australian maximum temperature data. *arXiv:1710.02961.*

Prangle, D., Fearnhead, P., Cox, M. P., Biggs, P. J., and French, N. P. (2014). Semi-automatic selection of summary statistics for ABC model choice. *Statistical Applications in Genetics and Molecular Biology, 13*, 67–82.