MSIAM M2 Thesis: Extreme quantile Bayesian inference

Tony Zheng

27th July 2020

1 Introduction

The study of environmental conditions is crucial in the development of large scale construction projects. In particular, the study of long-term behaviour of environmental variables is necessary to understand the risks of hazardous meteorological events such as floods, storms, and droughts. This can be done using the well-established theory of extreme values, using asymptotic models such as the Process point process characterisation of extremes described in Coles 2001. However, the analysis of extreme events is often hindered by a scarcity of data, and an inadequate treatment of model and prediction uncertainties. To overcome these problems, a Bayesian methodology has been proposed. This approach allows us to exploit prior information, and provides a framework for taking into account the various uncertainties in the prediction process.

A review of the use of Bayesian methods in extreme value theory has been carried out in Coles and Powell 1996, and analyses have been performed in both Stuart G. Coles, Pericchi, and Sisson 2003 and Coles and J. Tawn 1996 to predict annual maximum rainfall.

In this report, we will propose prior elicitation strategies with the principle of maximum entropy in mind. We will consider appropriate parameters of the model for expert specification, the specification of distributions for these parameters, and the number of distribution specified. In particular, we will investigate a novel reparametrisation which uses the maximum entropy order statistics copula derived in Butucea, Delmas, Dutfoy, and Fischer 2018. We hope that the resulting prior distributions will be able to be specified more easily by an expert, and better account for uncertainty due to lack of information in the model.

In § 2, we will outline the Poisson point process model. In § 3, we will describe in detail the prior distributions which we will be testing. In § 4, we will explain how Markov chain Monte Carlo (MCMC) methods can be used in the analysis to avoid intractable calculations. In § 5, we will implement and compare our prior distributions in a simulation study and on real data. Finally, in the Appendix we have supplementary calculations and details of our MCMC implementations.

2 Likelihood model

Extreme values are often modelled using the generalised extreme value (GEV) distribution. Given a sequence of i.i.d. variables $X_1, \dots, X_n$, under certain conditions, its maximum

$$M_n \coloneqq \max\{X_1, \dots, X_n\}$$

asymptotically follows a GEV distribution as $n \to +\infty$, with CDF

$$F(x) = \begin{cases} \exp\left(-\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}_+^{-\frac{1}{\xi}}\right) & \text{if } \xi \neq 0\,, \\ \exp(-\exp(-\frac{x-\mu}{\sigma})) & \text{if } \xi = 0\,, \end{cases} \tag{1}$$

where $\{x\}_+ \coloneqq \max\{0, x\}$, with parameters $\theta \coloneqq (\mu, \sigma, \xi)$ defined in the set

$$\Theta \coloneqq \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \,. \tag{2}$$

The Poisson point process characterisation of extremes (Coles 2001) is an alternative model which takes into account all extreme observations in the data, i.e. all observations which are greater than some threshold u. Suppose that we have observations of X which are divided into M blocks. Denote by $\mathbf{x} = (x_1, \dots, x_{n_u})$ the exceedances of our data by some threshold u. The likelihood of the data is then a non-homogeneous Poisson point process

$$L(\theta \mid \mathbf{x}) = \exp\left(-M\Lambda[u,\infty)\right) \prod_{i=1}^{n_u} \lambda(x_i) \,, \tag{3}$$

with intensity function

$$\lambda(x) \coloneqq \begin{cases} \frac{1}{\sigma}\left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}_+^{-\frac{\xi+1}{\xi}} & \text{if } \xi \neq 0 \,, \\ \frac{1}{\sigma}\exp(-\frac{x-\mu}{\sigma}) & \text{if } \xi = 0 \,, \end{cases} \tag{4}$$

and

$$\Lambda[u,\infty) \coloneqq \int_u^\infty \lambda(x)\,\mathrm{d}x \tag{5}$$

$$= \begin{cases} \left\{1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right\}_+^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \,, \\ \exp(-\frac{u-\mu}{\sigma}) & \text{if } \xi = 0 \,. \end{cases} \tag{6}$$

This is an extension of the GEV model, as the block maxima asymptotically follow a GEV distribution with parameters θ as $n \to +\infty$. Inverting (1), we obtain the quantile function

$$q(p \mid \mu, \sigma, \xi) = \begin{cases} \mu + \frac{\sigma}{\xi}(\exp(-\log(-\log(p))\xi) - 1) & \text{if } \xi \neq 0 \,, \\ \mu - \sigma\log(-\log(p)) & \text{if } \xi = 0 \,. \end{cases} \tag{7}$$

When M is the number of observations in a year, this gives the quantiles of the annual maxima. The *return level* of a *return period* r in years is defined as the quantile $q(1 - 1/r \mid \theta)$. Plotting the return level against the return period allows us to visualise the extreme behaviour of the random variable.

3 Eliciting prior information

Bayesian inference is centred around the use of Bayes' theorem to update prior belief when new data is observed. This prior belief is incorporated into the model in the form of a prior distribution for the parameters. In our case, we assume that this information arises from the opinion of an expert. To choose a prior using expert opinion, we will make use of the principle of maximum entropy, which was introduced by Jaynes 1957. Entropy is a measure of how much information we have about a distribution. It is defined by Shannon 1948, for a continuous probability density p, as

$$\mathcal{E}(p) \coloneqq -\int_{x\,:\,p(x)>0} \log(p(x))\,\mathrm{d}x \,.$$

According to the principle of maximum entropy, if we are to choose a prior from a class of distributions satisfying certain constraints, then we should choose a distribution which maximises the entropy in that class, as it will be the least informative.

3.1 Reparametrisation

Unfortunately, it is not feasible for an expert to specify joint distributions for the parameters μ, σ, ξ. Instead, we will consider prior distributions for quantiles of the annual maxima.

Let $1 > p_1 > p_2 > p_3 > 0$ be a decreasing sequence of probabilities, and for $i = 1, 2, 3$, consider the corresponding $(1 - p_i)$-quantiles

$$q_i \coloneqq q(1 - p_i \mid \theta)\,,$$

where q is defined in (7). We are assuming that $\xi \neq 0$. This defines an invertible transformation

$$\begin{aligned} g_3 \colon \Theta &\to \{(x, y, z) \in \mathbb{R}^3 \colon x < y < z\} \\ \theta &\mapsto (q_1, q_2, q_3)\,, \end{aligned}$$

which allows us to convert a prior for (q_1, q_2, q_3) into a prior for θ.

We will also consider the case when an expert specifies a prior on only two quantiles, q_1 and q_2. As σ and ξ are correlated, we will put an uninformative uniform prior on $\log \sigma$ and use the invertible transformation

$$\begin{aligned} g_2 \colon \Theta &\to \mathbb{R}^+ \times \{(x, y) \in \mathbb{R}^2 \colon x < y\} \\ \theta &\mapsto (\sigma, q_1, q_2)\,. \end{aligned}$$

The determinants of g_3 and g_2 are derived in Appendix 7.1.

We will denote the number of quantiles specified by k. In the absence of any prior information, one possible prior is the Jeffreys prior, which is derived in Appendix 7.2.

3.2 The order constraint on quantiles

Without loss of generality, we suppose that $k = 3$. As the data are assumed to be strictly positive, the support of the quantiles (q_1, q_2, q_3) implies that a prior is valid if and only if

$$\Pr(0 < q_1 < q_2 < q_3) = 1\,.$$

In order to automatically satisfy this constraint, Coles and J. Tawn 1996 specify positive marginal priors for the three quantile differences

$$\begin{aligned} \tilde{q}_1 &\coloneqq q_1\,, \\ \tilde{q}_2 &\coloneqq q_2 - q_1\,, \\ \tilde{q}_3 &\coloneqq q_3 - q_2\,, \end{aligned}$$

and construct a joint prior for the quantile differences using an independent copula, i.e. by assuming that the marginals are independent.

Alternatively, we would like to consider marginal priors directly on the quantiles q_1, q_2, q_3, which may be more convenient for an expert to specify, and apply a copula which satisfies the order constraint $\Pr(q_1 < q_2 < q_3) = 1$. Furthermore, we could choose a copula which results in a joint distribution of maximum entropy given these constraints. In Butucea, Delmas, Dutfoy, and Fischer 2018, it is shown that if a sequence of marginals with PDFs $(f_i)_{1 \geq i \geq d}$ and CDFs $(F_i)_{1 \geq i \geq d}$ satisfy the condition

$$\forall 1 \leq i \leq d - 1\, \forall x \in \{x \colon 1 > F_i(x), F_{i+1}(x) > 0\}\, (F_i(x) > F_{i+1}(x))\,, \tag{8}$$

then this maximum entropy distribution is uniquely defined by the PDF

$$f(q_1, q_2, q_3) = f_1(q_1) \prod_{i=2}^{3} h_i(q_i) \exp\left(-\int_{q_{i-1}}^{q_i} h_i(s)\, \mathrm{d}s\right) \mathbb{1}_{q_1 \leq q_2 \leq q_3}\,, \tag{9}$$

where f_i are the PDFs of the marginals, and

$$h_i(x) = \frac{f_i(x)}{F_{i-1}(x) - F_i(x)}$$

if $x > 0$, and $h(x) = 0$ otherwise. See Appendix 7.3 for more details on the condition on the marginals.

This distribution, as well as verification of the conditions on the marginals, can be calculated using the library OpenTURNS (Baudin, Dutfoy, Iooss, and Popelin 2015).

After a change of variables

$$\begin{aligned} y &= F_i(x) \\ \implies \mathrm{d}y &= f_i(x)\,\mathrm{d}x\,, \end{aligned}$$

the integral in 9 can be simplified to

$$\int_{1-p_i}^{1-p_i} \frac{1}{F_{i-1}(F_i^{-1}(y)) - y}\,\mathrm{d}y\,,$$

whose limits are contained in $[0,1]$.

3.3 Expert specification

As quantile differences are positive, we are looking for distributions with support $[0,+\infty)$, with constraints defined by certain quantities which are fixed by an expert. We will investigate the case where the mean and variance, or equivalently the first two moments, are fixed by an expert.

Let p be a probability density with support $[0,+\infty)$. We define $n+1$ constraints on p in the form

$$\int_0^{+\infty} p(x) f_i(x)\,\mathrm{d}x - c_i\,, \quad i = 0,\dots,n\,,$$

for measurable functions f_i and constants $c_i \in \mathbb{R}$, including the constraint $f_0 \equiv 1, c_i = 1$ to ensure that the distribution is well defined. Using Lagrange multipliers and calculus of variations, we obtain the objective function

$$L(p,\lambda) = -\int_0^{+\infty} p(x)\log(p(x))\,\mathrm{d}x + \sum_{i=0}^{n} \lambda_i \left(\int_0^{+\infty} p(x) f_i(x)\,\mathrm{d}x - c_i \right)\,,$$

which has partial derivatives

$$\begin{aligned} \frac{\partial L}{\partial p}(p) &= -\log(p(x)) - 1 + \sum_{i=0}^{n} \lambda_i f_i(x)\,, \\ \frac{\partial^2 L}{\partial p^2}(p) &= -\frac{1}{p(x)} < 0\,, \end{aligned}$$

and therefore

$$\begin{aligned} \frac{\partial L}{\partial p} = 0 \iff p(x) &= \exp\left(-1 + \sum_{i=0}^{n} \lambda_i f_i(x)\right)\,, \\ \implies p(x) &\propto \exp\left(\sum_{i=1}^{n} \lambda_i f_i(x)\right)\,. \end{aligned} \tag{10}$$

If we fix the first two moments,

$$\int_0^{+\infty} p(x)x \,\mathrm{d}x \quad \text{and} \quad \int_0^{+\infty} p(x)x^2 \,\mathrm{d}x \,,$$

this implies that

$$f_1(x) = x \quad \text{and} \quad f_2(x) = x^2 \,,$$

and therefore from (10),

$$p(x) \propto \exp\left(\lambda_1 x + \lambda_2 x^2\right) \,.$$

Therefore the maximum entropy distribution is a truncated normal distribution with support $[0, +\infty)$. It has PDF

$$f(x) = \frac{1}{\sigma'} \frac{\phi(\frac{x-\mu'}{\sigma})}{1 - \Phi(\frac{-\mu'}{\sigma'})} \mathbb{1}_{x>0} \,,$$

where ϕ and Φ are the PDF and CDF of the standard normal distribution respectively. The parameters $\mu' \in \mathbb{R}$ and $\sigma' \in \mathbb{R}^+$ are the mean and standard deviation of the parent normal distribution.

3.4 Construction of priors

We would like to compare specification of quantiles with independent copulas vs specification of quantile differences with maximum entropy copula, and $k = 2$ vs $k = 3$ distributions specified by the expert. This gives us a total of four possible priors.

In the absence of a real expert, we choose the mean and variance of the quantile differences $\tilde{q}_1, \tilde{q}_2, \tilde{q}_3$, for the independent copula, using the data. Specifically, our expert specification of the quantile differences will consist of four values $(\mu_1, \mu_2, \mu_3, \sigma^2)$, such that

$$\mathrm{E}(\tilde{q}_i) = \mu_i - \mu_{i-1} \,, \quad \mathrm{Var}(\tilde{q}_i) = \sigma^2 \quad \forall i = 2, 3 \,.$$

In order to specify distributions for the quantiles q_1, q_2, q_3, for the maximum entropy copula, we choose distributions which approximate the marginals induced by $\tilde{q}_1, \tilde{q}_2, \tilde{q}_3$ by minimisation of relative entropy. When the truncated normal distributions are similar to normal distributions, these approximations will be good. The parameters of a truncated normal distribution can be obtained from the mean and variance using an algorithm described in Appendix 7.4.

For expert specification of two distributions, we consider just q_1, q_2 and $\tilde{q}_1, \tilde{q}_2$.

In Appendix 7.5, we propose a variant of any other prior in which ξ has a non-zero probability to be zero.

4 Sampling using MCMC methods

We can sample from prior and posterior distributions of θ using a Metropolis-Within-Gibbs algorithm with independent symmetrical proposal distributions for each parameter. We can reparametrise to $(\mu, \log(\sigma), \xi)$, which allows us to use normal distributions for each parameter. The algorithm is described in Alg. 1. This algorithm is a special case of the Metropolis–Hastings algorithm, whose convergence is studied in Roberts and Smith 1994.

In order to improve convergence by reducing dependence between the parameters, we reparametrise the model in terms of the hyperparameter M, the number of blocks, as described in Sharkey and J. A. Tawn 2017. Specifically, a model with M_1 and parameters θ_1 is equivalent to a model with M_2 and parameters θ_2 through the reparametrisation

$$T_{M_1,M_2} : \theta_1 = (\mu_1, \sigma_1, \xi_1) \mapsto \theta_2 = (\mu_2, \sigma_2, \xi_2) \tag{11}$$

Algorithm 1: Metropolis-Within-Gibbs with normal proposal distributions.

Input: function $f : \mathbb{R}^3 \to \mathbb{R}$ such that exp f is proportional to the target density
proposal distribution standard deviations $\sigma_j \in \mathbb{R}^+$ for $i = 1, 2, 3$
initial state $\theta_0 \in \mathbb{R}^3$
number of iterations $n \in \mathbb{N}$
length of burn-in $b \in \mathbb{N}$
Output: samples $s^{(1)}, \dots, s^{(n-b)} \in \mathbb{R}^3$
begin
    $(\theta, y) \longleftarrow (\theta_0, f(\theta_0))$
    for $i \in 1, \dots, n$ **do**
        for $j \in 1, 2, 3$ **do**
            $\theta_j^* \sim \mathcal{N}(\theta_j, \sigma_j)$
            $y^* \longleftarrow f(\theta^*)$
            $u \sim \mathcal{U}(0, 1)$
            if $y^* - y > \log(u)$ **then**
                $(\theta, y) \longleftarrow (\theta^*, y^*)$
        if $i > b$ **then**
            $s^{(i-b)} \longleftarrow \theta$

given by

$$\mu_2 = \mu - \frac{\sigma_1}{\xi_1}\left(1 - \left(\frac{M_2}{M_1}\right)^{-\xi_1}\right) ,$$
$$\sigma_2 = \sigma_1 \left(\frac{M_2}{M_1}\right)^{-\xi_1} ,$$
$$\xi_2 = \xi_1 .$$

Supposing that M is the number of years of data, we can choose a value $\tilde{M}$ which leads to better MCMC convergence, reparametrise the prior from θ to $\tilde{\theta}$ using the determinant of the inverse of $T_{\tilde{M},M}$, obtain a sample of $\tilde{\theta}$ using MCMC, and finally transform the sample using $T_{\tilde{M},M}$ to obtain a sample of θ.

Once we have samples of θ, we can obtain samples of the three quantiles (q_1, q_2, q_3). We can then use kernel density estimation with Gaussian kernels to visualise the marginals of θ and (q_1, q_2, q_3) using line and contour plots. We can use the samples of the posterior of θ to plot the mean return level against the return period, as well as 95% credible intervals.

5 Results

In order to compare the priors, we performed analyses on two datasets. The first was created by simulating the exceedances of a certain threshold from a Poisson point process. The second is real data consisting of observed daily average wind speed at Tours, France from 1931 to 2011 during the months November to May. These datasets are illustrated in Fig. 1. We fixed $p = (0.1, 0.01, 0.001)$ for both studies.

The analyses are coded in Python, using the libraries and packages NumPy (Oliphant 2006), Matplotlib (Hunter 2007), SciPy (Virtanen et al. 2020), and OpenTURNS (Baudin, Dutfoy, Iooss, and Popelin 2015). The code is available on GitHub. [1] Details of the Metropolis algorithm

[1] https://github.com/tzhg/extreme-bayes

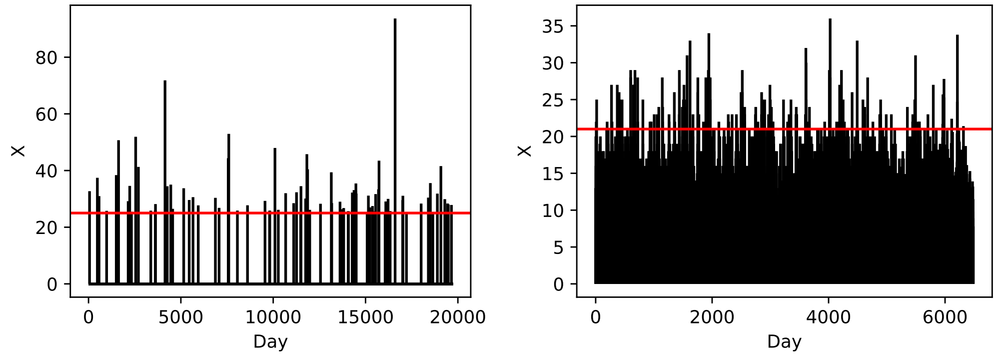


Figure 1: Simulation study (left) and wind speed (right) datasets, with chosen thresholds in red.

implementations are tabulated in Appendix 7.6.

5.1 Simulation study

According to the model described in § 2, the exceedances of the threshold u approximately follow a non-homogeneous Poisson point process with intensity function

$$\lambda(x) = \frac{1}{\sigma}\left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}_{+}^{-\frac{\xi+1}{\xi}} ,$$

which is a Generalised Pareto distribution. We simulated data for which this approximation is exact, with known parameters. Consistent with the study in Coles and J. Tawn 1996, we generated 19710 observations, $n_u = 86$ of which were above some threshold u. We chose $u = 25$ and $\theta = (25, 5, 0.2)$, and set $M = n_u$, which implied that $u = \mu$. We simulated 86 observations from a Generalised Pareto distribution with parameters θ, and distributed them uniformly over the total time period of 19710 observations, setting the remaining observations to 0.

The above construction of the data ensures that block maxima approximately follow a GEV distribution whose parameters can be determined by (11). A Q-Q plot of the fit of annual maxima is shown in Fig. 2. The asymptotic quantiles q_1, q_2, q_3 were calculated using (7) as

$$\hat{q}_1 = 39.211\,, \quad \hat{q}_2 = 62.734\,, \quad \hat{q}_3 = 99.517\,,$$

and we chose the expert specification

$$(\hat{q}_1, \hat{q}_2, \hat{q}_3, 27)\ .$$

The marginals of θ and (q_1, q_2, q_3) for the different priors are illustrated in Tables 3 and 4. The posteriors of μ and $\log\sigma$ are very similar, and very concentrated compared to the priors, whereas the posterior of ξ varies considerably, with $k = 2$ and maximum entropy copula priors being more spread out than $k = 3$ and independent priors.

We plotted the mean return level compared to the asymptotic return level in Fig. 5. The return levels of the $k = 2$ posteriors are almost identical, with the maximum entropy copula posteriors having a slightly wider credible interval. The $k = 3$ posteriors are higher and more variable, with the maximum entropy copula posterior in particular having a very large credible interval.

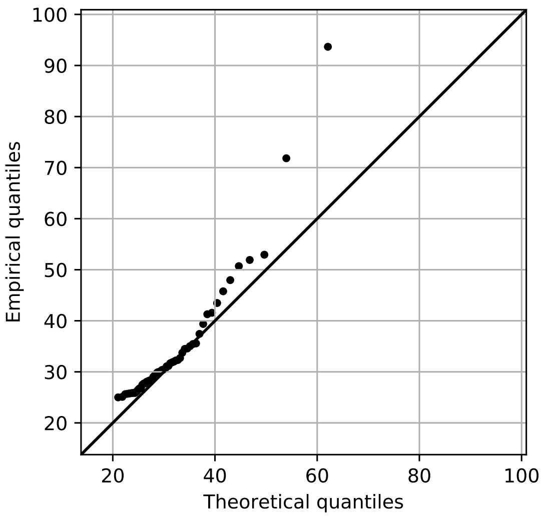


Figure 2: Q-Q plot of GEV distribution fit for simulation study.

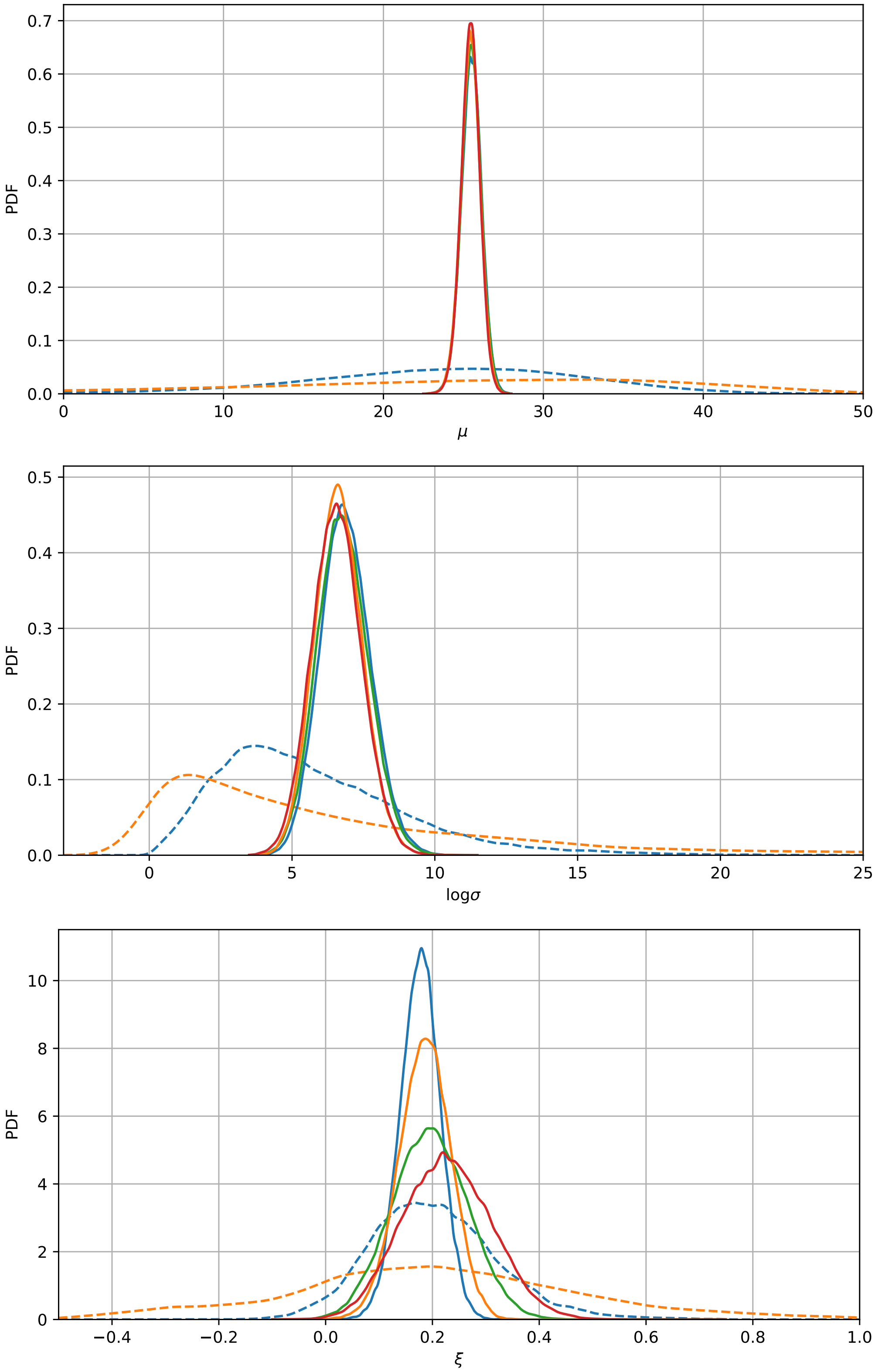


Figure 3: Prior (dashed) and posterior (solid) marginals of $(\mu, \log\sigma, \xi)$, for priors with $k = 3$ and independent copula **(blue)**, $k = 2$ and independent copula **(green)**, $k = 3$ and maximum entropy copula **(orange)**, and $k = 2$ and maximum entropy copula **(red)** for simulation study.

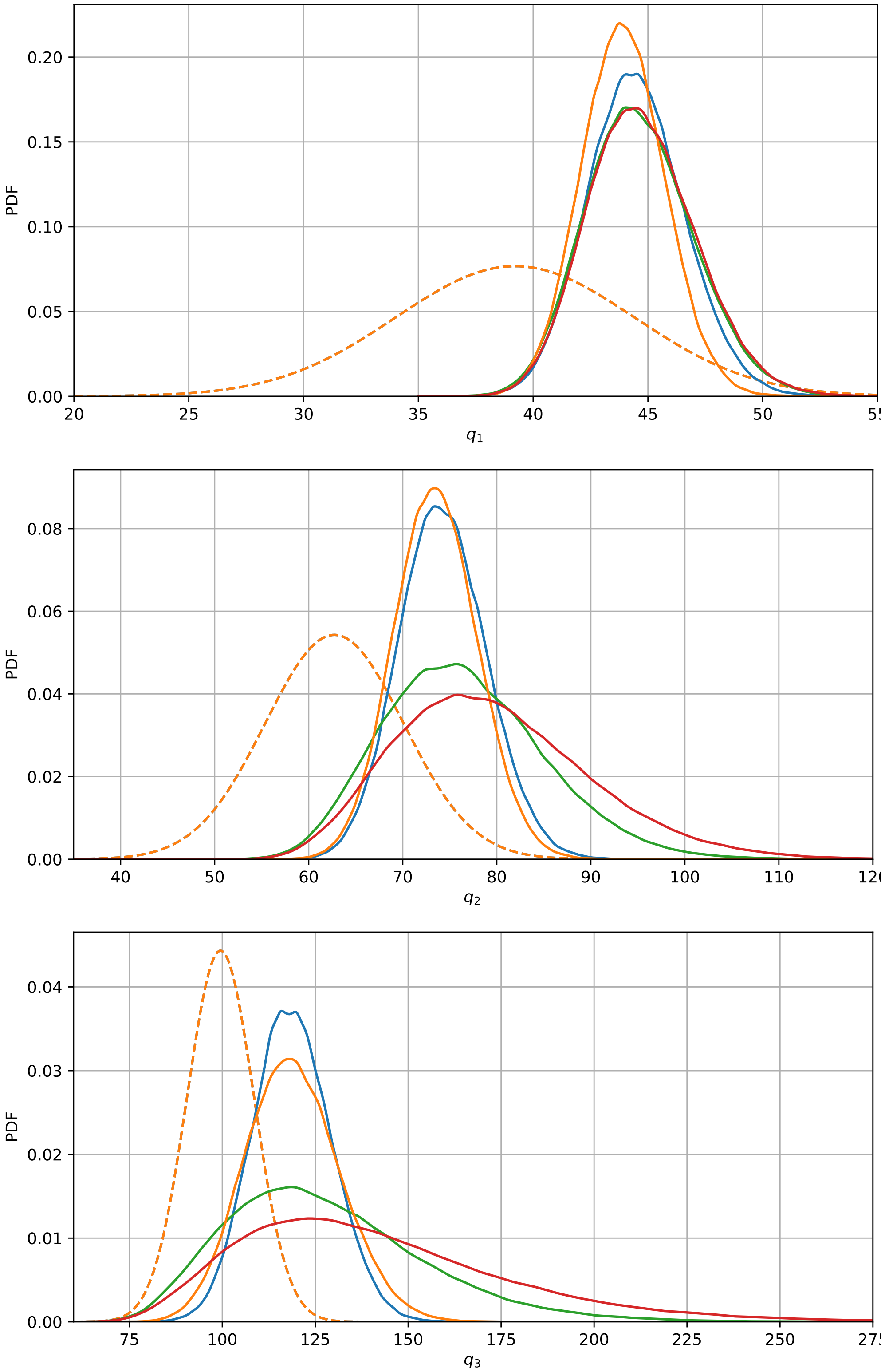


Figure 4: Prior (dashed) and posterior (solid) marginals of (q_2, q_2, q_3), for priors with $k = 3$ and independent copula **(blue)**, $k = 2$ and independent copula **(green)**, $k = 3$ and maximum entropy copula **(orange)**, and $k = 2$ and maximum entropy copula **(red)** for simulation study.

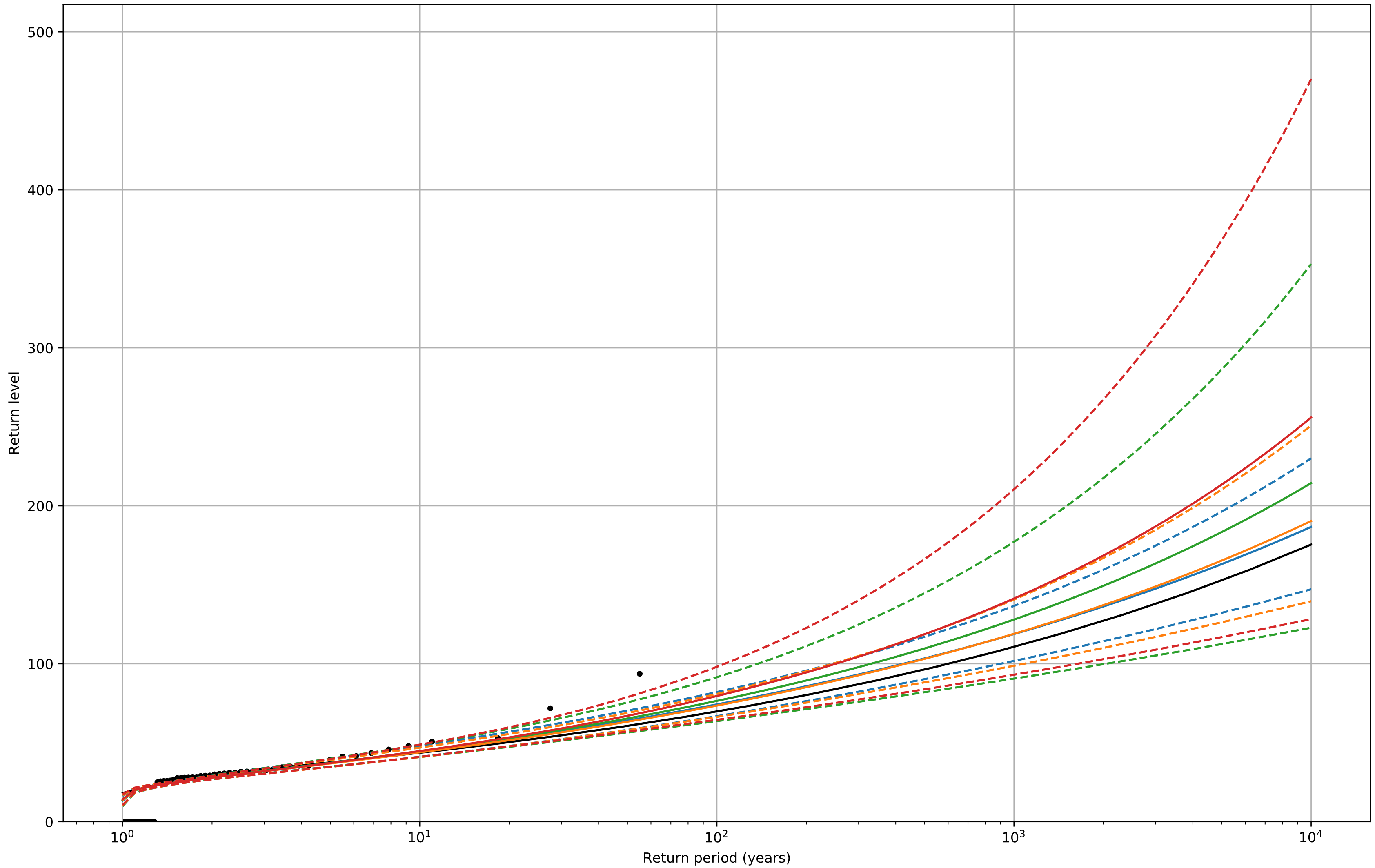


Figure 5: Mean return level with 95% credible intervals estimated using posterior distributions of priors with $k = 3$ and independent copula **(blue)**, $k = 2$ and independent copula **(green)**, $k = 3$ and maximum entropy copula **(orange)**, and $k = 2$ and maximum entropy copula **(red)**, with asymptotic return level **(black solid)** and empirical quantiles **(black dots)** for simulation study.

5.2 Wind speed data

This dataset consists of observations of average wind speed at Tours, France over a period of 30.34 years, from 1981 to 2011. We restricted the data to the period November to May to reduce non-stationarity. Violin plots of the data are shown in Fig. 6. We chose a threshold of $u = 21$ with 214 exceedances. The choice of threshold is investigated in Coles 2001, and will not be considered here.

In order to construct the priors, we first fitted a GEV distribution on the block maxima of the data using maximum likelihood, with the number of blocks equal to the number of exceedances. The fitted parameters were $(19.065, 3.851, -0.088)$, and a Q-Q plot of the fit is shown in Fig. 7. The quantiles q_1, q_2, q_3 were calculated using (7) as

$$\hat{q}_1 = 26.927\,, \quad \hat{q}_2 = 33.636\,, \quad \hat{q}_3 = 39.004$$

and we chose the expert specification

$$(26.45, 33.636, 48, 27)\ .$$

In order to satisfy the condition (8), we did not choose the exact fitted quantiles as means for the distributions on the quantile differences.

The univariate and bivariate marginals of θ and (q_1, q_2, q_3) for the different priors are illustrated in Tables 8 and 9. Compared to the simulation study, the posterior distribution of the parameters μ and $\log\sigma$ are more variable, and the $k = 3$ maximum entropy copula is more spread out for $\log\sigma$, and more concentrated for ξ.

We plotted the mean return level in Fig. 10 along with 95% credible intervals. The results are similar to the simulation study, except the return level for the $k = 3$ maximum entropy copula posterior is significantly higher and more variable compared to the $k = 3$ independent copula posterior.

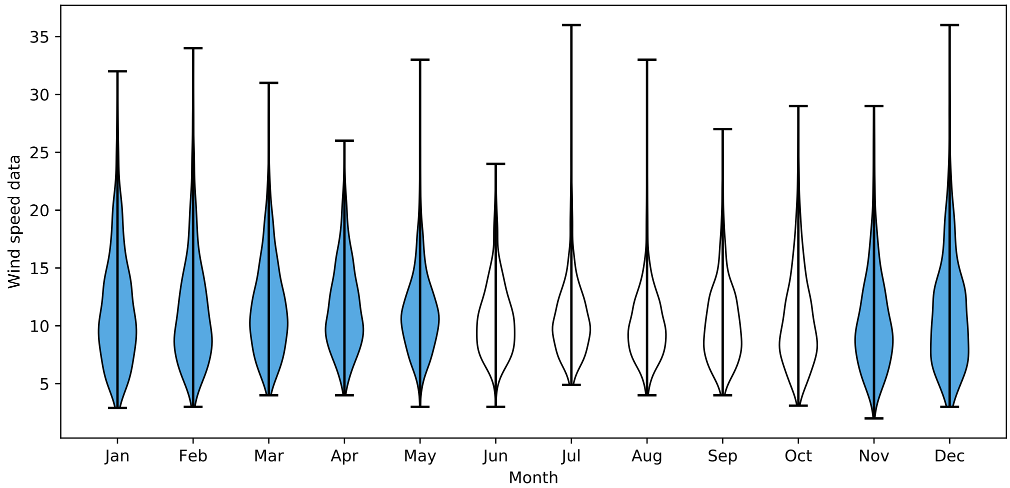


Figure 6: Violin plots of wind speed data by month.

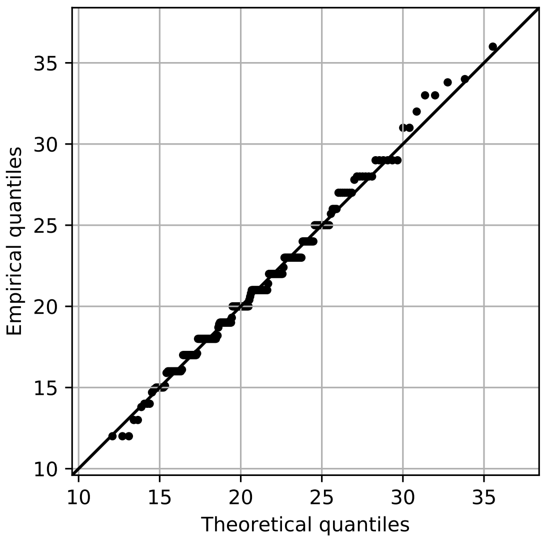


Figure 7: Q-Q plot of GEV distribution fit for wind speed data.

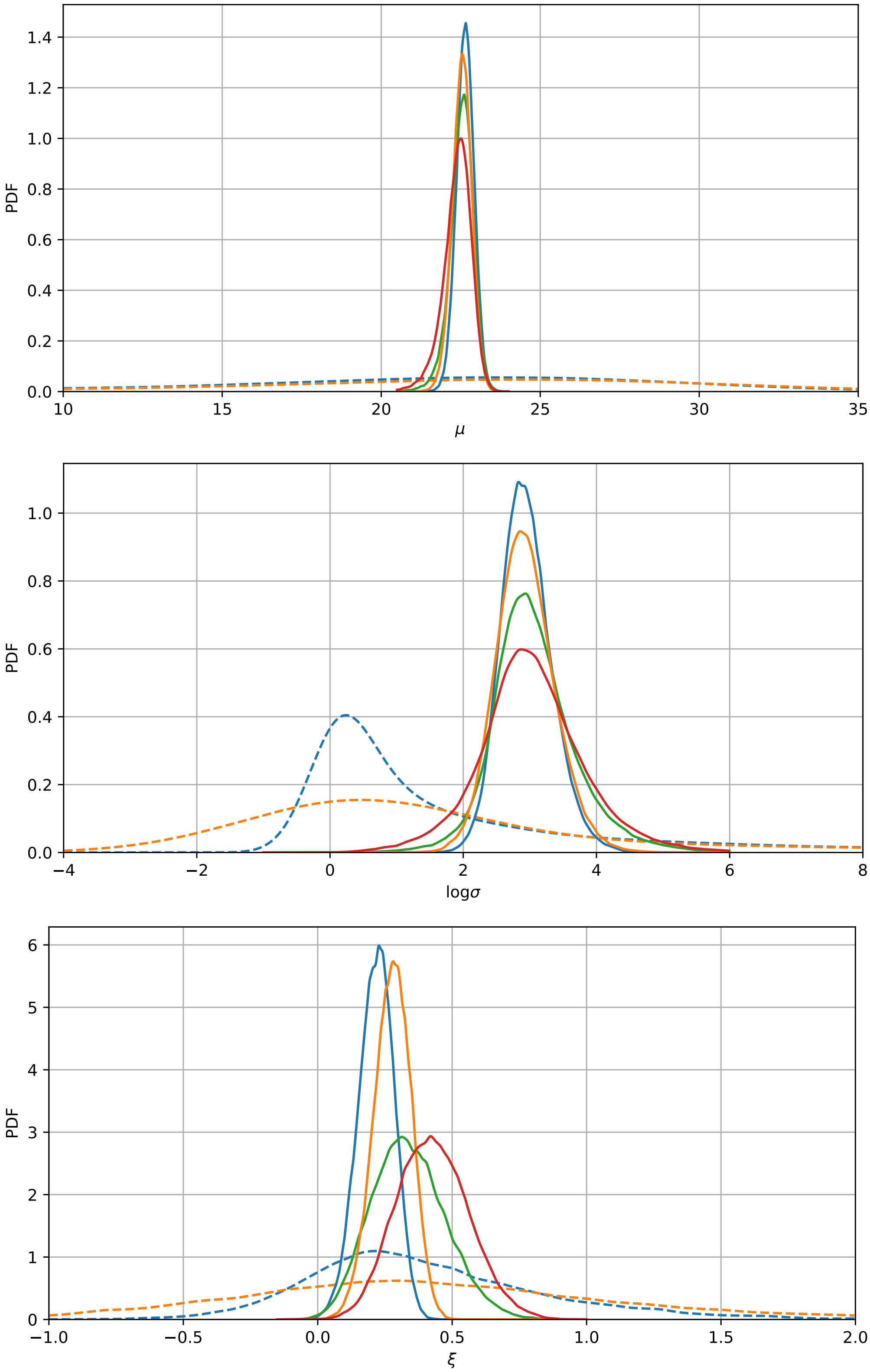


Figure 8: Prior (dashed) and posterior (solid) marginals of $(\mu, \log \sigma, \xi)$, for priors with $k = 3$ and independent copula **(blue)**, $k = 2$ and independent copula **(green)**, $k = 3$ and maximum entropy copula **(orange)**, and $k = 2$ and maximum entropy copula **(red)** for wind speed data.

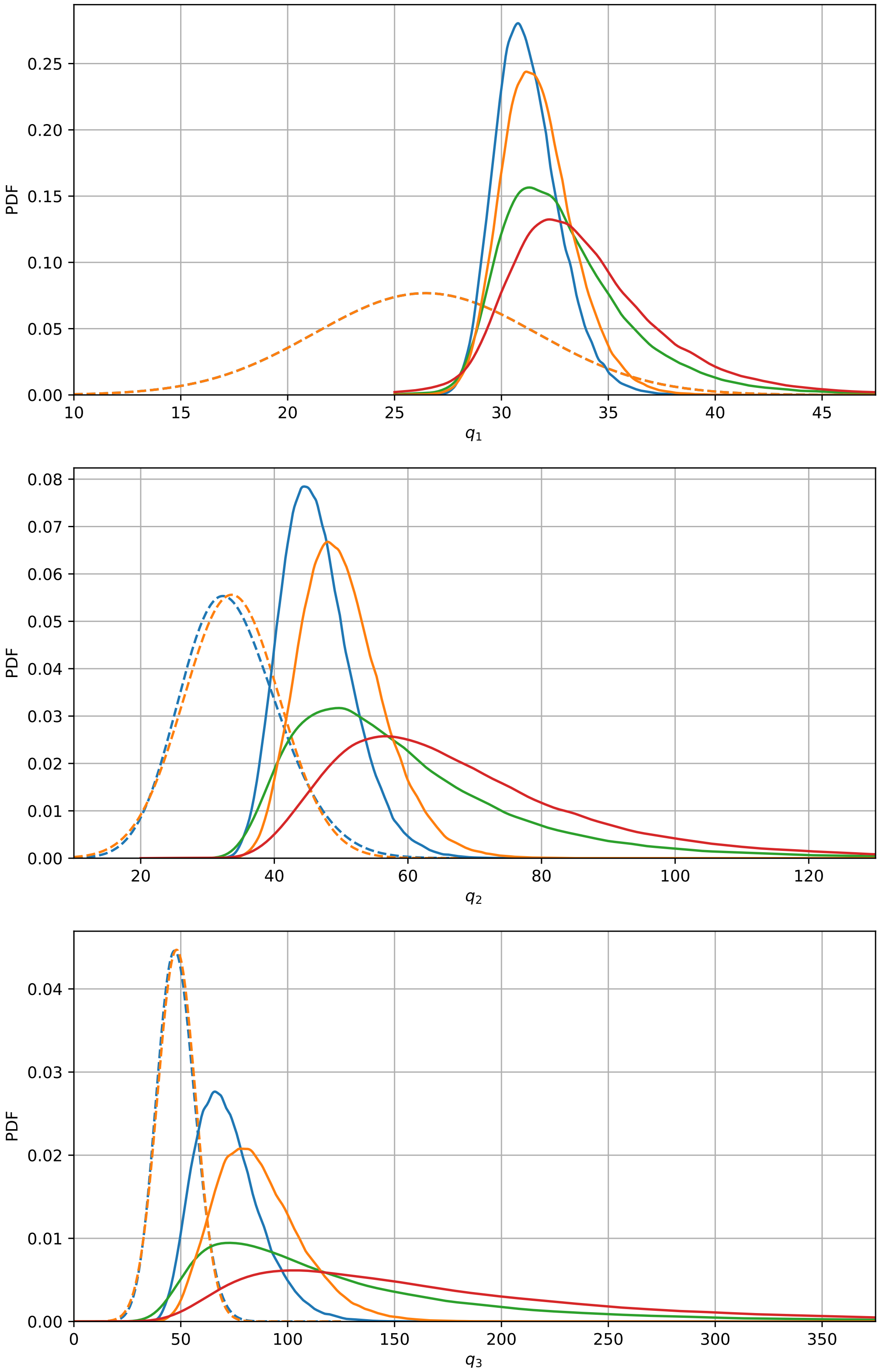


Figure 9: Prior (dashed) and posterior (solid) marginals of (q_2, q_2, q_3), for priors with $k = 3$ and independent copula **(blue)**, $k = 2$ and independent copula **(green)**, $k = 3$ and maximum entropy copula **(orange)**, and $k = 2$ and maximum entropy copula **(red)** for wind speed data.

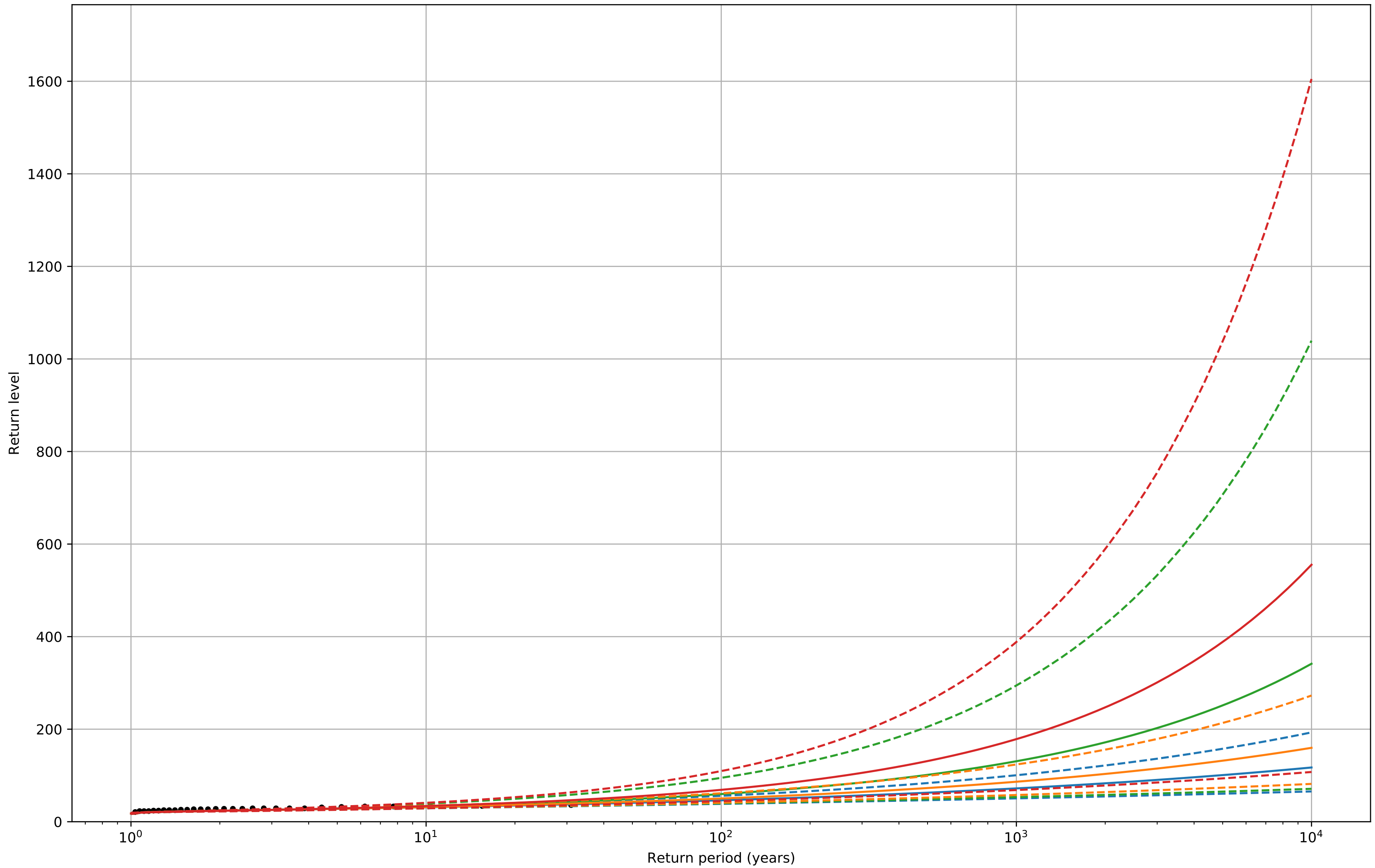


Figure 10: Mean return level with 95% credible intervals estimated using posterior distributions of priors with $k = 3$ and independent copula **(blue)**, $k = 2$ and independent copula **(green)**, $k = 3$ and maximum entropy copula **(orange)**, and $k = 2$ and maximum entropy copula **(red)**, with asymptotic return level **(black solid)** and empirical quantiles **(black dots)** for wind speed data.

6 Discussion

We have proposed different priors for the estimation of extreme quantiles, and compared them on two datasets in § 5.

Based on the results, the choice of $k = 2$ or $k = 3$ distributions specified is very significant in estimating the return level. Using $k = 3$ priors will considerably reduce the credibility intervals, at the expense of an extra distribution to specify. Also, the use of $k = 2$ priors is unjustified as we have not yet shown that the resulting posteriors are proper. Northrop and Attalides 2016 show that for least 4 observations, for a uniform prior on $(\mu, \log\sigma, \xi)$, i.e. $k = 0$, the resulting posterior is proper.

The choice of independent or maximum entropy copula is also significant. The maximum entropy copula gives wider return level credible intervals for the same amount of expert information provided. As well as better taking into account the uncertainty in the model, the greater ease of expert specification of quantiles rather than the difference of quantiles is a major advantage of these priors.

In order to better compare these priors, it may be valuable to consider quantitative measures of prior informativeness. For example, Reimherr, Meng, and Nicolae 2014 derived measures of prior sensitivity and prior informativeness and Reimherr, Meng, and Nicolae 2014 introduced an extension of prior sample size to non-conjugate priors.

7 Appendix

7.1 Determinants of reparametrisations of GEV model

Three quantile reparametrisation

The transformation is

$$\begin{aligned} g_3 \colon \Theta &\to \{(x,y,z) \in \mathbb{R}^3 \colon x < y < z\} \\ \theta &\mapsto (q_1, q_2, q_3)\,, \end{aligned}$$

given by

$$q_i = \mu + \sigma \frac{\exp(s_i\xi) - 1}{\xi}\,, \quad i = 1, 2, 3$$

where $s_i \coloneqq -\log(-\log(1 - p_i))$. We assume that

$$p_1 < 1 - \mathrm{e}^{-1} \approx 0.63\,,$$

which implies that $s_i > 0$ for $i = 1, 2, 3$. The partial derivatives are

$$\frac{\mathrm{d}q_i}{\mathrm{d}\mu}(\theta) = 1\,, \tag{12}$$

$$\frac{\mathrm{d}q_i}{\mathrm{d}\sigma}(\theta) = \frac{\exp(s_i\xi) - 1}{\xi}\,, \tag{13}$$

$$\frac{\mathrm{d}q_i}{\mathrm{d}\xi}(\theta) = \frac{\sigma(s_i\xi\exp(s_i\xi) - \exp(s_i\xi) + 1)}{\xi^2}\,, \tag{14}$$

and so the determinant of the Jacobian is

$$\det J(g_3)(\theta) = \frac{\sigma}{\xi^3} \begin{vmatrix} 1 & 1 \\ \exp(s_1\xi) - 1 & \exp(s_2\xi) - 1 \\ s_1\xi\exp(s_1\xi) - \exp(s_1\xi) + 1 & s_2\xi\exp(s_2\xi) - \exp(s_2\xi) + 1 \end{vmatrix}$$

$$
\begin{array}{c|}
1 \\
\exp(s_3\xi) - 1 \\
s_3\xi\exp(s_3\xi) - \exp(s_3\xi) + 1
\end{array}
$$

$$
\begin{aligned}
&= \frac{\sigma}{\xi^3}\begin{vmatrix} 1 & 1 & 1 \\ \exp(s_1\xi) & \exp(s_2\xi) & \exp(s_3\xi) \\ s_1\xi\exp(s_1\xi) & s_2\xi\exp(s_2\xi) & s_3\xi\exp(s_3\xi)\end{vmatrix} \\
&= \exp\left(\xi\sum_{i=1}^{3} s_i\right)\frac{\sigma}{\xi^3}\begin{vmatrix} \exp(-s_1\xi) & \exp(-s_2\xi) & \exp(-s_3\xi) \\ 1 & 1 & 1 \\ s_1\xi & s_2\xi & s_3\xi\end{vmatrix} \\
&= -\exp\left(\xi\sum_{i=1}^{3} s_i\right)\frac{\sigma}{\xi^2}\begin{vmatrix} 1 & 1 & 1 \\ \exp(-s_1\xi) & \exp(-s_2\xi) & \exp(-s_3\xi) \\ s_1 & s_2 & s_3\end{vmatrix} \\
&= -\exp\left(\xi\sum_{i=1}^{3} s_i\right)\frac{\sigma}{\xi^2}(\exp(-s_2\xi)s_3 - \exp(-s_3\xi)s_2 \\
&\quad + \exp(-s_3\xi)s_1 - \exp(-s_1\xi)s_3 \\
&\quad + \exp(-s_1\xi)s_2 - \exp(-s_2\xi)s_1)\,.
\end{aligned}
$$

Two quantile reparametrisation

The transformation is

$$
\begin{aligned}
g_2 \colon \Theta &\to \mathbb{R}^+ \times \{(x,y)\in\mathbb{R}^2 \colon x < y\} \\
\theta &\mapsto (\sigma, q_1, q_2)
\end{aligned}
$$

given by

$$
q_i = \mu + \sigma\frac{\exp(s_i\xi) - 1}{\xi}\,, \quad i = 1, 2\,.
$$

From the partial derivatives in (12)–(14), the determinant of the Jacobian is

$$
\begin{aligned}
\det J(g_2)(\theta) &= \frac{\sigma}{\xi^2}\begin{vmatrix} 0 & 1 & 1 \\ 1 & (\exp(s_1\xi)-1)\xi^{-1} & (\exp(s_2\xi)-1)\xi^{-1} \\ 0 & s_1\xi\exp(s_1\xi) - \exp(s_1\xi) + 1 & s_2\xi\exp(s_2\xi) - \exp(s_2\xi) + 1\end{vmatrix} \\
&= \frac{\sigma}{\xi^2}\begin{vmatrix} 1 & 1 \\ s_2\xi\exp(s_2\xi) - \exp(s_2\xi) + 1 & s_1\xi\exp(s_1\xi) - \exp(s_1\xi) + 1\end{vmatrix} \\
&= \frac{\sigma}{\xi^2}((s_1\xi - 1)\exp(s_1\xi) - (s_2\xi - 1)\exp(s_2\xi))\,.
\end{aligned}
$$

7.2 The Jeffreys prior

Let l be the log-likelihood of the model (3),

$$
l(\theta \mid \mathbf{x}) = -M\left\{1 + \xi\frac{u-\mu}{\sigma}\right\}_+^{-\frac{1}{\xi}} - n_u\log\sigma - \left(\frac{1}{\xi} + 1\right)\sum_{i=1}^{n_u}\log\left\{1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right\}_+\,,
$$

and define the Fisher information matrix $I(\theta)$ with entries

$$
I(\theta)_{ij} = \mathrm{E}\left(-\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}\right)\,.
$$

In the situation where we do not have access to any prior information, Jeffreys 1946 proposes an uninformative prior which is invariant to reparameterisation. It is proportional to the square root of the determinant of the Fisher information matrix:

$$\pi \propto \sqrt{\det I(\phi)}\,.$$

Using the Fisher information matrix derived in Sharkey and J. A. Tawn 2017, if we let

$$\begin{aligned}
a &:= \frac{(\xi+1)^2}{2\xi+1}\,,\\
b &:= -w_1(\xi)2v + (\xi+1)v^2 - 1 + 2(\xi+1)w_1(\xi)w_2(\xi) - \xi w_4(\xi)\,,\\
c &:= \left\{\frac{v^2}{\xi w_1(\xi)^2} - \frac{2\log w_1(\xi)}{\xi^3} + \frac{2v}{\xi^2 w_1(\xi)} + \frac{1}{\xi^2}\left(\frac{\log w_1(\xi)}{\xi} - \frac{1}{w_1(\xi)}\right)^2\right\}_+\\
&\quad + \frac{2\{\xi + \log w_1(\xi)\}_+}{M\xi^3} - \frac{2w_2(\xi)}{\xi^2 w_1(\xi)} + \frac{w_4(\xi)}{\xi w_1(\xi)^2}\\
d &:= (\xi+1)v - \xi w_3(\xi)\,,\\
e &:= \frac{1}{\xi^2}\left\{\log w_1(\xi) - \frac{v\xi(\xi+1)}{w_1(\xi)}\right\}_+ - \frac{1}{\xi+1} + \frac{w_3(\xi)}{w_1(\xi)}\,,\\
f &:= \frac{v}{\xi^2}\left\{\log w_1(\xi) + \frac{v\xi(\xi+1)}{w_1(\xi)}\right\}_+ - w_2(\xi) + \frac{w_4(\xi)}{w_1(\xi)}\,,
\end{aligned}$$

where

$$\begin{aligned}
v &:= \frac{u-\mu}{\sigma}\,,\\
w_1(\xi) &:= \{1+\xi v\}_+\,,\\
w_2(\xi) &:= \frac{\{1+(\xi+1)v\}_+}{\xi+1}\,,\\
w_3(\xi) &:= \frac{\{1+(2\xi+1)v\}_+}{2\xi+1}\,,\\
w_4(\xi) &:= \frac{\{(2\xi^2+3\xi+1)v^2+(4\xi+2)v+2\}_+}{2\xi+1}\,,
\end{aligned}$$

then Jeffreys prior is proportional to

$$\pi(\theta) \propto \frac{\sigma^2}{w_1(\xi)^2}\sqrt{a(df-e^2)+b(ce-bf)+c(be-dc)}\,.$$

Fig. 11 shows the Jeffreys prior as a function of ξ, with $(\mu,\sigma)=(0,1)$, $u=\mu$, and $M=1$.

7.3 Maximum entropy copula: condition on marginals

The construction of the maximum entropy distribution in § 3.2 requires the following condition on the marginals: for a sequence of distributions $(F_i)_{1\geq i\geq d}$,

$$\forall 1\leq i\leq d-1\ \forall x\in\{x\colon 1>F_i(x), F_{i+1}(x)>0\}\ (F_i(x)>F_{i+1}(x))\,. \tag{15}$$

We will investigate this condition for Weibull and gamma distributed marginals.

Weibull distribution

Let $1\leq i\leq d-1$. If the marginals are Weibull distributed, then we have that

$$1>F_i(x), F_{i+1}(x)>0 \iff x>0\,.$$

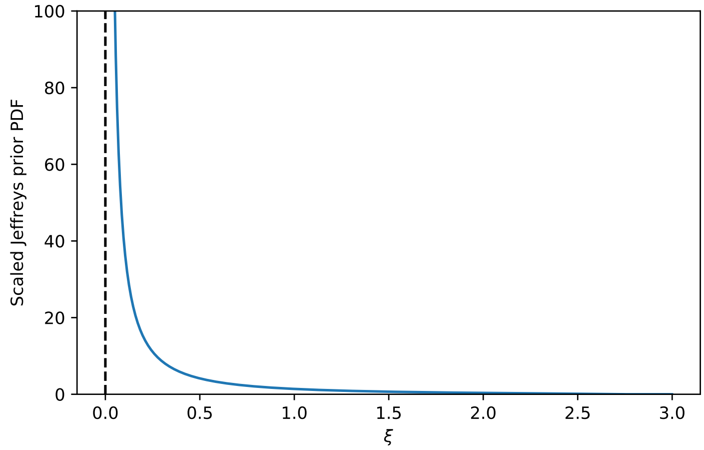


Figure 11: Scaled Jeffreys prior PDF against ξ.

The CDF when $x > 0$ is

$$F_i(x) = 1 - \exp\left(-\left(\frac{x}{\lambda_i}\right)^{k_i}\right),$$

with $k_i, \lambda_i > 0$. Let $x > 0$. Then (15) is equivalent to

$$\begin{aligned}
1 - \exp\left(-\left(\frac{x}{\lambda_i}\right)^{k_i}\right) &> 1 - \exp\left(-\left(\frac{x}{\lambda_{i+1}}\right)^{k_{i+1}}\right) \\
\iff \left(\frac{x}{\lambda_i}\right)^{k_i} &> \left(\frac{x}{\lambda_{i+1}}\right)^{k_{i+1}}.
\end{aligned}$$

If $k_i = k_{i+1} = k$,

$$\begin{aligned}
\left(\frac{x}{\lambda_i}\right)^{k} > \left(\frac{x}{\lambda_{i+1}}\right)^{k} &\iff \frac{x}{\lambda_i} > \frac{x}{\lambda_{i+1}} \\
&\iff \frac{1}{\lambda_i} > \frac{1}{\lambda_{i+1}} \\
&\iff \lambda_i < \lambda_{i+1}.
\end{aligned}$$

If $k_i \neq k_{i+1}$,

$$\begin{aligned}
F_i(x) = F_{i+1}(x) &\iff \left(\frac{x}{\lambda_i}\right)^{k_i} = \left(\frac{x}{\lambda_{i+1}}\right)^{k_{i+1}} \\
&\iff \exp\left(k_i(\log x - \log \lambda_i)\right) = \exp\left(k_{i+1}(\log x - \log \lambda_{i+1})\right) \\
&\iff k_i(\log x - \log \lambda_i) = k_{i+1}(\log x - \log \lambda_{i+1}) \\
&\iff k_i \log x - k_i \log \lambda_i = k_{i+1} \log x - k_{i+1} \log \lambda_{i+1} \\
&\iff k_i \log x - k_{i+1} \log x = k_i \log \lambda_i - k_{i+1} \log \lambda_{i+1} \\
&\iff (k_i - k_{i+1}) \log x = k_i \log \lambda_i - k_{i+1} \log \lambda_{i+1} \\
&\iff \log x = \frac{k_i \log \lambda_i - k_{i+1} \log \lambda_{i+1}}{k_i - k_{i+1}} \\
&\iff x = \underbrace{\exp\left(\frac{k_i \log \lambda_i - k_{i+1} \log \lambda_{i+1}}{k_i - k_{i+1}}\right)}_{=: h(k_i, \lambda_i, k_{i+1}, \lambda_{i+1})}.
\end{aligned}$$

Therefore the two CDFs intersect, and so (15) cannot be satisfied. In conclusion, the condition is equivalent to

$$(k_i = k_{i+1}) \wedge (\lambda_i < \lambda_{i+1}) \quad \forall 1 \leq i \leq d-1\,.$$

In practice however, as $y \coloneqq h(k_i, \lambda_i, k_{i+1}, \lambda_{i+1}) \to \pm\infty$, the intersection will tend to the bounds of the set $\{x\colon 1 > F_i(x), F_{i+1}(x) > 0\}$ and so (15) will be asymptotically satisfied. When does this happen? Consider a reparametrisation

$$(k_i, \lambda_i, k_{i+1}, \lambda_{i+1}) \mapsto (k_i, \lambda_i, k_i - k_{i+1}, \lambda_{i+1})\,,$$

which implies that

$$h(k_i, \lambda_i, z_i, \lambda_{i+1}) = \frac{k_i \log \lambda_i + (z_i - k_i) \log \lambda_{i+1}}{z_i}\,, \quad z_i < k_i,\ z_i \neq 0\,.$$

Then

- As $z_i \to 0$, $y \to +\infty$,
- As $k_i \to +\infty$, $y \to \pm\infty$, with sign depending on the sign of $\log(\frac{\lambda_i}{\lambda_{i+1}})$,
- As $\lambda_i \to +\infty$ or $\lambda_{i+1} \to 0$, $y \to +\infty$,
- As $\lambda_i \to 0$ or $\lambda_{i+1} \to +\infty$, $y \to -\infty$.

Gamma distribution

Let $1 \leq i \leq d-1$. If the marginals are gamma distributed, then we have that

$$1 > F_i(x), F_{i+1}(x) > 0 \iff x > 0\,.$$

Suppose that $x > 0$, and that we have two such distributions,

$$F_1 \sim \Gamma(\alpha_1, \beta_1) \quad \text{and} \quad F_2 \sim \Gamma(\alpha_2, \beta_2)\,,$$

with PDFs f_1 and f_2 given by

$$f_i(x) = \frac{\beta_i^{\alpha_i} x^{\alpha_i - 1} \exp(-\beta_i x)}{\Gamma(\alpha_i)} \quad i = 1, 2\,.$$

Let $x > 0$, and define

$$\phi(x) = F_2(x) - F_1(x)\,,$$

so that if $(i, i+1) = (1, 2)$, (15) is equivalent to $\phi(x) < 0$, and if $(i, i+1) = (2, 1)$, (15) is equivalent to $\phi(x) > 0$. We have that

$$\begin{aligned}
\phi'(x) &= f_2(x) - f_1(x) \\
&= f_1(x) \left(\frac{f_2(x)}{f_1(x)} - 1 \right) \\
&= f_1(x) \left(\frac{\Gamma(\alpha_1)\beta_2^{\alpha_2}}{\Gamma(\alpha_2)\beta_1^{\alpha_1}} x^{\alpha_2 - \alpha_1} \exp((\beta_1 - \beta_2)x) - 1 \right)\,.
\end{aligned}$$

Let

$$C \coloneqq \frac{\Gamma(\alpha_1)\beta_2^{\alpha_2}}{\Gamma(\alpha_2)\beta_1^{\alpha_1}} > 0$$

and

$$f(x) := Cx^{\alpha_2-\alpha_1}\exp((\beta_1-\beta_2)x) - 1\,.$$

Then

$$\begin{aligned} f'(x) &= C\left[(\alpha_2-\alpha_1)x^{\alpha_2-\alpha_1-1}\exp((\beta_1-\beta_2)x) + x^{\alpha_2-\alpha_1}\exp((\beta_1-\beta_2)x)(\beta_1-\beta_2)\right] \\ &= \underbrace{Cx^{\alpha_2-\alpha_1-1}\exp((\beta_1-\beta_2)x)}_{>0}\left[(\alpha_2-\alpha_1)+x(\beta_1-\beta_2)\right] \end{aligned}$$

Case 1: $(\alpha_1 \le \alpha_2 \wedge \beta_1 \ge \beta_2) \wedge \neg(\alpha_1 = \alpha_2 \wedge \beta_1 = \beta_2)$

We have that $\alpha_2-\alpha_1 \ge 0$ and $\beta_1-\beta_2 \ge 0$, and therefore

$$\begin{aligned} (\alpha_2-\alpha_1)+x(\beta_1-\beta_2) &> 0 \\ \implies f'(x) &> 0\,. \end{aligned}$$

This implies that f is strictly increasing on $\mathbb{R}^+$.

1. If $\alpha_1 < \alpha_2$, we have that

$$\lim_{x\to 0^+} f(x) = -1 \quad \text{and} \quad \lim_{x\to+\infty} f(x) = +\infty\,.$$

2. If $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 > \beta_2$, we have that

$$\lim_{x\to 0^+} f(x) = \left(\frac{\beta_2}{\beta_1}\right)^{\alpha} - 1 < 0 \quad \text{and} \quad \lim_{x\to+\infty} f(x) = +\infty\,.$$

Therefore in both cases, there exists a unique $z_0 > 0$ such that $f(z_0) = 0$. This implies that

$$\begin{cases} \phi'(x) < 0 & \text{if} \quad x < z_0\,, \\ \phi'(x) = 0 & \text{if} \quad x = z_0\,, \\ \phi'(x) > 0 & \text{if} \quad x > z_0\,. \end{cases}$$

In order to proceed we will need the following Lemma.

Lemma 7.1. *1. Let $f\colon (a,b] \to \mathbb{R}$ be continuous on $(a,b]$ and differentiable on (a,b), with $a \in \mathbb{R}\cup\{+\infty\}, b \in \mathbb{R}$. If for all $x \in (a,b)$, $f'(x) > 0$ (resp. $<$) and $\lim_{x\to a^+} f(x) = L \in \mathbb{R}$, then for all $z \in (a,b]$, $f(z) > L$ (resp. $<$).*

2. Let $f\colon [a,b) \to \mathbb{R}$ be continuous on $[a,b)$ and differentiable on (a,b), with $a \in \mathbb{R}, b \in \mathbb{R}\cup\{+\infty\}$. If for all $x \in (a,b)$, $f'(x) > 0$ (resp. $<$) and $\lim_{x\to b^-} f(x) = L \in \mathbb{R}$, then for all $z \in [a,b)$, $f(z) < L$ (resp. $>$).

Proof. We will prove the first case, for $f'(x) > 0$, as both cases are symmetrical. Suppose that $z \in (a,b)$. We need to show that $f(z), f(b) > L$. Let $y \in (a,z)$. As f' is strictly positive on (a,b), f is strictly increasing (by the mean value theorem), and so $f(z) - f(y) = c > 0$. Therefore $f(z) > f(y) + \frac{c}{2}$. Then $f(z) = \lim_{y\to a^+} f(z) \ge \lim_{y\to a^+} f(y) + \frac{c}{2} = L + \frac{c}{2} > L$. Furthermore, $f(b) = \lim_{z\to b^-} f(z) \ge \lim_{x\to b^-}(L + \frac{c}{2}) = L + \frac{c}{2} > L$. $\square$

By Lemma 7.1, for all $z \in (0, z_0]$, $\phi(z) < \lim_{x\to 0^+}\phi(x) = 0$, and for all $z \in [z_0, +\infty)$, $\phi(z) < \lim_{x\to+\infty}\phi(x) = 0$. Therefore for all $x > 0$, $\phi(x) < 0$, and so if $(i, i+1) = (1,2)$, (15) is satisfied, and if $(i, i+1) = (2,1)$, (15) is violated.

Case 2: $\alpha_1 < \alpha_2,\ \beta_1 < \beta_2$

Since

$$\lim_{x \to 0^+} f(x) = -1\,,$$

there exists a $\delta > 0$ such that for all $z \in (0, \delta]$, $f(z) = \phi'(z) < 0$, and so by Lemma 7.1, $\phi(z) < \lim_{x\to 0^+} \phi(x) = 0$. Since

$$\lim_{x \to +\infty} f(x) = -1\,,$$

there exists a $\delta' > 0$ such that for all $z \in [\delta', +\infty)$, $f(z) = \phi'(z) < 0$, and so by Lemma 7.1, $\phi(z) > \lim_{x\to 0^+} \phi(x) = 0$. Therefore ϕ is neither strictly positive nor strictly negative, and so if either $(i, i+1) = (1, 2)$ or $(i, i+1) = (2, 1)$, (15) is violated.

In conclusion, we have shown that the condition is equivalent to

$$(\alpha_i \leq \alpha_{i+1} \wedge \beta_i \geq \beta_{i+1}) \wedge \neg(\alpha_i = \alpha_{i+1} \wedge \beta_i = \beta_{i+1}) \quad \forall 1 \leq i \leq d-1\,.$$

Like in the Weibull case, the condition can be asymptotically satisfied. For instance, if the means $\frac{\alpha}{\beta}$ are sufficiently separated, and the variances $\frac{\alpha}{\beta^2}$ are small, there will be less significant overlap between the CDFs.

7.4 Obtaining truncated normal parameters from mean and variance

Suppose that we are given the mean μ^* and variance $(\sigma^*)^2$ of a random variable X which follows a truncated normal distribution with support parameters $a = 0$ and $b = +\infty$, and we would like to determine the remaining parameters μ and σ.

This means that we need to invert the equations

$$\begin{aligned}
\mathrm{E}(X) &= \mu + \frac{\phi(\alpha)}{1 - \Phi(\alpha)}\sigma \\
\mathrm{Var}(X) &= \sigma^2 \left(1 + \frac{\alpha\phi(\alpha)}{1 - \Phi(\alpha)} - \left(\frac{\phi(\alpha)}{1 - \Phi(\alpha)}\right)^2\right)\,,
\end{aligned}$$

where ϕ and Φ are the PDF and CDF of the standard normal distribution respectively and

$$\alpha := -\frac{\mu}{\sigma}\,. \tag{16}$$

Setting

$$H(x) := \frac{\phi(x)}{1 - \Phi(x)}\,,$$

which is the hazard function of the standard normal distribution, we can simplify the equations to

$$\mathrm{E}(X) = \mu + H(\alpha)\sigma \tag{17}$$
$$\mathrm{Var}(X) = \sigma^2 \left(1 + \alpha H(\alpha) - (H(\alpha))^2\right)\,. \tag{18}$$

Therefore,

$$\begin{aligned}
\left(\frac{\sigma^*}{\mu^*}\right)^2 &= \frac{\sigma^2 \left(1 + \alpha H(\alpha) - (H(\alpha))^2\right)}{(\mu + H(\alpha)\sigma)^2} \\
&= \frac{1 + \alpha H(\alpha) - (H(\alpha))^2}{(H(\alpha) - \alpha)^2} =: f(\alpha)\,,
\end{aligned}$$

and once we have α, from 17 we have

$$\frac{\mu^*}{\sigma} = H(\alpha) - \alpha \implies \sigma = \frac{\mu^*}{H(\alpha) - \alpha}$$

and from 16,

$$\alpha = -\frac{\mu}{\sigma} \implies \sigma = -\alpha\mu \,.$$

It remains to solve $f(\alpha) = \left(\frac{\sigma^*}{\mu^*}\right)^2$. This can be done numerically using Newton's method with the derivative of f. Since

$$\phi'(x) = -x\phi(x)\,,$$

the derivative of H is

$$\begin{aligned}
H'(x) &= \frac{(1-\Phi(x))x(-\phi(x)) - \phi(x)(-\phi(x))}{(1-\Phi(x))^2} \\
&= \phi(x)\frac{-(1-\Phi(x))x + \phi(x)}{(1-\Phi(x))^2} \\
&= H(x)\frac{-(1-\Phi(x))x + \phi(x)}{1-\Phi(x)} \\
&= H(x)\left(\frac{-(1-\Phi(x))x}{1-\Phi(x)} + H(x)\right) \\
&= H(x)\left(H(x) - x\right)\,.
\end{aligned}$$

Therefore the derivative of f is

$$\begin{aligned}
f'(x) &= \Big((H(x)-x)^2\left(H(x) + xH(x)\left(H(x)-x\right) - 2(H(x))^2\left(H(x)-x\right)\right) \\
&\quad - \left(1 + xH(x) - (H(x))^2\right) 2\left(H(x)-x\right)\left(H(x)\left(H(x)-x\right) - 1\right)\Big)\left(H(x)-x\right)^{-4} \\
&= \frac{(2 + H(x)(H(x)-x)(-3 + (H(x)-x)x))}{(H(x)-x)^3}\,.
\end{aligned}$$

7.5 Spike-and-slab prior distribution for ξ

This is a prior with a non-zero probability mass (spike) at $\xi = 0$, and a flat slab elsewhere.

We use the method proposed by Kuo and Mallick 1998. We introduce an indicator random variable γ such that $\gamma = 0$ when ξ is "in" the spike and $\gamma = 1$ when ξ is "in" the slab. The probability that ξ is non-zero can then be calculated as the expectation of γ. We also need a random variable β such that $\xi = \beta\gamma$, which determines the distribution of ξ in the slab. Our model now has parameters $\eta \coloneqq (\mu, \sigma, \beta, \gamma)$, and we suppose that γ does not depend on the other parameters.

If we have a prior π_θ for θ, then we can set

$$\begin{aligned}
&\pi_{\mu,\sigma,\beta} \coloneqq \pi_\theta \\
&\pi_\gamma(\gamma) \coloneqq \alpha^\gamma(1-\alpha)^{1-\gamma} \\
\implies\ &\pi_\eta(\eta) = \pi_\theta(\mu,\sigma,\beta)\alpha^\gamma(1-\alpha)^{1-\gamma}\,,
\end{aligned}$$

with $\alpha \in [0,1]$. The posterior is then proportional to

$$\pi_{\eta|\mathbf{x}}(\eta \mid \mathbf{x}) \propto L(\mu,\sigma,\beta\gamma \mid \mathbf{x})\pi_\theta(\mu,\sigma,\beta)\alpha^\gamma(1-\alpha)^{1-\gamma}\,.$$

Since, up to the same constant of proportionality, we have

$$\pi_{\gamma|\mu,\sigma,\beta,\mathbf{x}}(0 \mid \mu,\sigma,\beta,\mathbf{x}) \propto (1-\alpha)L(\mu,\sigma,0 \mid \mathbf{x})$$

and

$$\pi_{\gamma|\mu,\sigma,\beta,\mathbf{x}}(1 \mid \mu,\sigma,\beta,\mathbf{x}) \propto \alpha L(\mu,\sigma,\beta \mid \mathbf{x})\,,$$

the full conditional of γ is a Bernoulli distribution with probability of success

$$\frac{\alpha L(\mu,\sigma,\beta \mid \mathbf{x})}{\alpha L(\mu,\sigma,\beta \mid \mathbf{x}) + (1-\alpha)L(\mu,\sigma,0 \mid \mathbf{x})}\,.$$

We can sample γ from its full conditional in the Metropolis–Hastings algorithm.

7.6 Metropolis-Hastings algorithm implementation details

In Tables 4–24, we show details on our implementations of the Metropolis-Hastings algorithm, including the proposal distributions, initialisation, sample size, traceplots, histograms, and acceptance rates.

Table 1: Metropolis-Hastings algorithm for simulation study with $k = 3$ and independent copula prior.

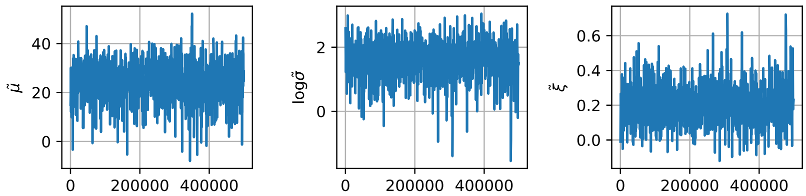
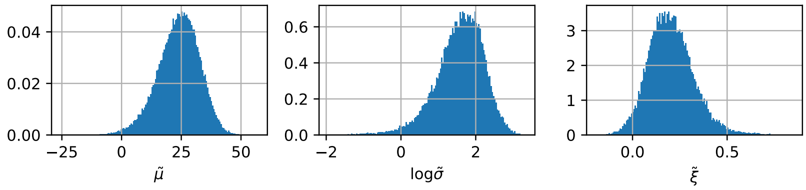
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 25^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 1^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.2^2)$
Initialisation	25, 2, 0.2
Sample size	500,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.251, 0.138, 0.14

Table 2: Metropolis algorithm for simulation study with $k = 3$ and independent copula posterior.

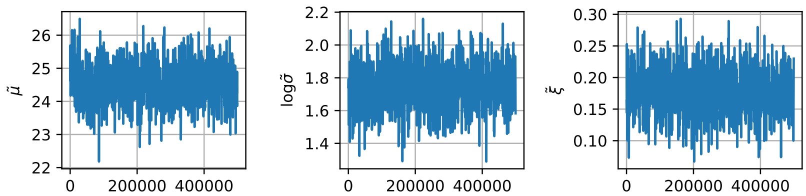
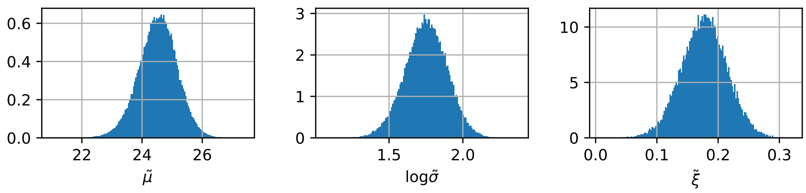
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 3^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 0.45^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.25^2)$
Initialisation	25, 2, 0.2
Sample size	500,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.252, 0.227, 0.111

Table 3: Metropolis-Hastings algorithm for simulation study with $k = 2$ and independent copula posterior.

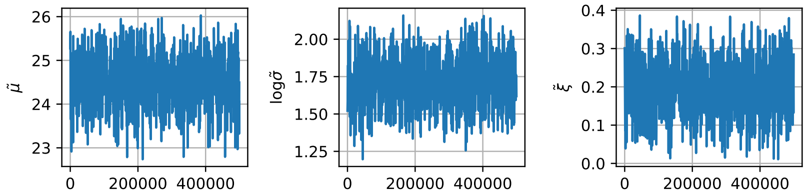
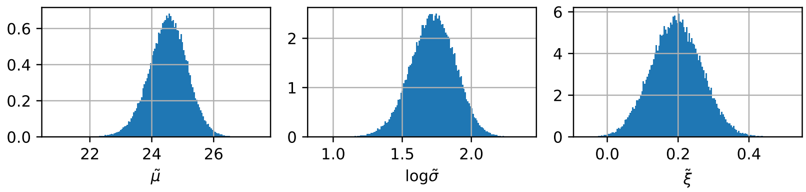
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 3^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 0.5^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.4^2)$
Initialisation	25, 2, 0.2
Sample size	$500{,}000$ (after burn-in period of $1{,}000$)
Traceplots	
Histograms	
Acceptance rates	0.247, 0.246, 0.141

Table 4: Metropolis-Hastings algorithm for simulation study with $k = 3$ and maximum entropy copula prior.

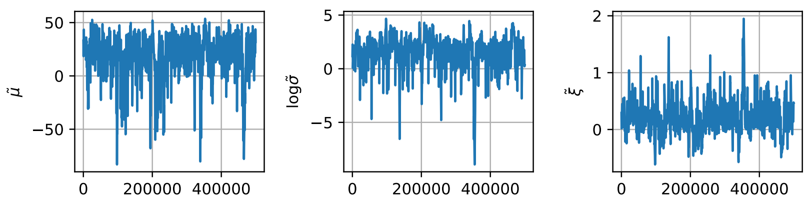
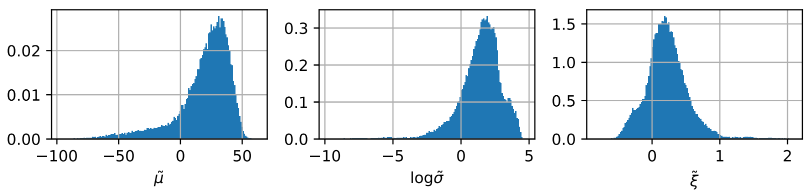
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 25^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 0.5^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.1^2)$
Initialisation	25, 2, 0
Sample size	$500{,}000$ (after burn-in period of $1{,}000$)
Traceplots	
Histograms	
Acceptance rates	0.191, 0.232, 0.302

Table 5: Metropolis algorithm for simulation study with $k = 3$ and maximum entropy copula posterior.

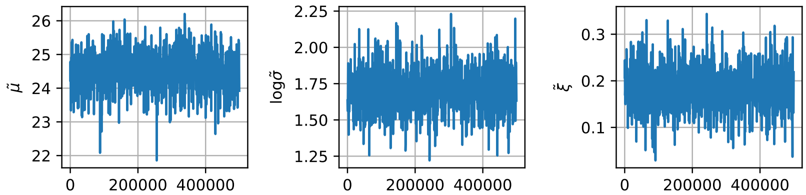
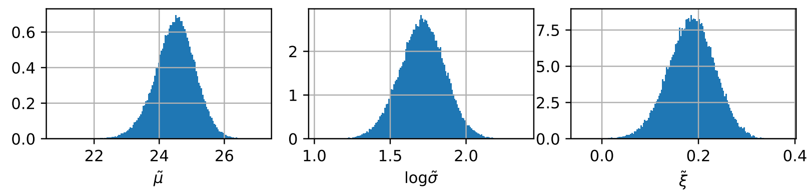
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 3^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 0.4^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.15^2)$
Initialisation	25, 2, 0.2
Sample size	500,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.242, 0.232, 0.204

Table 6: Metropolis-Hastings algorithm for simulation study with $k = 2$ and maximum entropy copula posterior.

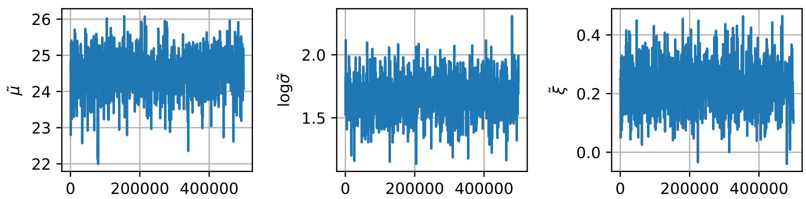
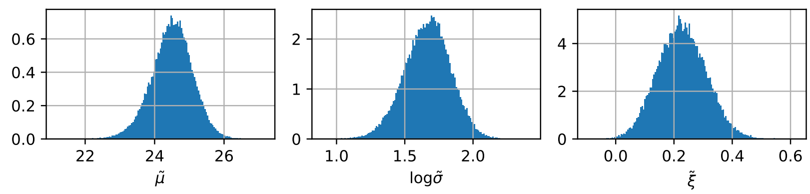
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 3^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 0.5^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.4^2)$
Initialisation	25, 2, 0.2
Sample size	500,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.234, 0.238, 0.152

Table 7: Metropolis-Hastings algorithm for wind speed data with $k = 3$ and independent copula prior.

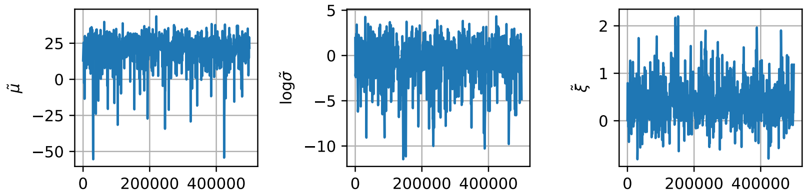
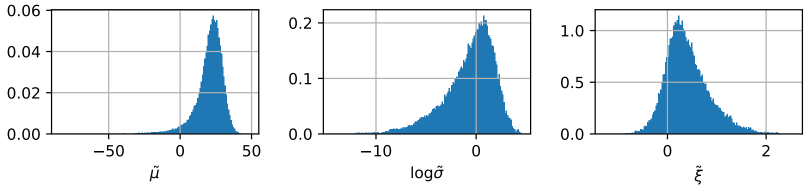
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 25^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 2^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.6^2)$
Initialisation	25, 2, 0
Sample size	$500{,}000$ (after burn-in period of $1{,}000$)
Traceplots	
Histograms	
Acceptance rates	0.25, 0.18, 0.124

Table 8: Metropolis algorithm for wind speed data with $k = 3$ and independent copula posterior.

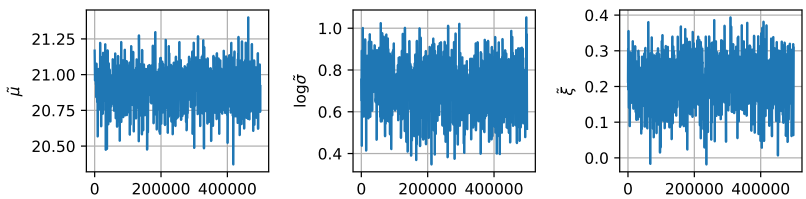
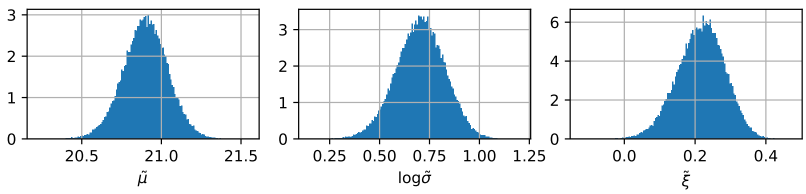
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 1^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 0.45^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.25^2)$
Initialisation	22.5, 1, 0.2
Sample size	$500{,}000$ (after burn-in period of $1{,}000$)
Traceplots	
Histograms	
Acceptance rates	0.172, 0.219, 0.217

Table 9: Metropolis-Hastings algorithm for wind speed data with $k = 2$ and independent copula posterior.

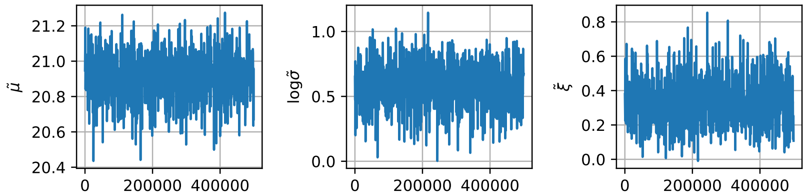
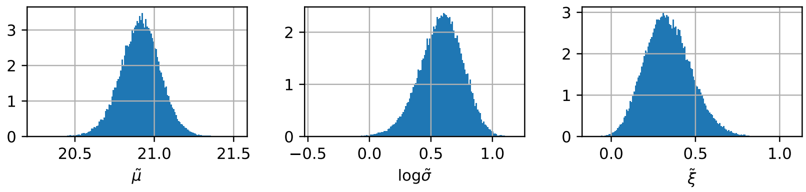
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 1^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 0.5^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.4^2)$
Initialisation	22.5, 1, 0.3
Sample size	$500{,}000$ (after burn-in period of $1{,}000$)
Traceplots	
Histograms	
Acceptance rates	0.152, 0.221, 0.218

Table 10: Metropolis-Hastings algorithm for wind speed data with $k = 3$ and maximum entropy copula prior.

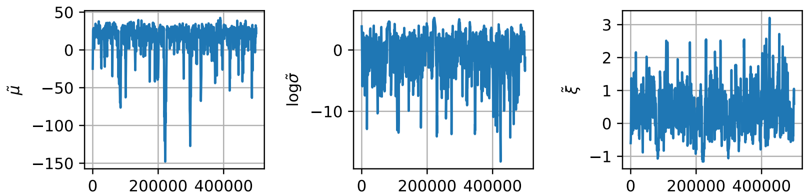
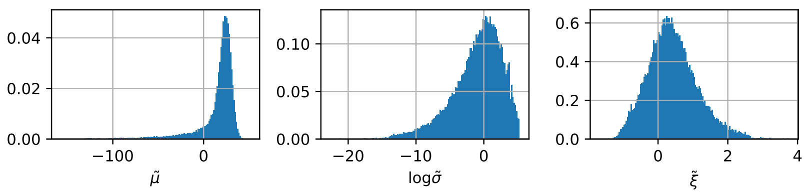
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 25^2)$, $\log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 2^2)$, $\tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.8^2)$
Initialisation	25, 2, 0
Sample size	$500{,}000$ (after burn-in period of $1{,}000$)
Traceplots	
Histograms	
Acceptance rates	0.231, 0.194, 0.119

Table 11: Metropolis algorithm for wind speed data with $k = 3$ and maximum entropy copula posterior.

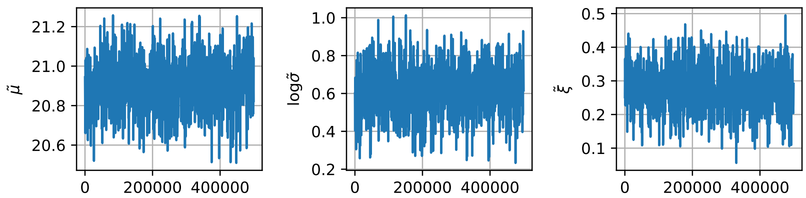
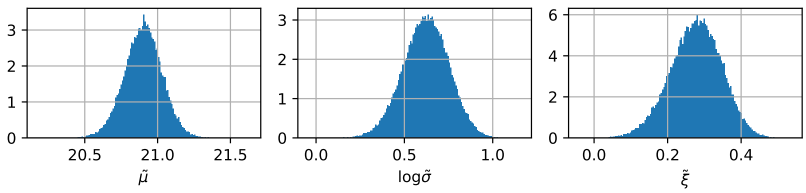
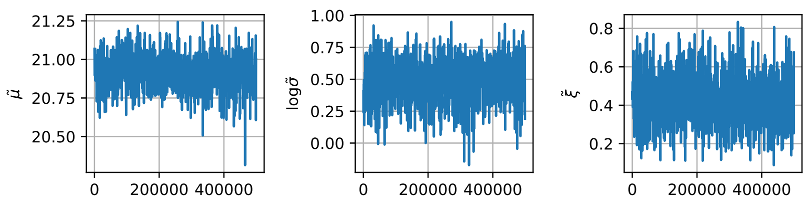
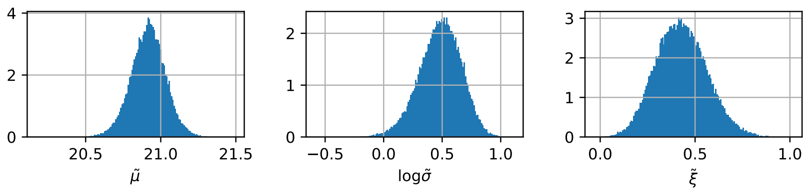
Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 1^2), \quad \log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 0.4^2), \quad \tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.3^2)$
Initialisation	$22.5, \quad 1, \quad 0.1$
Sample size	$500{,}000$ (after burn-in period of $1{,}000$)
Traceplots	
Histograms	
Acceptance rates	$0.159, \quad 0.246, \quad 0.184$

Table 12: Metropolis-Hastings algorithm for wind speed data with $k = 2$ and maximum entropy copula posterior.

Proposal distributions	$\tilde{\mu}^* \sim \mathcal{N}(\tilde{\mu}, 1^2), \quad \log\tilde{\sigma}^* \sim \mathcal{N}(\log\tilde{\sigma}, 0.5^2), \quad \tilde{\xi}^* \sim \mathcal{N}(\tilde{\xi}, 0.4^2)$
Initialisation	$22.5, \quad 1, \quad 0.4$
Sample size	$500{,}000$ (after burn-in period of $1{,}000$)
Traceplots	
Histograms	
Acceptance rates	$0.138, \quad 0.23, \quad 0.22$

References

[1] Michaël Baudin, Anne Dutfoy, Bertrand Iooss, and Anne-Laure Popelin. *OpenTURNS: An industrial software for uncertainty quantification in simulation.* 2015.

[2] Cristina Butucea, Jean-François Delmas, Anne Dutfoy, and Richard Fischer. "Maximum entropy distribution of order statistics with given marginals". *Bernoulli* 24.1 (2018). DOI: 10.3150/16-BEJ868.

[3] Stuart G. Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics. London: Springer, 2001. ISBN: 978-1-84996-874-4.

[4] Stuart G. Coles and Elwyn A. Powell. "Bayesian Methods in Extreme Value Modelling: A Review and New Developments". *International Statistical Review / Revue Internationale de Statistique* 64.1 (1996), pp. 119–136. DOI: 10.2307/1403426.

[5] Stuart G. Coles and Jonathan A. Tawn. "A Bayesian Analysis of Extreme Rainfall Data". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 45.4 (1996), pp. 463–478. DOI: 10.2307/2986068.

[6] John D. Hunter. "Matplotlib: A 2D graphics environment". *Computing in science & engineering* 9.3 (2007), pp. 90–95.

[7] Edwin T. Jaynes. "Information Theory and Statistical Mechanics". *Phys. Rev.* 106 (4 1957), pp. 620–630. DOI: 10.1103/PhysRev.106.620.

[8] Harold Jeffreys. "An invariant form for the prior probability in estimation problems". *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 186.1007 (1946), pp. 453–461. DOI: 10.1098/rspa.1946.0056.

[9] Lynn Kuo and Bani Mallick. "Variable Selection for Regression Models". *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 60.1 (1998), pp. 65–81.

[10] Ulrich K. Müller. "Measuring prior sensitivity and prior informativeness in large Bayesian models". *Journal of Monetary Economics* 59.6 (2012), pp. 581–597. DOI: 10.1016/j.jmoneco.2012.09.003.

[11] Paul J. Northrop and Nicolas Attalides. "Posterior propriety in Bayesian extreme value analyses using reference priors". *Statistica Sinica* 26.2 (2016), pp. 721–743. DOI: 10.2307/24721296.

[12] Travis E. Oliphant. *A guide to NumPy.* Vol. 1. Trelgol Publishing USA, 2006.

[13] Matthew Reimherr, Xiao-Li Meng, and Dan L. Nicolae. *Prior sample size extensions for assessing prior informativeness and prior–likelihood discordance.* 2014. arXiv: 1406.5958 [stat.ME].

[14] Gareth O. Roberts and Adrian F. M. Smith. "Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms". *Stochastic Processes and their Applications* 49.2 (1994), pp. 207–216. DOI: 10.1016/0304-4149(94)90134-1.

[15] Claude E. Shannon. "A Mathematical Theory of Communication". *Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[16] Paul Sharkey and Jonathan A. Tawn. "A Poisson process reparameterisation for Bayesian inference for extremes". *Extremes* 20 (2017), pp. 239–263. DOI: 10.1007/s10687-016-0280-2.

[17] Stuart G. Coles, Luis Pericchi, and Scott Sisson. "A fully probabilistic approach to extreme rainfall modeling". *Journal of Hydrology* 273 (2003), pp. 35–50. DOI: 10.1016/S0022-1694(02)00353-0.

[18] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.