Report version 11

Tony Zheng

1ˢᵗ 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 Tawn 1996 to predict annual maximum rainfall. In the latter, Coles and Tawn employ the Poisson point process model, and elicit prior information specified by an expert in the domain. In this paper, we will attempt to replicate the analysis of Coles and Tawn 1996 using alternative prior distributions. We will construct these distributions with the principle of maximum entropy in mind, which results in more uninformative priors. This will account for uncertainty due to lack of information in the model. In particular, we will propose a novel technique in prior elicitation in which a maximum entropy joint distribution is constructed from given marginals.

In § 2, we will define the distributions which we will use throughout the report. In § 3, we will outline the Poisson point process model. In § 4, we will describe the original prior distribution used by Coles and Tawn 1996, and propose alternative prior distributions. In § 5, we will explain how Markov chain Monte Carlo (MCMC) methods can be used in the analysis to avoid intractable calculations. In § 6, we will implement and compare our priors in simulation studies and on real wind speed data. Finally, in the Appendix we have supplementary calculations and details of our MCMC implementations.

2 Named distributions

2.1 Generalised extreme value distribution

The generalised extreme value (GEV) has 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 $\{x\}_+ := \max\{0, x\}$, with parameters $\theta := (\mu, \sigma, \xi)$ defined in the set

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

2.2 Gamma distribution

A random variable X follows a Gamma distribution with shape $\alpha \in \mathbb{R}^+$ and rate $\beta \in \mathbb{R}^+$ if it has PDF

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)} \mathbb{1}_{x>0} \,,$$

where Γ is the gamma function. This is denoted $X \sim \Gamma(\alpha, \beta)$.

2.3 Truncated normal distribution

Suppose that Y is a random variable which follows a normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma \in \mathbb{R}^+$. For $\infty \leq a < b \leq \infty$, the random variable $X := Y \mid Y \in (a, b)$ follows a truncated normal distribution with PDF

$$f(x) = \frac{1}{\sigma} \frac{\phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \mathbb{1}_{a<x<b} \,,$$

where ϕ and Φ are the PDF and CDF of the standard normal distribution respectively. This is denoted $X \sim \mathrm{TN}(\mu, \sigma, a, b)$. The parameters are μ and σ, the parent mean and standard deviation respectively, and the (possibly infinite) bounds a and b.

2.4 Exponential distribution

A random variable X follows an exponential distribution with rate $\lambda \in \mathbb{R}^+$ if it has PDF

$$f(x) = \lambda \exp(-\lambda x) \mathbb{1}_{x>0}$$

This is denoted $X \sim \mathrm{Exp}(\lambda)$.

3 Likelihood model

Extreme values are often modelled using the GEV distribution. Given a sequence of i.i.d. variables $X_1, \ldots, X_n$, under certain conditions, its maximum

$$M_n := \max\{X_1, \ldots, X_n\}$$

asymptotically follows a GEV distribution as $n \to +\infty$.

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. We will ignore the asymptotic nature of the model. Suppose that we have daily observations of X over M years. Denote by $\mathbf{x} = (x_1, \ldots, x_{n_u})$ 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

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

and

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

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

This is an extension of the GEV model, as the annual maxima M_{365} follow a GEV distribution with parameters θ.

Inverting (1), we obtain the quantile function

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

Of particular interest to us are the quantiles corresponding to very small probabilities. The *return level* of a *return period* r in years is defined as the quantile $q(1 - 1/r \mid \theta)$.

4   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. Unfortunately, it is not feasible to specify joint distributions for the parameters θ. 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

$$\begin{aligned} q_i &:= q(1 - p_i \mid \theta), \\ &= \mu + \frac{\sigma}{\xi}(\exp(s_i \xi) - 1), \end{aligned} \tag{8}$$

where q is defined in (7) and $s_i := -\log(-\log(1 - p_i))$. We are assuming that $\xi \neq 0$. We also assume that

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

which implies that $s_i > 0$ for $i = 1, 2, 3$. 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.

As the data are assumed to be strictly positive, the support of the quantiles implies that a distribution for (q_1, q_2, q_3) is a valid prior if and only if

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

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

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

and construct a joint prior for the quantile differences by assuming that the marginals are independent. In particular, for each $\tilde{q}_i$, the expert provided 0.5-quantile and 0.9-quantile estimates of the marginal distribution, and Gamma distributions were fitted to these specifications. We will denote this prior π^{G3}.

The choice of a distribution for the quantile differences can be motivated using 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

$$\mathcal{E}(p) = -\int_{x:\,p(x)>0} \log_2(p(x))\,\mathrm{d}x\,.$$

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. As quantile differences are positive, we are looking for distributions with support $(0, +\infty)$, and functional constraints defined by certain quantities which are fixed by an expert. The existence of maximum entropy distributions for various fixed quantities is shown in Appendix A, and the results are summarised below.

Fixed quantities	Maximum entropy distribution on $(0, +\infty)$
Any number of quantiles	Does not exist
1ˢᵗ moment	Exponential
1ˢᵗ, 2ⁿᵈ moments	Truncated normal

Specifying 0.5 and 0.9-quantiles, as in π^{G3}, does not have a maximum entropy solution. However, specifying the mean leads to an exponential distribution, and specifying the mean and variance leads to a truncated normal distribution. We will denote these alternative priors π^{E} and π^{TN} respectively.

Although we have thus far been considering an independent copula, the choice of copula is also one which can be made using the principle of maximum entropy. We will consider a prior which is the joint distribution with maximum entropy for fixed marginals q_1, q_2, q_3 satisfying $\Pr(q_1 < q_2 < q_3) = 1$. This is equivalent to using a copula with maximum entropy, and we will denote this prior π^{MEC} for Maximum Entropy Copula (Butucea, Delmas, Dutfoy, and Fischer 2018). The order constraint allows us to specify priors on the quantiles instead of the quantile differences, which may be more convenient for the expert.

In the case that we are only able to elicit information on two quantile differences, we will consider a prior π^{G2} which includes an improper uniform prior on $\log(\sigma)$, as well as two Gamma distributed priors on the quantile differences. With only a single quantile difference, or equivalently a single quantile, we will consider a prior π^{G1} which includes an improper uniform prior on $\log(\sigma)$, an improper uniform prior on μ, as well as a Gamma distributed prior on the quantile. As the maximum entropy distribution with a compact support $[a, b]$ is uniform, an improper uniform prior may be thought of as the maximum entropy distribution when a and b tend to $-\infty$ and $+\infty$ respectively.

In order to compare these different priors, we will first construct π^{G3} by placing priors on the quantile differences $(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3)$. As we do not have access to expert specification of the 0.5 and 0.9-quantiles, we will specify the Gamma distributions directly, for instance by first estimating the quantile differences and then centring Gamma distributions around these estimations. For π^{MEC}, Gamma prior distributions on the quantiles can be obtained by approximating the marginals of π_q^{G3} with Gamma distributions. This can be done using numerical minimisation of the Kullback–Leibler divergence. The construction of a maximum entropy copula requires a certain condition $(C2)$ to be satisfied (see Appendix C) during this minimisation. For π^{G2}

and π^{G1}, we will consider only the first two quantile differences. The mean and variance of the priors for the quantile differences will be used to construct π^{E} and π^{TN}. The mean of an exponential distribution with rate parameter λ is simply λ^{-1}. Thus, λ can easily be obtained from μ. In Appendix B, we show how to obtain the parameters of a truncated distribution numerically from the mean and variance.

A summary of the various priors is tabulated in Table 1, and in the rest of this section we will derive the PDFs of each prior.

In Appendix D.1, we propose a variant of G3 in which ξ has a non-zero probability to be zero.

4.1   G3: Gamma prior distributions for three quantile differences

Gamma prior distributions are specified for three quantile differences

$$\tilde{q}_i \sim \Gamma(\alpha_i, \beta_i), \quad i = 1, 2, 3,$$

and the joint distribution $\pi^{\text{G3}}_{\tilde{q}}$ is formed by assuming that the quantile differences are independent. This is then converted into a prior for the quantiles π^{G3}_q using the transformation

$$(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3) \mapsto (\tilde{q}_1, \tilde{q}_1 + \tilde{q}_2, \tilde{q}_1 + \tilde{q}_2 + \tilde{q}_3)$$
$$= (q_1, q_2, q_3).$$

Explicitly, the prior for (q_1, q_2, q_3) is given by

$$\pi^{\text{G3}}_q(q_1, q_2, q_3) = C q_1^{\alpha_1 - 1} \exp(-\beta_1 q_1)$$
$$\times \prod_{i=2}^{3} (q_i - q_{i-1})^{\alpha_i - 1} \exp(-\beta_i(q_i - q_{i-1})) \mathbb{1}_{0 < q_1 \leq q_2 \leq q_3}, \quad (9)$$

where

$$C := \prod_{i=1}^{3} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}.$$

Next, we convert this into a prior for (μ, σ, ξ) using the invertible transformation

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

given by (8):

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

	Marginal distributions of parameters	Expert specification of $\tilde{q}_i / q_i$	Copula	
G3	$\tilde{q}_1, \tilde{q}_2, \tilde{q}_3 \sim \Gamma$	Two quantiles	Independent	§ 4.1
G2	$\tilde{q}_1, \tilde{q}_2 \sim \Gamma, \quad \log \sigma \propto 1$	Two quantiles	Independent	§ 4.2
G1	$q_1 \sim \Gamma, \quad \mu, \log \sigma \propto 1$	Two quantiles	Independent	§ 4.3
MEC	$q_1, q_2, q_3 \sim \Gamma$	Two quantiles	Maximum entropy	§ 4.4
E	$\tilde{q}_1, \tilde{q}_2, \tilde{q}_3 \sim \text{Exp}$	Mean	Independent	§ 4.5
TN	$\tilde{q}_1, \tilde{q}_2, \tilde{q}_3 \sim \text{TN}$	Mean and variance	Independent	§ 4.6

Table 1: Summary of the various priors

We have that

$$\frac{\mathrm{d}q_i}{\mathrm{d}\mu}(\theta) = 1 \,, \tag{10}$$

$$\frac{\mathrm{d}q_i}{\mathrm{d}\sigma}(\theta) = \frac{\exp(s_i\xi) - 1}{\xi} \,, \tag{11}$$

$$\frac{\mathrm{d}q_i}{\mathrm{d}\xi}(\theta) = \frac{\sigma(s_i\xi\exp(s_i\xi) - \exp(s_i\xi) + 1)}{\xi^2} \,, \tag{12}$$

and so the determinant of the Jacobian is

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

$$\begin{matrix} 1 \\ \exp(s_3\xi) - 1 \\ s_3\xi\exp(s_3\xi) - \exp(s_3\xi) + 1 \end{matrix}$$

$$= \frac{\sigma}{\xi^3} \begin{vmatrix} 1 & 1 & 1 \\ \exp(s_1\xi) & \exp(s_2\xi) & \exp(s_3\xi) \\ s_1\xi\exp(s_1\xi) & s_2\xi\exp(s_2\xi) & s_3\xi\exp(s_3\xi) \end{vmatrix}$$

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

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

$$= -\exp\left(\xi\sum_{i=1}^{3} s_i\right) \frac{\sigma}{\xi^2} (\exp(-s_2\xi)s_3 - \exp(-s_3\xi)s_2$$

$$+ \exp(-s_3\xi)s_1 - \exp(-s_1\xi)s_3$$

$$+ \exp(-s_1\xi)s_2 - \exp(-s_2\xi)s_1) \,.$$

If we set

$$\Psi_3(\xi) := \frac{|\det J(g_3)(\theta)|}{\sigma} \,, \tag{13}$$

then using the change of variables formula,

$$\pi_\theta^{\mathrm{G3}}(\theta) = \pi_q^{\mathrm{G3}}(g_3(\theta))\sigma\Psi_3(\xi)\mathbb{1}_{\sigma>0} \,. \tag{14}$$

Substituting (9) into (14) gives us

$$\pi_\theta^{\mathrm{G3}}(\theta) = C(q_1(\theta))^{\alpha_1-1}\exp\left(-\beta_1 q_1(\theta)\right) \prod_{i=2}^{3} \left(\sigma\frac{\exp(s_i\xi) - \exp(s_{i-1}\xi)}{\xi}\right)^{\alpha_i-1}$$

$$\times \exp\left(-\beta_i\sigma\frac{\exp(s_i\xi) - \exp(s_{i-1}\xi)}{\xi}\right)\sigma\Psi_3(\xi)\mathbb{1}_{q_1(\theta)>0}\mathbb{1}_{\sigma>0}$$

$$= C(q_1(\theta))^{\alpha_1-1}\exp\left(-\beta_1 q_1(\theta)\right)\sigma^{\alpha_2+\alpha_3-1}\prod_{i=2}^{3} \left(\frac{\exp(s_i\xi) - \exp(s_{i-1}\xi)}{\xi}\right)^{\alpha_i-1}$$

$$\times \exp\left(-\frac{\sigma}{\xi}\sum_{i=2}^{3}\beta_i(\exp(s_i\xi) - \exp(s_{i-1}\xi))\right)\Psi_3(\xi)\mathbb{1}_{q_1(\theta)>0}\mathbb{1}_{\sigma>0}$$

$$= C(q_1(\theta))^{\alpha_1-1} \exp\left(-\beta_1 q_1(\theta)\right) \sigma^{\alpha_2+\alpha_3-1} \Phi_3(\xi) \exp\left(-\sigma\Omega_3(\xi)\right) \mathbb{1}_{q_1(\theta)>0} \mathbb{1}_{\sigma>0}, \quad (15)$$

where

$$\Phi_3(\xi) := \prod_{i=2}^{3} \left(\frac{\exp(s_i\xi) - \exp(s_{i-1}\xi)}{\xi}\right)^{\alpha_i-1} \Psi_3(\xi),$$

$$\Omega_3(\xi) := \frac{1}{\xi} \sum_{i=2}^{3} \beta_i(\exp(s_i\xi) - \exp(s_{i-1}\xi)) > 0.$$

The marginal of (σ, ξ) is

$$\pi_{\sigma,\xi}^{\mathrm{G3}}(\sigma, \xi) = \int_{-\infty}^{\infty} \pi_{\theta}^{\mathrm{G3}}(\mu, \sigma, \xi) \, \mathrm{d}\mu$$

$$= C \int_{0}^{\infty} q_1^{\alpha_1-1} \exp\left(-\beta_1 q_1\right) \sigma^{\alpha_2+\alpha_3-1} \Phi_3(\xi) \exp\left(-\sigma\Omega_3(\xi)\right) \mathbb{1}_{\sigma>0} \, \mathrm{d}q_1$$

$$= C\sigma^{\alpha_2+\alpha_3-1} \Phi_3(\xi) \exp\left(-\sigma\Omega_3(\xi)\right) \beta_1^{-\alpha_1}$$

$$\times \int_{0}^{\infty} (\beta_1 q_1)^{\alpha_1-1} \exp\left(-\beta_1 q_1\right) \beta_1 \, \mathrm{d}q_1 \mathbb{1}_{\sigma>0}$$

$$= C\sigma^{\alpha_2+\alpha_3-1} \Phi_3(\xi) \exp\left(-\sigma\Omega_3(\xi)\right) \beta_1^{-\alpha_1} \Gamma(\alpha_1) \mathbb{1}_{\sigma>0}$$

$$= C_2 \sigma^{\alpha_2+\alpha_3-1} \Phi_3(\xi) \exp\left(-\sigma\Omega_3(\xi)\right) \mathbb{1}_{\sigma>0}, \quad (16)$$

where

$$C_2 = \frac{\beta_2^{\alpha_2} \beta_3^{\alpha_3}}{\Gamma(\alpha_2)\Gamma(\alpha_3)}.$$

Thus,

$$\pi_{\sigma|\xi}^{\mathrm{G3}}(\sigma \mid \xi) \sim \Gamma(\alpha_2 + \alpha_3, \Omega_3(\xi)),$$

and the marginal of ξ is

$$\pi_{\xi}^{\mathrm{G3}}(\xi) = \int_{-\infty}^{\infty} \pi_{\sigma,\xi}^{\mathrm{G3}}(\sigma, \xi) \, \mathrm{d}\sigma$$

$$= C_2 \int_{0}^{\infty} \sigma^{\alpha_2+\alpha_3-1} \Phi_3(\xi) \exp\left(-\sigma\Omega_3(\xi)\right) \, \mathrm{d}\sigma$$

$$= C_2 \Phi_3(\xi) \Omega_3(\xi)^{-\alpha_2-\alpha_3}$$

$$\times \int_{0}^{\infty} (\sigma\Omega_3(\xi))^{\alpha_2+\alpha_3-1} \exp\left(-\sigma\Omega_3(\xi)\right) \Omega_3(\xi) \, \mathrm{d}\sigma$$

$$= C_2 \Phi_3(\xi) \Omega_3(\xi)^{-\alpha_2-\alpha_3} \Gamma(\alpha_2 + \alpha_3)$$

$$= C_3 \frac{\Phi_3(\xi)}{\Omega_3(\xi)^{\alpha_2+\alpha_3}}, \quad (17)$$

where

$$C_3 = \frac{\Gamma(\alpha_2 + \alpha_3)\beta_2^{\alpha_2} \beta_3^{\alpha_3}}{\Gamma(\alpha_2)\Gamma(\alpha_3)}.$$

4.2   G2: Gamma prior distributions for two quantile differences

Gamma prior distributions are specified for two quantile differences

$$\tilde{q}_i \sim \Gamma(\alpha_i, \beta_i), \quad i = 1, 2,$$

and an improper uniform prior is put on $\log(\sigma)$. That is,

$$\pi_\sigma^{\mathrm{G2}}(\sigma) \propto \frac{1}{\sigma} \mathbb{1}_{\sigma>0}\,,$$

$$\pi_{q_1,q_2|\sigma}^{\mathrm{G2}}(q_1, q_2 \mid \sigma) \propto q_1^{\alpha_1-1} \exp(-\beta_1 q_1)$$
$$\times (q_2 - q_1)^{\alpha_2-1} \exp(-\beta_2(q_2 - q_1)) \mathbb{1}_{0<q_1\le q_2}$$
$$\implies \pi_{\sigma,q_1,q_2}^{\mathrm{G2}}(\sigma, q_1, q_2) \propto q_1^{\alpha_1-1} \exp(-\beta_1 q_1)$$
$$\times (q_2 - q_1)^{\alpha_2-1} \exp(-\beta_2(q_2 - q_1)) \frac{1}{\sigma} \mathbb{1}_{\sigma>0} \mathbb{1}_{0<q_1\le q_2}\,. \tag{18}$$

We convert this into a prior for θ using the invertible transformation

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

given by (8):

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

From the partial derivatives in (10)–(12), the determinant of the Jacobian is

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

$$= \frac{\sigma}{\xi^2} \begin{vmatrix} 1 & 1 \\ s_2\xi\exp(s_2\xi) - \exp(s_2\xi) + 1 & s_1\xi\exp(s_1\xi) - \exp(s_1\xi) + 1 \end{vmatrix}$$

$$= \frac{\sigma}{\xi^2} ((s_1\xi - 1)\exp(s_1\xi) - (s_2\xi - 1)\exp(s_2\xi))\,.$$

If we set

$$\Psi_2(\xi) := \frac{|\det J(g_2)(\theta)|}{\sigma}\,,$$

then using the change of variables formula,

$$\pi_\theta^{\mathrm{G2}}(\theta) = \pi_q(g_2(\theta))\sigma\Psi_2(\xi) \mathbb{1}_{\sigma>0}\,.$$

Substituting in (18),

$$\pi_\theta^{\mathrm{G2}}(\theta) \propto (q_1(\theta))^{\alpha_1-1} \exp(-\beta_1 q_1(\theta)) \sigma^{\alpha_2-1} (\Omega_2(\xi))^{\alpha_2-1}$$
$$\times \exp\left(-\beta_2\sigma\Omega_2(\xi)\right) \Psi_2(\xi) \mathbb{1}_{\sigma,q_1(\theta)>0}\,.$$

where

$$\Omega_2(\xi) := \frac{\exp(s_2\xi) - \exp(s_1\xi)}{\xi}\,.$$

The marginal of (σ, ξ) is

$$\pi_{\sigma,\xi}^{\mathrm{G2}}(\sigma, \xi) = \int_{-\infty}^{\infty} \pi_\theta^{\mathrm{G2}}(\mu, \sigma, \xi)\,\mathrm{d}\mu$$

$$\propto \int_0^{\infty} q_1^{\alpha_1-1} \exp\left(-\beta_1 q_1\right) \sigma^{\alpha_2-1} (\Omega_2(\xi))^{\alpha_2-1} \exp\left(-\beta_2\sigma\Omega_2(\xi)\right) \Psi_2(\xi) \mathbb{1}_{\sigma>0}\,\mathrm{d}q_1$$

$$\propto \sigma^{\alpha_2-1} \left(\Omega_2(\xi)\right)^{\alpha_2-1} \exp\left(-\beta_2\sigma\Omega_2(\xi)\right)$$
$$\times \int_0^\infty \left(\beta_1 q_1\right)^{\alpha_1-1} \exp\left(-\beta_1 q_1\right) \beta_1 \, \mathrm{d}q_1 \, \Psi_2(\xi) \mathbb{1}_{\sigma>0}$$
$$\propto \left(\sigma\Omega_2(\xi)\right)^{\alpha_2-1} \exp\left(-\beta_2\sigma\Omega_2(\xi)\right) \Psi_2(\xi) \mathbb{1}_{\sigma>0}\,, \tag{19}$$

Thus,

$$\pi_{\sigma|\xi}^{\mathrm{G2}}(\sigma \mid \xi) \sim \Gamma(\alpha_2, \beta_2\Omega_3(\xi))\,,$$

and the marginal of ξ is

$$\pi_\xi^{\mathrm{G2}}(\xi) = \int_{-\infty}^\infty \pi_{\sigma,\xi}^{\mathrm{G2}}(\sigma,\xi) \, \mathrm{d}\sigma$$
$$\propto \int_0^\infty \left(\sigma\Omega_2(\xi)\right)^{\alpha_2-1} \exp\left(-\beta_2\sigma\Omega_2(\xi)\right) \Psi_2(\xi) \, \mathrm{d}\sigma$$
$$\propto \Psi_2(\xi) \left(\sigma\Omega_2(\xi)\right)^{-1} \int_0^\infty \left(\beta_2\Omega_2(\xi)\right)^{\alpha_2-1} \exp\left(-\beta_2\Omega_2(\xi)\right) \beta_2\sigma\Omega_2(\xi) \, \mathrm{d}\sigma$$
$$\propto \Psi_2(\xi) \left(\Omega_2(\xi)\right)^{-1}$$
$$= \frac{\xi}{\xi^2} \frac{\left|(s_2\xi-1)\exp(s_2\xi) - (s_1\xi-1)\exp(s_1\xi)\right|}{\exp(s_2\xi) - \exp(s_1\xi)}$$
$$= \frac{1}{\xi} \frac{\left|s_2\xi\exp(s_2\xi) - s_1\xi\exp(s_1\xi) - \exp(s_2\xi) + \exp(s_1\xi)\right|}{\exp(s_2\xi) - \exp(s_1\xi)}$$
$$= \left|\frac{s_2\xi\exp(s_2\xi) - s_1\xi\exp(s_1\xi) - \exp(s_2\xi) + \exp(s_1\xi)}{\xi(\exp(s_2\xi) - \exp(s_1\xi))}\right|$$
$$= \left|\frac{s_2\exp(s_2\xi) - s_1\exp(s_1\xi)}{\exp(s_2\xi) - \exp(s_1\xi)} - \frac{1}{\xi}\right|\,.$$

Since

$$0 < s_1 < s_2\,,$$

as $\xi \to +\infty$,

$$\pi_\xi^{\mathrm{G2}}(\xi) \propto \left|\frac{s_2 - s_1\exp((s_1-s_2)\xi)}{1 - \exp((s_1-s_2)\xi)} - \frac{1}{\xi}\right|$$
$$\to s_2\,.$$

Similarly, as $\xi \to +\infty$, $\pi_\xi^{\mathrm{G2}}(\xi) \to C + s_1$ for some constant C. Therefore the prior is indeed improper.

The posterior, with a single observation x, is written

$$\pi_{\theta|x}^{\mathrm{G2}}(\theta \mid x) \propto \left(q_1(\theta)\right)^{\alpha_1-1} \exp(-\beta_1 q_1(\theta)) \sigma^{\alpha_2-1} \left(\Omega_2(\xi)\right)^{\alpha_2-1}$$
$$\times \exp\left(-\beta_2\sigma\Omega_2(\xi)\right) \Psi_2(\xi)$$
$$\times \exp\left(-M\left\{1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right\}_+^{-\frac{1}{\xi}}\right)$$
$$\times \frac{1}{\sigma}\left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}_+^{-\frac{\xi+1}{\xi}} \mathbb{1}_{\sigma, q_1(\theta)>0}\,.$$

In order to show that this is proper, we need to show that the following is integrable:

$$\pi_{\sigma|x,\mu,\xi}^{\mathrm{G2}}(\sigma \mid x, \mu, \xi) \propto \left(q_1(\theta)\right)^{\alpha_1-1} \exp(-\beta_1 q_1(\theta)) \sigma^{\alpha_2-2} \exp\left(-\beta_2\sigma(\Omega_2(\xi)\right)$$

$$\times \exp\left(-M\left\{1 + \xi\left(\frac{u - \mu}{\sigma}\right)\right\}_+^{-\frac{1}{\xi}}\right)$$

$$\times \left\{1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right\}_+^{-\frac{\xi+1}{\xi}} \mathbb{1}_{\sigma, q_1(\theta) > 0}.$$

Northrop and Attalides 2016 show that for a GEV likelihood model with at least 2 observations, if uniform priors are specified for μ and $\log\sigma$, and a proper prior is placed on ξ, then the joint prior will yield a proper posterior.

4.3   G1: Gamma prior distribution for one quantile

Gamma prior distributions are specified for a quantile q_1

$$\tilde{q}_1 \sim \Gamma(\alpha_1, \beta_1),$$

and improper uniform priors are put on μ and $\log(\sigma)$. That is,

$$\pi_\mu^{\mathrm{G1}}(\mu) \propto 1,$$
$$\pi_\sigma^{\mathrm{G1}}(\sigma) \propto \frac{1}{\sigma}\mathbb{1}_{\sigma > 0},$$
$$\pi_{q_1|\mu,\sigma}^{\mathrm{G1}}(q_1 \mid \mu, \sigma) \propto q_1^{\alpha_1 - 1} \exp(-\beta_1 q_1)\mathbb{1}_{0 < q_1}$$
$$\implies \pi_{\mu,\sigma,q_1}^{\mathrm{G1}}(\mu, \sigma, q_1) \propto q_1^{\alpha_1 - 1} \exp(-\beta_1 q_1)\frac{1}{\sigma}\mathbb{1}_{\sigma > 0}\mathbb{1}_{0 < q_1 \leq q_2}. \tag{20}$$

We convert this into a prior for θ using the inverse transformation

$$g_1 : \Theta \to \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}$$
$$g_1 : (\theta) \mapsto (\mu, \sigma, q_1)$$

given by (8):

$$q_1 = \mu + \sigma\frac{\exp(s_1\xi) - 1}{\xi}.$$

From the partial derivatives in (10)–(12), the determinant of the Jacobian is

$$\det J(g_1)(\theta) = \frac{\sigma}{\xi^2}\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & (\exp(s_1\xi) - 1)\xi^{-1} \\ 0 & 0 & s_2\xi\exp(s_1\xi) - \exp(s_1\xi) + 1 \end{vmatrix}$$
$$= \frac{\sigma}{\xi^2}((s_1\xi - 1)\exp(s_1\xi) + 1).$$

If we set

$$\Psi_1(\xi) := \frac{|\det J(g_1)(\theta)|}{\sigma},$$

then using the change of variables formula,

$$\pi_\theta^{\mathrm{G1}}(\theta) = \pi_q(g_1(\theta))\sigma\Psi_1(\xi)\mathbb{1}_{\sigma > 0}.$$

Substituting in (20),

$$\pi_\theta^{\mathrm{G1}}(\theta) \propto (q_1(\theta))^{\alpha_1 - 1}\exp(-\beta_1 q_1(\theta))\Psi_1(\xi)\mathbb{1}_{\sigma > 0}\mathbb{1}_{q_1(\theta) > 0}.$$

4.4   MEC: Gamma prior distributions for three quantiles with maximum entropy copula

Marginal prior distributions are specified for three quantiles q_1, q_2, q_3 directly. In particular, we choose Gamma distributions

$$q_i \sim \Gamma(\alpha_i, \beta_i), \quad i = 1, 2, 3.$$

Let f_i and F_i be their PDFs and CDFs respectively. We denote by γ the lower incomplete Gamma function given by

$$\gamma(s, x) = \int_0^x t^{s-1} \exp^{-t} \, dt.$$

Let

$$
\begin{aligned}
h_i(x) &= \frac{f_i(x)}{F_{i-1}(x) - F_i(x)} \\
&= \frac{\Gamma(\alpha_{i-1})\beta_i^{\alpha_i} x^{\alpha_i - 1} \exp(-\beta_i x)}{\gamma(\alpha_{i-1}, \beta_{i-1}x)\Gamma(\alpha_i) - \gamma(\alpha_i, \beta_i x)\Gamma(\alpha_{i-1})}
\end{aligned}
$$

if $x > 0$, and $h(x) = 0$ otherwise. According to Butucea, Delmas, Dutfoy, and Fischer 2018, under certain conditions on the F_i, the joint distribution of (q_1, q_2, q_3) with maximum entropy which satisfies $\Pr(q_1 \leq q_2 \leq q_3) = 1$ is uniquely defined by the density

$$
\begin{aligned}
\pi_q^{\mathrm{MEC}}(q_1, q_2, q_3) &= f_1(q_1) \prod_{i=2}^{3} h_i(q_i) \exp\left(-\int_{q_{i-1}}^{q_i} h_i(s)\,ds\right) \mathbb{1}_{q_1 \leq q_2 \leq q_3} \\
&= \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} q_1^{\alpha_1 - 1} \exp(-\beta_1 q_1) \prod_{i=2}^{3} h_i(q_i) \exp\left(-\int_{q_{i-1}}^{q_i} h_i(s)\,ds\right) \mathbb{1}_{0 < q_1 \leq q_2 \leq q_3}.
\end{aligned}
$$

See Appendix C for more detail on one of these conditions. This distribution can be obtained in Python using the class `MaximumEntropyOrderStatisticsDistribution` from the library OpenTURNS (Baudin, Dutfoy, Iooss, and Popelin 2015). Therefore, from equation (14), the corresponding prior on the GEV parameters becomes

$$
\begin{aligned}
\pi_\theta^{\mathrm{MEC}}(\theta) &= \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} (q_1(\theta))^{\alpha_1 - 1} \exp(-\beta_1 q_1(\theta)) \\
&\quad \times \prod_{i=2}^{3} h_i(q_i(\theta)) \exp\left(-\int_{(q_{i-1}(\theta))}^{q_i(\theta)} h_i(s)\,ds\right) \sigma \Psi_3(\xi) \mathbb{1}_{q_1(\theta)>0} \mathbb{1}_{\sigma>0}, \quad (21)
\end{aligned}
$$

with Ψ_3 defined in (13).

4.5   E: Exponential prior distributions for three quantile differences

As the exponential distribution is a special case of the Gamma distribution, this prior is a special case of the G3 prior with marginals

$$\tilde{q}_i \sim \mathrm{Exp}(\lambda_i) = \Gamma(1, \lambda_i), \quad i = 1, 2, 3.$$

Therefore, from (15),

$$\pi_\theta^{\mathrm{E}}(\theta) \propto \exp\left(-\lambda_1 q_1(\theta)\right) \exp\left(-\sigma \Omega_3(\xi)\right) \sigma \Psi_3(\xi) \mathbb{1}_{q_1(\theta)>0} \mathbb{1}_{\sigma>0},$$

where Ψ_3 is defined in (13), and

$$\Omega_3(\xi) = \frac{1}{\xi} \sum_{i=2}^{3} \lambda_i (\exp(s_i \xi) - \exp(s_{i-1}\xi)) > 0.$$

4.6   TN: Truncated normal prior distributions for three quantile differences

Truncated normal distributions with support $(0, +\infty)$ are specified for three quantile differences

$$\tilde{q}_i \sim \text{TN}(\mu'_i, \sigma'_i, 0, +\infty)\,, \quad i = 1, 2, 3\,,$$

and the joint distribution $\pi^{\text{TN}}_{\tilde{q}}$ is formed by assuming that the quantile differences are independent. This is then converted into a prior for the quantiles π^{TN}_q using the transformation

$$(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3) \mapsto (\tilde{q}_1, \tilde{q}_1 + \tilde{q}_2, \tilde{q}_1 + \tilde{q}_2 + \tilde{q}_3)$$
$$= (q_1, q_2, q_3)\,.$$

Explicitly, the prior for (q_1, q_2, q_3) is given by

$$\pi^{\text{TN}}_q(q_1, q_2, q_3) \propto \phi\left(\frac{q_1 - \mu'_1}{\sigma'_1}\right) \prod_{i=2}^{3} \phi\left(\frac{q_i - q_{i-1} - \mu'_i}{\sigma'_i}\right) \mathbb{1}_{0 < q_1 < q_2 < q_3} \tag{22}$$

$$\propto \exp\left(-\frac{1}{2}\left(\left(\frac{q_1 - \mu'_1}{\sigma'_1}\right)^2 + \sum_{i=2}^{3}\left(\frac{q_i - q_{i-1} - \mu'_i}{\sigma'_i}\right)^2\right)\right) \tag{23}$$

$$\times \mathbb{1}_{0 < q_1 < q_2 < q_3}\,. \tag{24}$$

This can be converted into a prior for θ using the change of variables formula (14):

$$\pi^{\text{TN}}_\theta(\theta) = \exp\left(-\frac{1}{2}\left(\left(\frac{q_1(\theta) - \mu'_1}{\sigma'_1}\right)^2 + \sum_{i=2}^{3}\left(\frac{\sigma \exp(s_i \xi) - \exp(s_{i-1}\xi) - \xi\mu'_i}{\xi\sigma'_i}\right)^2\right)\right) \tag{25}$$

$$\times \sigma \Psi_3(\xi) \mathbb{1}_{q_1(\theta),\sigma > 0}\,, \tag{26}$$

with Ψ_3 defined in (13).

5   Sampling using MCMC

For the priors π^{G3}, π^{MEC}, and π^{G2}, 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. For π^{SS}, as we have access to the full conditional for γ, we can replace the normal proposal density for γ by its full conditional. These algorithms are special cases of the Metropolis–Hastings algorithm, whose convergence is studied in Roberts and Smith 1994.

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. The smoothing from the density estimation leads to the constraints on the quantiles not always being satisfied. The marginals for π^{G3} and $\pi^{\text{G3}}_{\log(\sigma),\xi}$ do not have to be estimated as we can calculate them analytically. We can use the samples of the posterior of θ to plot the mean return level against the return period, as well as 95% credibility intervals.

6   Results

In order to compare the priors, we performed analyses on three datasets. The first, denoted $\mathbf{x}^{\text{PPP}}$ for Poisson Point Process, was created by simulating the exceedances of a certain threshold from

---

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

Input: function $f \colon \mathbb{R}^3 \to \mathbb{R}$ 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 $s^{(1)}, \ldots, s^{(n-b)} \in \mathbb{R}^3$

begin

$\quad (\theta, y) \longleftarrow (\theta_0, f(\theta_0))$

$\quad$ **for** $i \in 1, \ldots, n$ **do**

$\quad\quad$ **for** $j \in 1, 2, 3$ **do**

$\quad\quad\quad \theta_j^* \sim \mathcal{N}(\theta_j, \sigma_j)$

$\quad\quad\quad y^* \longleftarrow f(\theta^*)$

$\quad\quad\quad u \sim \mathcal{U}(0, 1)$

$\quad\quad\quad$ **if** $y^* - y > \log(u)$ **then**

$\quad\quad\quad\quad (\theta, y) \longleftarrow (\theta^*, y^*)$

$\quad\quad$ **if** $i > b$ **then**

$\quad\quad\quad s^{(i-b)} \longleftarrow \theta$

---

a Poisson point process. The second, denoted $\mathbf{x}^{\mathrm{PD}}$ for Pseudo-Data, was created by simulating all observations from a GEV distribution. The third, denoted $\mathbf{x}^{\mathrm{WS}}$ for Wind Speed, consists of observed daily average wind speed at Tours, France from 1931 to 2011. This datasets are illustrated in Fig. 1.

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 at `https://github.com/tzhg/extreme-bayes`. Details of the Metropolis algorithm implementations are tabulated in Appendix D.

6.1 PPP: Poisson point process data

According to the model described in § 3, the exceedances of the threshold u approximately follow a non-homogeneous Poisson point process with intensity function

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

We simulated data for which this approximation is exact, with known parameters. We set $M = 54$ and $n_u = 86$ to match the study in Coles and Tawn 1996. We choose parameters
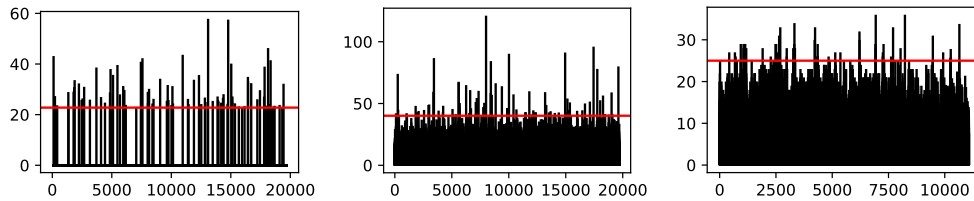
Figure 1: PPP (left), PD (middle), and WD (right) datasets, with chosen thresholds in red

$\theta = (25, 5, 0.2)$, and calculated the threshold u using the equation

$$M\Lambda[u, \infty) = n_u$$

$$\iff u = \mu + \sigma \frac{\left(\frac{n_u}{M}\right)^{-\xi} - 1}{\xi},$$

with Λ defined in (6). This resulted in a threshold of $u = 22.778$. The intensity function is equal to the PDF of a Generalised Pareto distribution, except with support $\{x \colon x > u\}$. This means that we can simulate from the intensity by simulating from a Generalised Pareto distribution with parameters (u, σ, ξ). We simulated 86 observations, and distributed them uniformly over the total time period of $365M$ observations, setting the remaining points to 0.

We constructed the priors with the knowledge of the true quantiles and quantile differences. Fixing $p = (0.1, 0.01, 0.001)$, then

$$\tilde{q}_1^* = 39.211, \quad \tilde{q}_2^* = 23.523, \quad \tilde{q}_3^* = 36.783$$

are the true quantile differences. To construct π_q^{G3}, we chose

$$\tilde{q}_i \sim \Gamma\left(\frac{(\tilde{q}_i^*)^2}{v}, \frac{\tilde{q}_i^*}{v}\right), \quad i = 1, 2, 3$$

$$\implies \mathrm{E}(\tilde{q}_i) = \tilde{q}_i^*, \quad \mathrm{Var}(\tilde{q}_i) = v$$

for $v = 5$. For π^{MEC}, the approximation of quantile marginals by Gamma distributions is shown in Fig. 2. For π^{TN}, the target mean and variance were $(36.783, 5)$, while the mean and variance of the constructed truncated normal distributions were $(36.783, 5)$.

The univariate and bivariate marginals of θ and q_1, q_2, q_3 for the different priors are illustrated in Table 2, Table 3, Table 4 and Table 5. We plotted the mean return level compared to the analytic true return level in Fig. 6. The dashed line was obtained by simulating 270,000 years of data and calculating the empirical quantiles.
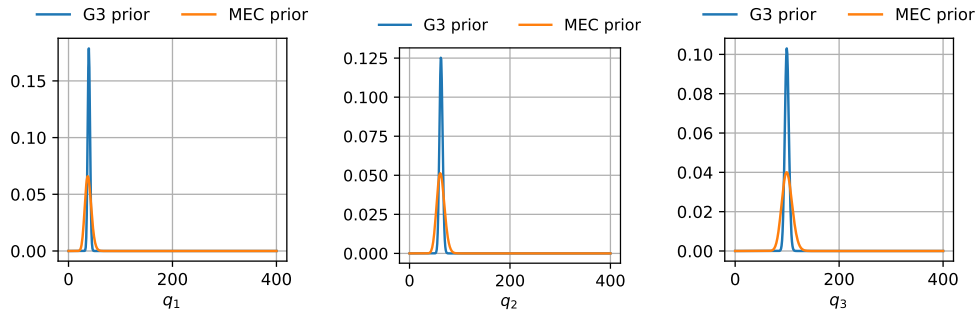
Figure 2: Comparison of marginals of π_q^{G3} and π_q^{MEC}
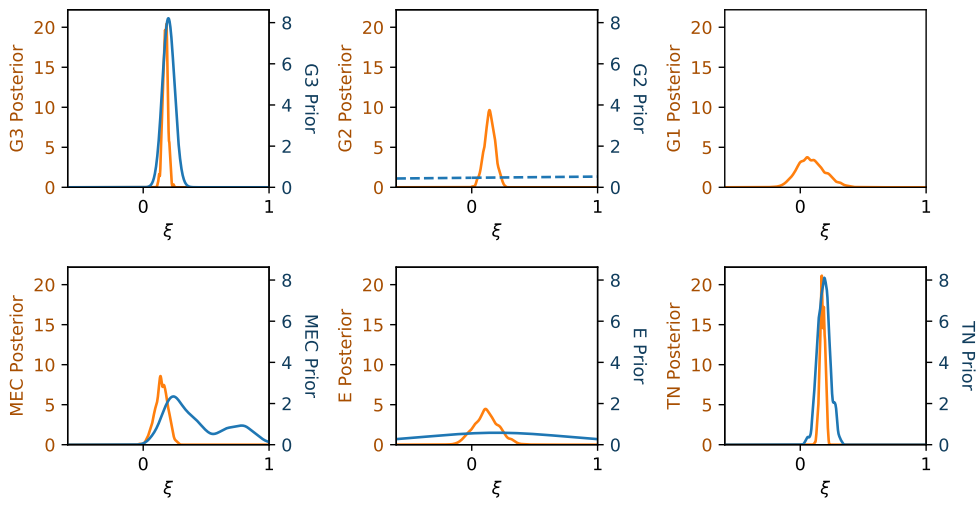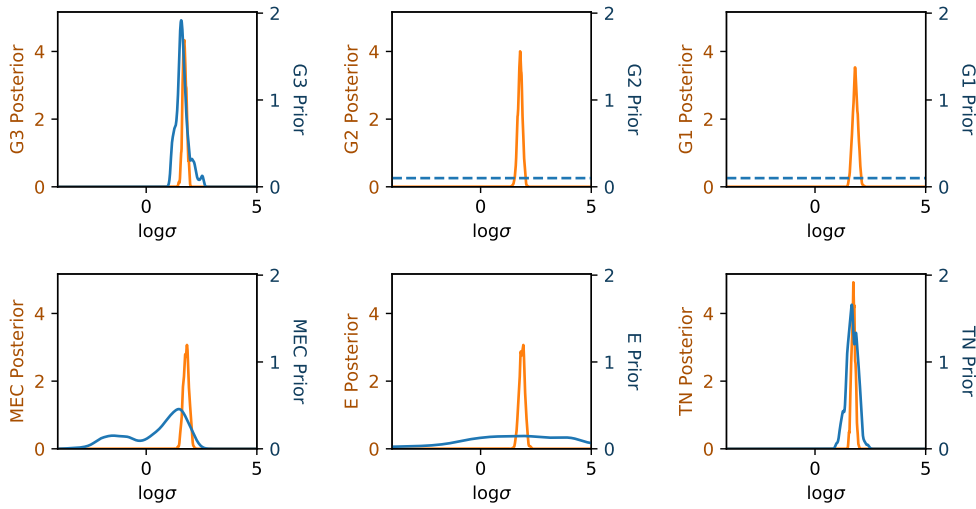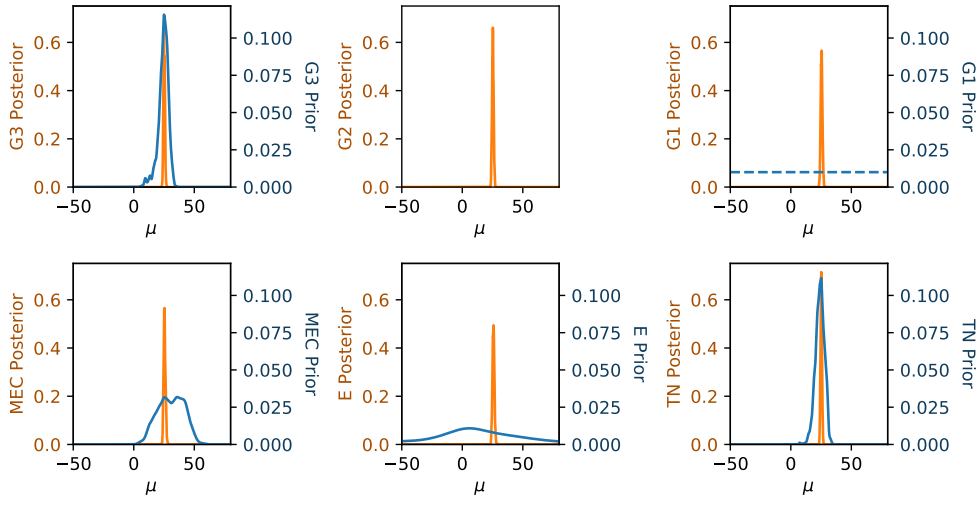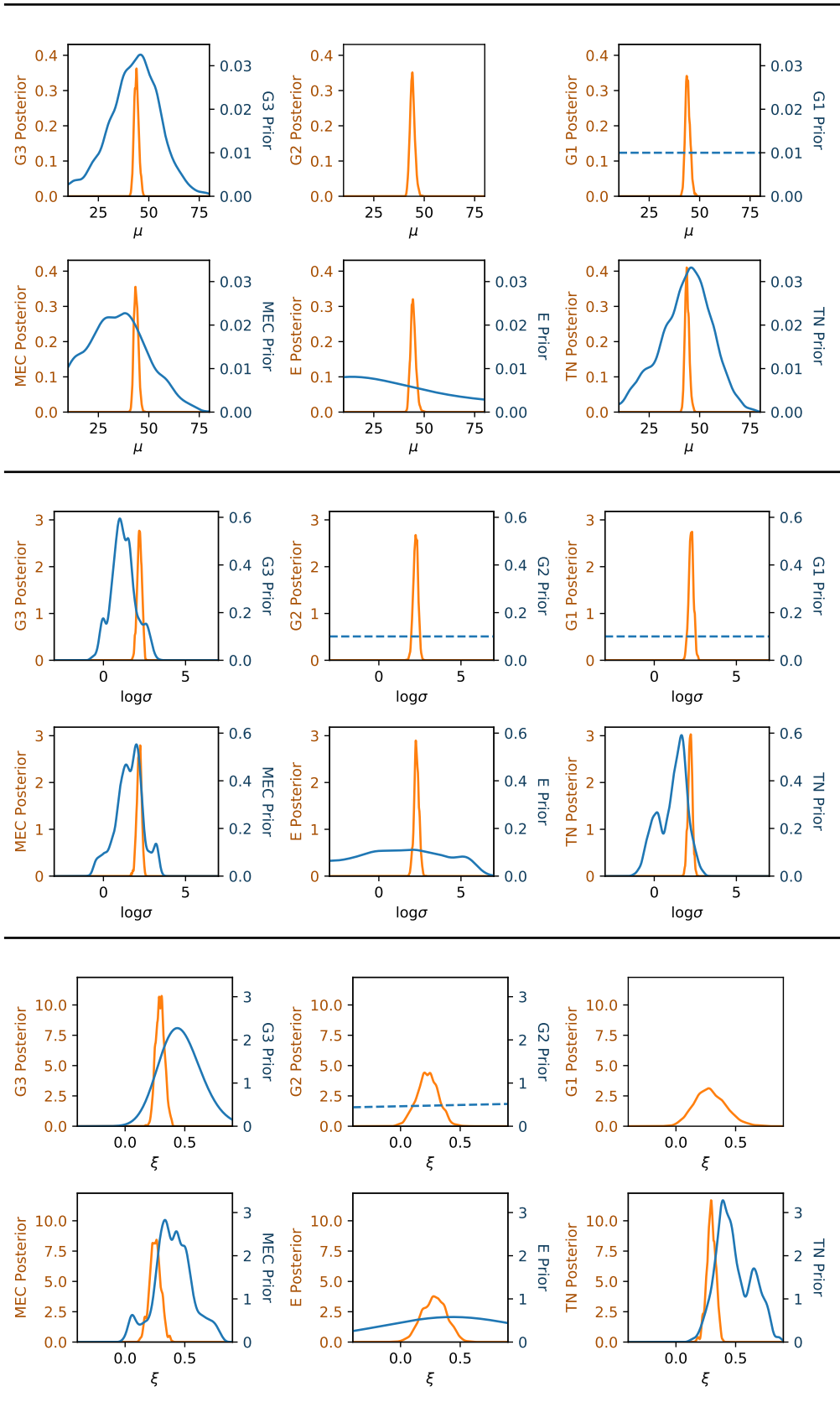
Table 2: Prior and posterior univariate marginals of θ for Poisson process simulation study, with dashed lines for improper priors
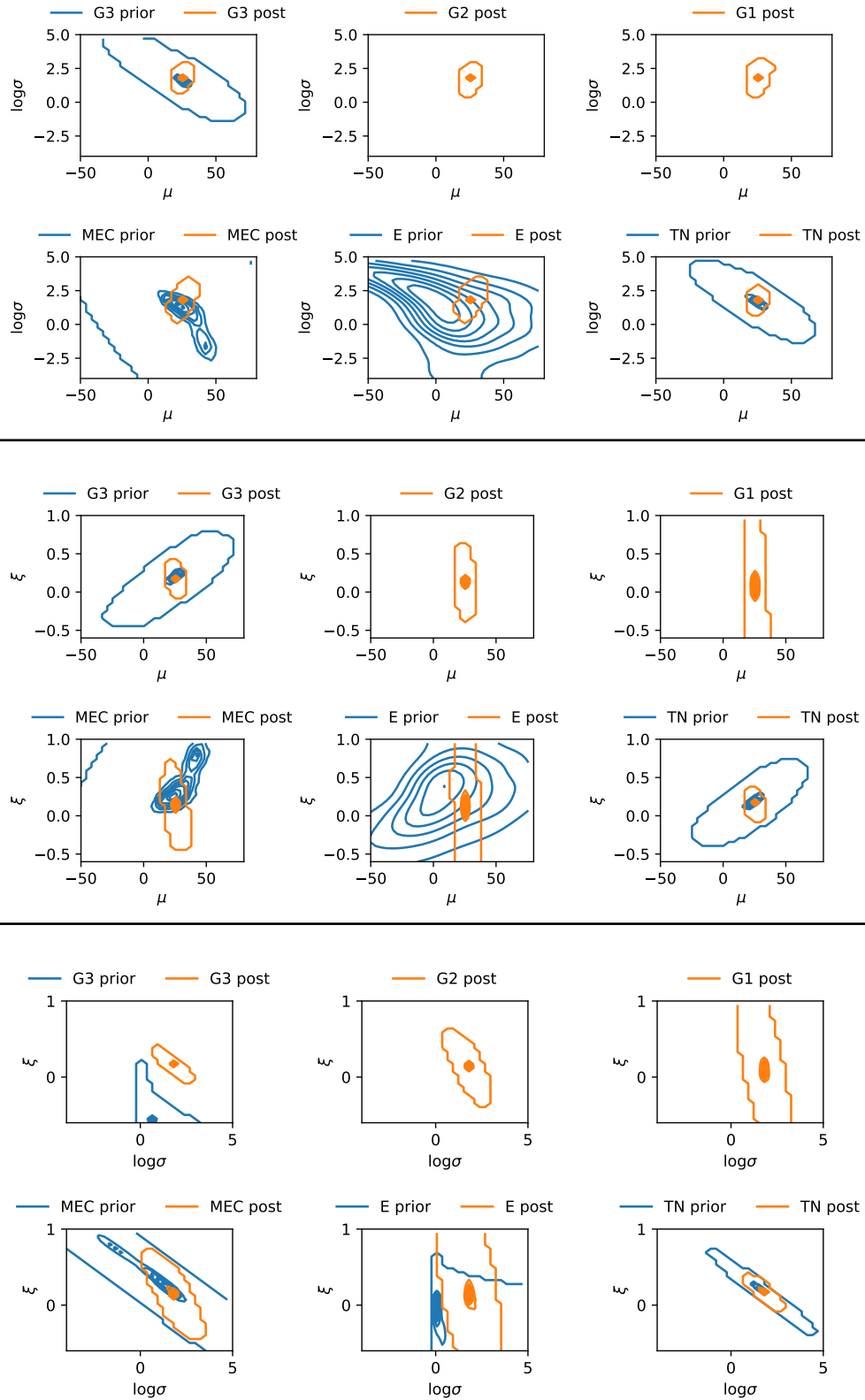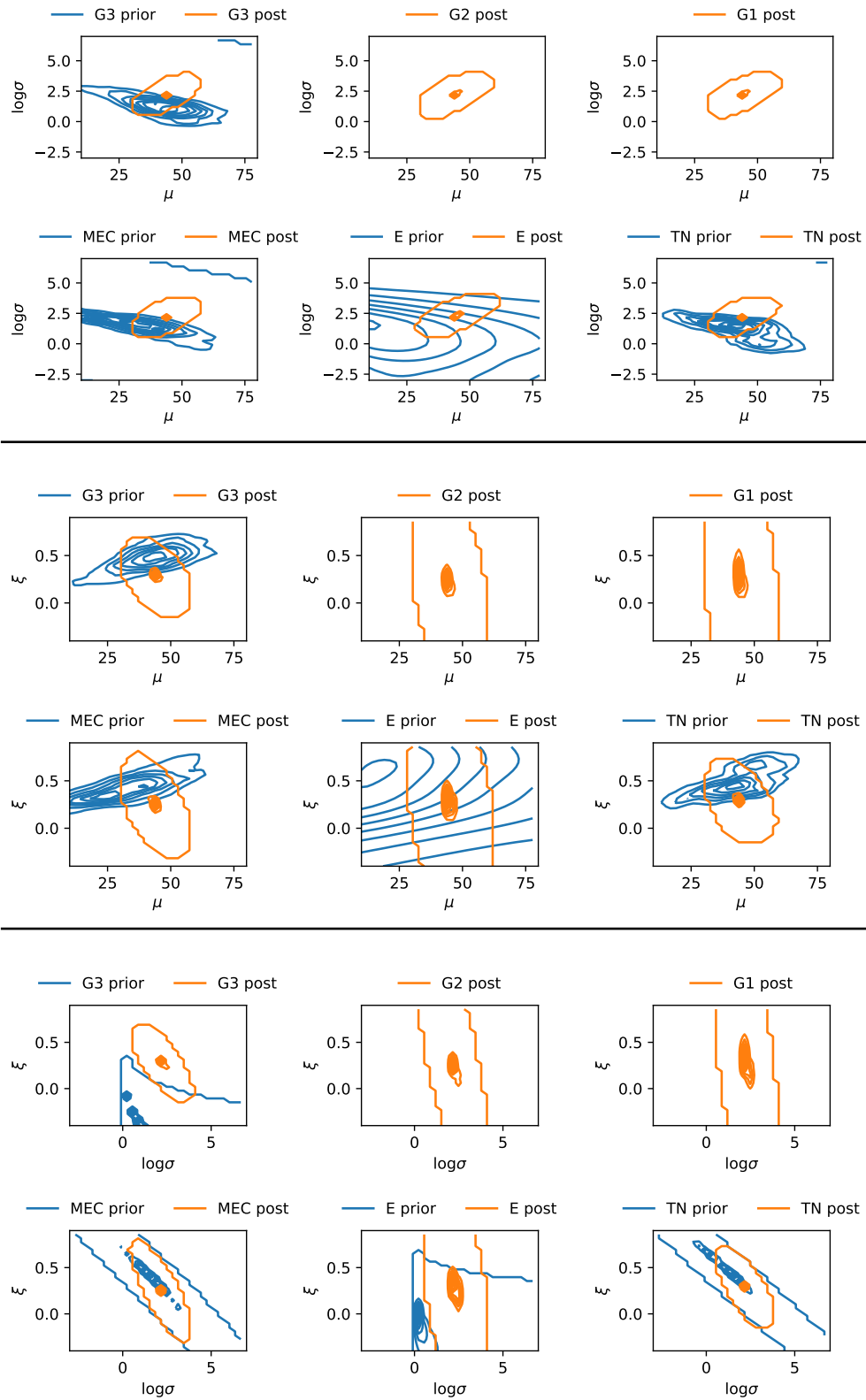
Table 3: Prior and posterior bivariate marginals of θ for Poisson process simulation study, with dashed lines for improper priors
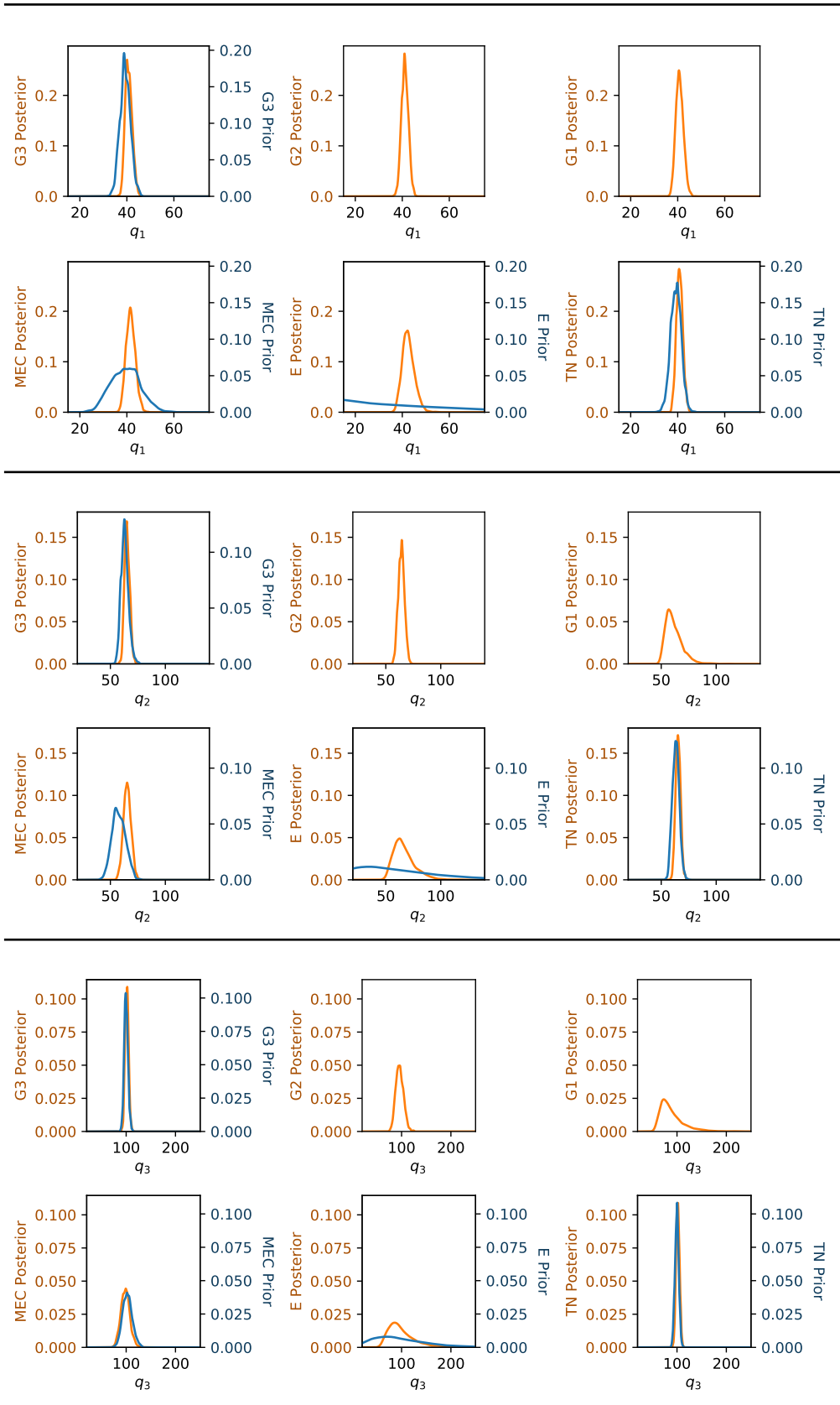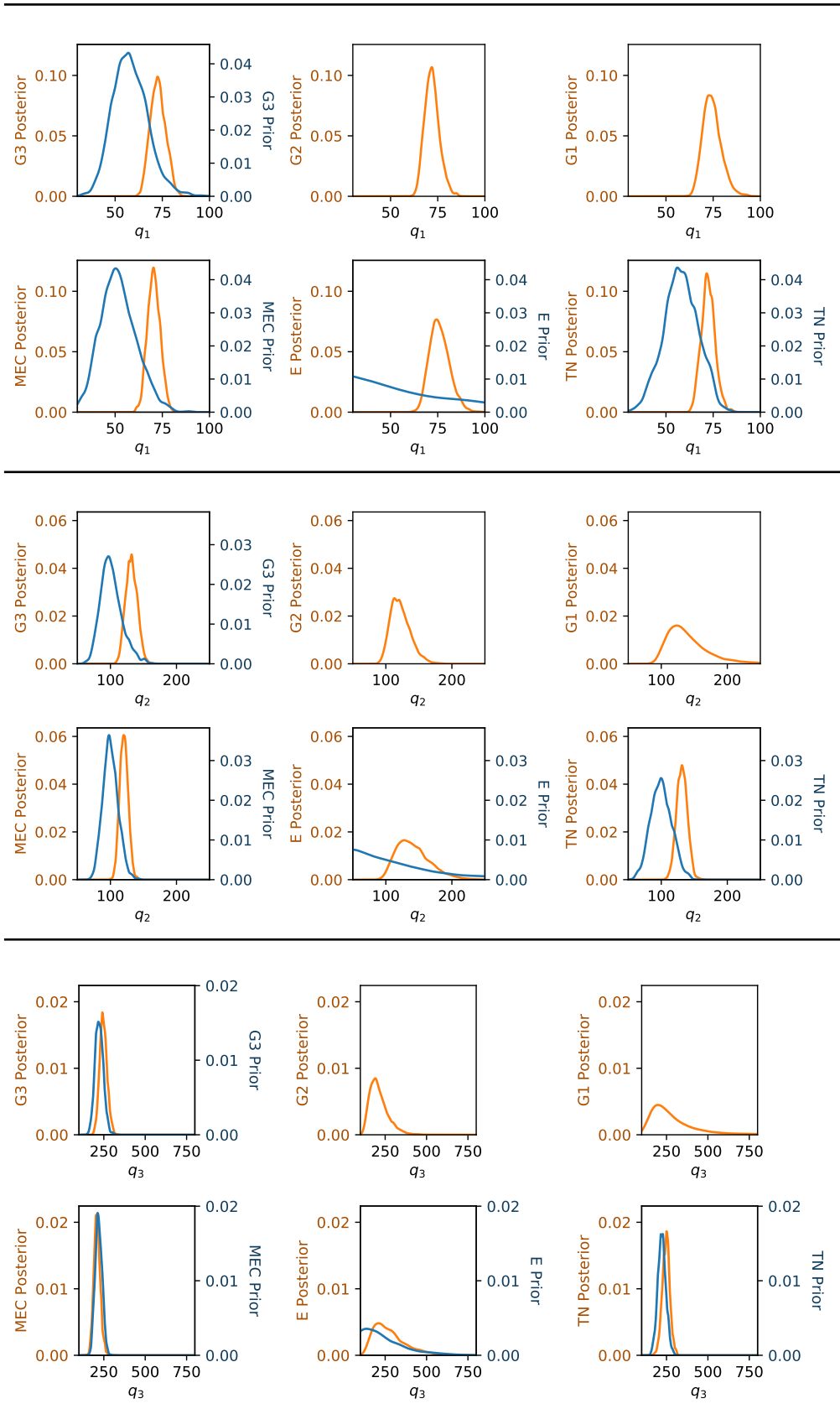
Table 4: Prior and posterior univariate marginals of (q_1, q_2, q_3) for Poisson process simulation study
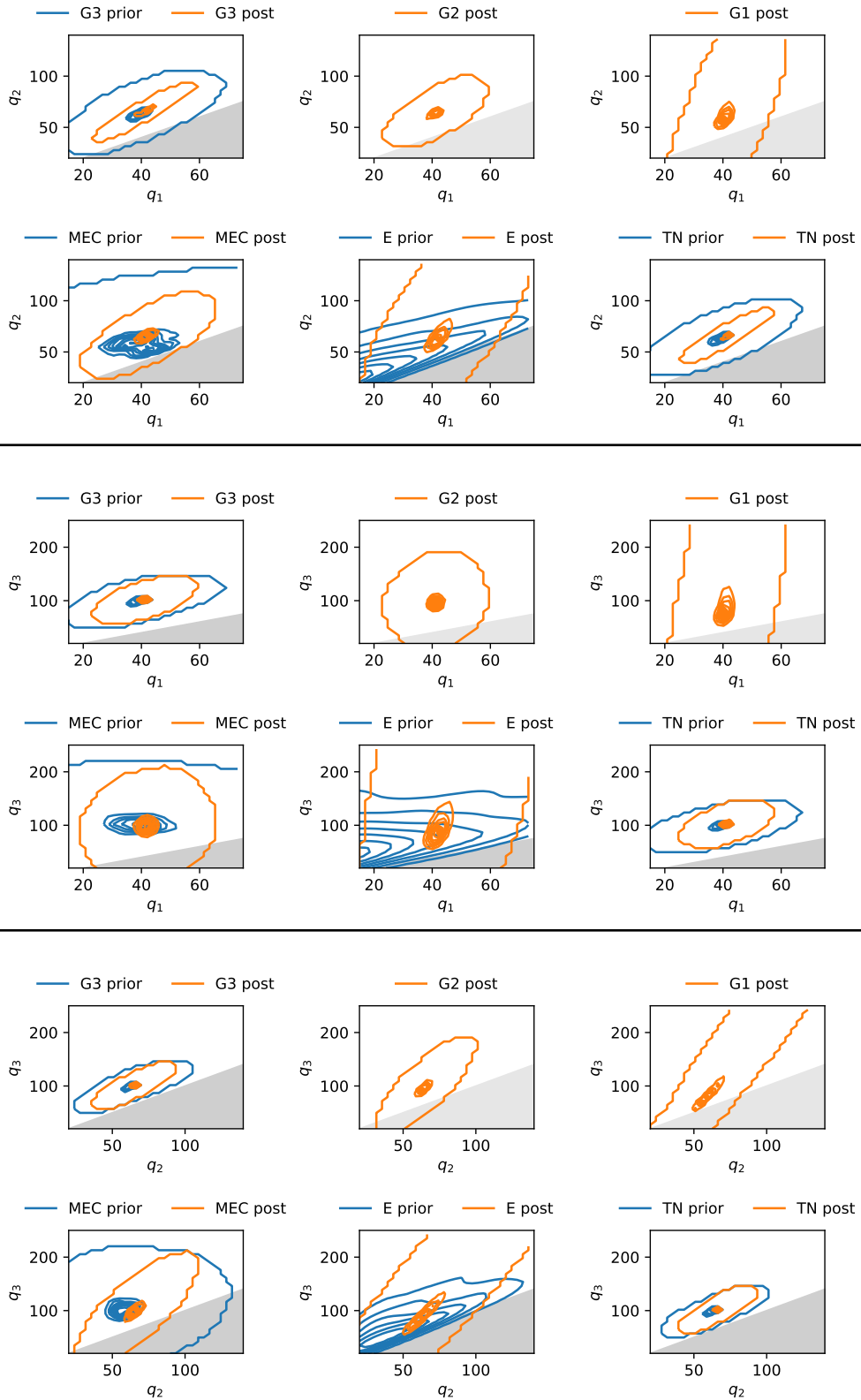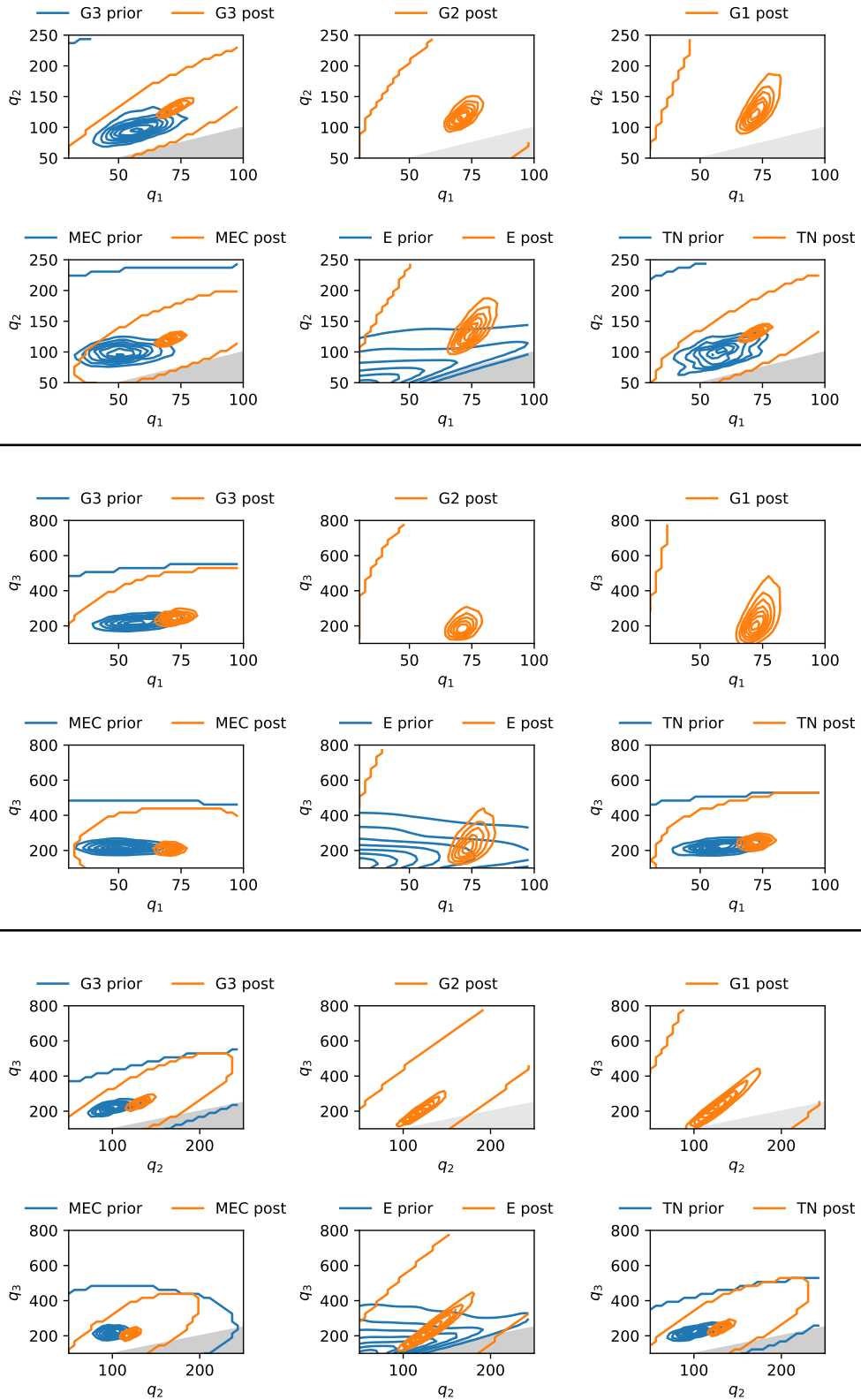
Table 5: Prior and posterior bivariate marginals of (q_1, q_2, q_3) for Poisson process simulation study, with grey shaded region $x < y$
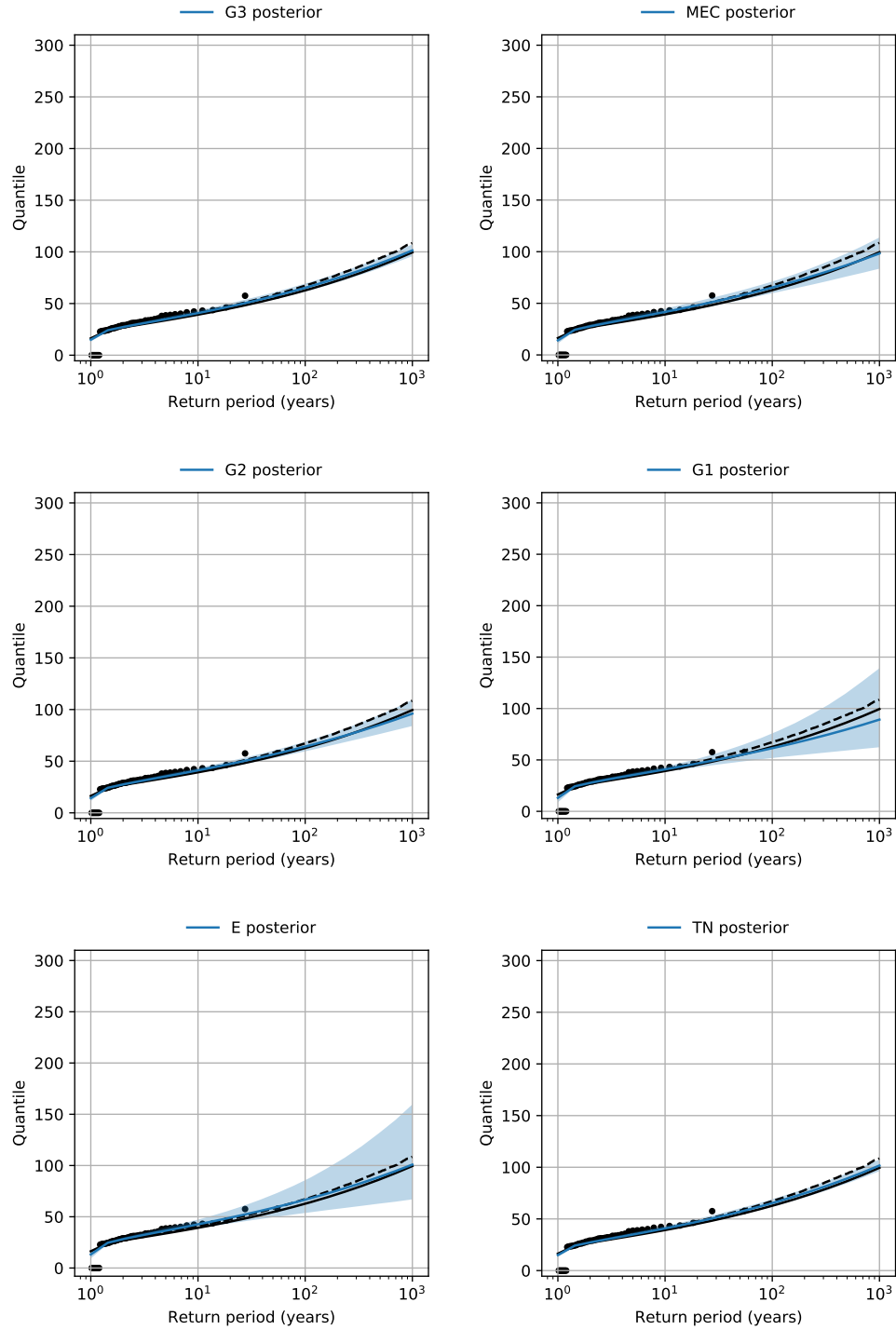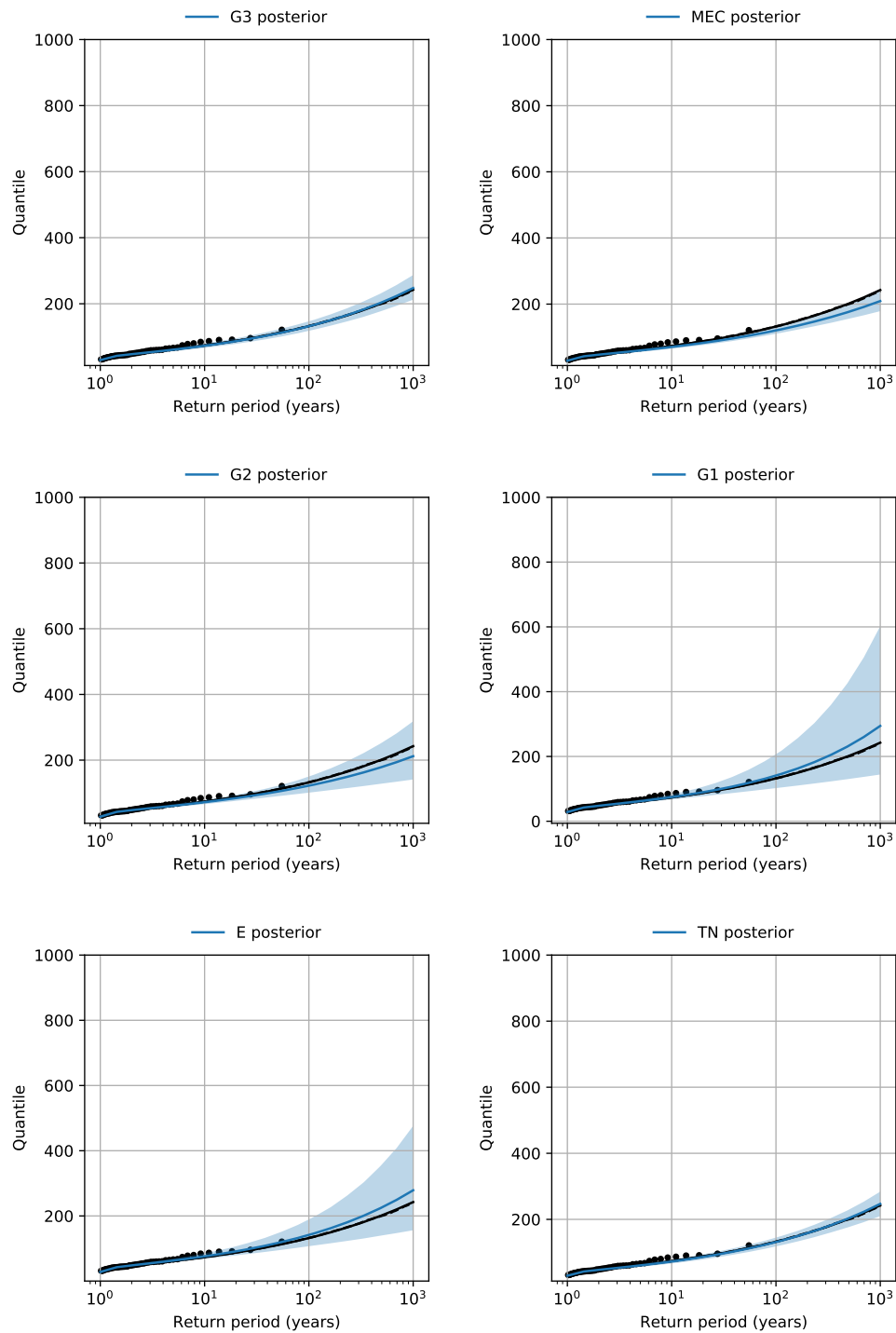
Table 6: Mean return level with 95% credibility intervals estimated using various posterior distributions, with analytic return level **(black solid)**, simulated return level **(black dashed)** and empirical quantiles **(black dots)** for Poisson process simulation study

6.2   PD: Pseudo-data

In order to compare our results to Coles and Tawn 1996, we generated data, denoted $\mathbf{x}^{\mathrm{PD}}$, from $\mathrm{GEV}(17, 1.8, 0.28)$ to resemble their data, such that:

	$\mathbf{x}^{\mathrm{PD}}$	Coles and Tawn 1996
M	54	54
u	40.109	40
n_u	86	86

We used the same prior $\pi_{\tilde{q}}^{\mathrm{G3}}$ for independent quantile differences as specified by Coles and Tawn 1996, given by

$$p = (0.1, 0.01, 0.001),$$
$$\tilde{q}_1 \sim \Gamma(38.9, 0.67),$$
$$\tilde{q}_2 \sim \Gamma(7.1, 0.16),$$
$$\tilde{q}_3 \sim \Gamma(47, 0.39).$$

For π^{MEC}, the approximation of quantile marginals by Gamma distributions is shown in Fig. 3. For π^{TN}, the target mean and variance were $(120.513, 309.007)$, while the mean and variance of the constructed truncated normal distributions were $(120.513, 309.007)$.

The univariate and bivariate marginals of θ and q_1, q_2, q_3 for the different priors are illustrated in Table 8, Table 9, Table 10 and Table 11. In Table 7, we compare the statistics of the quantile differences specified by the expert, to the statistics estimated using the MCMC samples of (μ, σ, ξ) for each of the priors. In Fig. 12, the mean return level is plotted with 95% credibility intervals and empirical quantiles. The dashed black line was obtained by simulating 270,000 years of data and calculating the empirical quantiles. We know that the annual maximum has CDF F^{365}, where F is the CDF of $\mathrm{GEV}(17, 1.8, 0.28)$, and the solid black line shows the quantiles calculated analytically from this CDF.

In Fig. 4, we varied the threshold, and plotted the number of exceedances against the return level for return periods 10^2, 10^3, and 10^4, all estimated using the prior $\pi_{\theta|\mathbf{x}^{\mathrm{PD}}}^{\mathrm{G3}}$.
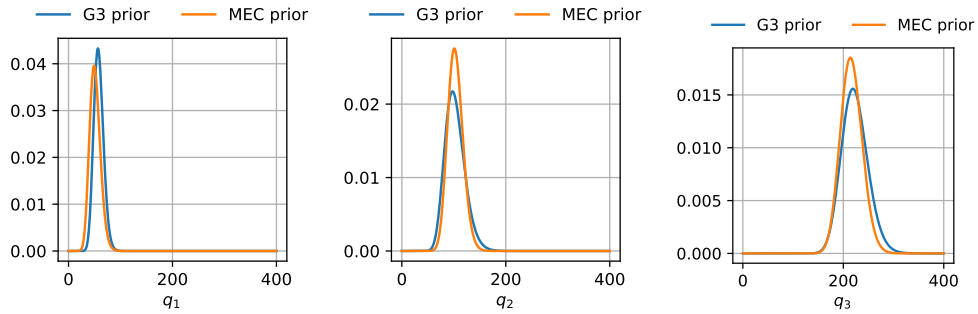
Figure 3: Comparison of marginals of π_q^{G3} and π_q^{MEC}

	$\tilde{q}_1$	$\tilde{q}_2$	$\tilde{q}_3$
Expert	$(57.563, 70.263)$	$(42.310, 66.805)$	$(119.09, 142.836)$
π_θ^{G3}	$(57.402, 69.767)$	$(41.141, 62.683)$	$(120.004, 144.618)$
$\pi_\theta^{\mathrm{MEC}}$	$(51.082, 64.635)$	$(47.966, 65.294)$	$(118.74, 147.957)$
π_θ^{E}	$(38.878, 135.433)$	$(27.553, 94.297)$	$(88.617, 279.682)$
π_θ^{TN}	$(57.888, 69.991)$	$(42.583, 58.529)$	$(122.371, 143.133)$

Table 7: Pairs of $(\text{median}, 0.9\text{-quantile})$ of quantile differences for pseudo-data simulation study
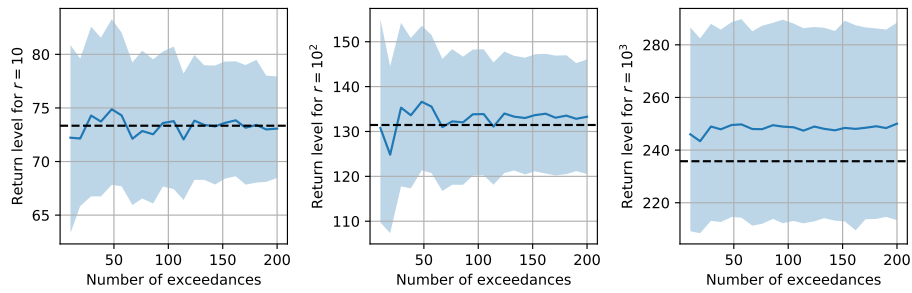
Figure 4: Mean return levels **(blue)** with 95% credibility intervals estimated using $\pi_{\theta|\mathbf{x}^{\mathrm{PD}}}^{\mathrm{G3}}$ for various thresholds, with simulated return levels **(black dashed)**

Table 8: Prior and posterior univariate marginals of θ for pseudo-data simulation study, with dashed lines for improper priors

Table 9: Prior and posterior bivariate marginals of θ for pseudo-data simulation study, with dashed lines for improper priors

Table 10: Prior and posterior univariate marginals of (q_1, q_2, q_3) for pseudo-data simulation study

Table 11: Prior and posterior bivariate marginals of (q_1, q_2, q_3) for pseudo-data simulation study, with grey shaded region $x < y$

Table 12: Mean return level with 95% credibility intervals estimated using various posterior distributions, with analytic return level **(black solid)**, simulated return level **(black dashed)** and empirical quantiles **(black dots)** for pseudo-data process simulation study

6.3   WS: Daily average wind speed

This dataset consists of observations of average wind speed at Tours, France over a period of 30.34 years, from 1981 to 2011. We chose a threshold of $u = 25$ with 76 exceedances.

In order to construct the priors, We first fitted a GEV distribution on the annual maxima of the data. The Q-Q plot is shown in Fig. 5. The estimated corresponding quantiles differences estimates were then obtained, and Gamma distributions centred on these estimates were constructed with variance 5. For π^{MEC}, the approximation of quantile marginals by Gamma distributions is shown in Fig. 6. For π^{TN}, the target mean and variance were $(2.968, 5)$, while the mean and variance of the constructed truncated normal distributions were $(2.968, 5)$.

The univariate and bivariate marginals of θ and q_1, q_2, q_3 for the different priors are illustrated in Table 13, Table 14, Table 15 and Table 16. We plotted the mean return level in Fig. 17.
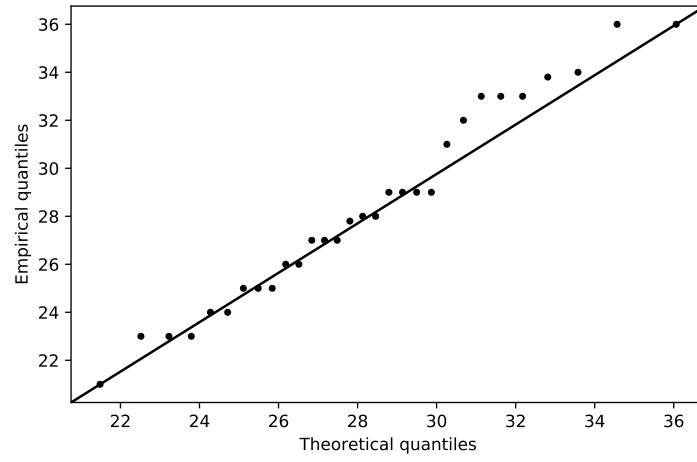
Figure 5: Q-Q plot of GEV distribution fit for wind speed data
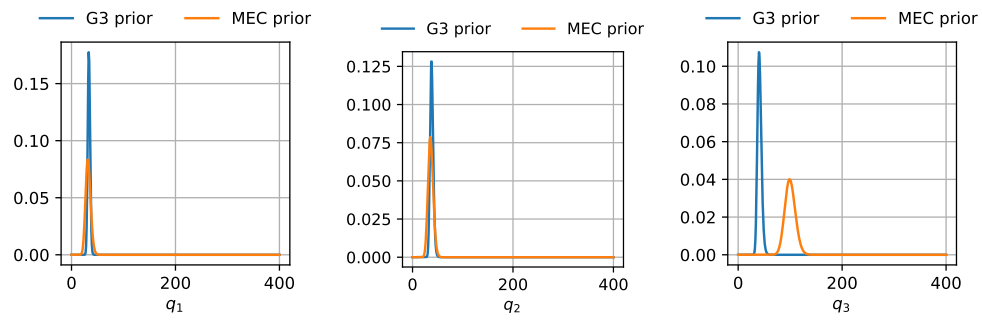
Figure 6: Comparison of marginals of π_q^{G3} and π_q^{MEC}
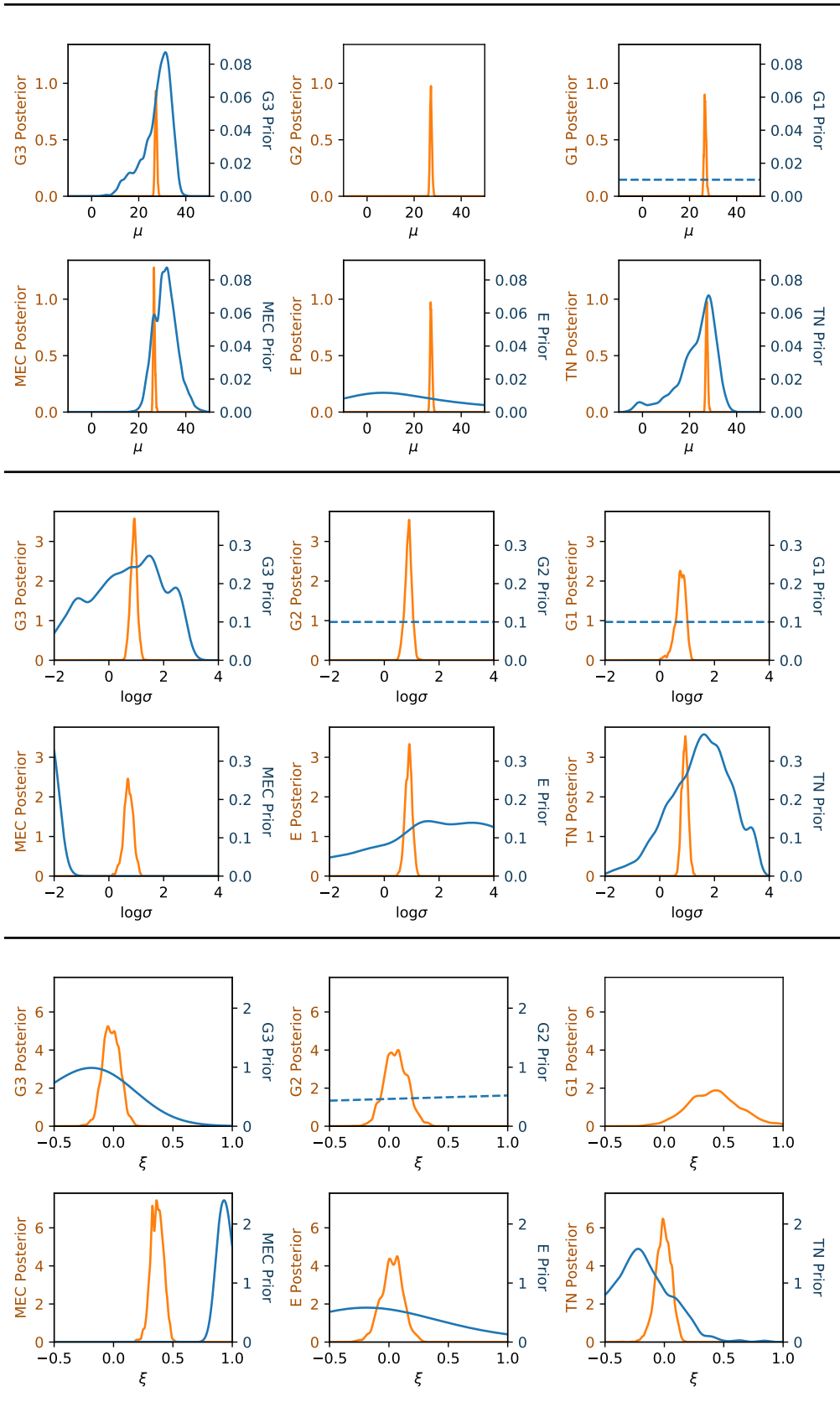
Table 13: Prior and posterior univariate marginals of θ for wind speed data, with dashed lines for improper priors
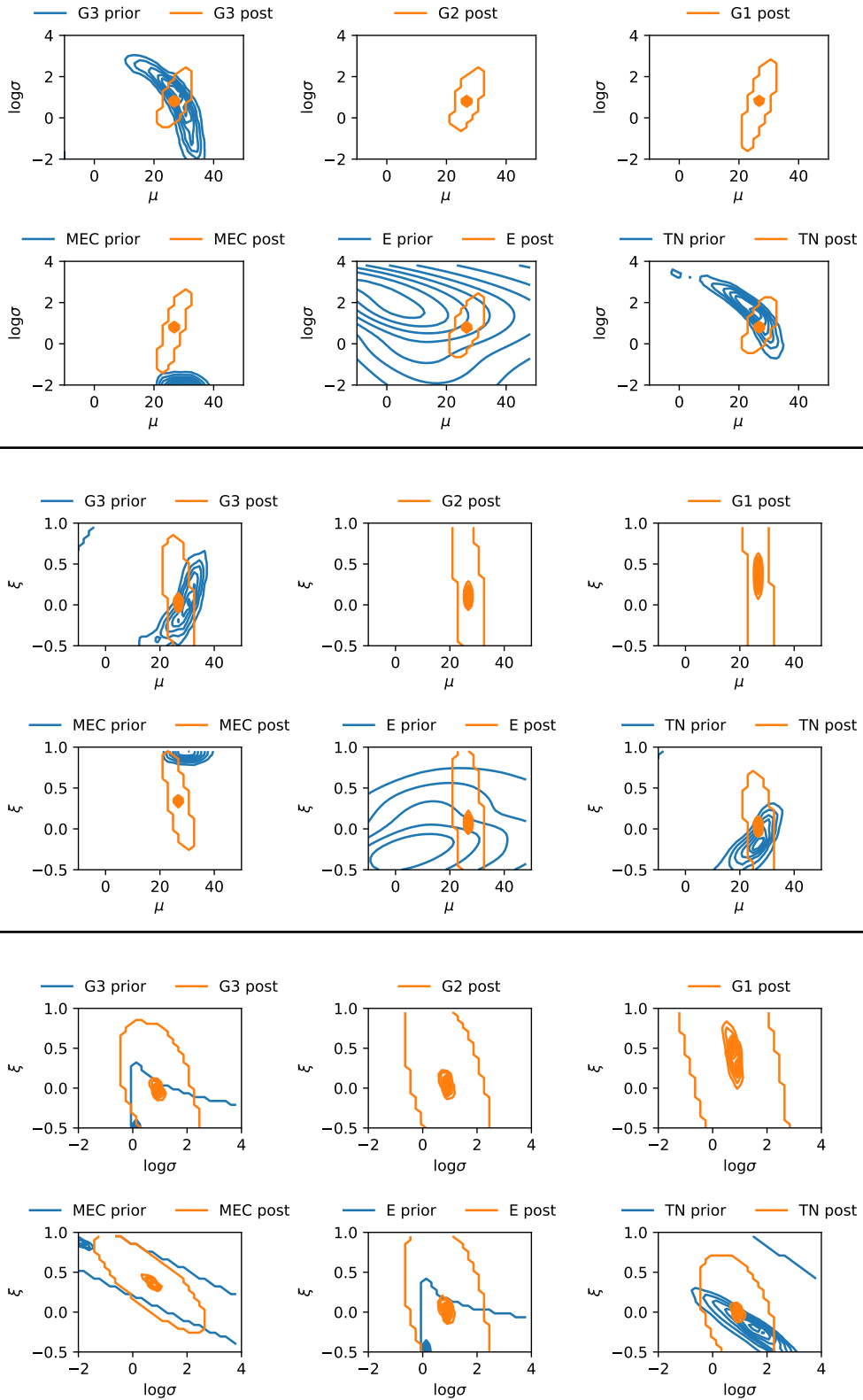
Table 14: Prior and posterior bivariate marginals of θ for wind speed data, with dashed lines for improper priors
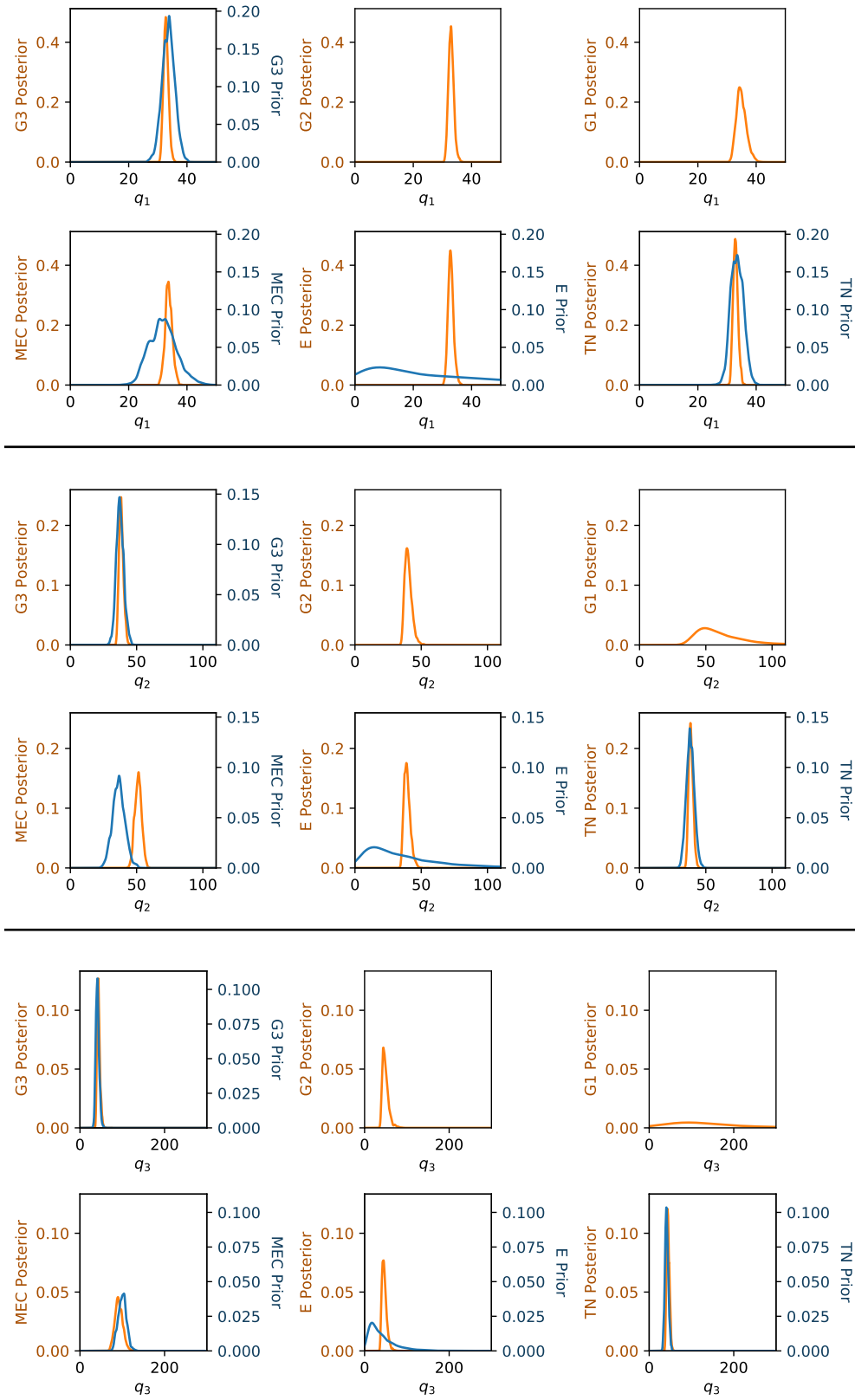
Table 15: Prior and posterior univariate marginals of (q_1, q_2, q_3) for wind speed data
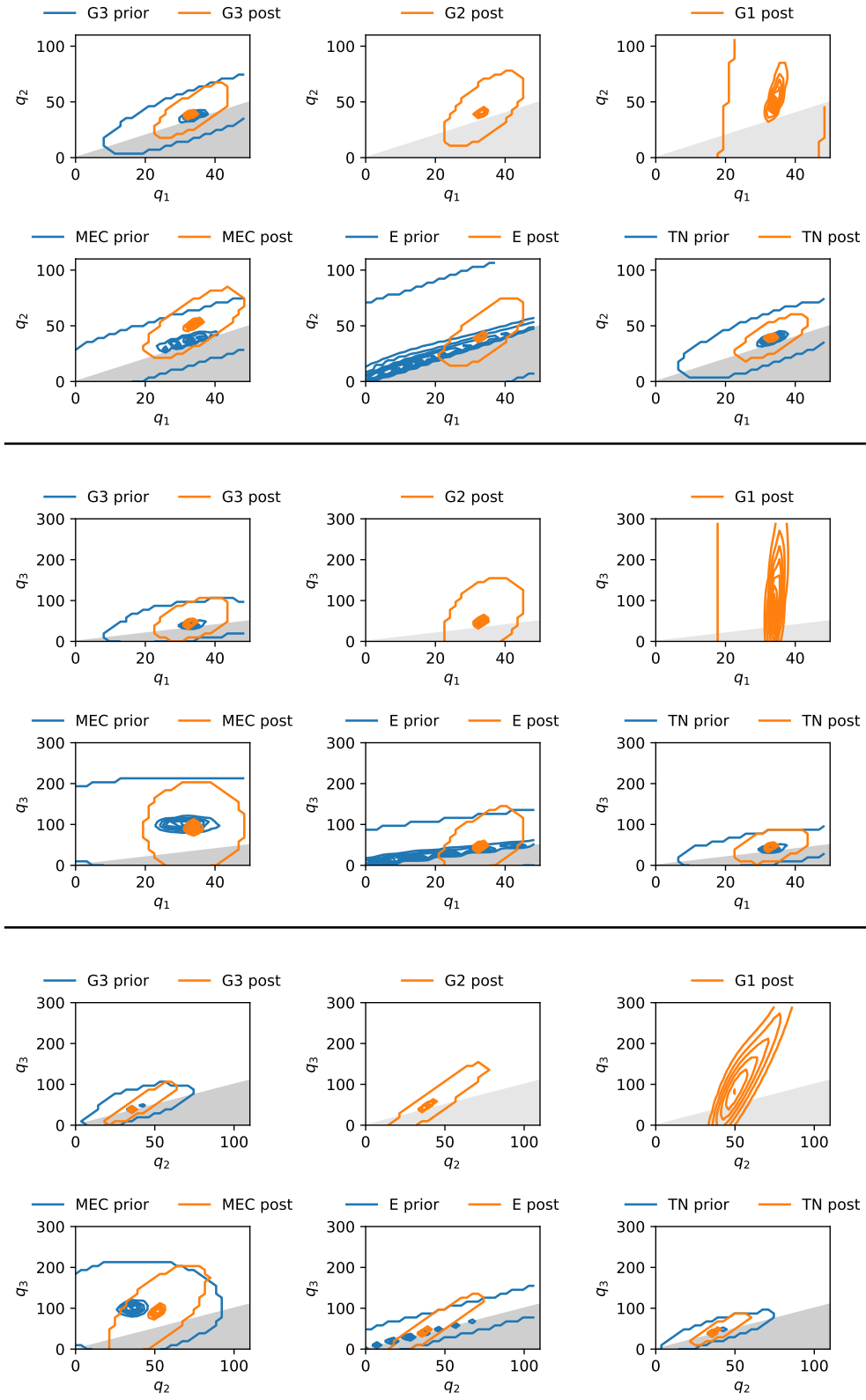
Table 16: Prior and posterior bivariate marginals of (q_1, q_2, q_3) for wind speed data, with grey shaded region $x < y$
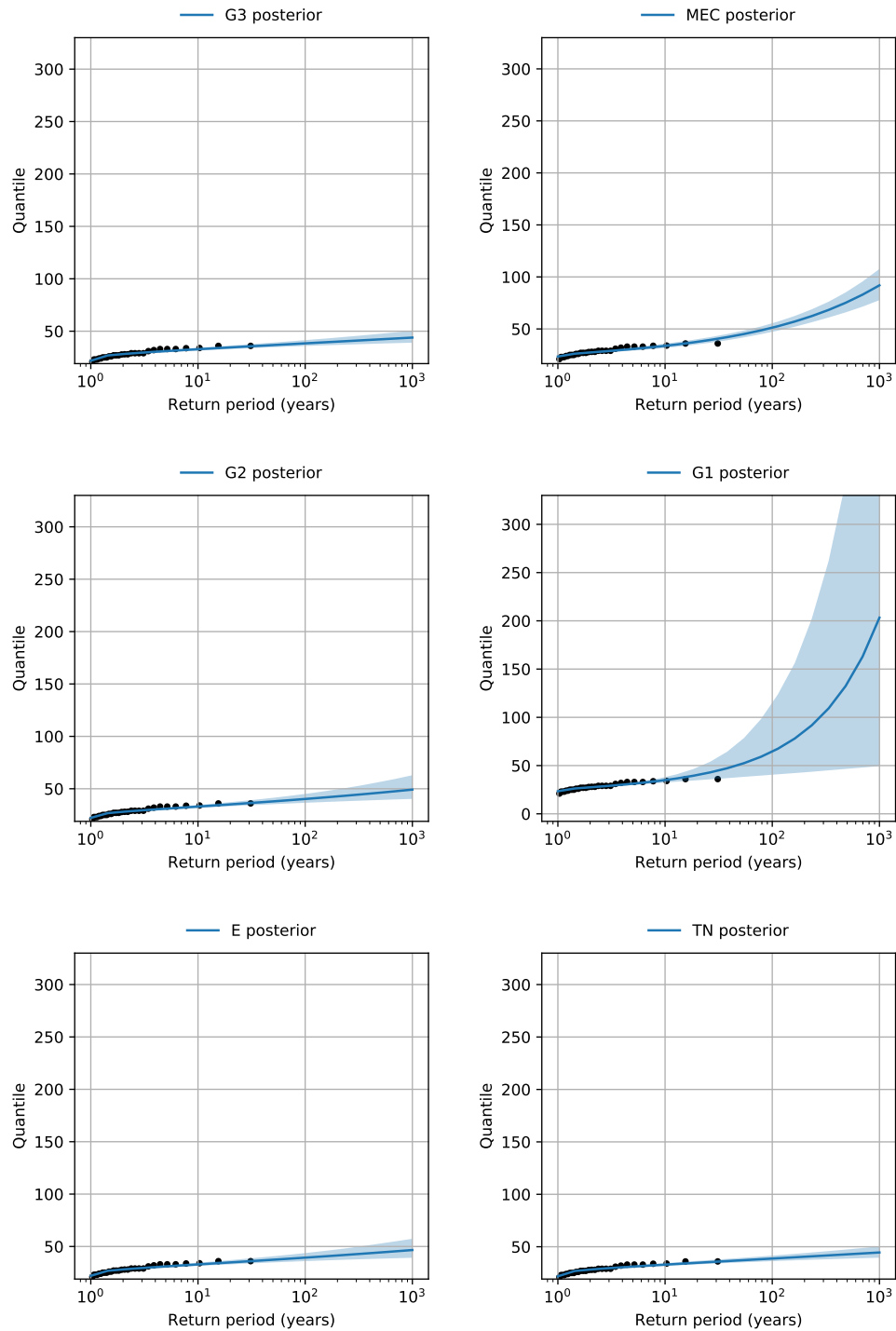
Table 17: Mean return level with 95% credibility intervals estimated using various posterior distributions, with analytic return level **(black solid)**, simulated return level **(black dashed)** and empirical quantiles **(black dots)** for wind speed data

A   Maximum entropy distributions

In order to find the distribution which maximises the entropy under certain constraints, we will use Lagrange multipliers and calculus of variations. Let p be a probability density with support $(0, +\infty)$. That is,

$$\int_0^{+\infty} p(x)\,\mathrm{d}x = 1\,,$$
$$p(x) > 0 \quad \forall x \in (0, +\infty)\,.$$

We define entropy as

$$\mathcal{E}(p) = -\int_0^{+\infty} p(x)\log(p(x))\,\mathrm{d}x\,,$$

where we take the natural logarithm instead of the base 2 logarithm, as this simplifies the calculation without changing the results. We define n constraints in the form

$$\int_0^{+\infty} p(x)f_i(x)\,\mathrm{d}x - c_i\,, \quad i = 1, \ldots, n\,.$$

for measurable functions f_i. Therefore, we obtain the objective function

$$L(p, \lambda) = -\int_0^{+\infty} p(x)\log(p(x))\,\mathrm{d}x + \lambda_0 \left(\int_0^{+\infty} p(x)\,\mathrm{d}x - 1 \right)$$
$$+ \sum_{i=1}^n \lambda_i \left(\int_0^{+\infty} p(x)f_i(x)\,\mathrm{d}x - c_i \right)\,.$$

It has partial derivatives

$$\frac{\partial L}{\partial p}(p) = -\log(p(x)) - 1 + \lambda_0 + \sum_{i=1}^n \lambda_i f_i(x)$$
$$\frac{\partial^2 L}{\partial p^2}(p) = -\frac{1}{p(x)}\,,$$

and setting $\frac{\partial L}{\partial p} = 0$, we get that for all $x \in (0, +\infty)$,

$$\frac{\partial L}{\partial p} = 0 \iff p(x) = \exp\left(1 - \lambda_0 + \sum_{i=1}^n \lambda_i f_i(x) \right)\,,$$
$$\implies p(x) \propto \exp\left(\sum_{i=1}^n \lambda_i f_i(x) \right)\,. \tag{27}$$

Fixed quantiles

Given n probabilities p_i, and their corresponding p_i-quantiles q_i, the constraints have the form

$$f_i(x) = \mathbb{1}_{(0, q_i)}(x)\,, \quad c_i = p_i$$

and therefore from (27),

$$p(x) \propto \exp\left(\sum_{i=1}^n \lambda_i \mathbb{1}_{(0, q_i)}(x) \right) = \sum_{i=1}^n \exp\left(\lambda_i \right) \mathbb{1}_{(0, q_i)}(x)\,.$$

This implies that the maximum entropy distribution is proportional to a piecewise uniform distribution on $(0, +\infty)$. However, such a distribution doesn't exist.

Fixed mean

Given mean μ, the constraint has the form

$$f_1(x) = x\,, \quad c_1 = \mu\,,$$

and therefore from (27),

$$p(x) \propto \exp\left(\lambda_1 x\right)\,.$$

This implies that the maximum entropy distribution is an exponential distribution.

Fixed mean and variance

Given mean μ and variance σ^2, the constraints have the form

$$
\begin{aligned}
f_1(x) &= x\,, \quad c_1 = \mu \\
f_2(x) &= x^2\,, \quad c_2 = \sigma^2 + \mu^2\,,
\end{aligned}
$$

and therefore from (27),

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

This implies that the maximum entropy distribution is a normal distribution, truncated to $(0, +\infty)$.

B 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

$$\alpha := -\frac{\mu}{\sigma}\,. \tag{28}$$

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

$$
\mathrm{E}(X) = \mu + H(\alpha)\sigma \tag{29}
$$

$$
\mathrm{Var}(X) = \sigma^2\left(1 + \alpha H(\alpha) - (H(\alpha))^2\right)\,. \tag{30}
$$

Therefore,

$$
\left(\frac{\sigma^*}{\mu^*}\right)^2 = \frac{\sigma^2\left(1 + \alpha H(\alpha) - (H(\alpha))^2\right)}{(\mu + H(\alpha)\sigma)^2}
$$

$$
= \frac{1 + \alpha H(\alpha) - (H(\alpha))^2}{(H(\alpha) - \alpha)^2} =: f(\alpha),
$$

and once we have α, from 29 we have

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

and from 28,

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

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

$$
H'(x) = \frac{(1 - \Phi(x))x(-\phi(x)) - \phi(x)(-\phi(x))}{(1 - \Phi(x))^2}
$$

$$
= \phi(x)\frac{-(1 - \Phi(x))x + \phi(x)}{(1 - \Phi(x))^2}
$$

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

$$
= H(x)\left(\frac{-(1 - \Phi(x))x}{1 - \Phi(x)} + H(x)\right)
$$

$$
= H(x)\left(H(x) - x\right).
$$

Therefore the derivative of f is

$$
f'(x) = \Big((H(x) - x)^2\left(H(x) + xH(x)\left(H(x) - x\right) - 2(H(x))^2\left(H(x) - x\right)\right)
$$

$$
- \left(1 + xH(x) - (H(x))^2\right)2\left(H(x) - x\right)\left(H(x)\left(H(x) - x\right) - 1\right)\Big)(H(x) - x)^{-4}
$$

$$
= \frac{(2 + H(x)(H(x) - x)(-3 + (H(x) - x)x))}{(H(x) - x)^3}.
$$

C Maximum entropy copula condition (C2)

The construction of the maximum entropy distribution in § 4.4 requires the following condition on the marginals: for a sequence of distributions $(F_i)_{1 \geq i \geq d}$,

$$
(C2) := \forall 1 \leq i \leq d - 1 \; \forall x \in \{x \colon 1 > F_i(x), F_{i+1}(x) > 0\} \; (F_i(x) > F_{i+1}(x)).
$$

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

Weibull distribution

Let $1 \leq i \leq d - 1$. We have that

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

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

$$\textit{(C2)} \iff 1 - \exp\left(-\left(\frac{x}{\lambda_i}\right)^{k_i}\right) > 1 - \exp\left(-\left(\frac{x}{\lambda_{i+1}}\right)^{k_{i+1}}\right)$$

$$\iff \left(\frac{x}{\lambda_i}\right)^{k_i} > \left(\frac{x}{\lambda_{i+1}}\right)^{k_{i+1}} \,.$$

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

$$\left(\frac{x}{\lambda_i}\right)^{k} > \left(\frac{x}{\lambda_{i+1}}\right)^{k} \iff \frac{x}{\lambda_i} > \frac{x}{\lambda_{i+1}}$$

$$\iff \frac{1}{\lambda_i} > \frac{1}{\lambda_{i+1}}$$

$$\iff \lambda_i < \lambda_{i+1} \,.$$

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

$$F_i(x) = F_{i+1}(x) \iff \left(\frac{x}{\lambda_i}\right)^{k_i} = \left(\frac{x}{\lambda_{i+1}}\right)^{k_{i+1}}$$

$$\iff \exp\left(k_i(\log x - \log \lambda_i)\right) = \exp\left(k_{i+1}(\log x - \log \lambda_{i+1})\right)$$

$$\iff k_i(\log x - \log \lambda_i) = k_{i+1}(\log x - \log \lambda_{i+1})$$

$$\iff k_i \log x - k_i \log \lambda_i = k_{i+1} \log x - k_{i+1} \log \lambda_{i+1}$$

$$\iff k_i \log x - k_{i+1} \log x = k_i \log \lambda_i - k_{i+1} \log \lambda_{i+1}$$

$$\iff (k_i - k_{i+1}) \log x = k_i \log \lambda_i - k_{i+1} \log \lambda_{i+1}$$

$$\iff \log x = \frac{k_i \log \lambda_i - k_{i+1} \log \lambda_{i+1}}{k_i - k_{i+1}}$$

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

Therefore the two CDFs intersect, and so *(C2)* cannot be satisfied. In conclusion,

$$\textit{(C2)} \iff (k_i = k_{i+1}) \wedge (\lambda_i < \lambda_{i+1}) \quad \forall 1 \leq i \leq d - 1 \,.$$

Gamma distribution

Let $1 \leq i \leq d - 1$. We have that

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

Suppose that we have two such distributions,

$$F_1(x) = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} x^{\alpha_1} \exp(\beta_1 x) \quad \text{and} \quad F_2(x) = \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} x^{\alpha_2} \exp(\beta_2 x),$$

with $\alpha_i, \beta_i > 0$ the shape and rate parameters. Denote their respective PDFs by f_1 and f_2. Let $x > 0$, and define

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

so that if $(i, i+1) = (1, 2)$,

$$\textit{(C2)} \iff \phi(x) < 0,$$

and if $(i, i+1) = (2, 1)$,

$$\textit{(C2)} \iff \phi(x) > 0.$$

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

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

and

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

Then

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

Case 1: $(\alpha_1 \leq \alpha_2 \wedge \beta_1 \geq \beta_2) \wedge \neg(\alpha_1 = \alpha_2 \wedge \beta_1 = \beta_2)$

We have that $\alpha_2 - \alpha_1 \geq 0$ and $\beta_1 - \beta_2 \geq 0$, 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 C.1.   *1. Let $f : (a, b] \to \mathbb{R}$ 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 $f : [a, b) \to \mathbb{R}$ 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 $f(z) = \lim_{y \to a^+} f(z) \geq \lim_{y \to a^+} f(y) + \frac{c}{2} = L + \frac{c}{2} > L$. Furthermore, $f(b) = \lim_{z \to b^-} f(z) \geq \lim_{x \to b^-} (L + \frac{c}{2}) = L + \frac{c}{2} > L$. $\qquad\square$

By Lemma C.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 $(i, i+1) = (1, 2)$, (C2) is satisfied, and if $(i, i+1) = (2, 1)$, (C2) 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 C.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 C.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 $(i, i+1) = (2, 1)$, (C2) is violated.

In conclusion, we have shown that

$$\text{(C2)} \iff (\alpha_i \leq \alpha_{i+1} \wedge \beta_i \geq \beta_{i+1}) \wedge \neg(\alpha_i = \alpha_{i+1} \wedge \beta_i = \beta_{i+1}) \quad \forall 1 \leq i \leq d - 1 \,.$$

D   Metropolis-Hastings algorithm implementation details

In Tables 28-47, we show details on our implementations of the Metropolis-Hastings algorithm, including the proposal distributions, initialisation, sample size, traceplots, histograms, and acceptance rates.
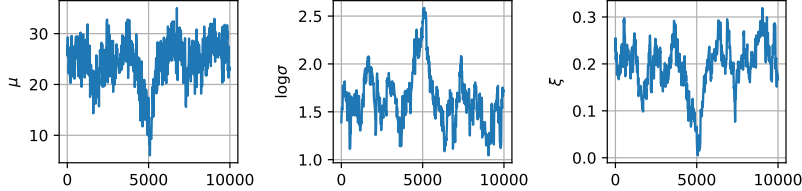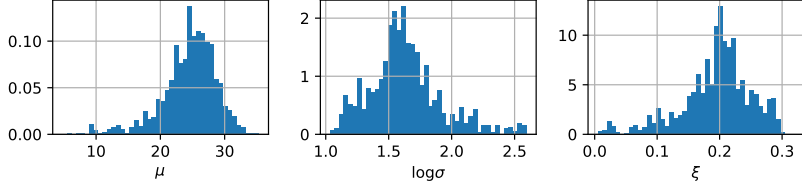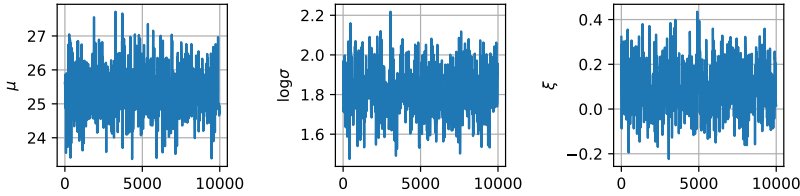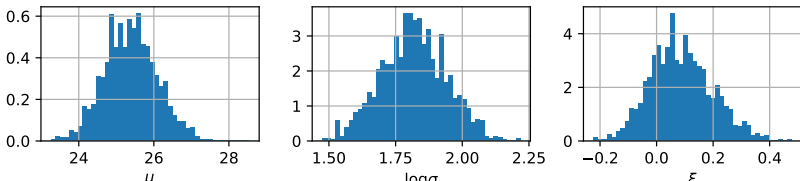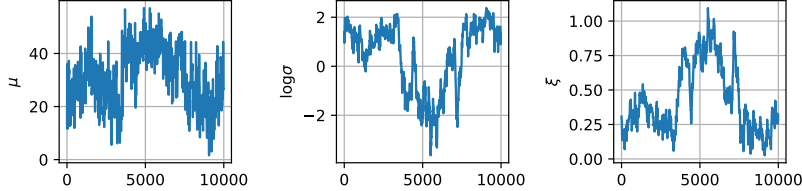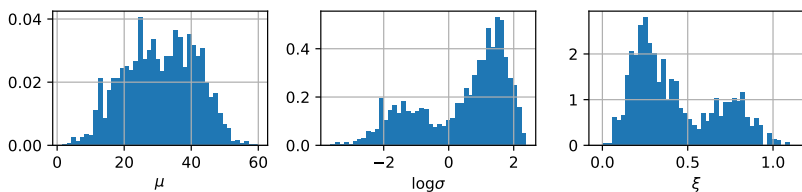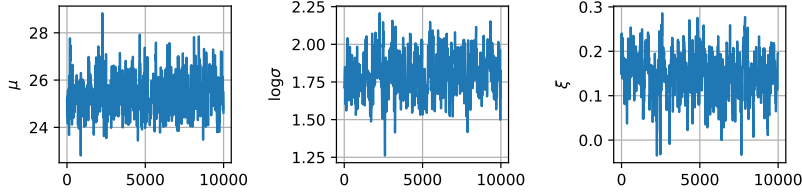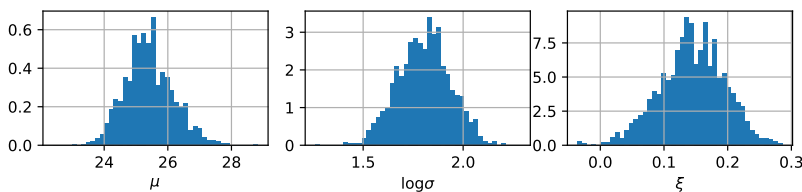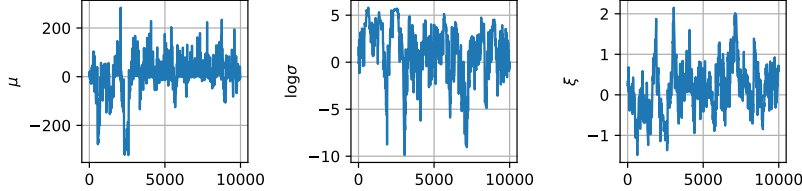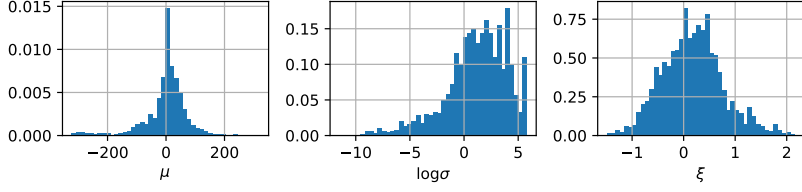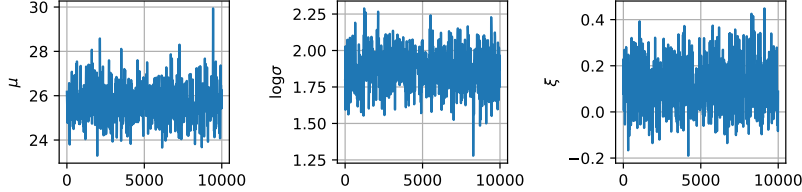
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 20^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.1^2)$
Initialisation	25, 0.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.137, 0.124, 0.121

Table 18: Metropolis-Hastings algorithm for π_θ^{G3} for Poisson process simulation study
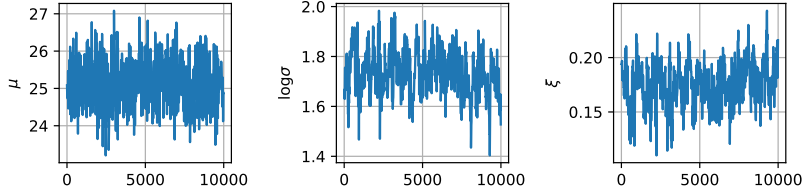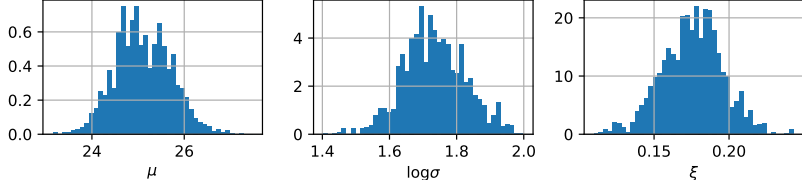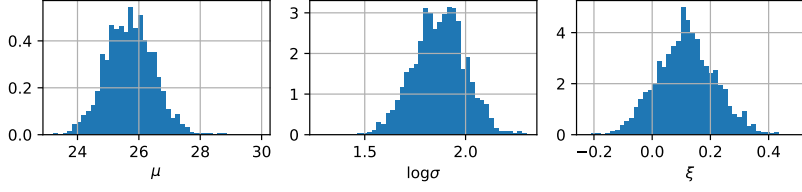
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.1^2)$
Initialisation	26, 1.8, 0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.172, 0.108, 0.13

Table 19: Metropolis algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{PPP}}}^{\mathrm{G3}}$ for Poisson process simulation study
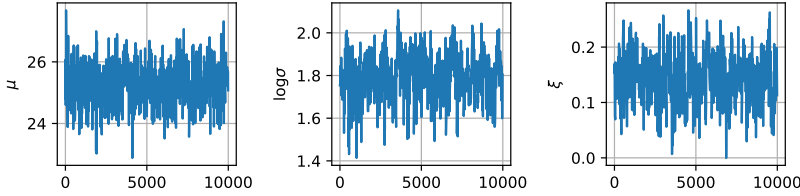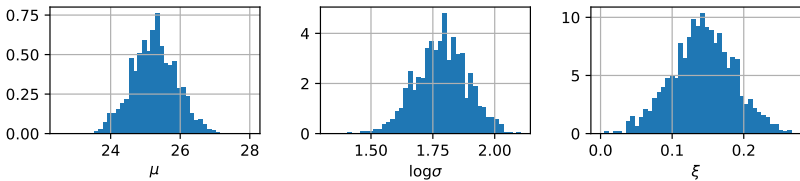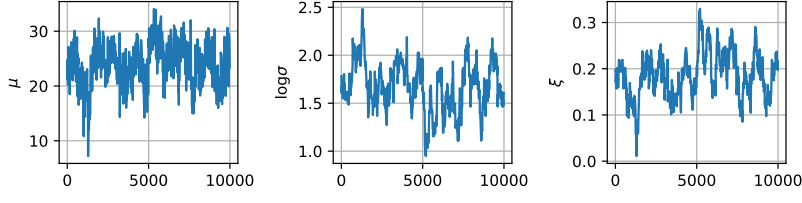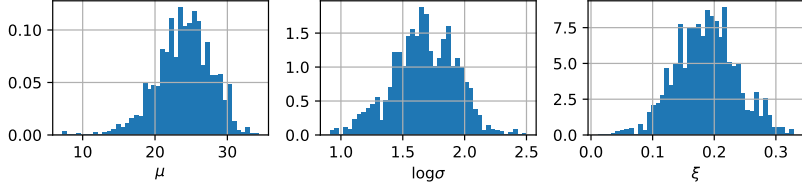
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log\sigma^* \sim \mathcal{N}(\log\sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.2^2)$
Initialisation	26,　1.8,　0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.183,　0.158,　0.163

Table 20: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{PPP}}}^{\mathrm{G2}}$ for Poisson process simulation study
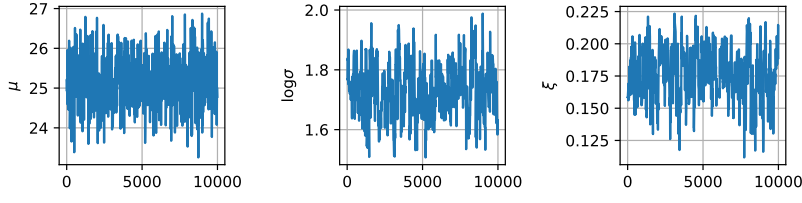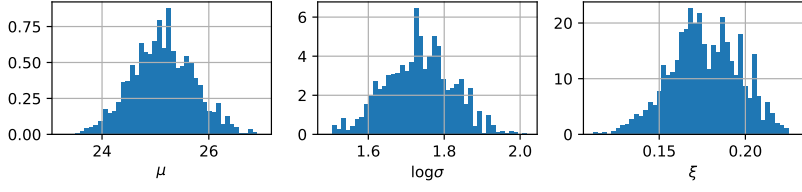
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log\sigma^* \sim \mathcal{N}(\log\sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.4^2)$
Initialisation	26,　1.8,　0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.197,　0.201,　0.238

Table 21: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{PPP}}}^{\mathrm{G1}}$ for Poisson process simulation study
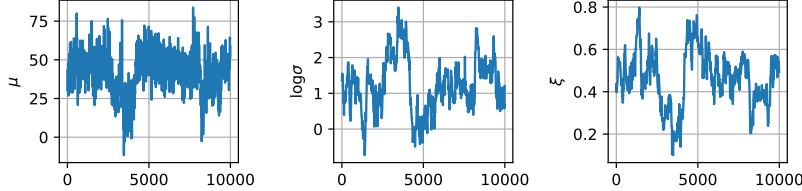
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 40^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 1^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.3^2)$
Initialisation	25, 0.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.138, 0.16, 0.117

Table 22: Metropolis-Hastings algorithm for $\pi_\theta^{\mathrm{MEC}}$ for Poisson process simulation study
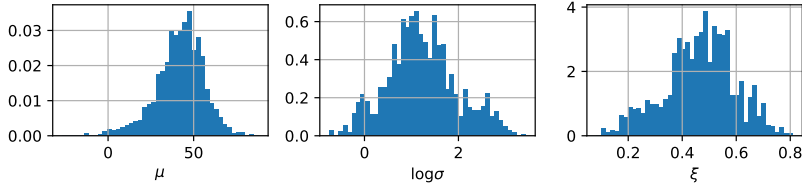
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.3^2)$
Initialisation	26, 1.8, 0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.196, 0.185, 0.127

Table 23: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{PPP}}}^{\mathrm{MEC}}$ for Poisson process simulation study
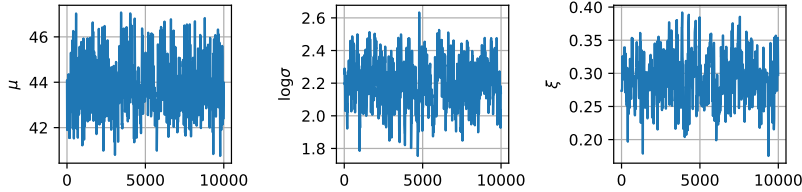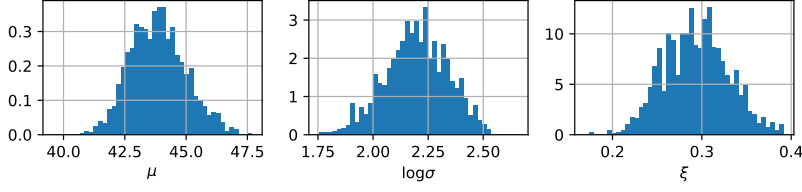
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 100^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 2.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.8^2)$
Initialisation	25, 0.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.283, 0.264, 0.222

Table 24: Metropolis-Hastings algorithm for π_θ^{E} for Poisson process simulation study
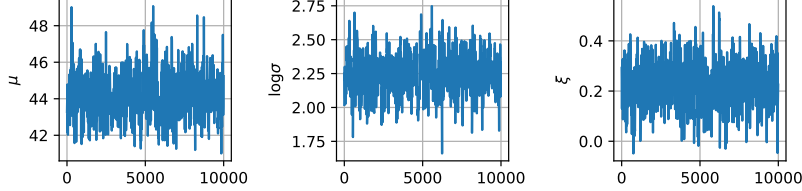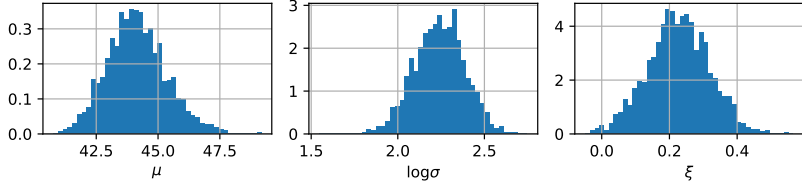
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.4^2)$
Initialisation	26, 1.8, 0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.21, 0.24, 0.274

Table 25: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{PPP}}}^{\mathrm{E}}$ for Poisson process simulation study

Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 20^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.1^2)$
Initialisation	25, 0.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.144, 0.119, 0.124

Table 26: Metropolis-Hastings algorithm for π_θ^{TN} for Poisson process simulation study
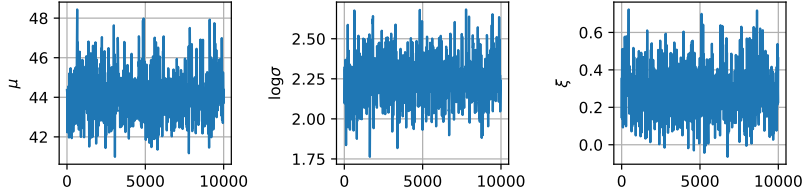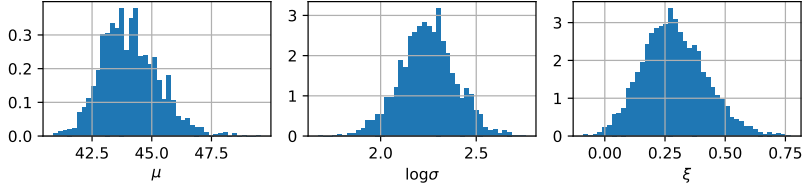
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.1^2)$
Initialisation	26, 1.8, 0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.172, 0.104, 0.123

Table 27: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\text{PPP}}}^{\text{TN}}$ for Poisson process simulation study
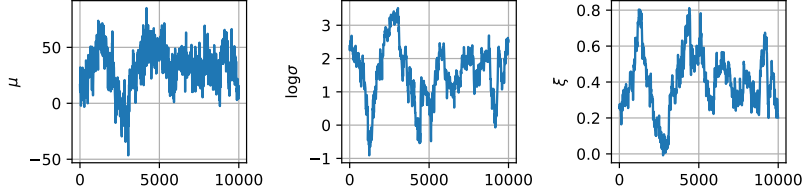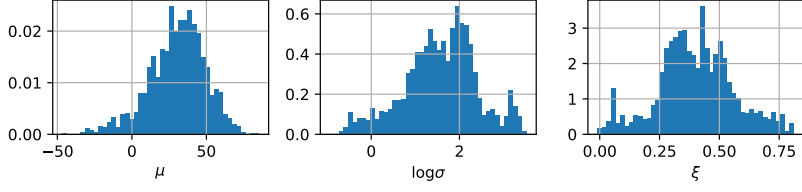
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 40^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 1.2^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.3^2)$
Initialisation	50, 1.5, 0.4
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.273, 0.138, 0.099

Table 28: Metropolis-Hastings algorithm for π_θ^{G3} for pseudo-data simulation study
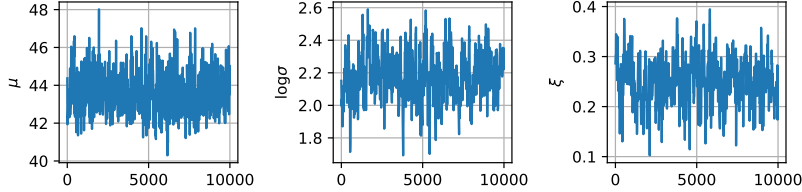
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.3^2)$
Initialisation	43, 2.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.258, 0.197, 0.098

Table 29: Metropolis algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{PD}}}^{\mathrm{G3}}$ for pseudo-data simulation study
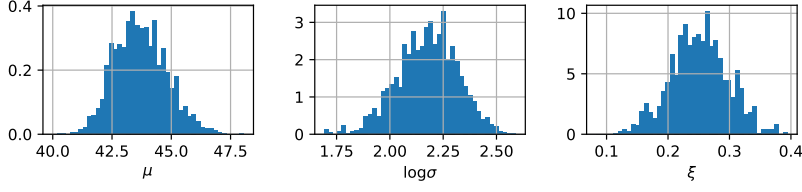
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.3^2)$
Initialisation	43, 2.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.274, 0.221, 0.275

Table 30: Metropolis-Hastings algorithm for $\pi^{\mathrm{G2}}_{\theta|\mathbf{x}^{\mathrm{PD}}}$ for pseudo-data simulation study
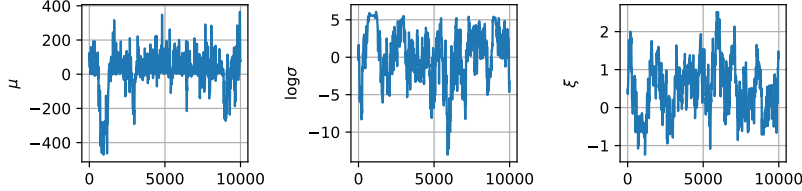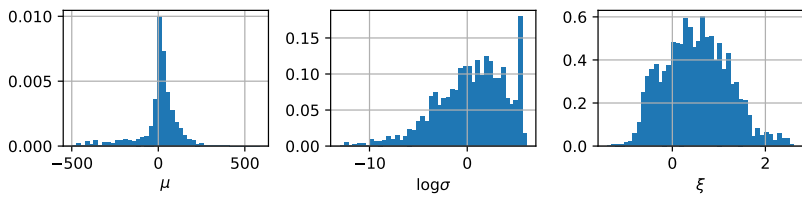
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 5^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.15^2)$
Initialisation	43, 2.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.221, 0.241, 0.611

Table 31: Metropolis-Hastings algorithm for $\pi^{\mathrm{G1}}_{\theta|\mathbf{x}^{\mathrm{PD}}}$ for pseudo-data simulation study
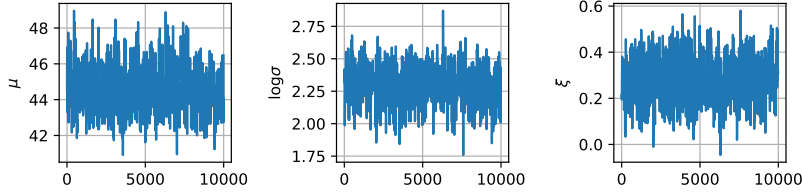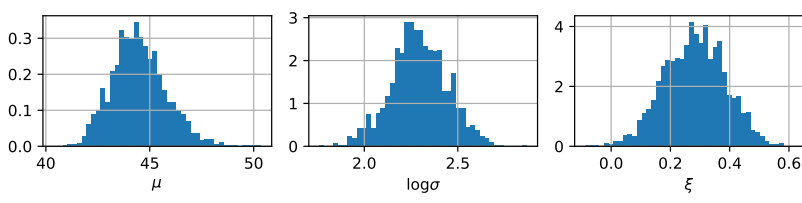
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 40^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.6^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.2^2)$
Initialisation	50, 1.5, 0.4
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.235, 0.204, 0.146

Table 32: Metropolis-Hastings algorithm for $\pi_\theta^{\mathrm{MEC}}$ for pseudo-data simulation study
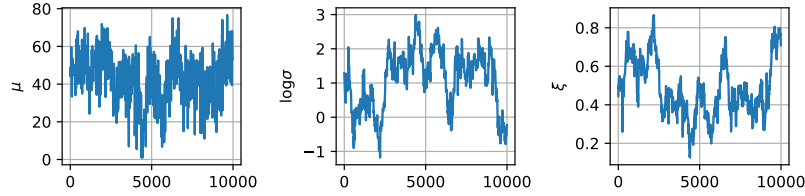
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.3^2)$
Initialisation	43, 2.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.255, 0.184, 0.11

Table 33: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{PD}}}^{\mathrm{MEC}}$ for pseudo-data simulation study
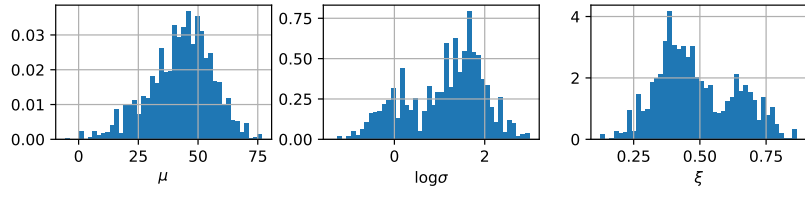
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 150^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 3^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.6^2)$
Initialisation	50, 1.5, 0.4
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.278, 0.248, 0.282

Table 34: Metropolis-Hastings algorithm for π_θ^{E} for pseudo-data simulation study
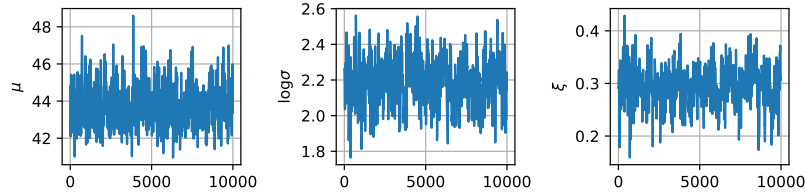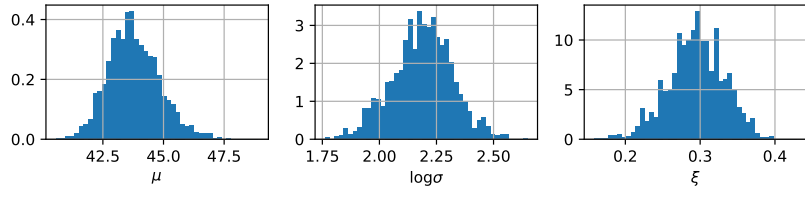
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.4^2)$
Initialisation	43, 2.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.279, 0.246, 0.272

Table 35: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{PD}}}^{\mathrm{E}}$ for pseudo-data simulation study

Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 100^2), \quad \log\sigma^* \sim \mathcal{N}(\log\sigma, 1.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.25^2)$
Initialisation	50, 1.5, 0.4
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.111, 0.11, 0.119

Table 36: Metropolis-Hastings algorithm for π_θ^{TN} for pseudo-data simulation study

Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 4^2), \quad \log\sigma^* \sim \mathcal{N}(\log\sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.25^2)$
Initialisation	43, 2.2, 0.25
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.248, 0.199, 0.12

Table 37: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\text{PD}}}^{\text{TN}}$ for pseudo-data simulation study
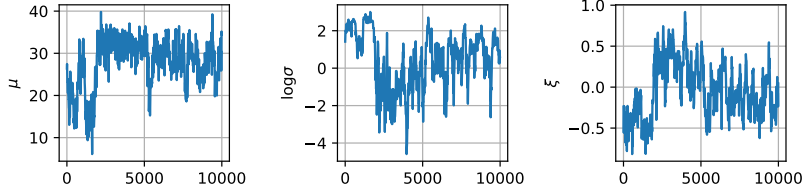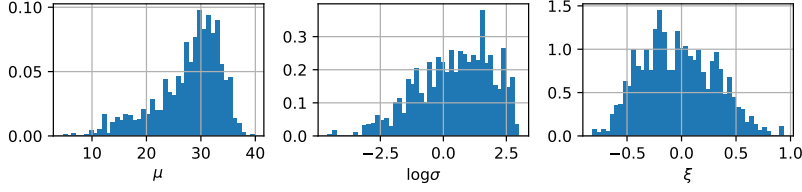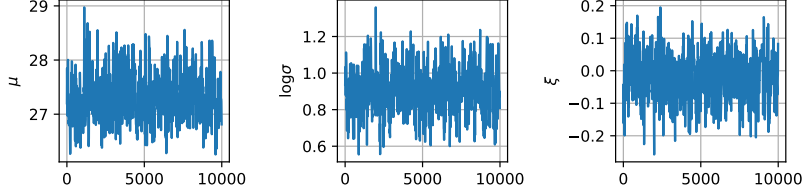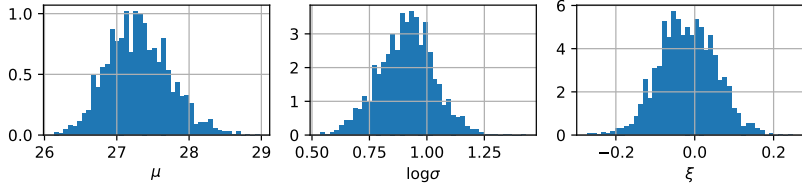
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 25^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 3^2), \quad \xi^* \sim \mathcal{N}(\xi, 1^2)$
Initialisation	27, 3.7, -0.12
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.112, 0.126, 0.116

Table 38: Metropolis-Hastings algorithm for π_θ^{G3} for wind speed data
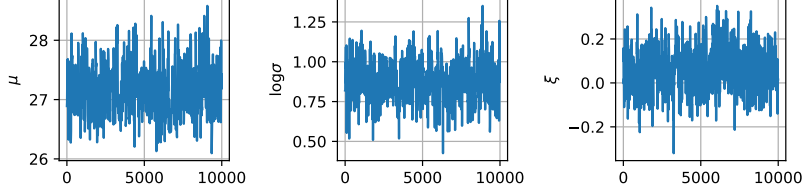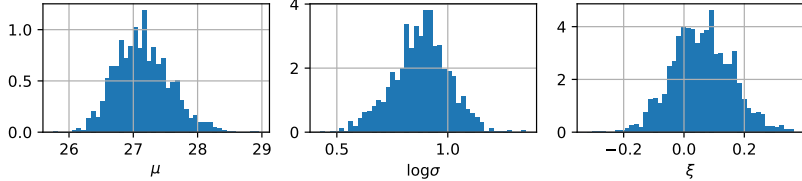
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 2^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.5^2)$
Initialisation	27, 0.8, 0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.182, 0.198, 0.149

Table 39: Metropolis algorithm for $\pi_{\theta|\mathbf{x}^{\text{ws}}}^{\text{G3}}$ for wind speed data
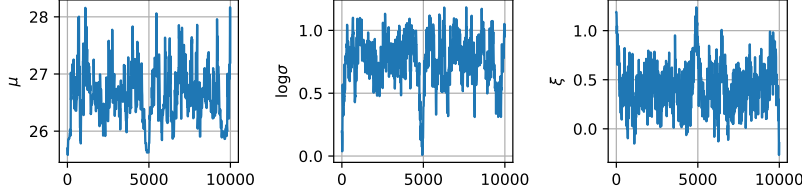
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 2^2), \quad \log\sigma^* \sim \mathcal{N}(\log\sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.6^2)$
Initialisation	27, 0.8, 0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.163, 0.204, 0.169

Table 40: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{ws}}}^{\mathrm{G2}}$ for wind speed data
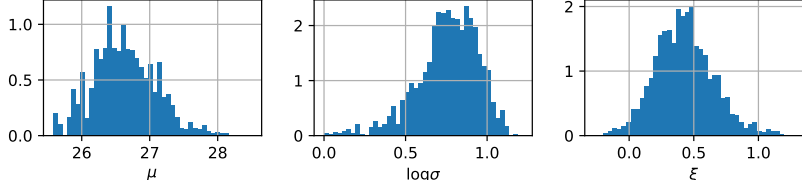
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 2^2), \quad \log\sigma^* \sim \mathcal{N}(\log\sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.6^2)$
Initialisation	27, 0.8, 0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.103, 0.186, 0.267

Table 41: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{ws}}}^{\mathrm{G1}}$ for wind speed data
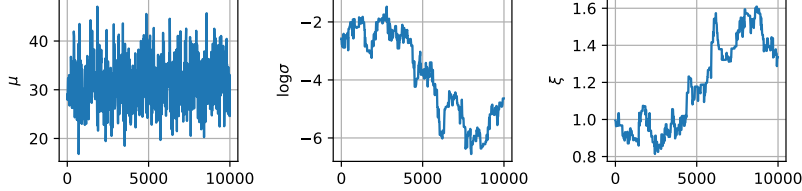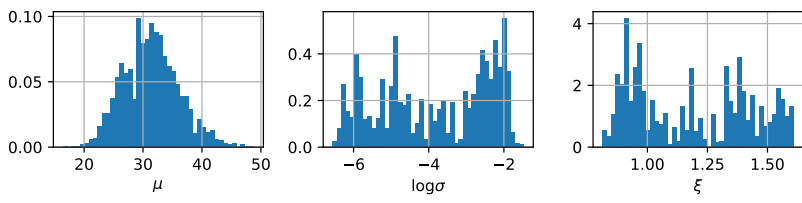
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 30^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 5^2), \quad \xi^* \sim \mathcal{N}(\xi, 1^2)$
Initialisation	27, 3.7, -0.12
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.171, 0.036, 0.027

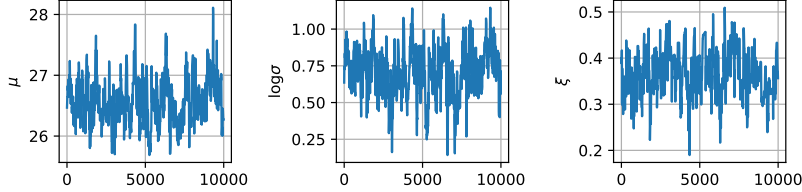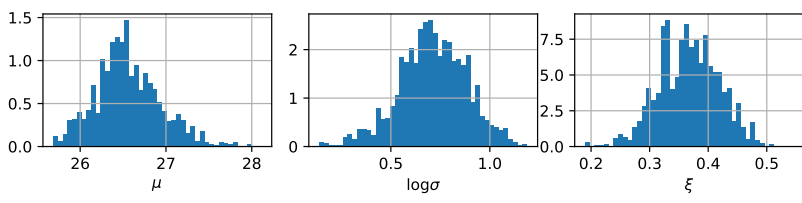Table 42: Metropolis-Hastings algorithm for π_θ^{MEC} for wind speed data

Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 2^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.5^2)$
Initialisation	27, 0.8, 0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.112, 0.175, 0.072

Table 43: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\text{ws}}}^{\text{MEC}}$ for wind speed data
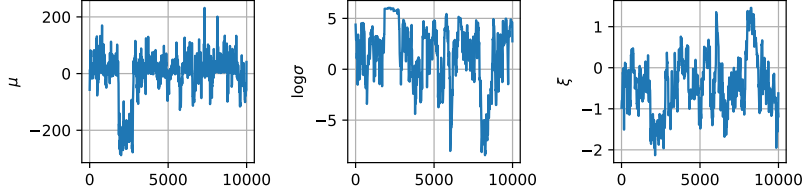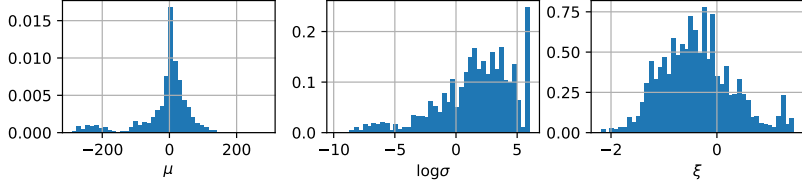
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 100^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 5^2), \quad \xi^* \sim \mathcal{N}(\xi, 1.5^2)$
Initialisation	$27, \quad 3.7, \quad -0.12$
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.238, 0.142, 0.128

Table 44: Metropolis-Hastings algorithm for π_θ^{E} for wind speed data

Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 2^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.5^2)$
Initialisation	$27, \quad 0.8, \quad 0.15$
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.17, 0.2, 0.197

Table 45: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{ws}}}^{\mathrm{E}}$ for wind speed data
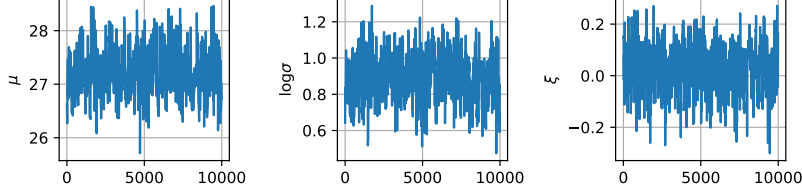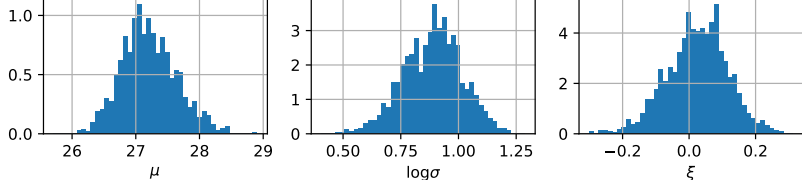
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 20^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 2^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.8^2)$
Initialisation	27, 3.7, -0.12
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.131, 0.146, 0.142

Table 46: Metropolis-Hastings algorithm for π_θ^{TN} for wind speed data
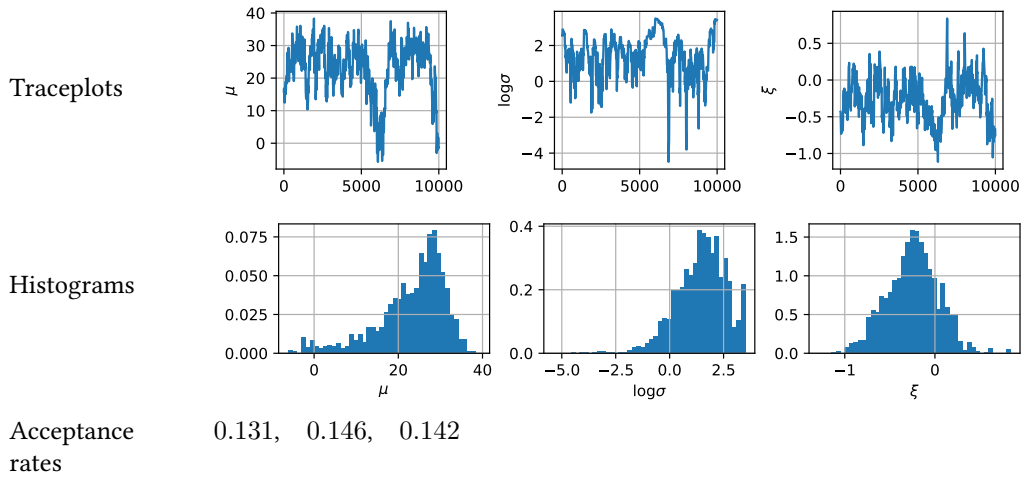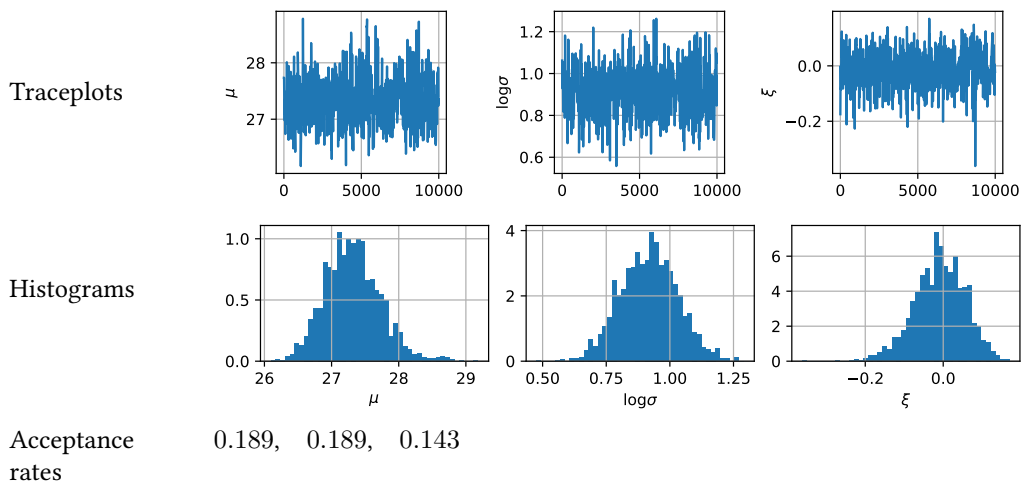
Proposal distributions	$\mu^* \sim \mathcal{N}(\mu, 2^2), \quad \log \sigma^* \sim \mathcal{N}(\log \sigma, 0.5^2), \quad \xi^* \sim \mathcal{N}(\xi, 0.5^2)$
Initialisation	27, 0.8, 0.15
Sample size	10,000 (after burn-in period of 1,000)
Traceplots	
Histograms	
Acceptance rates	0.189, 0.189, 0.143

Table 47: Metropolis-Hastings algorithm for $\pi_{\theta|\mathbf{x}^{\mathrm{ws}}}^{\mathrm{TN}}$ for wind speed data

D.1   SS: 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 nonzero can then be calculated as the mean 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 $\eta := (\mu, \sigma, \beta, \gamma)$, and we suppose that γ does not depend on the other parameters.

If we have a prior π_θ for (μ, σ, ξ), then we can set

$$\pi^{\text{SS}}_{\mu,\sigma,\beta} := \pi_\theta$$
$$\pi^{\text{SS}}_\gamma(\gamma) := \alpha^\gamma(1-\alpha)^{1-\gamma}$$
$$\implies \pi^{\text{SS}}_\eta(\eta) = \pi_\theta(\mu, \sigma, \beta)\alpha^\gamma(1-\alpha)^{1-\gamma},$$

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

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

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

$$\pi^{\text{SS}}_{\gamma|\mu,\sigma,\beta,\mathbf{x}}(0 \mid \mu, \sigma, \beta, \mathbf{x}) \propto (1-\alpha)L(\mu, \sigma, 0 \mid \mathbf{x})$$

and

$$\pi^{\text{SS}}_{\gamma|\mu,\sigma,\beta,\mathbf{x}}(1 \mid \mu, \sigma, \beta, \mathbf{x}) \propto \alpha L(\mu, \sigma, \beta \mid \mathbf{x}),$$

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 set $\pi_\theta = \pi^{\text{G3}}_\theta$, and for an uninformative prior on γ, we set $\alpha = 0.5$.

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]   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.

[9]   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.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.