Supplementary Material

S1. Invariance of the estimators to the selection of the reference locations

S1.1. Proof

To prove the invariance of the estimators to the selection of the reference locations, it is sufficient to show that, given two data transformations $\boldsymbol{Z}_a$ and $\boldsymbol{Z}_b$ (see the three cases below for details; in the third case, they are denoted as $\boldsymbol{Z}_{a,b}$ and $\boldsymbol{Z}_{c,d}$), the function $\boldsymbol{h}_{a\to b} : \boldsymbol{Z}_a \mapsto \boldsymbol{Z}_b$ is a diffeomorphism, so that the corresponding likelihood functions are proportional to each other, with proportionality factor given by the Jacobian term (see Davison, 2003, 4.1.2). In the sequel $\mathbf{1}_d$ and $\mathbf{0}_d$ will denote the ones vector and the zero vector in $\mathbb{R}^d$, respectively. The matrix $\mathbf{I}_d$ is the $d \times d$ identity matrix. Let $\boldsymbol{W} = (W_i,\, i \in M)$, $M = \{1, \ldots, m\}$, be the m-dimensional centred Gaussian vector with positive definite covariance matrix $\boldsymbol{\Sigma}$.

Gaussian location mixtures

For $k \in M = \{1, \ldots, m\}$, let $\boldsymbol{Z}_k = \left(Z_i^{(k)}\right)_{i \in M \setminus \{k\}}$, with $Z_i^{(k)} = W_i - W_k$. We study the transformation $\boldsymbol{h}_{a\to b} : \boldsymbol{Z}_a \mapsto \boldsymbol{Z}_b$, for $a, b \in M$. The component-wise map $\boldsymbol{h}_{a\to b}$ is given by

$$\begin{aligned} Z_a^{(b)} &= W_a - W_b = -(W_b - W_a) = -Z_b^{(a)}, \\ Z_i^{(b)} &= W_i - W_b = (W_i - W_a) - (W_b - W_a) = Z_i^{(a)} - Z_b^{(a)}, \quad i \in M \setminus \{a, b\}. \end{aligned}$$

The inverse map $\boldsymbol{h}_{a\to b}^{-1}$ from $\boldsymbol{Z}_b$ to $\boldsymbol{Z}_a$ is given by

$$\begin{aligned} Z_b^{(a)} &= W_b - W_a = -(W_a - W_b) = -Z_a^{(b)}, \\ Z_i^{(a)} &= W_i - W_a = (W_i - W_b) - (W_a - W_b) = Z_i^{(b)} - Z_a^{(b)}, \quad i \in M \setminus \{a, b\}. \end{aligned}$$

The map $\boldsymbol{h}_{a\to b}$ is a diffeomorphism, and we derive the Jacobian matrix and its determinant. We order the coordinates of $\boldsymbol{z}_a$ as $(z_b^{(a)}, \boldsymbol{z}_D^{(a)})$, where $\boldsymbol{z}_D^{(a)} \in \mathbb{R}^{m-2}$ and $D = M \setminus \{a, b\}$. Similarly we order the coordinates of $\boldsymbol{z}_b$ as $(z_a^{(b)}, \boldsymbol{z}_D^{(b)})$, $\boldsymbol{z}_D^{(b)} \in \mathbb{R}^{m-2}$. The inverse map can be written in block form as $z_b^{(a)} = -z_a^{(b)}$ and $\boldsymbol{z}_D^{(a)} = \boldsymbol{z}_D^{(b)} - z_a^{(b)} \mathbf{1}_{m-2}$. The Jacobian matrix is

$$\boldsymbol{J} = \frac{\partial\, \boldsymbol{h}_{a\to b}^{-1}(\boldsymbol{z}_b)}{\partial \boldsymbol{z}_b} = \frac{\partial(z_b^{(a)}, \boldsymbol{z}_D^{(a)})}{\partial(z_a^{(b)}, \boldsymbol{z}_D^{(b)})} = \begin{pmatrix} -1 & \mathbf{0}_{m-2}^\top \\ -\mathbf{1}_{m-2} & \mathbf{I}_{m-2} \end{pmatrix}.$$

Since this matrix is block lower triangular, we have that $|\det(\boldsymbol{J})| = 1$. Therefore, $f_{\boldsymbol{Z}_a}(\boldsymbol{z}_a; \boldsymbol{\Sigma}) = f_{\boldsymbol{Z}_b}(\boldsymbol{z}_b; \boldsymbol{\Sigma})$, for any a and b, i.e. the likelihood is the same, namely

$$\mathcal{L}(\boldsymbol{\Sigma}; \boldsymbol{z}_b) = f_{\boldsymbol{Z}_b}(\boldsymbol{z}_b; \boldsymbol{\Sigma}) = f_{\boldsymbol{Z}_a}(\boldsymbol{h}_{a\to b}^{-1}(\boldsymbol{z}_b); \boldsymbol{\Sigma}) = f_{\boldsymbol{Z}_a}(\boldsymbol{z}_a; \boldsymbol{\Sigma}) = \mathcal{L}(\boldsymbol{\Sigma}; \boldsymbol{z}_a).$$

Gaussian scale mixtures

For $k \in M = \{1, \ldots, m\}$, let $\boldsymbol{Z}_k = \left(Z_i^{(k)}\right)_{i \in M \setminus \{k\}}$, with $Z_i^{(k)} = W_i / W_k$. We study the transformation $\boldsymbol{h}_{a \to b} : \boldsymbol{Z}_a \mapsto \boldsymbol{Z}_b$, for $a, b \in M$. The component-wise map $\boldsymbol{h}_{a \to b}$ is given by

$$\begin{aligned} Z_a^{(b)} &= W_a / W_b = \frac{1}{W_b / W_a} = \frac{1}{Z_b^{(a)}}, \\ Z_i^{(b)} &= W_i / W_b = \frac{W_i / W_a}{W_b / W_a} = \frac{Z_i^{(a)}}{Z_b^{(a)}}, \quad i \in M \setminus \{a, b\}. \end{aligned}$$

The inverse map, $\boldsymbol{h}_{a \to b}^{-1}$, from $\boldsymbol{Z}_b$ to $\boldsymbol{Z}_a$ is given by

$$\begin{aligned} Z_b^{(a)} &= W_b / W_a = \frac{1}{W_a / W_b} = \frac{1}{Z_a^{(b)}}, \\ Z_i^{(a)} &= W_i / W_a = \frac{W_i / W_b}{W_a / W_b} = \frac{Z_i^{(b)}}{Z_a^{(b)}}, \quad i \in M \setminus \{a, b\}. \end{aligned}$$

Therefore the map $\boldsymbol{h}_{a \to b}$ is a diffeomorphism on the open set $\{W_a W_b \neq 0, \text{ a.s.}\}$. We calculate the Jacobian determinant of the inverse transformation. Once again we order the coordinates of $\boldsymbol{z}_a$ as $(z_b^{(a)}, \boldsymbol{z}_D^{(a)})$ and the coordinates of $\boldsymbol{z}_b$ as $(z_a^{(b)}, \boldsymbol{z}_D^{(b)})$, with $\boldsymbol{z}_D^{(a)}, \boldsymbol{z}_D^{(b)} \in \mathbb{R}^{m-2}$. The inverse map can be written in block form as $z_b^{(a)} = (z_a^{(b)})^{-1}$ and $\boldsymbol{z}_D^{(a)} = (z_a^{(b)})^{-1} \boldsymbol{z}_D^{(b)}$. The Jacobian matrix is

$$\boldsymbol{J} = \frac{\partial\, \boldsymbol{h}_{a \to b}^{-1}(\boldsymbol{z}_b)}{\partial \boldsymbol{z}_b} = \frac{\partial (z_b^{(a)}, \boldsymbol{z}_D^{(a)})}{\partial (z_a^{(b)}, \boldsymbol{z}_D^{(b)})} = \begin{pmatrix} -(z_a^{(b)})^{-2} & \boldsymbol{0}_{m-2}^\top \\ -(z_a^{(b)})^{-2} \boldsymbol{y}_D & (z_a^{(b)})^{-1} \mathbf{I}_{m-2} \end{pmatrix}.$$

Since this matrix is block lower triangular, $|\det(\boldsymbol{J})| = |-(z_a^{(b)})^{-2} (z_a^{(b)})^{-(m-2)}| = |z_a^{(b)}|^{-m}$. Therefore

$$\mathcal{L}(\boldsymbol{\Sigma}; \boldsymbol{z}_b) = f_{\boldsymbol{Z}_b}(\boldsymbol{z}_b; \boldsymbol{\Sigma}) = |z_a^{(b)}|^{-m} f_{\boldsymbol{Z}_a}(\boldsymbol{h}_{a \to b}^{-1}(\boldsymbol{z}_b); \boldsymbol{\Sigma}) = |z_a^{(b)}|^{-m} f_{\boldsymbol{Z}_a}(\boldsymbol{z}_a; \boldsymbol{\Sigma}) \propto \mathcal{L}(\boldsymbol{\Sigma}; \boldsymbol{z}_a),$$

so the likelihoods for $\boldsymbol{\Sigma}$ are proportional, with proportionality factor $|z_a^{(b)}|^{-m} = |w_a / w_b|^{-m}$, that depends only on the observation $\boldsymbol{w} = (w_1, \ldots, w_m)^T$, not on $\boldsymbol{\Sigma}$.

Gaussian location-scale mixtures

For $j, k \in M = \{1, \ldots, m\}$, $j \neq k$, let $\boldsymbol{Z}_{j,k} = \left(Z_i^{(j,k)}\right)_{i \in M \setminus \{j,k\}}$ with $Z_i^{(j,k)} = (W_i - W_j)/(W_k - W_j)$. We study the transformation $\boldsymbol{h}_{(a,b) \to (c,d)} : \boldsymbol{Z}_{a,b} \mapsto \boldsymbol{Z}_{c,d}$, for $a, b, c, d \in M$. It is convenient to introduce the extended components

$$\widetilde{Z}_j^{(j,k)} = 0, \qquad \widetilde{Z}_k^{(j,k)} = 1, \qquad \widetilde{Z}_i^{(j,k)} = Z_i^{(j,k)} \quad (i \neq j, k),$$

such that, for all $i \in M$,

$$\widetilde{Z}_i^{(j,k)} = \frac{W_i - W_j}{W_k - W_j}.$$

Assume $W_b \neq W_a$ and $W_d \neq W_c$ a.s.. For any $i \in M \setminus \{c, d\}$, the component-wise map $\boldsymbol{h}_{(a,b) \to (c,d)}$ is given by

$$\widetilde{Z}_i^{(c,d)} = \frac{W_i - W_c}{W_d - W_c} = \frac{(W_i - W_a) - (W_c - W_a)}{(W_d - W_a) - (W_c - W_a)}$$
$$= \frac{\dfrac{W_i - W_a}{W_b - W_a} - \dfrac{W_c - W_a}{W_b - W_a}}{\dfrac{W_d - W_a}{W_b - W_a} - \dfrac{W_c - W_a}{W_b - W_a}} = \frac{\widetilde{Z}_i^{(a,b)} - \widetilde{Z}_c^{(a,b)}}{\widetilde{Z}_d^{(a,b)} - \widetilde{Z}_c^{(a,b)}},$$

valid whenever $\widetilde{Z}_d^{(a,b)} \neq \widetilde{Z}_c^{(a,b)}$ a.s. (or equivalently $W_d \neq W_c$ a.s.). By symmetry, swapping the roles of (a,b) and (c,d), the inverse map $\boldsymbol{h}^{-1}_{(a,b)\to(c,d)}$ is

$$\widetilde{Z}_i^{(a,b)} = \frac{\widetilde{Z}_i^{(c,d)} - \widetilde{Z}_a^{(c,d)}}{\widetilde{Z}_b^{(c,d)} - \widetilde{Z}_a^{(c,d)}}, \qquad i \in M \setminus \{a,b\}, \tag{S1}$$

valid whenever $\widetilde{Z}_b^{(c,d)} \neq \widetilde{Z}_a^{(c,d)}$ a.s. (or equivalently $W_b \neq W_a$ a.s.). We remark that the formulas above remain correct even if some indices overlap (e.g. $c = a$), provided one uses the extended values $\widetilde{Z}_a^{(a,b)} = 0$, $\widetilde{Z}_b^{(a,b)} = 1$, $\widetilde{Z}_c^{(c,d)} = 0$, $\widetilde{Z}_d^{(c,d)} = 1$, and restricts to the events where the relevant denominators are non-zero.

Finally, we show that the map $\boldsymbol{h}_{(a,b)\to(c,d)}$ is a diffeomorphism, starting from the case in which the four indices are all distinct: $\{a,b\} \cap \{c,d\} = \varnothing$. In this case, $\boldsymbol{Z}_{c,d}$ contains the components $Z_a^{(c,d)}$ and $Z_b^{(c,d)}$. The inverse map $\boldsymbol{h}^{-1}_{(a,b)\to(c,d)}$ can be written for all $i \in M \setminus \{a,b\}$ as in (S1). Here $\widetilde{Z}_d^{(c,d)} = 1$ and $\widetilde{Z}_c^{(c,d)} = 0$ are constants; all other $\widetilde{Z}_i^{(c,d)}$ with $i \notin \{c,d\}$ are coordinates of $\boldsymbol{Z}_{(c,d)}$. We conclude that the map $\boldsymbol{h}_{(a,b)\to(c,d)}$ is a diffeomorphism on the open set $\{(W_b - W_a)(W_d - W_c) \neq 0,$ a.s.$\}$ and we derive the Jacobian matrix and its determinant.

Consider the Jacobian matrix of $\boldsymbol{h}^{-1}_{(a,b)\to(c,d)}$ with respect to the $(m-2)$ variables $(Z_i^{(c,d)})_{i\in S\setminus\{c,d\}}$ (which include $Z_a^{(c,d)}$ and $Z_b^{(c,d)}$). A direct determinant computation yields

$$\det\left(\frac{\partial (Z_i^{(a,b)})_{i\neq a,b}}{\partial (Z_u^{(c,d)})_{u\neq c,d}}\right) = \frac{1}{(Z_b^{(c,d)} - Z_a^{(c,d)})^{\,m-1}}.$$

Therefore the absolute Jacobian determinant of the inverse map is $|Z_b^{(c,d)} - Z_a^{(c,d)}|^{-(m-1)}$.

If the four indices are not all distinct (e.g. $a = c$), the same component-wise formulas for the map and inverse still hold, but the Jacobian expression must be written using a pair of variable coordinates present in the source vector (one typically chooses two indices whose extended values are known constants in the target coordinate system). The generic formula above is the cleanest closed form.

Therefore

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\Sigma}; \boldsymbol{z}_{c,d}) &= f_{\boldsymbol{Z}_{c,d}}(\boldsymbol{z}_{c,d}; \boldsymbol{\Sigma}) \\
&= |Z_b^{(c,d)} - Z_a^{(c,d)}|^{-(m-1)} f_{\boldsymbol{Z}_{a,b}}(\boldsymbol{h}^{-1}_{(a,b)\to(c,d)}(\boldsymbol{z}_{c,d}); \boldsymbol{\Sigma}) \\
&= |Z_b^{(c,d)} - Z_a^{(c,d)}|^{-(m-1)} f_{\boldsymbol{Z}_{a,b}}(\boldsymbol{z}_{a,b}; \boldsymbol{\Sigma}) \propto \mathcal{L}(\boldsymbol{\Sigma}; \boldsymbol{z}_{a,b}),
\end{aligned}$$

so the likelihoods for $\boldsymbol{\Sigma}$ are proportional, with proportionality factor $|Z_b^{(c,d)} - Z_a^{(c,d)}|^{-(m-1)} = |(w_b - w_a)/(w_d - w_c)|^{-(m-1)}$, that depends only on the observation $\boldsymbol{w} = (w_1, \ldots, w_m)^T$, not on $\boldsymbol{\Sigma}$.

S1.2. Numerical example

Let

$$(X_1, X_2, X_3)^\top = R\,(W_1, W_2, W_3)^\top$$

be a trivariate Gaussian scale mixture, with $R \sim \text{Gamma}(1,1)$ and

$$\begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} \sim \mathcal{N}_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_{\theta_W} = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix} \right).$$

This non-spatial toy example can be used to show that the selection of the reference locations (or of the reference variable, in this case) has no effect on the estimated parameter values. The following R code produces the maximum likelihood estimates for $\theta_W = (\rho_1, \rho_2, \rho_3)$ by selecting all the possible elements X_i, $i = 1, 2, 3$, as the reference one when computing the ratios that cancel the random scale R. Note that the value of the reference random variable has to be positive by construction. For example, if X_1 is selected as reference element (`index=1`), the vector of ratios is

$$\mathbf{Z} = \left(\frac{X_2}{X_1}, \frac{X_3}{X_1} \right)$$

and its density function is

$$f_{\mathbf{Z}}(\mathbf{z};\, \theta_W) = \; \pi^{-m/2} |\Sigma_{\theta_W}|^{-1/2} \Gamma\,(m/2) \left(\dot{\mathbf{z}}^\top \Sigma_{\theta_W}^{-1} \dot{\mathbf{z}} \right)^{-m/2},$$

with $m = 3$ and $\dot{\mathbf{z}} = (1, x_2/x_1, x_3/x_1)^\top$ (referred to as `z` in the code). Note that $\pi^{-m/2}$ and $\Gamma\,(m/2)$ are constants that can be excluded from the likelihood computations. By choosing `index=1,2,3` it is possible to see that the resulting estimated parameter values are numerically equivalent.

```
set.seed(123)
cor1 = 0.7
cor2 = 0.5
cor3 = 0.3
Sigma = matrix(c(1,cor1,cor2,cor1,1,cor3,cor2,cor3,1), nrow=3)

library(mvtnorm)
n = 1000
W = rmvnorm(n, sigma = Sigma)
R = rgamma(n, rate=1, shape=1)
X = R * W

nll = function(par, x, index=1){
  if(min(par) < 0) return(Inf)
  if(max(par) > 1) return(Inf)

  m = ncol(x)
  S = matrix(c(1,par[1],par[2],par[1],1,par[3],par[2],par[3],1), nrow=3)

  # vector of ratios z
  z = x/x[,index]

  # computation of z^T Sigma^{-1} z
  L = chol(S)
```

```
  y = forwardsolve(t(L), t(z))
  zT_Sigma_z = colSums(y^2)

  # negative log-likelihood (excluding constants)
  -sum(-0.5*log(det(S)) -m/2*log(zT_Sigma_z))
}

fit1 = optim(c(0.5,0.5,0.5), nll, method="Nelder-Mead", x=X, index=1)
fit2 = optim(c(0.5,0.5,0.5), nll, method="Nelder-Mead", x=X, index=2)
fit3 = optim(c(0.5,0.5,0.5), nll, method="Nelder-Mead", x=X, index=3)
fit1$par
fit2$par
fit3$par
```

S2. Identifiability of Gaussian location-scale mixtures

In the general case of Gaussian location-scale mixtures, i.e., when in equation (2) in the main text both S and R are non-degenerate random variables, identifiability of the different models and of the parameter $\theta_{S,R}$ deserves some discussions.

S2.1. Parameter identifiability

A first issue may be the identifiability of $\theta_{S,R}$ from the univariate distribution of $\bar{X}$, used in Step 2 of the proposed inferential method, as detailed in Section 5.2 in the main text. To check that the parameter of a specific model is identifiable, it is possible to compute the characteristic function of $\bar{X}$, keeping in mind that $\bar{X} = S + R\,\bar{W}$, $\bar{W} \sim \mathcal{N}\left(0, \sigma^2_{\bar{W}}\right)$ and $\sigma^2_{\bar{W}}$ is a known variance that depends on the value of θ_W estimated in Step 1. Then, $R\,\bar{W} \mid R = r \sim \mathcal{N}\left(0, r^2\sigma^2_{\bar{W}}\right)$, and the characteristic function of $\bar{X}$ is

$$\varphi_{\bar{X}}(t) = \mathbb{E}\left[\exp(itS + itR\bar{W})\right] = \mathbb{E}\left[\exp(itS)\right] \cdot \mathbb{E}\left[\exp\left(-\frac{1}{2}t^2R^2\sigma^2_{\bar{W}}\right)\right] = \varphi_S(t) \cdot M_{R^2}\left(-\frac{t^2\sigma^2_{\bar{W}}}{2}\right),$$

where M_{R^2} is the moment generating function of R^2. While the identifiability of $\theta_{S,R}$ is not guaranteed in general, we can show that it is ensured in the case for the two location-scale mixture models presented in this work, LSM1 and LSM2. Recall that, if $Y \sim \mathrm{E}(\lambda)$, then $\varphi_Y(t) = \lambda/(\lambda - it)$ and $M_Y(t) = \lambda/(\lambda - t)$. In model LSM1, $S \sim \mathrm{E}(\lambda)$ and $R^2 \sim \mathrm{E}(1/2)$, so

$$\varphi_{\bar{X}}(t) = \varphi_S(t) \cdot M_{R^2}\left(-\frac{t^2\sigma^2_{\bar{W}}}{2}\right) = \frac{\lambda}{\lambda - it} \cdot \frac{1}{1 + t^2\sigma^2_{\bar{W}}}.$$

In model LSM2, $S = S_1 - S_2$, with $S_1 \sim \mathrm{E}(\lambda_1)$ and $S_2 \sim \mathrm{E}(\lambda_2)$, and $R^2 \sim \mathrm{E}(1/2)$, so

$$\begin{aligned}\varphi_{\bar{X}}(t) &= \varphi_S(t) \cdot M_{R^2}\left(-\frac{t^2\sigma^2_{\bar{W}}}{2}\right) = \mathbb{E}\left[\exp(itS_1 - itS_2)\right] \cdot M_{R^2}\left(-\frac{t^2\sigma^2_{\bar{W}}}{2}\right) \\ &= \varphi_{S_1}(t) \cdot \varphi_{S_2}(-t) \cdot M_{R^2}\left(-\frac{t^2\sigma^2_{\bar{W}}}{2}\right) = \frac{\lambda_1}{\lambda_1 - it} \cdot \frac{\lambda_2}{\lambda_2 + it} \cdot \frac{1}{1 + t^2\sigma^2_{\bar{W}}}.\end{aligned}$$

The identifiability of $\theta_{S,R}$ may not be guaranteed for more complex models involving a higher-dimensional parameter. The study of the characteristic function of $\bar{X}$ is a tool that can help in these cases.

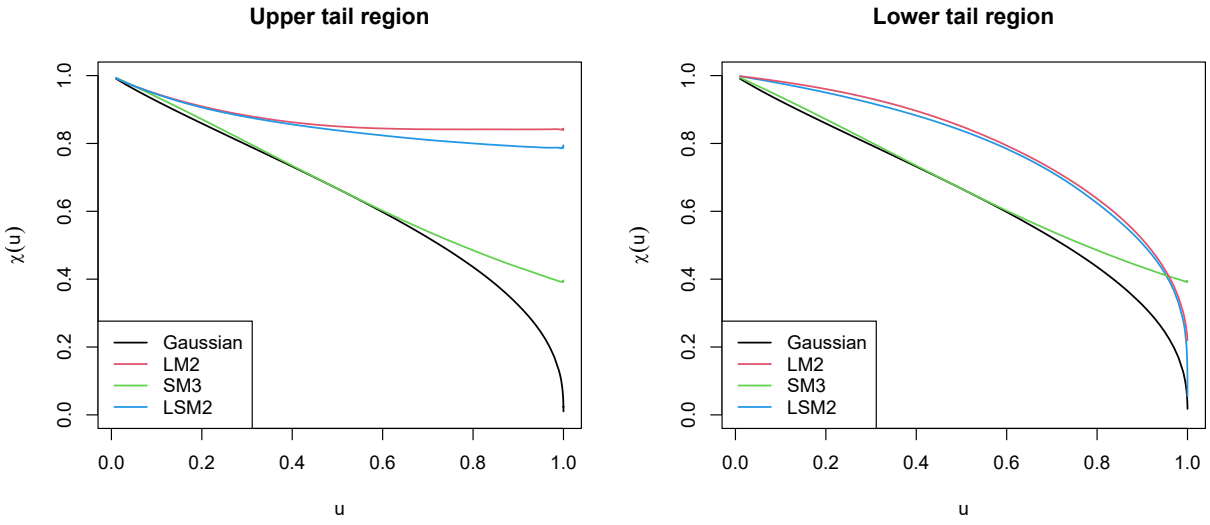


Figure S1: Values of $\chi(u)$, for $u \in (0,1)$, for the upper and lower tail regions of the four models described in Section S2.2.

S2.2. Model identifiability

A separate issue is the identifiability of the model itself, as different pairs (S, R) may potentially lead to similar finite-dimensional distributions. For simplicity, we limit our discussion here to the bivariate case shown in Figure 1 in the main text and described in Section 3.1. Recall that $(W_1, W_2)^T$ is a bivariate Gaussian and, for $i = 1, 2$,

(LM2) $X_i = S + W_i$, with $S = E_1 - E_2$, $E_1 \sim \mathrm{E}(0.4)$, $E_2 \sim \mathrm{E}(2.5)$;

(SM3) $X_i = R\, W_i$ with $R = 1/\sqrt{E_3}$, $E_3 \sim \mathrm{E}(1)$;

(LSM2) $X_i = S + R\, W_i$ with $S = E_1 - E_2$ and $R = \sqrt{E_4}$, $E_4 \sim \mathrm{E}(1/2)$.

Figure S1 shows the values of $\chi(u)$, for $u \in (0,1)$, for the upper and lower tail regions of the four models, fixing $\mathrm{cor}(W_1, W_2) = 0.5$. A first difference we can observe is that asymmetry between the upper and lower tail regions is possible only when S is not degenerate, while Gaussian models and Gaussian scale-mixture models are symmetric by construction. The dependence structure looks very similar for models LM2 and LSM2, although the former has a slight stronger dependence and is AD in both tail regions, while the latter is AD in the upper tail region and AI in the lower tail region. This difference, however, may not be evident when looking to the sub-asymptotic dependence structures of the two models. For this reason, we study here the bivariate characteristic functions of LM2 and LSM2, to inspect the difference between their distributions.

Let $\mathbf{t} = (t_1, t_2)^\top$ and $\mathbf{W} = (W_1, W_2)^\top \sim (\mathbf{0}, \Sigma_2)$. Note that the characteristic function of $\mathbf{W}$ is $\varphi_{\mathbf{W}}(\mathbf{t}) = \exp(-\mathbf{t}^\top \Sigma_2 \mathbf{t}/2)$. Then, the characteristic function of a general bivariate Gaussian location-scale mixture $\mathbf{X} = (X_1, X_2)^\top$ is

$$\begin{aligned}\varphi_{\mathbf{X}}(\mathbf{t}) &= \mathbb{E}\left[\exp\{-i(t_1 X_1 + t_2 X_2)\}\right] = \mathbb{E}\left[\exp\{-i(t_1 S + t_1 R W_1 + t_2 S + t_2 R W_2)\}\right] \\ &= \varphi_S(t_1 + t_2) \cdot \mathbb{E}_R\left[\varphi_{\mathbf{W}}(R\,\mathbf{t})\right] = \varphi_S(t_1 + t_2) \cdot \mathbb{E}_R\left[\exp\left(-\frac{1}{2} R^2\, \mathbf{t}^\top \Sigma_2 \mathbf{t}\right)\right].\end{aligned}$$

In the case of model LM2, $R = 1$, so

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_S(t_1 + t_2) \cdot \exp\left(-\frac{1}{2}\mathbf{t}^\top \Sigma_2 \mathbf{t}\right).$$

In the case of model LSM2, since $R^2 \sim \mathrm{E}(1/2)$ (see also Section S2.1),

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_S(t_1 + t_2) \cdot M_{R^2}\left(-\frac{1}{2}\mathbf{t}^\top \Sigma_2 \mathbf{t}\right) = \varphi_S(t_1 + t_2) \cdot \frac{1}{1 + \mathbf{t}^\top \Sigma_2 \mathbf{t}}.$$

The bivariate distributions of the two models are different because the presence of the random variable R modifies the effect of the dependence of the bivariate Gaussian $\mathbf{W}$, with respect to the random variable S, common to the two models.

In general, the study of bivariate or multivariate characteristic functions can help to check the differences between the distributions of models characterized by different pairs (S, R), in case these differences are not clear when visualizing their dependence structure through some graphics.

S3. Consistency of simulated minimum distance estimator

Let $X_1, \ldots, X_n$ be i.i.d. observations from a continuous $F_0 = F_{\theta_0}$, and define $Y_b(\theta) = F_\theta^{-1}(U_b)$, $b = 1, \ldots, B$, where $U_1, \ldots, U_B$ are i.i.d. uniform variables independent of the observations. We remark that the same uniforms are used for every θ. Consider the following estimator

$$\widehat{\theta}_{n,B} = \arg\min_\theta \frac{1}{n} \sum_{i=1}^{n} \left[\frac{i - 1/2}{n} - \widehat{F}_B(X_{(i)})\right]^2.$$

Here $\widehat{F}_B$ is the empirical cdf of $Y_b(\theta)$ and $X_{(i)}$, $i = 1, \ldots, n$, are the ordered values of X_i, $i = 1, \ldots, n$.

Proposition 1. *We assume that*

1. $\Theta \subset \mathbb{R}^p$ *is compact and* $\theta_0 \in \mathrm{int}(\Theta)$.
2. F_0 *is continuous and strictly increasing. The functions* $G_\theta(t) = F_\theta(F_0^{-1}(t))$, $0 < t < 1$, *extend continuously to* $[0, 1]$, *and* $\theta \mapsto G_\theta$ *is continuous in the uniform norm.*
3. *The model is identified by* $\int_0^1 [G_\theta(t) - t]^2\, dt > 0$ *whenever* $\theta \neq \theta_0$.

Then

$$\widehat{\theta}_{n,B} \xrightarrow{p} \theta_0 \quad \text{for } n \to \infty, B = B_n \to \infty.$$

Proof. The $Z_i = F_0(X_i)$, $i = 1, \ldots, n$, are i.i.d. uniform random variables on $[0, 1]$. Write $E_n(t) = n^{-1} \sum_{i=1}^n \mathbf{1}\{Z_i \le t\}$ and $S_B(t) = B^{-1} \sum_{b=1}^B \mathbf{1}\{U_b \le t\}$. Inverse-transform simulation gives $F_{B,\theta}(X_i) = S_B(G_\theta(Z_i))$. Because the observations have no ties almost surely, $E_n(Z_{(i)}) = i/n$. Thus the contrast function can be written as

$$Q_{n,B}(\theta) = \int_0^1 \left[E_n(t) - S_B(G_\theta(t)) - \frac{1}{2n}\right]^2 dE_n(t).$$

Let $\widehat{\theta}_{n,B}$ be any measurable sequence satisfying

$$Q_{n,B}(\widehat{\theta}_{n,B}) \le \inf_{\theta \in \Theta} Q_{n,B}(\theta) + \frac{1}{n}.$$

Set $e_n(t) = E_n(t) - t$, $s_B(t) = S_B(t) - t$. The Glivenko–Cantelli theorem gives $\|e_n\|_\infty + \|s_B\|_\infty \xrightarrow{p} 0$. Moreover, $\sup_{\theta\in\Theta} \|S_B \circ G_\theta - G_\theta\|_\infty \le \|s_B\|_\infty$. Hence

$$\sup_{\theta\in\Theta} \sup_{t\in[0,1]} \left| E_n(t) - S_B(G_\theta(t)) - \frac{1}{2n} - \{t - G_\theta(t)\} \right| \le \|e_n\|_\infty + \|s_B\|_\infty + \frac{1}{2n} \xrightarrow{p} 0.$$

Since the functions involved are uniformly bounded, this implies

$$\sup_{\theta\in\Theta} \left| Q_{n,B}(\theta) - \int_0^1 [t - G_\theta(t)]^2 \, dE_n(t) \right| \xrightarrow{p} 0.$$

The class $\mathcal{F} = \{t \mapsto [t - G_\theta(t)]^2 : \theta \in \Theta\}$ is compact in the uniform norm. Indeed, it is a continuous image of compact Θ. Covering $\mathcal{F}$ by a finite uniform ε-net and applying the law of large numbers to its elements yields

$$\sup_{\theta\in\Theta} \left| \int [t - G_\theta(t)]^2 \, dE_n(t) - \int [t - G_\theta(t)]^2 \, dt \right| \xrightarrow{p} 0.$$

Therefore,

$$\sup_{\theta\in\Theta} |Q_{n,B}(\theta) - Q(\theta)| \xrightarrow{p} 0, \qquad Q(\theta) = \int_0^1 [t - G_\theta(t)]^2 \, dt.$$

For every $\varepsilon > 0$ for which the set below is nonempty, compactness, continuity, and identification imply $c_\varepsilon = \inf_{\|\theta-\theta_0\|\ge\varepsilon} Q(\theta) > 0$. On the event $\sup_\theta |Q_{n,B}(\theta) - Q(\theta)| < c_\varepsilon/3$, every θ outside the ε-ball has a larger empirical criterion than θ_0. Hence

$$\mathbb{P}(\|\widehat{\theta}_{n,B} - \theta_0\| \ge \varepsilon) \to 0.$$

□

S4. Additional simulation studies

S4.1. Estimation of different models

This simulation study aims to show the consistency of the two-step method for inference explained in Section 5 (in the main text) as we increase m and n, i.e. the number of spatial locations and data replications. In all the following cases, inference is performed on 100 simulated datasets for each of the configurations A, B, C and D (see Section 6 in the main text). This simulation study covers the three classes of Gaussian location, scale and location-scale mixtures described in the main text; for reference about the names of the models, see Section 3.2 and Table 1. In particular, Figure S2 shows the estimation results for the parameters of Gaussian location mixture 1, with $S \sim \text{Exp}(\lambda)$; Figure S3 is about the estimation of the parameters of Gaussian scale mixture 4, with $R = \sqrt{E}/G$, $E \sim \text{Exp}(1/2)$ and $G \sim \text{Gamma}(1/\xi, 1/\xi)$; Figure S4 refers to Gaussian location-scale mixture 1, with $S \sim \text{Exp}(\lambda)$ and $R = \sqrt{E}$, $E \sim \text{Exp}(1/2)$; finally, Figure S5 shows the estimation results for Gaussian location mixture 2, with $S = S_1 - S_2$, $S_1 \sim \text{Exp}(\lambda_1)$ and $S_2 \sim \text{Exp}(\lambda_2)$, i.e. a case in which S has two parameters. In all these cases, the estimators seem to be unbiased and

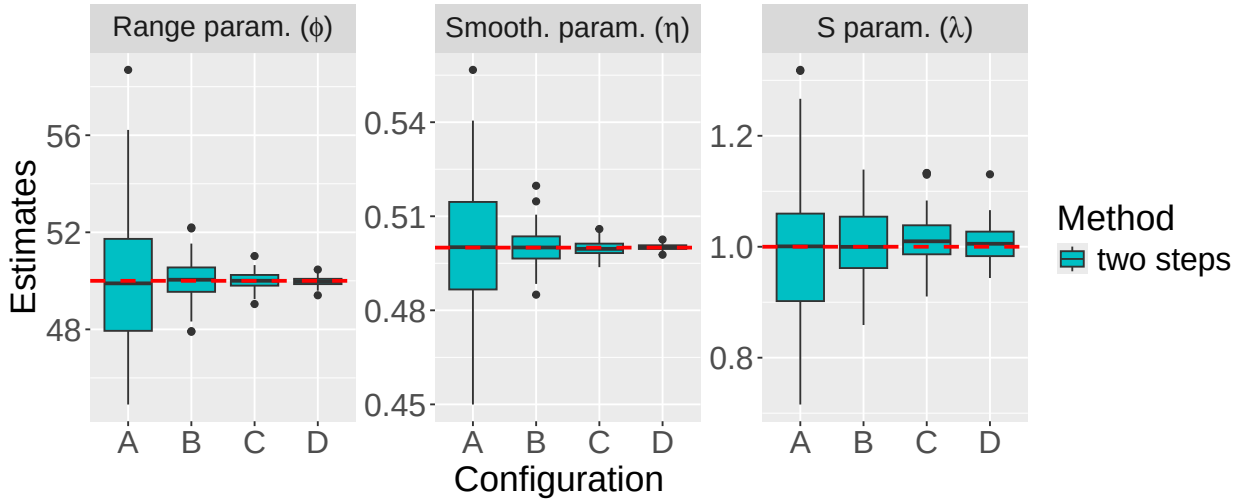


Figure S2: Estimation of Gaussian location mixture 1, with $S \sim \text{Exp}(\lambda)$, on 100 simulated datasets for each configuration, with true values $\varphi = 50$, $\eta = 0.5$ and $\lambda = 1$ (red lines). The boxplots refer to the two-step solution for inference explained in Section 5.

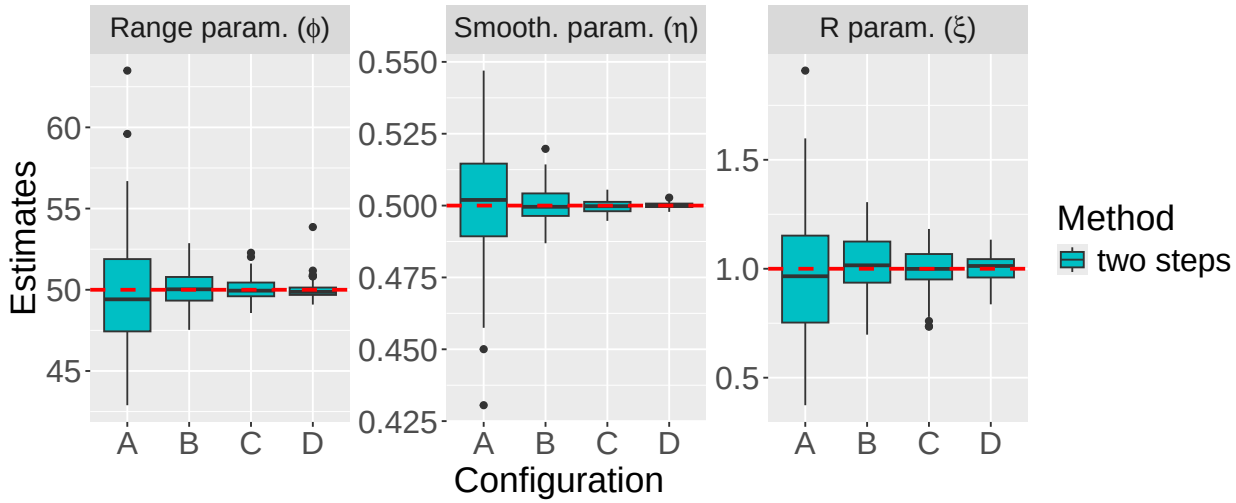


Figure S3: Estimation of Gaussian scale mixture 4, with $R = \sqrt{E}/G$, $E \sim \text{Exp}(1/2)$ and $G \sim \text{Gamma}(1/\xi, 1/\xi)$, on 100 simulated datasets for each configuration, with true values $\varphi = 50$, $\eta = 0.5$ and $\lambda = 1$ (red lines). The boxplots refer to the two-step solution for inference explained in Section 5.

consistent for all parameters. Nevertheless, the statistical efficiency is higher for the correlation parameter $\theta_W = (\varphi, \eta)$ of the Gaussian process $\mathbf{W}$, while the estimators for the parameter $\theta_{S,R}$ of the random variables S and R display higher variability, although they still appear to be empirically unbiased and consistent.

S4.2. Comparison with numerical integration

As stated in Section 5 in the main text, the standard solution for inference on Gaussian location and scale mixtures involves the numerical integration of density functions (6). Note that in the

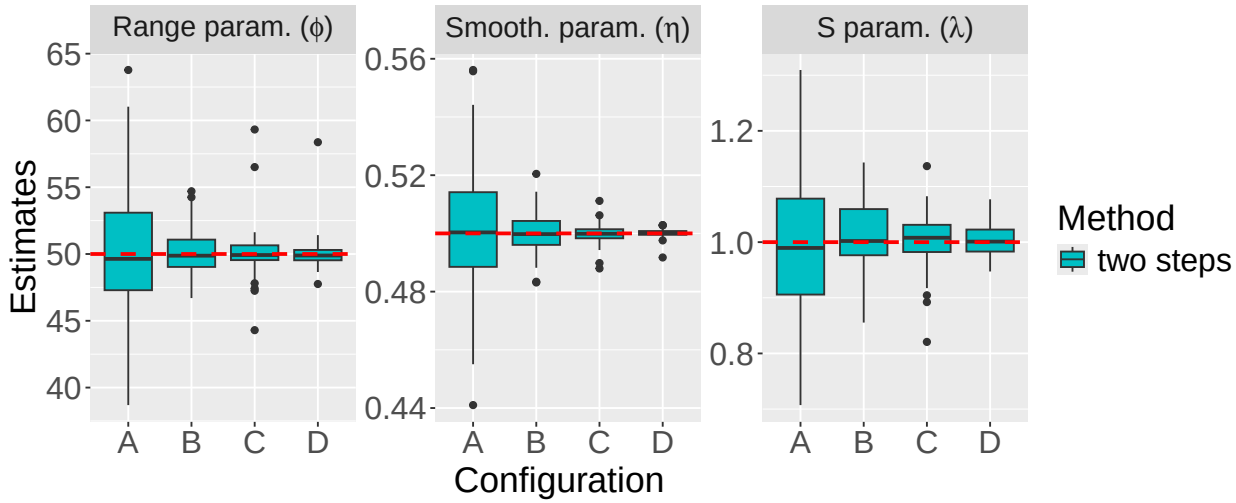


Figure S4: Estimation of Gaussian location-scale mixture 1, with $S \sim \text{Exp}(\lambda)$ and $R = \sqrt{E}$, $E \sim \text{Exp}(1/2)$, on 100 simulated datasets for each configuration, with true values $\varphi = 50$, $\eta = 0.5$ and $\lambda = 1$ (red lines). The boxplots refer to the two-step solution for inference explained in Section 5.

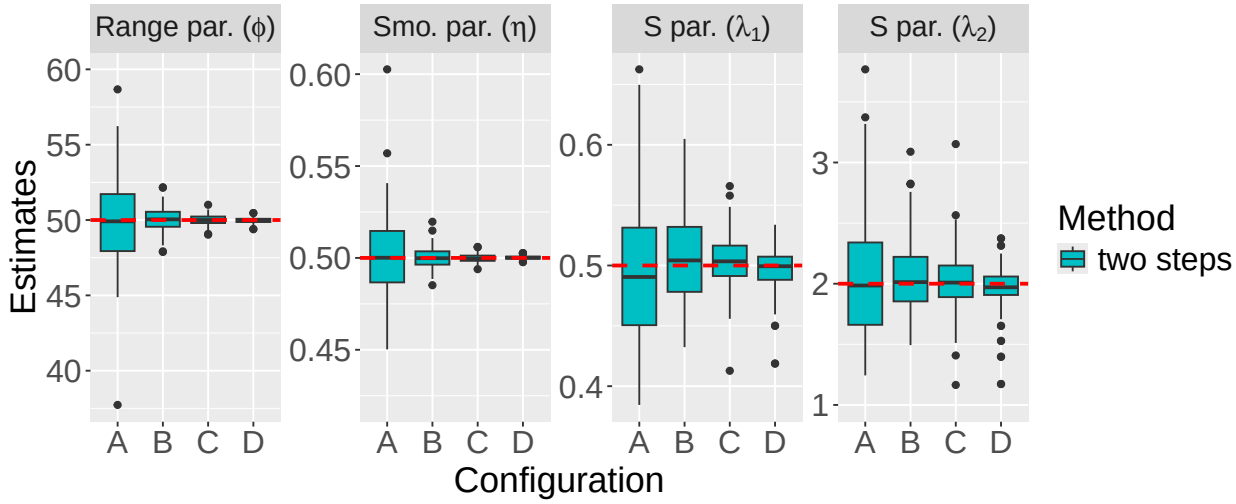


Figure S5: Estimation of Gaussian location mixture 2, with $S = S_1 - S_2$, $S_1 \sim \text{Exp}(\lambda_1)$ and $S_2 \sim \text{Exp}(\lambda_2)$, on 100 simulated datasets for each configuration, with true values $\varphi = 50$, $\eta = 0.5$, $\lambda_1 = 0.5$ and $\lambda_2 = 2$ (red lines). The boxplots refer to the two-step solution for inference explained in Section 5.

cases of Gaussian location mixtures and Gaussian scale mixtures, the formula (6) reduces to a single integral with respect to S or R, while a double integral must be evaluated in the case of Gaussian location-scale mixtures. In this section, we estimate two models for which a single integral is sufficient and compare the results with the newly proposed method and the closed-form maximum likelihood estimates, when available. All estimations are performed on 50 datasets and only for Configurations A and B, to reduce the computational cost of the integration-based method. Integration points are chosen adaptively by the R function `integrate()` within the interval $(0, 30)$.

Figure S6 shows estimates for a Laplace process (model SM1 in Table 1), comparing the new

method (in blue) with the maximum likelihood in closed form (in orange) and the maximum likelihood obtained by numerical integration (in green), similarly to Figure 2 in the main text. Figure S7 refers to Gaussian location mixture 1, similarly to Figure S2, and in this case the maximum likelihood is not computed in closed form. In both cases, the new estimation method seems to perform as well as the classical ones. However, the estimation based on numerical integration of densities (6) has a much higher computational cost. A comparison of computation times is provided in Section 6.2 in the main text.

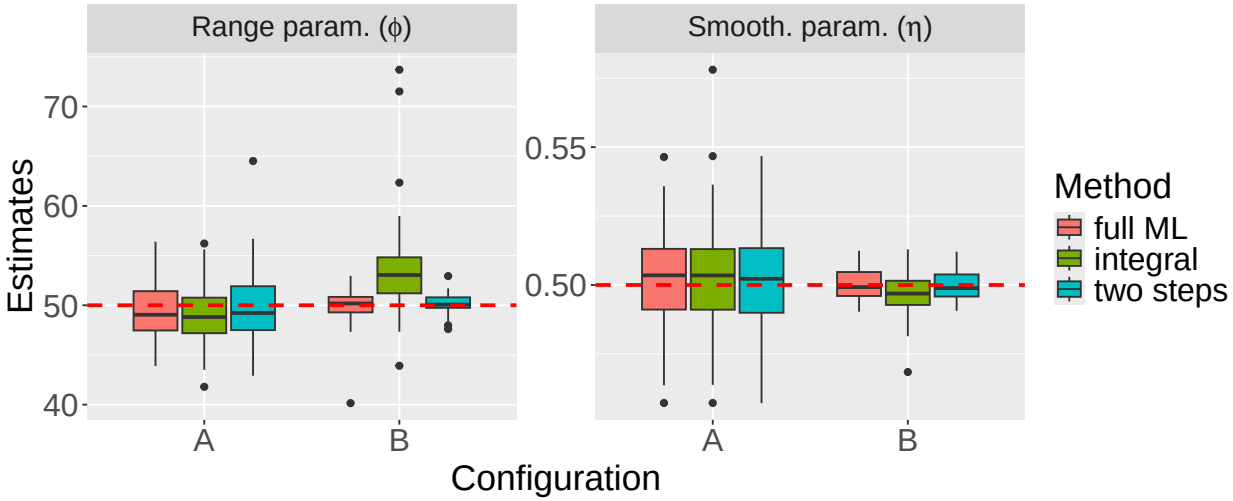


Figure S6: Estimation of Laplace process on 50 simulated datasets for each configuration, with true values $\varphi = 50$ and $\eta = 0.5$ (red lines). Orange boxplots refer to full maximum likelihood estimates, while green boxplots refer to estimates obtained via numerical integration and blue boxplots refer to the two-step solution for inference explained in Section 5.

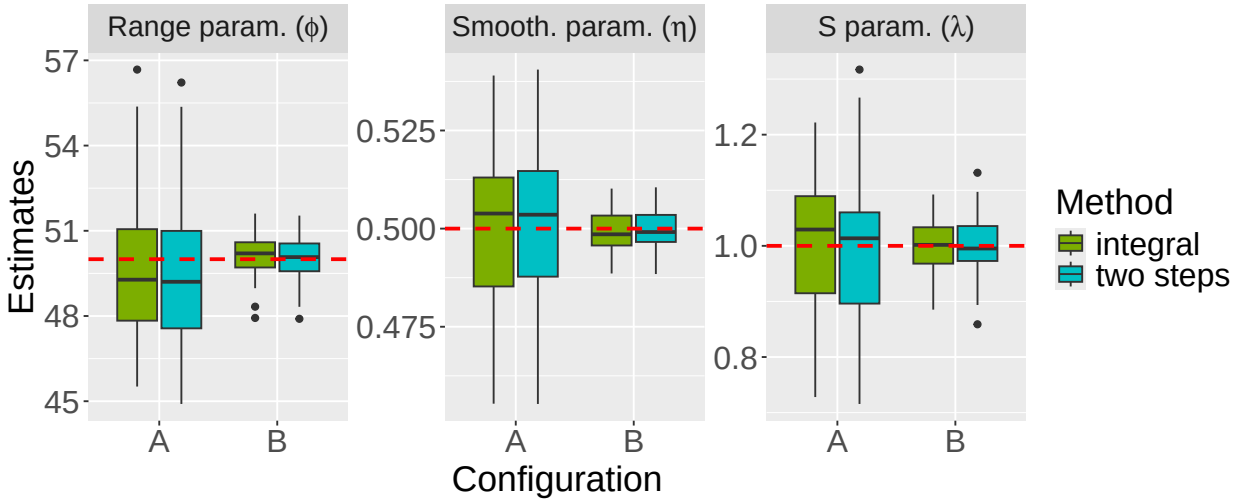


Figure S7: Estimation of Gaussian location mixture 1, with $S \sim \text{Exp}(\lambda)$, on 50 simulated datasets for each configuration, with true values $\varphi = 50$, $\eta = 0.5$ and $\lambda = 1$ (red lines). Green boxplots refer to estimates obtained via numerical integration, while blue boxplots refer to the two-step solution for inference explained in Section 5.

S4.3. Estimation with copula

The simulation studies in Section S4.1 and in Section 6.1 in the main text focus on the estimation of the parameters of Gaussian location-scale mixtures when data are simulated directly from these models. However, as explained in Section 5.3, data are often assumed to be observed in another scale, $Y(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$; after marginal transformations, a vector of uniforms $\mathbf{U}$ can be reconstructed as in equation (13) in the main text. Typically, a copula approach involves estimating the parameters of $\mathbf{X}$ on the vector $\mathbf{U}$; this approach is followed in the simulation studies presented in this section. To construct the vector $\mathbf{U}$, a non-parametric marginal transformation is applied to the data vector $\mathbf{X}$, simulated from the Gaussian location-scale mixture models. In particular, a location-wise rank transformation is employed. Then, the new estimation method is applied to $\mathbf{U}$ as described in Section 5.3 in the main text.

Figures S8 and S9 show analogue simulations to those in Section 6.1, for which a comparison with the full-likelihood case is possible. The estimators of $\theta_W = (\varphi, \eta)$ show some bias in the lower-dimensional configurations, in particular for the smoothness parameter η. However, the bias reduces as m and n are increased, and it appears to be the same for the two estimation methods (see Figure S8). This bias seems to be caused by the non-parametric rank transformation mentioned above, and it is not present when parametric marginal transformations are employed, as is the case in real data applications. In the three-parameter case of Figures S9, the new estimators of θ_W are again comparable to those based on the full likelihood, but the estimator of the third parameter $\theta_d = \nu$ displays higher variability than that based on the full likelihood, as in the non-copula case. Again, the consistency of the new method has to be shown on each of the three classes of models.

Figures S10, S11 and S12 display analogue simulations to those of Section S4.1, to show the consistency of the new estimation method on models belonging to the classes of Gaussian location, scale and location-scale mixtures, respectively. As above, a systematic bias is present due to the

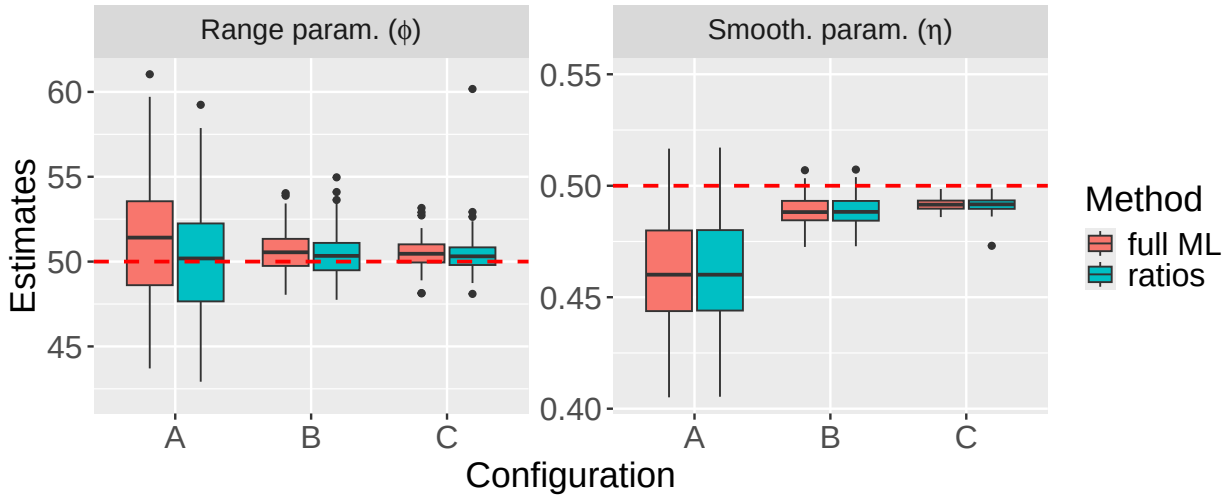


Figure S8: Estimation of Laplace process with copula on 100 simulated datasets for each configuration, with true values $\varphi = 50$ and $\eta = 0.5$ (red lines). Orange boxplots refer to full maximum likelihood estimates, while blue boxplots refer to the estimates based on the likelihood of ratios explained in Section 5.1.

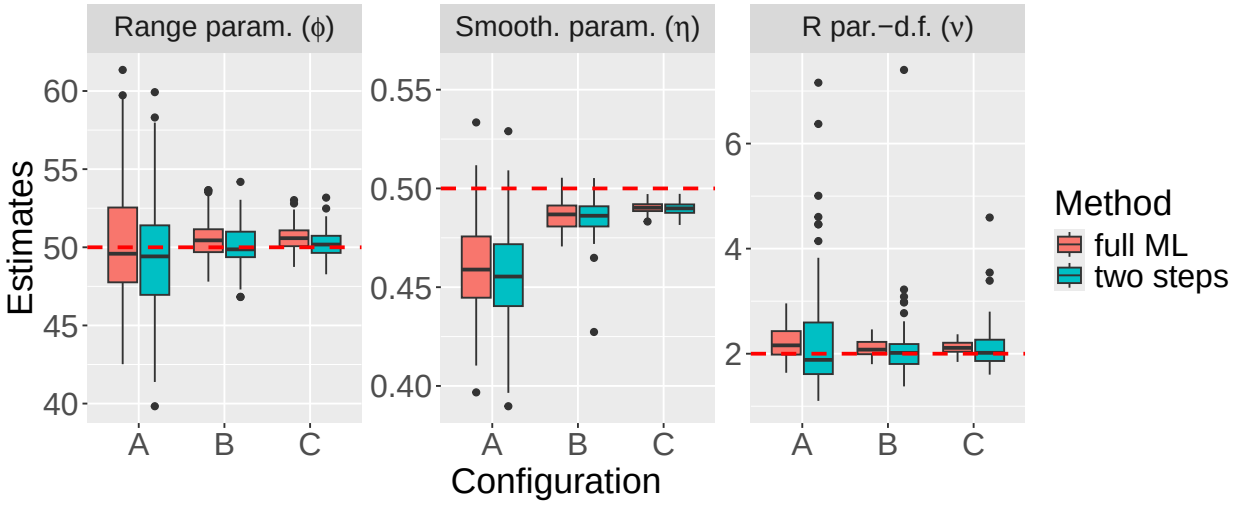


Figure S9: Estimation of Student's t process with copula on 100 simulated datasets for each configuration, with true values $\varphi = 50$, $\eta = 0.5$ and $\nu = 2$ (red lines). Orange boxplots refer to full maximum likelihood estimates, while blue boxplots refer to the two-step solution for inference explained in Section 5.

employed rank transformation, in particular for the correlation smoothness parameter η, although this bias reduces as m and n are increased, e.g. in Configuration C. The higher-dimensional Configuration D is not studied in these copula-based simulations, due to the increasing computational costs. An overview of the computation times for each data configuration and parameter settings is provided in Section 6.2 in the main text.

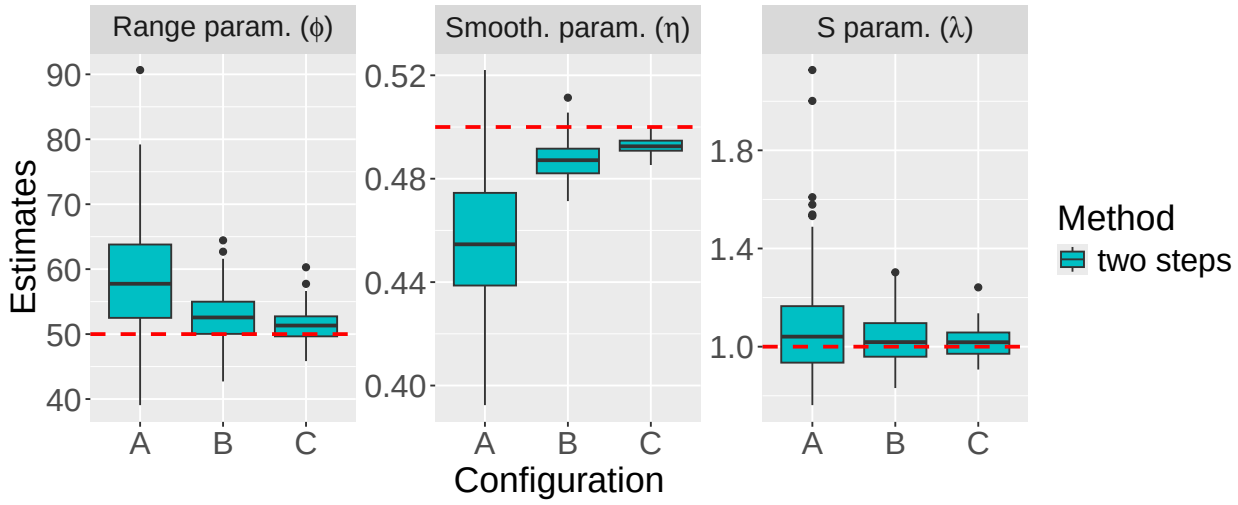


Figure S10: Estimation of Gaussian location mixture 1 with copula, with $S \sim \text{Exp}(\lambda)$, on 100 simulated datasets for each configuration, with true values $\varphi = 50$, $\eta = 0.5$ and $\lambda = 1$ (red lines). The boxplots refer to the two-step solution for inference explained in Section 5.

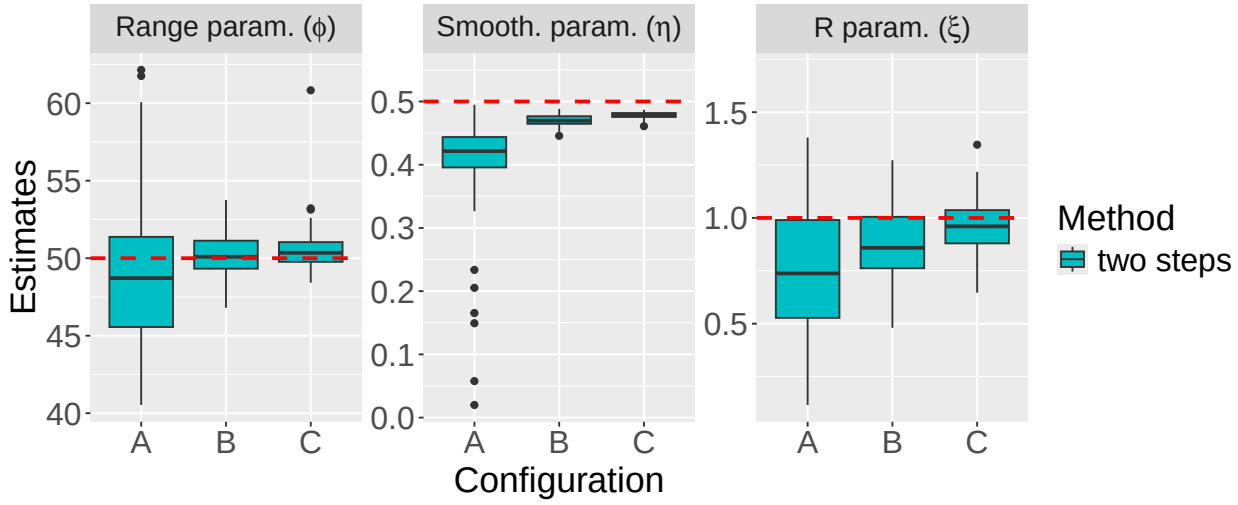


Figure S11: Estimation of Gaussian scale mixture 4 with copula, with $R = \sqrt{E}/G$, $E \sim \text{Exp}(1/2)$ and $G \sim \text{Gamma}(1/\xi, 1/\xi)$, on 100 simulated datasets for each configuration, with true values $\varphi = 50$, $\eta = 0.5$ and $\lambda = 1$ (red lines). The boxplots refer to the two-step solution for inference explained in Section 5.

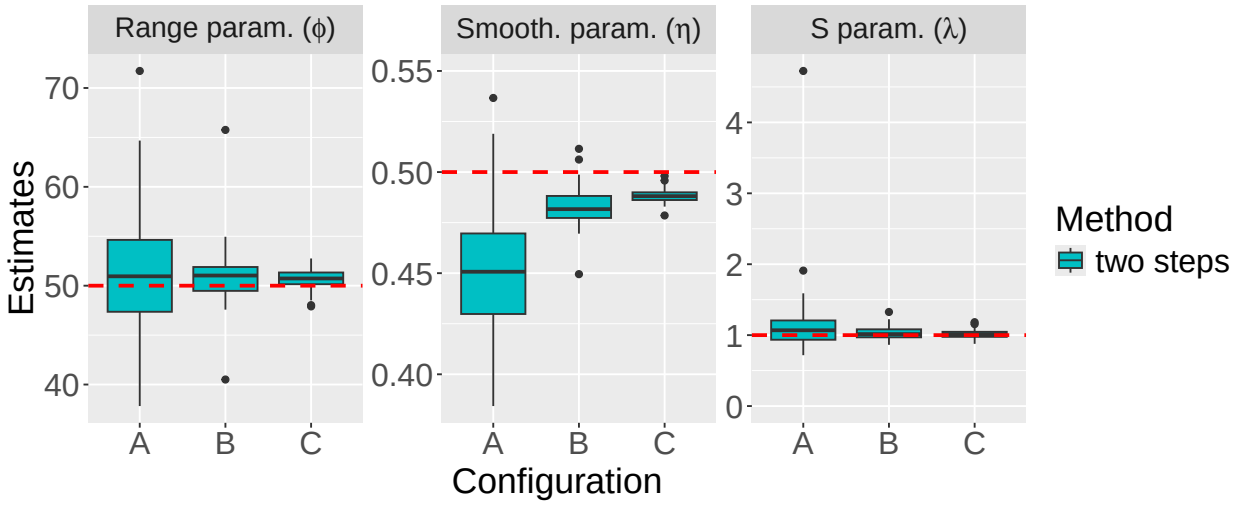


Figure S12: Estimation of Gaussian location-scale mixture 1 with copula, with $S \sim \text{Exp}(\lambda)$ and $R = \sqrt{E}$, $E \sim \text{Exp}(1/2)$, on 100 simulated datasets for each configuration, with true values $\varphi = 50$, $\eta = 0.5$ and $\lambda = 1$ (red lines). The boxplots refer to the two-step solution for inference explained in Section 5.

S5. Data application

Figure S13 shows the division of the spatial locations into north and south of Portugal. The rest of the analysis focuses on the north for the reasons explained in Section 7.1 in the main text.

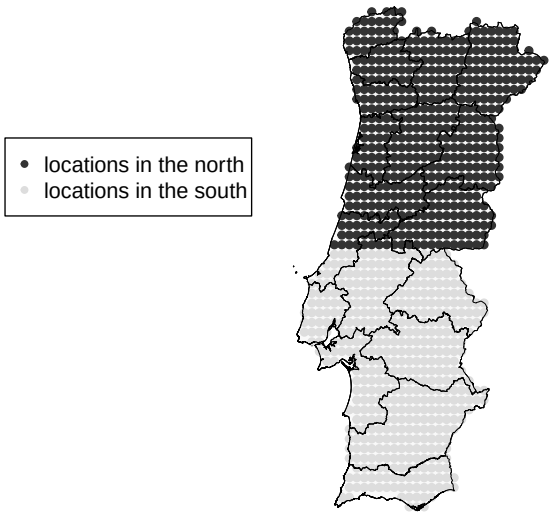


Figure S13: Division of the spatial locations into north and south of Portugal.

Figure S14 compares empirical and fitted quantiles of order $u \in \{0.5, 0.75, 0.9, 0.95\}$ for all the locations in the north of Portugal.

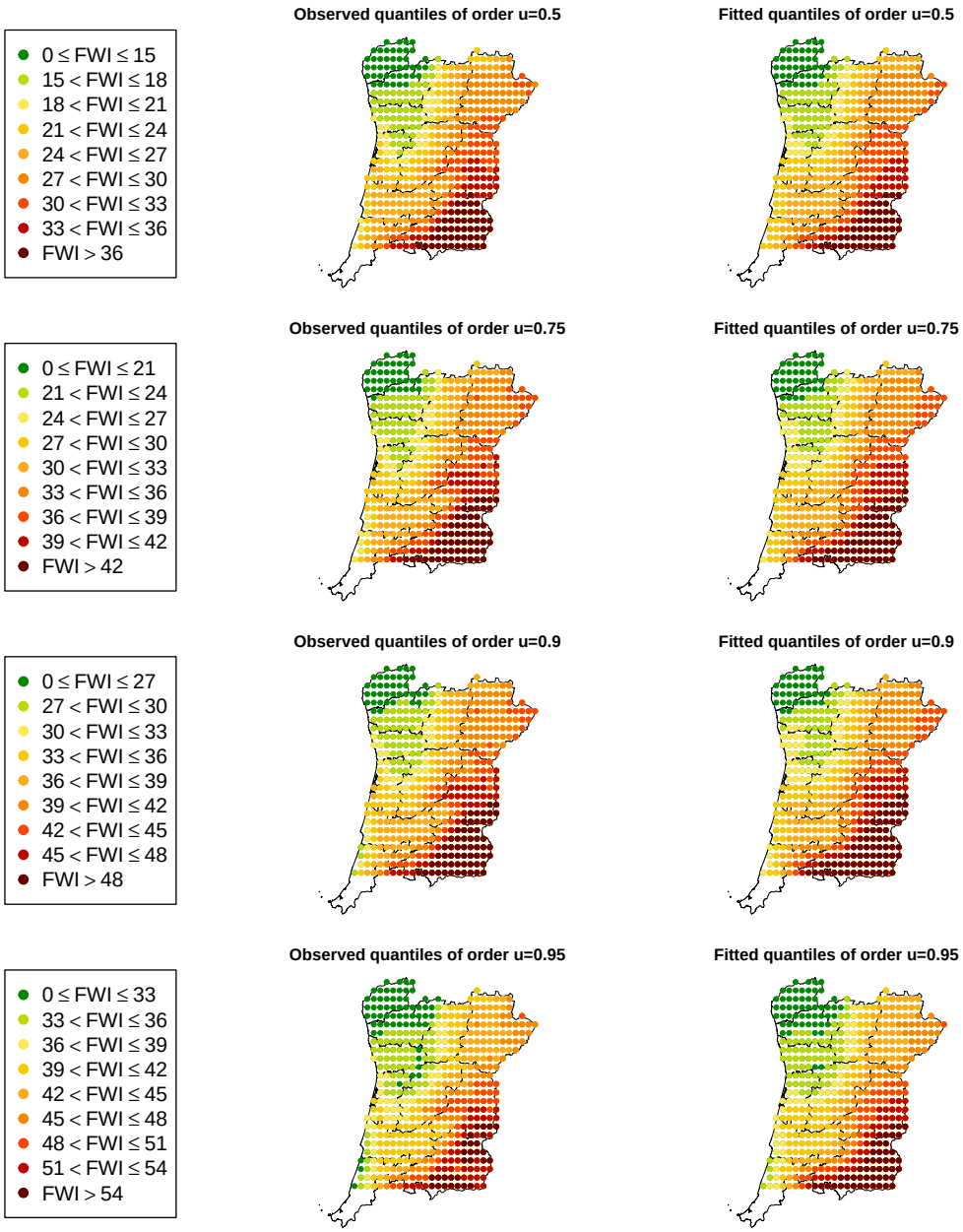


Figure S14: Comparison between empirical and fitted quantiles in the north of Portugal. Note that the colour legends are different in each row.

S5.1. Alternative dependence models

For comparison, two additional copula models are fitted on the data (and displayed in Figure 6 in the main text):

- A Huser-Wadsworth random-scale model (Huser and Wadsworth, 2019):

$$X(\mathbf{s}) = R^{\delta} \times W(\mathbf{s})^{1-\delta}, \quad \mathbf{s} \in \mathcal{S},$$

  where $\delta \in [0, 1]$, R is a standard Pareto random variable, and $W(\mathbf{s})$ Gaussian process with Matérn correlation function, transformed marginally into the standard Pareto scale. Due to the unfeasibility of likelihood evaluations for this model on $m = 491$ spatial locations, inference is performed through a neural estimator (Zammit-Mangion et al., 2025), adapting the methodology of Dell'Oro and Gaetan (2025) by replacing the grids of empirical χ statistics with five vectors of $\widehat{\chi}(u)$, computed for $u \in \{0.5, 0.75, 0.9, 0.95, 0.99\}$ on 24 regular bins covering distances from 0 to 240 km. The estimated parameter values are $\widehat{\delta} = 0.492$ and, in the Matérn correlation function, $\widehat{\varphi} = 137.7$ and $\widehat{\eta} = 1.057$. The value of δ, close to 0.5, confirms the ambiguity commented above between asymptotic dependence and independence, but the model seems to overestimate the χ values for the highest quantiles;

- A max-stable Brown-Resnick process (Brown and Resnick, 1977), estimated with pairwise likelihood using the R package `SpatialExtremes` (Ribatet, 2022), with covariance model `"brown"`. The estimated range parameter is 163.0 and the estimated smoothness is 1.130. This model overestimates the sub-asymptotic extremal dependence for high quantiles and seems too rigid to capture the decay of χ values for increasing spatial distances.