Supplemental material for
"Heteroscedastic asymmetric spatial processes"

Jiangyong Yin[a] , Peter F. Craigmile[a*]

[a] Department of Statistics, The Ohio State University, Columbus, OH 43210, USA

[*] pfc@stat.osu.edu

S1 Proofs of the results, and supplemental results

Theorem 1 can be proven with the well-known Stein's Identity.

Lemma S1 (Stein's Identity) *(Stein, 1981): If X is a normal random variable with mean μ and variance σ^2 and g is a real-valued function with a Lebesgue measurable derivative function g', then $E[(X-\mu)g(X)] = \sigma^2 E[g'(X)]$.*

Lemma S2 *For a normal random variable X with mean μ and variance σ^2, we have*

$$\begin{aligned} E[X^n e^{aX}] &= E[(X-\mu)X^{n-1}e^{aX}] + \mu E[X^{n-1}e^{aX}] \\ &= \sigma^2 E[aX^{n-1}e^{aX} + (n-1)X^{n-2}e^{aX}] + \mu E[X^{n-1}e^{aX}] \\ &= (\mu + a\sigma^2)E[X^{n-1}e^{aX}] + (n-1)\sigma^2 E[X^{n-2}e^{aX}]. \end{aligned}$$

In particular, for $n = 1, 2, 3$ and 4, we have

$$E[Xe^{aX}] = (\mu + a\sigma^2)E[e^{aX}]; \tag{S1}$$

$$E[X^2e^{aX}] = [(\mu + a\sigma^2)^2 + \sigma^2]E[e^{aX}]; \tag{S2}$$

$$E[X^3e^{aX}] = [(\mu + a\sigma^2)^3 + 3\sigma^2(\mu + a\sigma^2)]E[e^{aX}]; \tag{S3}$$

$$E[X^4e^{aX}] = [(\mu + a\sigma^2)^4 + 6\sigma^2(\mu + a\sigma^2)^2 + 3\sigma^4]E[e^{aX}]. \tag{S4}$$

Proof of Theorem 1 (a) Using (S1) and (S2) as well as the fact that

$$\epsilon(\boldsymbol{s})|\alpha(\boldsymbol{s}) \sim N(\alpha(\boldsymbol{s})\varrho/\tau, 1-\varrho^2),$$

and

$$\alpha(\boldsymbol{s})|\epsilon(\boldsymbol{s}) \sim N(\tau\varrho\epsilon(\boldsymbol{s}), \tau^2(1-\varrho^2)),$$

we can prove the aforementioned properties of the process $Z(\boldsymbol{s}) = e^{\alpha(\boldsymbol{s})/2}\epsilon(\boldsymbol{s})$. We have

$$E[Z(\boldsymbol{s})] = E[E[e^{\alpha(\boldsymbol{s})/2}\epsilon(\boldsymbol{s})|\alpha(\boldsymbol{s})]] = \frac{\varrho}{\tau}E[\alpha(\boldsymbol{s})e^{\alpha(\boldsymbol{s})/2}] = \frac{1}{2}\tau\varrho E[e^{\alpha(\boldsymbol{s})/2}] = \frac{1}{2}\tau\varrho e^{\tau^2/8},$$

and

$$\begin{aligned}
\text{var}(Z(\boldsymbol{s})) &= E[Z(\boldsymbol{s})^2] - E[Z(\boldsymbol{s})]^2 \\
&= E[E[\epsilon^2(\boldsymbol{s})e^{\alpha(\boldsymbol{s})}|\alpha(\boldsymbol{s})]] - \frac{1}{2}\tau\varrho E[e^{\alpha(\boldsymbol{s})/2}] \\
&= E[(1-\varrho^2+\frac{\varrho^2}{\tau^2}\alpha^2(\boldsymbol{s}))e^{\alpha(\boldsymbol{s})}] - \frac{1}{2}\tau\varrho E[e^{\alpha(\boldsymbol{s})/2}] \\
&= (1+\tau^2\varrho^2)E[e^{\alpha(\boldsymbol{s})}] - \frac{1}{2}\tau\varrho E[e^{\alpha(\boldsymbol{s})/2}] \\
&= (1+\tau^2\varrho^2)e^{\tau^2/2} - \frac{1}{4}\tau^2\varrho^2 e^{\tau^2/4}.
\end{aligned}$$

(b) We can still use the conditional expectation approach as above to find the third moment of $Z(\boldsymbol{s})$, but a much simpler approach is to use the properties of a bivariate normal random variable and express $\alpha(\boldsymbol{s})$ as

$$\alpha(\boldsymbol{s}) = \tau\varrho\epsilon(\boldsymbol{s}) + \tau\sqrt{1-\varrho^2}\eta(\boldsymbol{s}),$$

where $\boldsymbol{s}$ is a fixed spatial location, and $\eta(\boldsymbol{s})$ is a normal random variable with zero mean and unit variance and is independent from $\epsilon(\boldsymbol{s})$. Note that the original bivariate process $\{[\alpha(\boldsymbol{s}), \epsilon(\boldsymbol{s})]^T : \boldsymbol{s} \in \mathcal{D}\}$ is not equivalent to the bivariate process

$$\{[\tau\varrho\epsilon(\boldsymbol{s}) + \tau\sqrt{1-\varrho^2}\eta(\boldsymbol{s}), \epsilon(\boldsymbol{s})]^T : \boldsymbol{s} \in \mathcal{D}\}$$

since the cross covariance functions of these two processes are not necessarily the same. However, for a single fixed spatial location $\boldsymbol{s}$, the two bivariate normal random vectors $[\alpha(\boldsymbol{s}), \epsilon(\boldsymbol{s})]^T$ and $[\tau\varrho\epsilon(\boldsymbol{s}) + \tau\sqrt{1-\varrho^2}\eta(\boldsymbol{s}), \epsilon(\boldsymbol{s})]^T$ are indeed equivalent. Hence the random variables based on a measurable transformation of these two bivariate random vectors will have the same distribution and thus the same moments. Thus, we have

$$\begin{aligned}
E[Z^3(\boldsymbol{s})] &= E[e^{3\tau\varrho\epsilon(\boldsymbol{s})/2}\epsilon^3(\boldsymbol{s})]E\left[e^{3\tau\sqrt{1-\varrho^2}\eta(\boldsymbol{s})/2}\right] \\
&= \left[\frac{27}{8}\tau^3\varrho^3 + \frac{9}{2}\tau\varrho\right] E\left[e^{3(\tau\varrho\epsilon(\boldsymbol{s})+\tau\sqrt{1-\varrho^2}\eta(\boldsymbol{s}))/2}\right] \\
&= \left[\frac{27}{8}\tau^3\varrho^3 + \frac{9}{2}\tau\varrho\right] E\left[e^{3\alpha(\boldsymbol{s})/2}\right] \\
&= \left[\frac{27}{8}\tau^3\varrho^3 + \frac{9}{2}\tau\varrho\right] e^{9\tau^2/8}.
\end{aligned}$$

and the skewness of $Z(\boldsymbol{s})$ is given by

$$\begin{aligned}
&\frac{E[(Z(\boldsymbol{s}) - E[Z(\boldsymbol{s})])^3]}{[\mathrm{var}(Z(\boldsymbol{s}))]^{3/2}} \\
=&\frac{E[Z(\boldsymbol{s})^3] - 3E[Z(\boldsymbol{s})]E[Z(\boldsymbol{s})^2] + 2E[Z(\boldsymbol{s})]^3}{[\mathrm{var}(Z(\boldsymbol{s}))]^{3/2}} \\
=&\frac{(\frac{27}{8}\tau^3\varrho^3 + \frac{9}{2}\tau\varrho)E[e^{3\alpha(\boldsymbol{s})/2}] - \frac{3}{2}\tau\varrho(1+\tau^2\varrho^2)E[e^{\alpha(\boldsymbol{s})/2}]E[e^{\alpha(\boldsymbol{s})}] + \frac{1}{4}\tau^3\varrho^3 E[e^{\alpha(\boldsymbol{s})/2}]^3}{[\mathrm{var}(Z(\boldsymbol{s}))]^{3/2}} \\
=&\tau\varrho\frac{(\frac{27}{8}\tau^2\varrho^2 + \frac{9}{2})E[e^{3\alpha(\boldsymbol{s})/2}] - \frac{3}{2}(1+\tau^2\varrho^2)E[e^{\alpha(\boldsymbol{s})/2}]E[e^{\alpha(\boldsymbol{s})}] + \frac{1}{4}\tau^2\varrho^2 E[e^{\alpha(\boldsymbol{s})/2}]^3}{[\mathrm{var}(Z(\boldsymbol{s}))]^{3/2}} \\
=&\tau\varrho\frac{A}{[\mathrm{var}(Z(\boldsymbol{s}))]^{3/2}},
\end{aligned}$$

where $A = (\frac{27}{8}\tau^2\varrho^2 + \frac{9}{2})E[e^{3\alpha(\boldsymbol{s})/2}] - \frac{3}{2}(1+\tau^2\varrho^2)E[e^{\alpha(\boldsymbol{s})/2}]E[e^{\alpha(\boldsymbol{s})}] + \frac{1}{4}\tau^2\varrho^2 E[e^{\alpha(\boldsymbol{s})/2}]^3$. Using the fact that $E[XY] > E[X]E[Y]$ for positively correlated random variables X and Y, we can see that the numerator A satisfies $A > \{(\frac{27}{8}\tau^2\varrho^2 + \frac{9}{2}) - \frac{3}{2}(1+\tau^2\varrho^2)\}E[e^{\alpha(\boldsymbol{s})/2}]E[e^{\alpha(\boldsymbol{s})}] > 0$. Therefore, the skewness of $Z(\boldsymbol{s})$ always takes the same sign as ϱ – the correlation coefficient between $\alpha(\boldsymbol{s})$ and $Z(\boldsymbol{s})$.

(c) Consider a fixed spatial location $\boldsymbol{s}$. As above we assume that

$$\begin{aligned}
E[Z^4(\boldsymbol{s})] &= E\left[e^{2\tau\varrho\epsilon(\boldsymbol{s})+2\tau\sqrt{1-\varrho^2}\eta(\boldsymbol{s})}\epsilon(\boldsymbol{s})^4\right] \\
&= E\left[e^{2\tau\varrho\epsilon(\boldsymbol{s})}\epsilon(\boldsymbol{s})^4\right] E\left[\epsilon 2\tau\sqrt{1-\varrho^2}\eta(\boldsymbol{s})\right] \\
&= (16\tau^4\varrho^4 + 24\tau^2\varrho^2 + 3)E\left[e^{2\tau\varrho\epsilon(\boldsymbol{s})+2\tau\sqrt{1-\varrho^2}\eta(\boldsymbol{s})}\right] \\
&= (16\tau^4\varrho^4 + 24\tau^2\varrho^2 + 3)E\left[e^{2\alpha(\boldsymbol{s})}\right] \;=\; (16\tau^4\varrho^4 + 24\tau^2\varrho^2 + 3)e^{2\tau^2}.
\end{aligned}$$

To prove the excess kurtosis of $Z(\boldsymbol{s})$, we only need to show that

$$E[(Z(\boldsymbol{s}) - E[Z(\boldsymbol{s})])^4] - 3\mathrm{var}(Z(\boldsymbol{s}))^2 > 0.$$

The calculation is straightforward. Let m_i denote the ith noncentral moment of $Z(\boldsymbol{s})$. Then

$$
\begin{aligned}
&E[(Z(\boldsymbol{s})-E[Z(\boldsymbol{s})])^4]-3\mathrm{var}(Z(\boldsymbol{s}))^2 \\
&\quad = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4 - 3m_2^2 + 6m_2m_1^2 - 3m_1^4 \\
&\quad = (16\tau^4\varrho^4 + 24\tau^2\varrho^2 + 3)e^{2\tau^2} - 4\left(\frac{27}{8}\tau^3\varrho^3 + \frac{9}{2}\tau\varrho\right)e^{9\tau^2/8}\left(\frac{1}{2}\tau\varrho e^{\tau^2/8}\right) \\
&\qquad + 12(1+\tau^2\varrho^2)e^{\tau^2/2}\left(\frac{1}{2}\tau\varrho e^{\tau^2/8}\right)^2 - 6\left(\frac{1}{2}\tau\varrho e^{\tau^2/8}\right)^4 - 3\left((1+\tau^2\varrho^2)e^{\tau^2/2}\right)^2 \\
&\quad = (16\tau^4\varrho^4 + 24\tau^2\varrho^2 + 3)e^{2\tau^2} - \left(\frac{39}{4}\tau^4\varrho^4 + 15\tau^2\varrho^2 + 3\right)e^{\tau^2} \\
&\qquad + (3\tau^4\varrho^4 + 3\tau^2\varrho^2)e^{3\tau^2/4} - \frac{3}{8}\tau^4\varrho^4 e^{\tau^2/2} \\
&\quad > (16\tau^4\varrho^4 + 24\tau^2\varrho^2 + 3)e^{\tau^2} - \left(\frac{39}{4}\tau^4\varrho^4 + 15\tau^2\varrho^2 + 3\right)e^{\tau^2} \\
&\qquad + (3\tau^4\varrho^4 + 3\tau^2\varrho^2)e^{\tau^2/2} - \frac{3}{8}\tau^4\varrho^4 e^{\tau^2/2} \\
&\quad = \left(\frac{25}{4}\tau^4\varrho^4 + 9\tau^2\varrho^2\right)e^{\tau^2} + \left(\frac{21}{8}\tau^4\varrho^4 + 3\tau^2\varrho^2\right)e^{\tau^2/2} \;\geq 0.
\end{aligned}
$$

The inequality still holds when $\varrho = 0$ and it is easy to see that the excess kurtosis is even larger when $\varrho \neq 0$. □

Graphical illustration of Theorem 1

Figure S1 shows the functional relationships between the mean, variance, skewness and excess kurtosis of the process $Z(\boldsymbol{s})$ with the parameters τ and ϱ, as well as the simulated density of the marginal distribution of $Z(\boldsymbol{s})$. Theorem 1 can be easily verified from the graphs. The plots also clearly show that for fixed co-locational correlation coefficient ϱ, the magnitude of the mean, variance, skewness and kurtosis of the process $Z(\boldsymbol{s})$ increases rapidly as τ increases; for fixed τ, the magnitude of the four central moments also increases as the absolute value of ϱ increases.

We need the following lemma for the proof of Theorem 2.

Lemma S3 *For a mean zero and second-order stationary Gaussian spatial process $\epsilon(\boldsymbol{s})$ with*

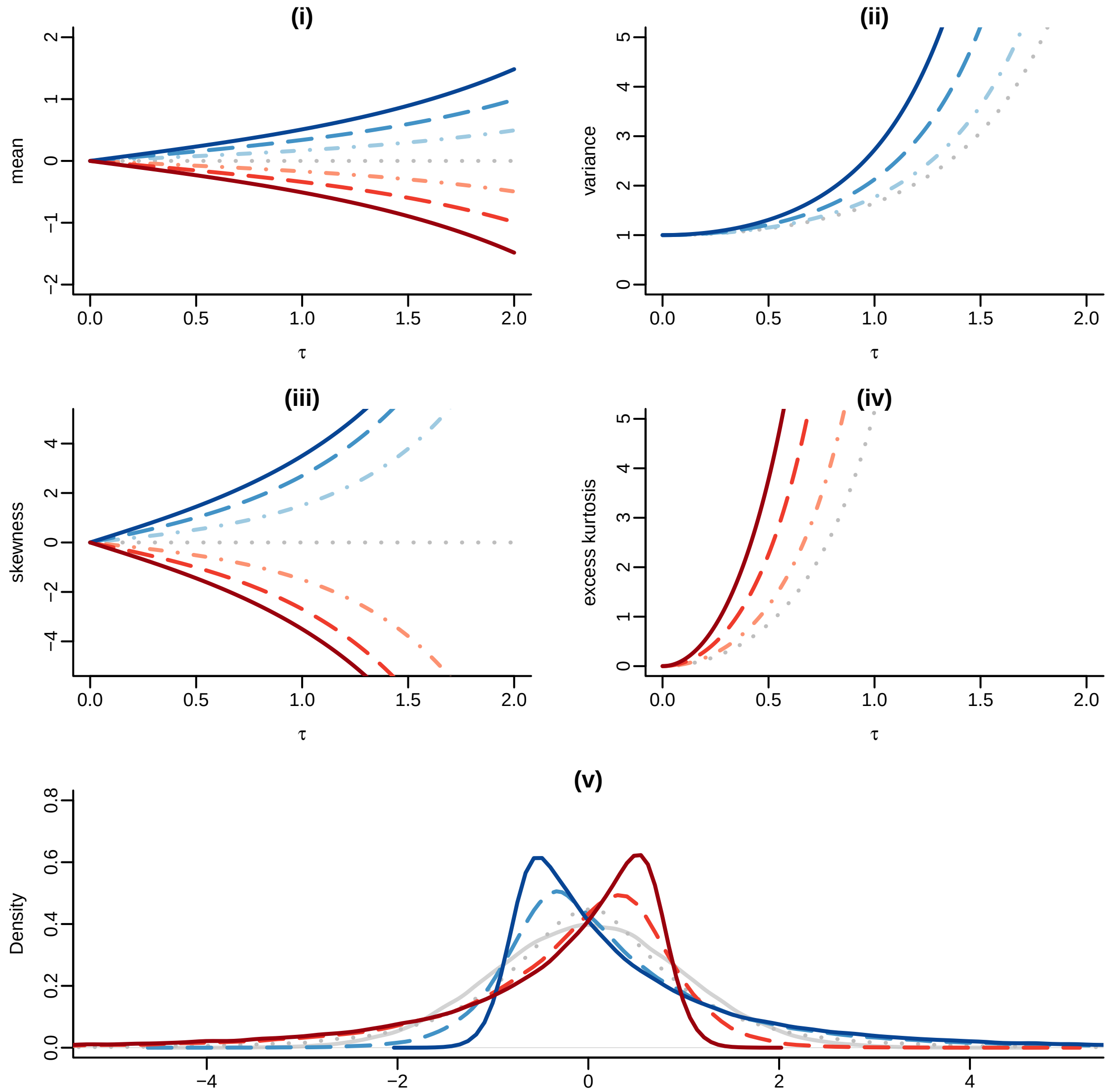

Figure S1: The plots (i) – (iv) show the mean, variance, skewness and excess kurtosis, respectively, of the process $Z(\boldsymbol{s})$ as a function of τ and ϱ. Plot (v) shows the simulated density function for the marginal distribution of $Z(\boldsymbol{s})$ for $\tau = 0.5$. The dotted line in each plot denotes the case where $\varrho = 0$, the dot-dash lines $\varrho = \pm 0.3$, the dashed lines $\varrho = \pm 0.6$ and the dark solid lines $\varrho = 0.9$. The grey solid line in plot (v) denotes the density function of a standard normal distribution.

correlation function $\rho_\epsilon(\boldsymbol{h})$, we have, for $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{s}'$,

$$E[\exp\{a(\epsilon(\boldsymbol{s}) + \epsilon(\boldsymbol{s}'))\}\epsilon(\boldsymbol{s})\epsilon(\boldsymbol{s}')] = [\rho_\epsilon(\boldsymbol{h}) + a^2(1 + \rho_\epsilon(\boldsymbol{h}))^2] \exp\{a^2(1 + \rho_\epsilon(\boldsymbol{h}))\}$$

Proof of Lemma S3 We prove the lemma using the conditional expectation approach, but note that the proof can be made simpler by applying the same linearization technique as in the proof of Theorem 1. We have

$$\begin{aligned}
&E[\exp\{a(\epsilon(\boldsymbol{s}) + \epsilon(\boldsymbol{s}'))\}\epsilon(\boldsymbol{s})\epsilon(\boldsymbol{s}')] \\
&=E[E[\exp\{a(\epsilon(\boldsymbol{s}) + \epsilon(\boldsymbol{s}'))\}\epsilon(\boldsymbol{s})\epsilon(\boldsymbol{s}')|\epsilon(\boldsymbol{s})]] \\
&=E[\epsilon(\boldsymbol{s}) \exp\{a\epsilon(\boldsymbol{s})\}E[\exp\{a\epsilon(\boldsymbol{s}')\}\epsilon(\boldsymbol{s}')|\epsilon(\boldsymbol{s})]] \\
&=E\left[\epsilon(\boldsymbol{s}) \exp\{a\epsilon(\boldsymbol{s})\}(\rho_\epsilon(\boldsymbol{h})\epsilon(\boldsymbol{s}) + a(1 - \rho_\epsilon^2(\boldsymbol{h})) \exp\left\{a\rho_\epsilon(\boldsymbol{h})\epsilon(\boldsymbol{s}) + \frac{a^2}{2}(1 - \rho_\epsilon^2(\boldsymbol{h}))\right\}\right] \\
&= \exp\left\{\frac{a^2}{2}(1 - \rho_\epsilon^2(\boldsymbol{h}))\right\} E\left[\{\rho_\epsilon(\boldsymbol{h})\epsilon^2(\boldsymbol{s}) + a(1 - \rho_\epsilon^2(\boldsymbol{h}))\epsilon(\boldsymbol{s})\} \exp\{a(1 + \rho_\epsilon(\boldsymbol{h}))\epsilon(\boldsymbol{s})\}\right].
\end{aligned}$$

Now applying (S1) and (S2), we obtain

$$\begin{aligned}
&E[\exp\{a(\epsilon(\boldsymbol{s}) + \epsilon(\boldsymbol{s}'))\}\epsilon(\boldsymbol{s})\epsilon(\boldsymbol{s}')] \\
&= \exp\left\{\frac{a^2}{2}(1 - \rho_\epsilon^2(\boldsymbol{h}))\right\} [\rho_\epsilon(\boldsymbol{h}) + \rho_\epsilon(\boldsymbol{h})a^2(1 + \rho_\epsilon(\boldsymbol{h}))^2 + a^2(1 + \rho_\epsilon(\boldsymbol{h}))(1 - \rho_\epsilon^2(\boldsymbol{h}))] \\
&\quad \times E[\exp\{a(1 + \rho_\epsilon(\boldsymbol{h}))\epsilon(\boldsymbol{s})\}] \\
&= \exp\left\{\frac{a^2}{2}(1 - \rho_\epsilon^2(\boldsymbol{h}))\right\} [\rho_\epsilon(\boldsymbol{h}) + a^2(1 + \rho_\epsilon(\boldsymbol{h}))^2]E[\exp\{a(1 + \rho_\epsilon(\boldsymbol{h}))\epsilon(\boldsymbol{s})\}] \\
&=[\rho_\epsilon(\boldsymbol{h}) + a^2(1 + \rho_\epsilon(\boldsymbol{h}))^2] \exp\left\{\frac{a^2}{2}(1 - \rho_\epsilon^2(\boldsymbol{h})) + \frac{a^2}{2}(1 + \rho_\epsilon(\boldsymbol{h}))^2\right\} \\
&=[\rho_\epsilon(\boldsymbol{h}) + a^2(1 + \rho_\epsilon(\boldsymbol{h}))^2] \exp\{a^2(1 + \rho_\epsilon(\boldsymbol{h}))\}.
\end{aligned}$$

□

Proof of Theorem 2 Since the latent bivariate process is assumed to be Gaussian and second-order stationary, we have

$$\begin{bmatrix} \alpha(\boldsymbol{s}) \\ \alpha(\boldsymbol{s}') \\ \epsilon(\boldsymbol{s}) \\ \epsilon(\boldsymbol{s}') \end{bmatrix} \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma} = \begin{bmatrix} \tau^2 \boldsymbol{P} & \tau\varrho \boldsymbol{R} \\ \tau\varrho\boldsymbol{R} & \boldsymbol{Q} \end{bmatrix}\right), \tag{S5}$$

where $\boldsymbol{P}$, $\boldsymbol{Q}$ and $\boldsymbol{R}$ are three 2-by-2 correlation matrices with correlation function $\rho_\alpha(\boldsymbol{h})$, $\rho_\epsilon(\boldsymbol{h})$ and $\rho_c(\boldsymbol{h})$, respectively. In general, the covariance matrix $\boldsymbol{\Sigma}$ is not necessarily non-negative definite. However, since we assumed that the bivariate process $(\alpha(\boldsymbol{s}), \epsilon(\boldsymbol{s}))^T$ has a valid second-order structure, $\boldsymbol{\Sigma}$ is indeed non-negative definite. Using the properties of multivariate normal distribution, we have

$$\begin{bmatrix} \alpha(\boldsymbol{s}) \\ \alpha(\boldsymbol{s}') \end{bmatrix} \Bigg| \begin{bmatrix} \epsilon(\boldsymbol{s}) \\ \epsilon(\boldsymbol{s}') \end{bmatrix} \sim N\left(\tau\varrho\boldsymbol{R}\boldsymbol{Q}^{-1}\begin{bmatrix} \epsilon(\boldsymbol{s}) \\ \epsilon(\boldsymbol{s}') \end{bmatrix}, \tau^2\boldsymbol{P} - \tau^2\varrho^2\boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R}\right),$$

and hence

$$\left.\frac{\alpha(\boldsymbol{s}) + \alpha(\boldsymbol{s}')}{2}\right| \epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{s}') \sim N(\mu_{\alpha|\epsilon}, \boldsymbol{\Sigma}_{\alpha|\epsilon}),$$

where

$$\mu_{\alpha|\epsilon} = \frac{\tau\varrho}{2}\boldsymbol{1}^T\boldsymbol{R}\boldsymbol{Q}^{-1}(\epsilon(\boldsymbol{s}) + \epsilon(\boldsymbol{s}')) = \frac{\tau\varrho}{2}\frac{1+\rho_c(\boldsymbol{h})}{1+\rho_\epsilon(\boldsymbol{h})}(\epsilon(\boldsymbol{s}) + \epsilon(\boldsymbol{s}'))$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_{\alpha|\epsilon} = \frac{\tau^2}{4}\boldsymbol{1}^T(\boldsymbol{P} - \varrho^2\boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R})\boldsymbol{1} &= \frac{\tau^2}{4}(\boldsymbol{1}^T\boldsymbol{P}\boldsymbol{1} - \varrho^2\boldsymbol{1}^T\boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R}\boldsymbol{1}) \\ &= \frac{\tau^2}{2}\left(1 + \rho_\alpha(\boldsymbol{h}) - \varrho^2\frac{(1+\rho_c(\boldsymbol{h}))^2}{1+\rho_\epsilon(\boldsymbol{h})}\right). \end{aligned}$$

Use the moment generating function of a normal random variable and Lemma S3, we have

$$\begin{aligned}
\operatorname{cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}')) =& E\left[E\left[\exp\left\{\frac{\alpha(\boldsymbol{s})+\alpha(\boldsymbol{s}')}{2}\right\}\epsilon(\boldsymbol{s})\epsilon(\boldsymbol{s}')\middle|\epsilon(\boldsymbol{s}),\epsilon(\boldsymbol{s}')\right]\right] - E[Z(\boldsymbol{s})]^2 \\
=& \exp\left\{\frac{\tau^2}{4}(1+\rho_\alpha(\boldsymbol{h})) - \frac{\tau^2\varrho^2}{4}\frac{(1+\rho_c(\boldsymbol{h}))^2}{1+\rho_\epsilon(\boldsymbol{h})}\right\}\times \\
& E\left[\epsilon(\boldsymbol{s})\epsilon(\boldsymbol{s}')\exp\left\{\frac{\tau\varrho}{2}\frac{1+\rho_c(\boldsymbol{h})}{1+\rho_\epsilon(\boldsymbol{h})}(\epsilon(\boldsymbol{s})+\epsilon(\boldsymbol{s}'))\right\}\right] - \frac{1}{4}\tau^2\varrho^2 e^{\tau^2/4} \\
=& \exp\left\{\frac{\tau^2}{4}(1+\rho_\alpha(\boldsymbol{h})) - \frac{\tau^2\varrho^2}{4}\frac{(1+\rho_c(\boldsymbol{h}))^2}{1+\rho_\epsilon(\boldsymbol{h})}\right\}\times \\
& \left(\rho_\epsilon(\boldsymbol{h}) + \frac{\tau^2\varrho^2}{4}(1+\rho_c(\boldsymbol{h}))^2\right)\exp\left\{\frac{\tau^2\varrho^2}{4}\frac{(1+\rho_c(\boldsymbol{h}))^2}{1+\rho_e(\boldsymbol{h})}\right\} - \frac{1}{4}\tau^2\varrho^2 e^{\tau^2/4} \\
=& \exp\left\{\frac{\tau^2}{4}(1+\rho_\alpha(\boldsymbol{h}))\right\}\left(\rho_\epsilon(\boldsymbol{h}) + \frac{\tau^2\varrho^2}{4}(1+\rho_c(\boldsymbol{h}))^2\right) - \frac{1}{4}\tau^2\varrho^2 e^{\tau^2/4}.
\end{aligned}$$

□

Proof of Theorem 3 It is easier to prove this result by definition than using Theorem 1. We need to establish that $\lim_{\Delta\boldsymbol{s}\to\boldsymbol{0}} E[(Z(\boldsymbol{s}+\Delta\boldsymbol{s})-Z(\boldsymbol{s}))^2]=0$. Let $\|\Delta\boldsymbol{s}\|=h$, $\mu=E[Z(\boldsymbol{s})]$ and $C(\boldsymbol{h})=\operatorname{cov}(Z(\boldsymbol{s}+\Delta), Z(\boldsymbol{s}))$. As has been proven, the process $Z(\boldsymbol{s})$ has finite mean, variance and a valid covariance function. Therefore,

$$\begin{aligned}
& E[(Z(\boldsymbol{s}+\Delta\boldsymbol{s})-Z(\boldsymbol{s}))^2] \\
&= E[Z^2(\boldsymbol{s}+\Delta\boldsymbol{s})] + E[Z^2(\boldsymbol{s})] - 2E[Z(\boldsymbol{s}+\Delta\boldsymbol{s})Z(\boldsymbol{s})] \\
&= 2E[Z^2(\boldsymbol{s})] - 2(C(\boldsymbol{h})+\mu^2) \\
&= 2(1+\tau^2\varrho^2)e^{\tau^2/2} - 2\left(\left[\rho_\epsilon(\boldsymbol{h}) + \frac{\tau^2\varrho^2}{4}(1+\rho_c(\boldsymbol{h}))^2\right]\exp\left[\frac{\tau^2}{4}(1+\rho_\alpha(\boldsymbol{h}))\right]\right).
\end{aligned}$$

It is easy to see that

$$\begin{aligned}
& \lim_{\Delta\boldsymbol{s}\to\boldsymbol{0}} E[(Z(\boldsymbol{s}+\Delta\boldsymbol{s})-Z(\boldsymbol{s}))^2] \\
&= \lim_{\boldsymbol{h}\to\boldsymbol{0}}\left\{2(1+\tau^2\varrho^2)e^{\tau^2/2} - 2\left(\left[\rho_\epsilon(\boldsymbol{h}) + \frac{\tau^2\varrho^2}{4}(1+\rho_c(\boldsymbol{h}))^2\right]\exp\left[\frac{\tau^2}{4}(1+\rho_\alpha(\boldsymbol{h}))\right]\right)\right\} \\
&= 2(1+\tau^2\varrho^2)e^{\tau^2/2} - 2(1+\tau^2\varrho^2)e^{\tau^2/2} \\
&= 0.
\end{aligned}$$

□

Theorem 4 follows directly from Pólya's Criterion (Pólya, 1949) as well as the following criteria of the Pólya type for the positive definiteness of radial functions (Gneiting, 2001). We present the latter result here first.

Theorem S4 *(Gneiting, 2001, Theorem 1.1) Let $\varphi : [0, \infty) \to \mathbb{R}$ be a continuous function with $\varphi(0) = 1$ and $\lim_{t\to\infty} \varphi(t) = 0$. Suppose that k and l are non-negative integers, at least one of which is strictly positive. Let*

$$\eta_1(t) = \left(-\frac{d}{du}\right)^k \varphi(\sqrt{u})\bigg|_{u=t^2}.$$

If there exists an $\alpha > 1/2$ so that

$$\eta_2(t) = \left(-\frac{d}{dt}\right)^{k+l-1} [-\eta_1'(t^\alpha)]$$

is convex for $t > 0$, then the radial function $\varphi(\|x\|)$, $x \in \mathbb{R}^n$ is positive definite for $n = 1, \ldots, 2l + 1$.

Proof of Theorem 4

(a) Use $\gamma(\boldsymbol{h})$ to denote the covariance function in Theorem 2, and then $\gamma(\boldsymbol{h})$ is continuous, $\gamma(0) = 1$ and $\lim_{h\to\infty} \gamma(\boldsymbol{h}) = 0$. In addition, given the conditions, $\gamma(\boldsymbol{h})$ is convex by the following properties of convex functions:

   (a) If $f(\boldsymbol{h})$ and $g(\boldsymbol{h})$ are convex, then $f(\boldsymbol{h}) + g(\boldsymbol{h})$ is also convex.

   (b) If $g(\boldsymbol{h})$ is convex, and $f(\boldsymbol{h})$ is convex and nondecreasing, then $f(g(\boldsymbol{h}))$ is convex.

   (c) If $f(\boldsymbol{h})$ and $g(\boldsymbol{h})$ are convex, non-negative and are either both non-increasing or both non-decreasing, then $f(\boldsymbol{h})g(\boldsymbol{h})$ is convex.

   Thus by Pólya's Criterion, $\gamma(\boldsymbol{h})$ is positive definite in $\mathbb{R}$.

(a) In Theorem S4, let $k = 0$, $l = 1$ and $\alpha = 1$. This special case of Theorem S4 (the result of Askey, 1973) dictates that $\gamma(\boldsymbol{h})$ is positive definite in $\mathbb{R}^3$ if $-\gamma'(\boldsymbol{h})$ is convex, where $-\gamma'(\boldsymbol{h})$ is given by

$$-\gamma'(\boldsymbol{h}) = -\, e^{\tau^2(1+\rho_\alpha(\boldsymbol{h}))/4}\left[\rho'_\epsilon(\boldsymbol{h}) + \rho'_c(\boldsymbol{h})\frac{\tau^2\varrho^2}{2}(1+\rho_c(\boldsymbol{h}))\right] - \frac{\tau^2}{4}\rho'_\alpha(\boldsymbol{h})e^{\tau^2(1+\rho_\alpha(\boldsymbol{h}))/4}\left[\rho_\epsilon(\boldsymbol{h}) + \frac{\tau^2\varrho^2}{2}(1+\rho_c(\boldsymbol{h}))^2\right].$$

Since $\rho'_\alpha(\boldsymbol{h})$, $\rho'_\epsilon(\boldsymbol{h})$ and $\rho'_c(\boldsymbol{h})$ are non-positive, non-decreasing and concave, then $-\rho'_\alpha(\boldsymbol{h})$, $-\rho'_\epsilon(\boldsymbol{h})$ and $-\rho'_c(\boldsymbol{h})$ are non-negative, non-increasing and convex. Therefore, by the same rules in part (a), $-\gamma'(\boldsymbol{h})$ is convex.

□

S2 A review of the multivariate Matérn process

Let $M(\boldsymbol{h}|\nu, a)$ denote the Matérn correlation function with smoothness parameter ν and scale parameter a:

$$M(\boldsymbol{h}|\nu, a) = \frac{2^{1-\nu}}{\Gamma(\nu)}(a\|\boldsymbol{h}\|)^\nu K_\nu(a\|\boldsymbol{h}\|), \tag{S6}$$

where $K_\nu(a\|\boldsymbol{h}\|)$ is a modified Bessel function of the second kind. If $\nu = 1/2$, $M(\boldsymbol{h}|\nu, a) = \exp(-a\|\boldsymbol{h}\|)$, the exponential correlation function and when $\nu = 3/2$, $M(\boldsymbol{h}|\nu, a) = (1 + a\|\boldsymbol{h}\|)\exp(-a\|\boldsymbol{h}\|)$. If general, if $\nu = 1/2 + n$, for $n = 0, 1, 2, \ldots$, then the Matérn function can be expressed as the product of an exponential function and a polynomial:

$$M\left(\boldsymbol{h}\,\middle|\, n + \frac{1}{2}, a\right) = \exp(-a\|\boldsymbol{h}\|)\sum_{n=0}^{n}\frac{(n+k)!}{(2n)!}\binom{n}{k}(2a\|\boldsymbol{h}\|)^{n-k}.$$

Then a multivariate process $\{\boldsymbol{U}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^p\}$ is a *multivariate Matérn processes* when the (cross)-covariance function at spatial lag $\boldsymbol{h}$ for $i, j = 1, \ldots, p$ is given by

$$C_{ij}(\boldsymbol{h}) = \begin{cases} \sigma_i^2 M(\boldsymbol{h}|\nu_i, a_i), & i = j; \\ \rho_{ij}\sigma_i\sigma_j M(\boldsymbol{h}|\nu_{ij}, a_{ij}), & \text{otherwise.} \end{cases} \tag{S7}$$

Theorems 1 and 4 of Gneiting et al. (2010) and Theorem 1 of Apanasovich et al. (2012) provide conditions on the parameters so that $\boldsymbol{U}(\boldsymbol{s})$ is a valid process. Gneiting et al. (2010) provides a sufficient (but not necessary) condition for the multivariate process $\boldsymbol{U}(\boldsymbol{s})$ to have valid second-order structure.

Theorem S5 *(Theorem 1 of Gneiting et al. (2010)) For $d \geq 1$, $p \geq 2$ and $1 \leq i \neq j \leq p$, suppose that*

$$\nu_{ij} = \frac{1}{2}(\nu_i + \nu_j),$$

and that the scale parameters satisfy

$$a_1 = \cdots = a_p = a_{ij} = a,$$

Then $\boldsymbol{U}(\boldsymbol{s})$ has a valid covariance structure in $\mathbb{R}^d$ defined by (S7) if the matrix $(\beta_{ij})_{i,i=1}^{p}$, with diagonal elements $\beta_{ii} = 1$ for $i = 1, \cdots, p$ and off-diagonal elements β_{ij} for $i \leq i \neq j \leq p$ given by

$$\beta_{ij} = \rho_{ij} \left[\frac{\Gamma(\nu_i + \frac{d}{2})^{1/2}}{\Gamma(\nu_i)^{1/2}} \frac{\Gamma(\nu_j + \frac{d}{2})^{1/2}}{\Gamma(\nu_j)^{1/2}} \frac{\Gamma(\frac{1}{2}(\nu_i + \nu_j))}{\Gamma(\frac{1}{2}(\nu_i + \nu_j) + \frac{d}{2})} \right]^{-1} \quad \text{for} \quad 1 \leq i \neq j \leq p,$$

is symmetric and non-negative definite.

Gneiting et al. (2010) argues that the conditions in Theorem S5 are not necessarily as restrictive as they seem. If $\nu = 1/2$, i.e., for the exponential covariance function, Ying (1991) showed that either a or σ^2 can be fixed arbitrarily and the composite quantity can still be estimated consistently and efficiently. More importantly, Zhang (2004) proved that in dimension $d \leq 3$, the parameters σ^2 and a of a Matérn covariance function with fixed smoothness parameter ν cannot be consistently estimated under an infill asymptotic, but that the composite quantity $\sigma^2 a^{2\nu}$ can be consistently estimated and that this quantity is more important for spatial prediction, as explained in the following theorem.

Theorem S6 *(Zhang (2004), Theorem 2) Let P_i, $i = 1, 2$, be two probability measures such that under P_i, the process $X(\boldsymbol{s})$, $\boldsymbol{s} \in \mathbb{R}^d$ is stationary Gaussian with mean 0 and an* *isotropic Matérn covariance function in* $\mathbb{R}^d$ *with a variance* σ_i^2*, a scale parameter* a_i *and the same smoothness parameter* ν*, where* $d = 1, 2, 3$*. For any bounded infinite set* $\mathcal{D}$*, the two probability measures* P_1 *and* P_2 *are equivalent on the paths of* $X(\boldsymbol{s})$*,* $\boldsymbol{s} \in \mathcal{D}$ *if and only if* $\sigma_1^2 a_1^{2\nu} = \sigma_2^2 a_2^{2\nu}$*.*

When $p = 2$, the restriction on ρ_{12} in Theorem S5 expressed in terms of the entries of some positive definite matrix can be rephrased as the following inequality

$$\rho_{12} \leq \frac{\Gamma(\nu_1 + \frac{d}{2})^{1/2}}{\Gamma(\nu_1)^{1/2}} \frac{\Gamma(\nu_2 + \frac{d}{2})^{1/2}}{\Gamma(\nu_2)^{1/2}} \frac{\Gamma(\frac{1}{2}(\nu_1 + \nu_2))}{\Gamma(\frac{1}{2}(\nu_1 + \nu_2) + \frac{d}{2})}.$$

This is trivial since for the matrix

$$\begin{pmatrix} 1 & \beta_{12} \\ \beta_{12} & 1 \end{pmatrix}$$

to be positive definite, we need to have $\beta_{12} \leq 1$. As it turns out, for $p = 2$ this condition on ρ_{12} is not only sufficient but also necessary, as is stated in the following theorem.

Theorem S7 *(Gneiting et al., 2010, Theorem 3) When* $p = 2$*, the full bivariate Matérn model described by (S7) is valid if and only if*

$$\rho_{12} \leq \frac{\Gamma(\nu_1 + \frac{d}{2})^{1/2}}{\Gamma(\nu_1)^{1/2}} \frac{\Gamma(\nu_2 + \frac{d}{2})^{1/2}}{\Gamma(\nu_2)^{1/2}} \frac{\Gamma(\frac{1}{2}(\nu_1 + \nu_2))}{\Gamma(\frac{1}{2}(\nu_1 + \nu_2) + \frac{d}{2})} \times \frac{a_1^{2\nu_1} a_2^{2\nu_2}}{a_{12}^{4\nu_{12}}} \inf_{t \geq 0} \frac{(a_{12}^2 + t^2)^{2\nu_{12} + d}}{(a_1^2 + t^2)^{\nu_1 + (d/2)} (a_2^2 + t^2)^{\nu_2 + (d/2)}}.$$

The above inequality implies the following important cases.

1. *If* $\nu_{12} < (\nu_1 + \nu_2)/2$*, the full bivariate Matérn model is valid if and only if* $\rho_{12} = 0$*. In other words,* $\nu_{12} \geq (\nu_1 + \nu_2)/2$ *is a necessary condition for a bivariate Matérn model to capture any cross-correlation in the multivariate spatial data.*

2. *If* $\nu_{12} = (\nu_1 + \nu_2)/2$ *and* $a_1 = a_2 = a_{12} = a$*, the full bivariate Matérn model is valid if and only if*

$$\rho_{12} \leq \frac{\Gamma(\nu_1 + \frac{d}{2})^{1/2}}{\Gamma(\nu_1)^{1/2}} \frac{\Gamma(\nu_2 + \frac{d}{2})^{1/2}}{\Gamma(\nu_2)^{1/2}} \frac{\Gamma(\frac{1}{2}(\nu_1 + \nu_2))}{\Gamma(\frac{1}{2}(\nu_1 + \nu_2) + \frac{d}{2})}.$$

As a special case, if $d = 2$ (e.g., if we focus only on the two-dimensional spatial domain), then the inequality condition on ρ_{12} can be simplified to

$$|\rho_{12}| \leq \frac{(\nu_1\nu_2)^{1/2}}{\frac{1}{2}(\nu_1+\nu_2)}.$$

3. *(Gneiting et al., 2010, Theorem 4) When $\nu_{12} \geq (\nu_1+\nu_2)/2$ and $a_{12}^2 \geq (a_1^2+a_2^2)/2$, a sufficient but not necessary condition for the full bivariate Matérn model to be valid is given by*

$$|\rho_{12}| \leq \frac{a_1^{\nu_1} a_2^{\nu^2}}{a_{12}^{2\nu_{12}}} \frac{\Gamma(\nu_{12})}{\Gamma(\nu_1)^{1/2}\Gamma(\nu_2)^{1/2}} \left(e\,\frac{a_{12}-(a_1^2+a_2^2)/2}{\nu_{12}-(\nu_1+\nu_2)/2}\right)^{\nu_{12}-(\nu_1+\nu_2)/end}.$$

Apanasovich et al. (2012) proved a more general result than Theorem S5, where the smoothness parameter in the cross-covariance function of two processes can be greater than or equal to the average of the smoothness parameters in the covariance functions of the corresponding two marginal processes, as opposed to the equality condition in Theorem S5.

Theorem S8 *(Apanasovich et al., 2012, Theorem 1) The Matérn model (S7) provides a valid structure if there exists Δ_A such that*

i) $\nu_{ij} - \frac{\nu_i+\nu_j}{2} = \Delta_A(1-A_{ij})$, $\;i,j = 1,\ldots,p$, *where $0 \leq A_{ij} \leq 1$ form a valid correlation matrix.*

ii) The collection $-a_{ij}^2$, $i,j = 1,\ldots,p$ form a conditional non-negative definite matrix. In other words, let $\boldsymbol{M}$ be the matrix formed by $-a_{ij}^2$, $i,j=1,\ldots,p$. Then $\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^p$ such that $\sum_{i=1}^p x_i = 0$.

iii) The collection,

$$\rho_{ij}\sigma_i\sigma_j a_{ij}^{2\Delta_A+\nu_i+\nu_j} \frac{\Gamma(\nu_{ij}+\frac{d}{2})}{\Gamma(\frac{\nu_i+\nu_j}{2}+\frac{d}{2})\Gamma(\nu_{ij})}, \quad i,j=1,\ldots,p,$$

form a non-negative definite matrix.

Condition (i) in Theorem S8 implies $\nu_{ij} \geq (\nu_i + \nu_j)$ since we must have $\Delta_A(1 - A_{ij}) \geq 0$. Examples of a collection of $\{a_{ij}\}_{i,j=1}^p$ that satisfy condition (ii) include $a_{ij}^2 = (a_i^2 + a_j^2)/2 + \tau(a_i - a_j)^2,\ 0 \leq \tau \leq \infty$ and $a_{ij} = \max\{a_i, a_j\}$ (see Apanasovich et al., 2012). In other words, Theorem S8 relaxes the restrictions on the smoothness parameters and the scale parameters in Theorem S5. However, as noted in Apanasovich et al. (2012), the constraints on the co-locational correlation coefficients $\rho_{ij},\ i,j = 1,\ldots,p$, still seem unavoidable. These constraints, i.e., condition (iii) in Theorem S8, can be reformulated in terms of upper bounds on ρ_{ij}, which depend on how much the smoothness and scale parameters deviate from the corresponding parameters of the two marginal processes:

$$\rho_{ij}^2 \leq \prod_{k=1}^{3} \tau_{ij}^{(k)} \leq 1, \qquad i,j = 1,\ldots,p,$$

$$\tau_{ij}^{(1)} = \frac{\mathcal{B}\left(\nu_{ij}, \frac{d}{2}\right)^2}{\mathcal{B}\left(\frac{\nu_i+\nu_j}{2}, \frac{d}{2}\right)^2}, \qquad \tau_{ij}^{(2)} = \left(\frac{a_i a_j}{a_{ij}^2}\right)^{2\Delta_A},$$

$$\text{and} \quad \tau_{ij}^{(3)} = \frac{\Gamma\left(\frac{\nu_i+\nu_j}{2}\right)^2}{\Gamma(\nu_i)\Gamma(\nu_j)} \frac{a_i^{2\nu_i} a_j^{2\nu_j}}{a_{ij}^{2(\nu_i+\nu_j)}},$$

where $\mathcal{B}(\cdot,\cdot)$ is the Beta function. See Apanasovich et al. (2012) for the proof. One implication of Theorems S5, S7 and S8 is that we need to place more stringent restrictions on the smoothness and scale parameters in the Matérn correlation functions in order to relax the constraints on the co-locational correlation parameters $\rho_{ij},\ 1 \leq i \neq j \leq p$. In particular, in Theorem S7, if all the smoothness parameters are equal and so are the scale parameters, then the restriction on ρ_{12} is no longer needed.

S3 A review of the linear model of co-regionalization

The linear model of co-regionalization (LMC) (Goulard and Voltz, 1992; Wackernagel, 2003; Schmidt and Gelfand, 2003; Zhang, 2007) is commonly used approach for obtaining valid covariance structure for multivariate spatial processes. The LMC expresses a p-dimensional multivariate process $\boldsymbol{U}(\boldsymbol{s})$ as a linear combination of r $(1 \leq r \leq p)$ independent spatial processes $\boldsymbol{W}(\boldsymbol{s}) = (w_1(\boldsymbol{s}), \ldots, w_r(\boldsymbol{s}))^T$, i.e.,

$$\boldsymbol{U}(\boldsymbol{s}) = \boldsymbol{A}\boldsymbol{W}(\boldsymbol{s}),$$

where $\boldsymbol{A}$ is a $p \times r$ full-rank coefficient matrix and $w_k(\boldsymbol{s})$ has zero mean, unit variance and stationary correlation function $\rho_k(\boldsymbol{h})$ for $1 \leq k \leq r$. The cross-covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{U}(\boldsymbol{s})$ is thus given by

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Theta}(\boldsymbol{h})\boldsymbol{A}^T,$$

where $\boldsymbol{\Theta}(\boldsymbol{h}) = \text{diag}\{\rho_1(\boldsymbol{h}), \ldots, \rho_r(\boldsymbol{h})\}$.

S4 A further discussion of parameterization of the correlation and covariance functions

We have discussed several possible choices for the correlation functions $\rho_\alpha(\boldsymbol{h})$, $\rho_\epsilon(\boldsymbol{h})$ and $\rho_c(\boldsymbol{h})$, but the question remains what restrictions we should embrace in practice. In addition to the modeling flexibility afforded by the assumptions on the correlation functions, we also need to consider the practicality of the model fitting procedures. As the marginal distribution is the same regardless of the choice of the correlation functions, we suggest the following two options that can make the model fitting elegantly tractable while, as the same time, still provide enough flexibility in the covariance structure.

1. As mentioned previously, in the Matérn model (S7), the constraint on ϱ depends on how much the smoothness and scale parameters of the cross-correlation function deviate from the corresponding parameters of the two marginal processes. Since a major feature of the HASP model is the ability to capture skewness in the process which depends on the co-locational correlation parameter ϱ, it is undesirable for us to place constraints on ϱ.

   According to Theorem S7, we can remove the restriction on ϱ by requiring both the $\alpha(\boldsymbol{s})$ and $\epsilon(\boldsymbol{s})$ processes have the same degree of smoothness, i.e., $\nu_1 = \nu_2$. In other words, there is no restriction on ϱ for the HASP model if we require

$$\rho_\alpha(\boldsymbol{h}) = \rho_c(\boldsymbol{h}) = \rho_\epsilon(\boldsymbol{h}) = \varrho(\boldsymbol{h}). \tag{S8}$$

In fact, it is easy to see that we can use any correlation function, not just Matérn functions, for $\varrho(\boldsymbol{h})$ and the bivariate Gaussian process is always valid. For $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{s}'$, the covariance function is

$$\operatorname{cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}')) = \left[\varrho(\boldsymbol{h}) + \frac{\tau^2\varrho^2}{4}(1+\varrho(\boldsymbol{h}))^2\right] \exp\left\{\frac{\tau^2}{4}(1+\varrho(\boldsymbol{h}))\right\} - \frac{1}{4}\tau^2\varrho^2 e^{\tau^2/4}. \tag{S9}$$

Under the assumption (S8), the three models (3), (8) and (9) are also equivalent to each other. The condition (S8) appears to be very restrictive, but as we have learned from our previous discussion, all the parameters in the covariance function, even in the more restrictive Gaussian process, cannot be consistently estimated at the same time (Zhang, 2004). In the case of the exponential covariance function, $\sigma^2 \exp(-ah)$, either a or σ^2 can be fixed arbitrarily and the composite quantity $a\sigma^2$ can still be estimated consistently and efficiently (Ying, 1991). As a result, the assumption (S8) might not be as restrictive as it seems and its impact on the model performance should be very limited.

2. A more flexible model is the LMC version (9). (9) is easier to work with than (8) since conditioning on $\boldsymbol{a}(\boldsymbol{s})$, $Y(\boldsymbol{s})$ is a Gaussian process. Regardless of the choice of correlation functions $\rho_\alpha(\boldsymbol{h})$ and $\rho_\xi(\boldsymbol{h})$, the model is always valid by construction. The covariance function of (9) is given by:

$$\begin{aligned}\operatorname{cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}')) = &\left[\varrho^2\rho_\alpha(\boldsymbol{h}) + (1-\varrho^2)\rho_\xi(\boldsymbol{h}) + \frac{\tau^2\varrho^2}{4}(1+\rho_\alpha(\boldsymbol{h}))^2\right] \times \\ &\exp\left\{\frac{\tau^2}{4}(1+\rho_\alpha(\boldsymbol{h}))\right\} - \frac{1}{4}\tau^2\varrho^2 e^{\tau^2/4}.\end{aligned} \tag{S10}$$

Plots of the two covariances are presented below to facilitate the understanding of their behaviors and differences. Figure S2 shows the covariance function (S9) for difference choices

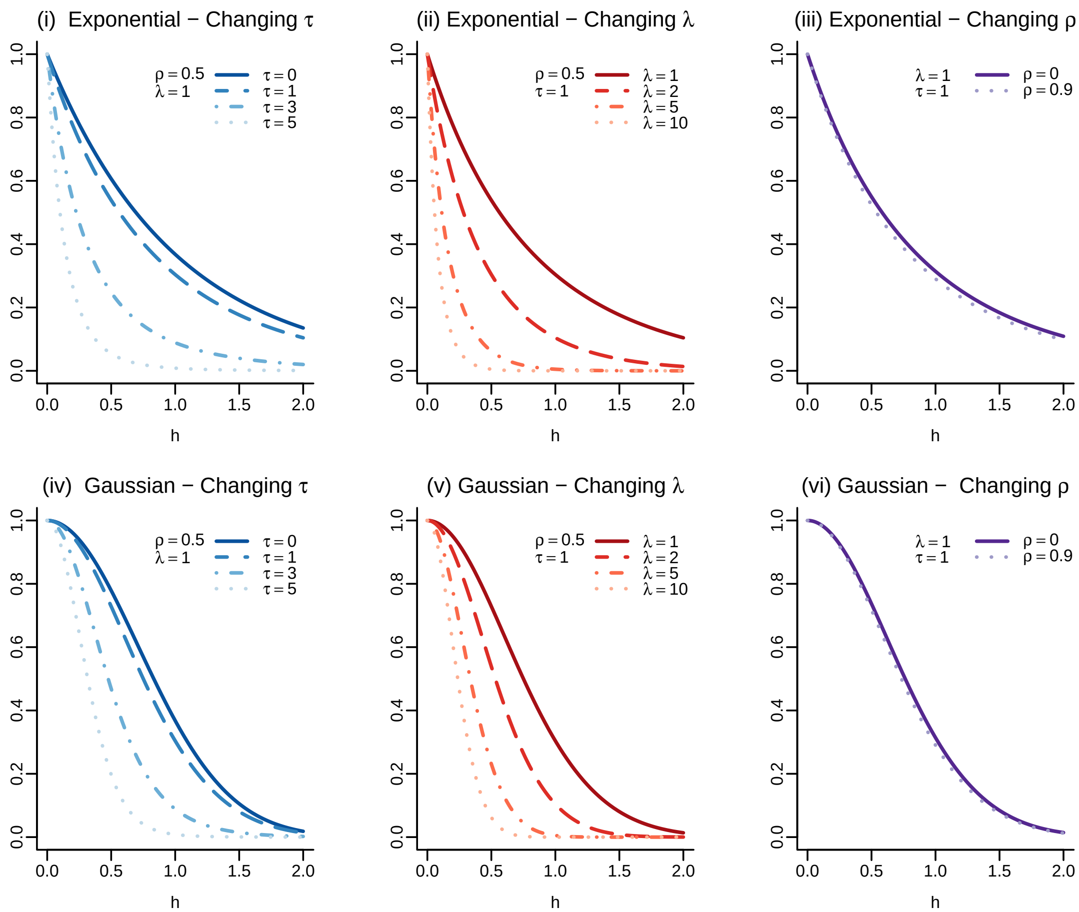

Figure S2: The shape of the correlation function implied by (S9) for different choices of $\varrho(\boldsymbol{h})$. Plot (i)–(iii) use the exponential correlation function for $\varrho(\boldsymbol{h})$, while plots (iv)–(vi) use the Gaussian correlation function. The values of the other parameters are shown in the plots.

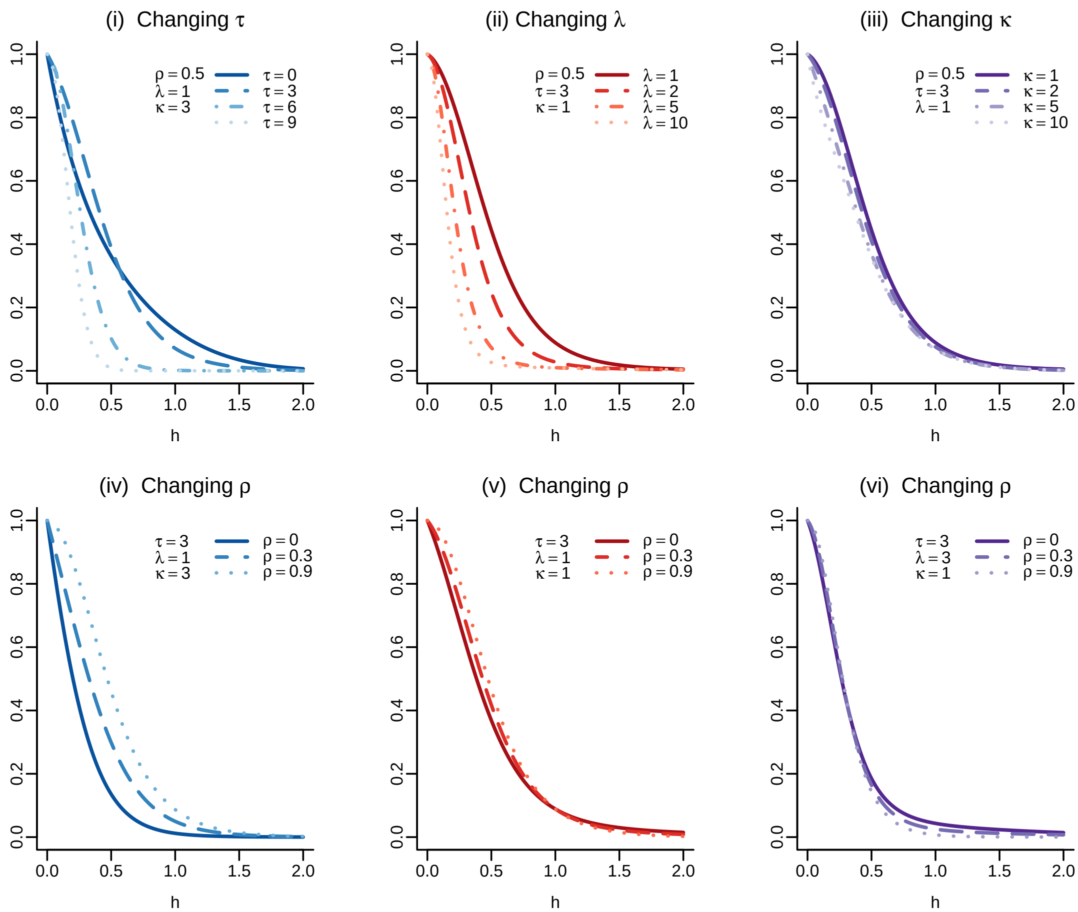

Figure S3: The shape of the correlation function implied by (S10) where $\rho_\alpha(\boldsymbol{h})$ assumes the form of a Gaussian correlation function and $\rho_\xi(\boldsymbol{h})$ the form of an exponential correlation function. The values of the other parameters are shown in the plots.

of $\varrho(\boldsymbol{h})$ and different values of other parameters. The exponential correlation function is used in plots (i)–(iii), i.e, $\varrho(\boldsymbol{h}) = \exp(-\lambda h)$. In plots (iv)–(vi) in Figure S2, the Gaussian correlation function is used, thus $\varrho(\boldsymbol{h}) = \exp(-\lambda h^2)$. The Gaussian (or double exponential) correlation function is chosen because it is the limit of the Matérn class (S6) as $\nu \to \infty$. The most important difference between a Gaussian correlation function and an exponential correlation function is the smoothness of the functions at the origin $h = 0$, which, in turn, determines the smoothness of the sample paths of the resulting spatial processes. For spatial prediction problems, the smoothness of a correlation function at the origin is of great importance. Stein (1999) argued that the smoothness of the sample path is equivalent to the rate of decay of the spectral density of the correlation function at high frequencies, and under infill asymptotics, the low frequency behavior of the spectral density has asymptotically negligible impact on spatial interpolation, while the high frequency behavior plays a pivotal role.

Figure S2 shows that the co-locational correlation parameter ϱ has little impact on the covariance function, while it has a big impact on the marginal distribution of the spatial process as can be seen in Figure S1. In addition, both τ and λ impact the covariance function in a similar fashion.

Due to the additional parameters, the covariance function (S10) is more flexible than (S9), which can be clearly seen in Figure S3. The plots in Figure S3 are based on the assumptions that $\rho_\alpha(\boldsymbol{h})$ takes the form of an exponential correlation function $\exp(-\lambda h)$, and that $\rho_\xi(\boldsymbol{h})$ takes the form of a Gaussian correlation function $\exp(-\kappa h^2)$ so that the process $\alpha(\boldsymbol{s})$ has a smoother sample path than $\epsilon(\boldsymbol{s})$. Same as in Figure S2, the impact of different values of ϱ on the shape of the covariance function in Figure S3 is not as large as the impacts of τ and λ. However, the smoothness of the covariance function at $h = 0$ as well as its effective correlation length does depend partially on ϱ. These impacts of ϱ are in a large part due to the relation

$$\rho_\epsilon(\boldsymbol{h}) = \varrho^2 \rho_\alpha(\boldsymbol{h}) + (1 - \varrho^2)\rho_\xi(\boldsymbol{h}).$$

The parameters τ and λ also have similar effects on the covariance function, suggesting that the consistent estimation of both parameters simultaneously might still be a problem.

It is ultimately up to the researcher to decide which of the above two strategies to choose. The covariance function (S10) is certainly more flexible than (S9) but in practice, we do not always have enough data to efficiently estimate it. When the data is not abundant or we have no need to estimate the smoothness of the underlying process by using a very general covariance function, the more restricted strategy (S9) might be a better choice. On the other hand, the model fitting procedures for these two strategies are very similar.

S5 Example HASP sample paths

Here we present several realizations from the heteroscedastic asymmetric spatial process and compare them with realizations from the SHP and the Gaussian process (GP). Figure S4 and S5 show the sample paths of the processes defined in $\mathbb{R}$, and the examples for the processes defined on $\mathbb{R}^2$ are presented in Figures S6 and S7. For Figures S4 and S6, an exponential function is used as the correlation and cross correlation functions of the (potentially latent) Gaussian processes, whereas a Gaussian correlation function is used for Figures S5 and S7.

As discussed earlier, the sample path of a Gaussian process with a Gaussian correlation function is smoother than that with an exponential correlation function, which can be clearly seen from the plots. From the time series literature, we know that a stationary and Markov Gaussian process defined on $\mathbb{R}^+$ with a continuous correlation function is necessarily an Orstein-Uhlenbeck process, which is the unique stationary solution to the differential equation

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t,\ t \geq 0,$$

where $\{W_t,\ t \geq 0\}$ is a Brownian motion with unit variance (see Uhlenbeck and Ornstein, 1930; Doob, 1942). The Orstein-Uhlenbeck process has a stationary and isotropic exponential correlation function and its Euler-Maruyama discretization is the discrete-time AR(1) process. In addition, the sample path of an Orstein-Uhlenbeck process is continuous but nowhere differentiable with probability 1.

Apart from the discrepancies in the smoothness of the sample paths caused by different correlation functions, we can also recognize the unique features of the curves or surfaces drawn from the non-Gaussian processes compared to those drawn from the Gaussian processes (which is set to have zero mean and unit variance). In Figures S4–S7, the curves or surfaces drawn from a Gaussian process mostly vary within -2 and 2. When we modulate the said Gaussian process with an independent exponential-Gaussian process and get the GLG/SHP model, the resulting sample paths end up with more fluctuations in terms of the ranges of the functions over the entire domain. Furthermore, when we correlate the Gaussian process and the modulating exponential-Gaussian process, i.e., for the HASP with positive or negative skewness, the curves or surfaces draw from the processes can have an even larger swing in one direction. This suggests that the HASP model (with SHP as a special case) is useful in modeling spatially indexed data when the underlying curve or surface has marked peaks or valleys which, if modeled by the Gaussian processes, might be overly shrunken toward the overall mean.

S6 An MCMC algorithm for fitting the HASP model

Let $\boldsymbol{\theta}$ denote all model parameters other than $\boldsymbol{\alpha}$, i.e., $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \phi, \psi^2, \tau^2, \lambda, \kappa, \tilde{\delta}\}$. For a single iteration in the MCMC algorithm, we simulate from the posterior distribution using the following steps.

1. Update the latent variables $\boldsymbol{\alpha}$ either one-by-one or through a block update as in Palacios and Steel (2006). One approach is to use a random walk Metropolis-Hastings step. Alternatively, we can consider a better proposal distribution by using techniques such as approximating the full conditional distribution as was done in Palacios and Steel (2006) or using the Metropolis-adjusted Langevin algorithm (Grenander and Miller,

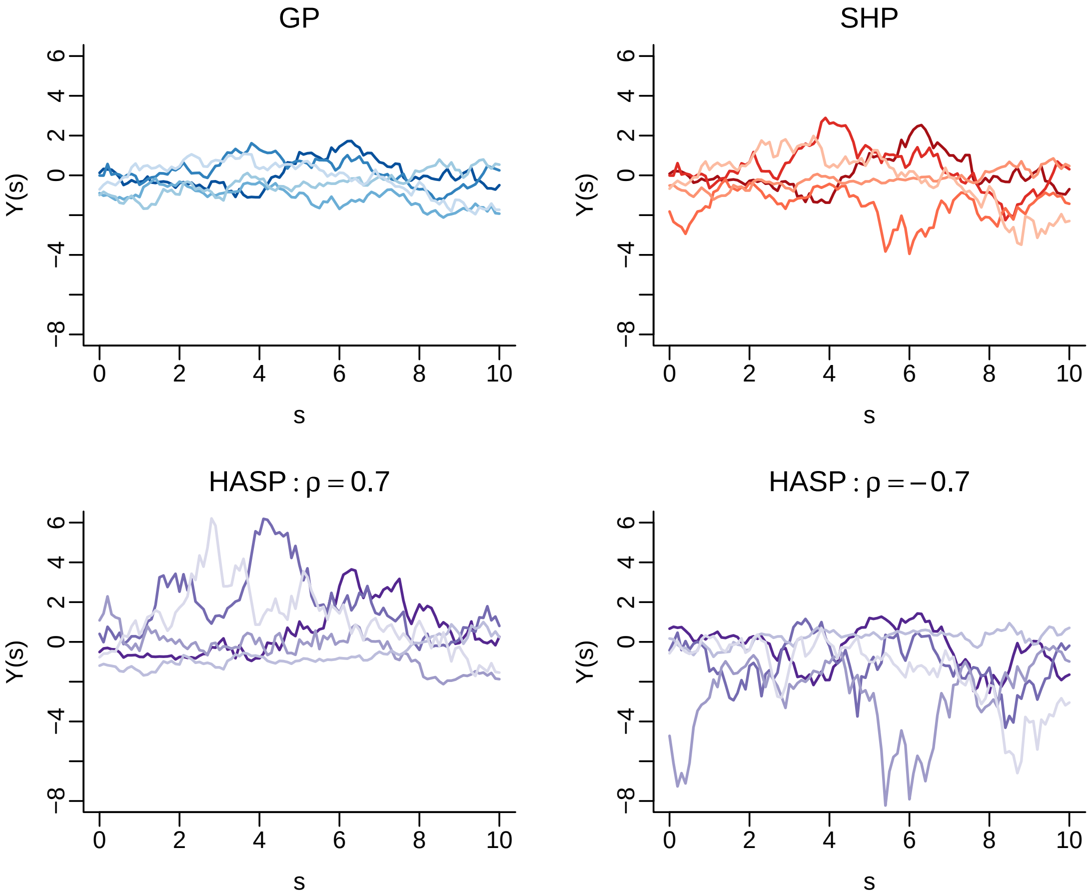

Figure S4: Five sample paths from each of the four spatial processes in $\mathbb{R}$: GP, GLG/SHP, HASP with positive skewness and HASP with negative skewness. An exponential correlation function is used for the GP. For the non-Gaussian processes, the same exponential correlation function is used for the marginal as well as the cross correlation functions of the latent multivariate Gaussian process. The sample paths for different processes bear resemblance to each other because the same seed is used for random number generation for easier comparison of the different processes. The sample paths are approximated based on a finite number of observations on a grid with an increment of 0.1.

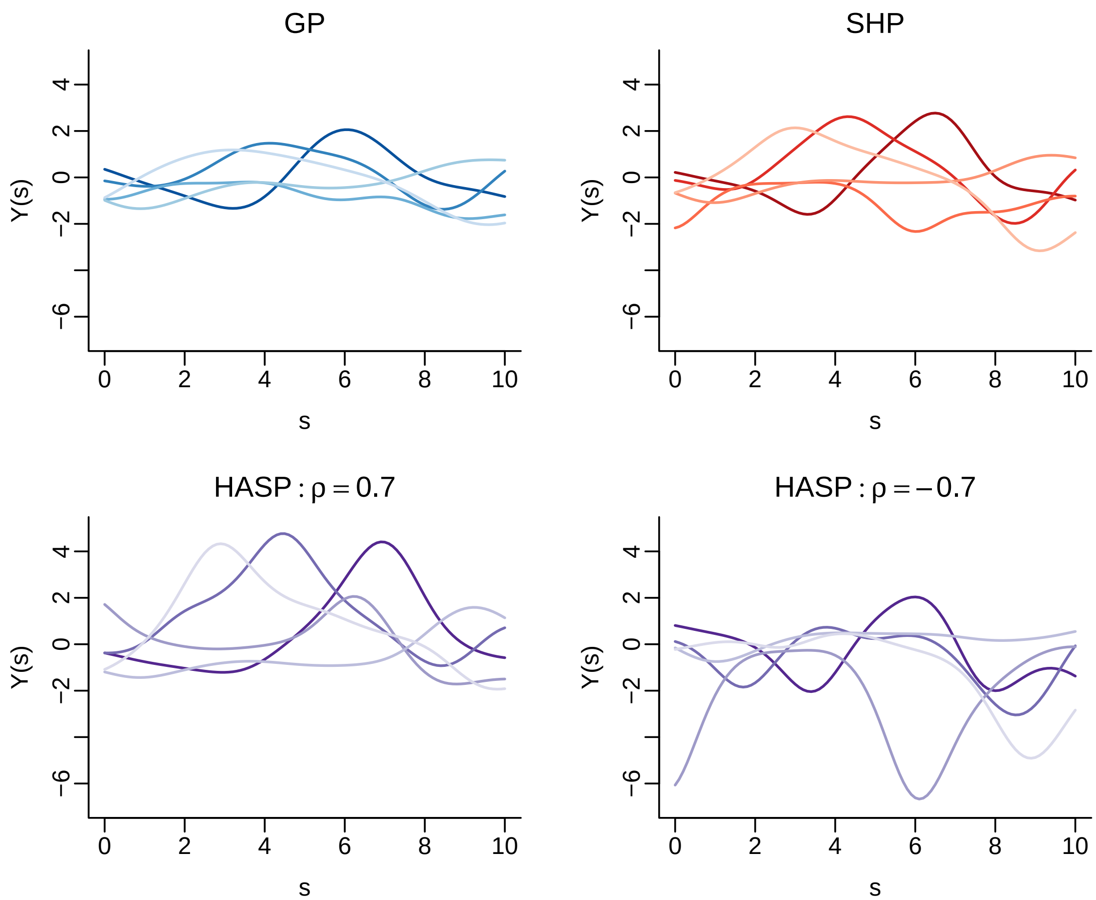

Figure S5: Five sample paths from each of the four spatial processes in $\mathbb{R}$: GP, GLG/SHP, HASP with positive skewness and HASP with negative skewness. A Gaussian (or double exponential) correlation function is used for the GP. For the non-Gaussian processes, the same Gaussian correlation function is used for the marginal as well as cross correlation functions of the latent multivariate Gaussian process. The sample paths are constructed in the same way as in Figure S4.

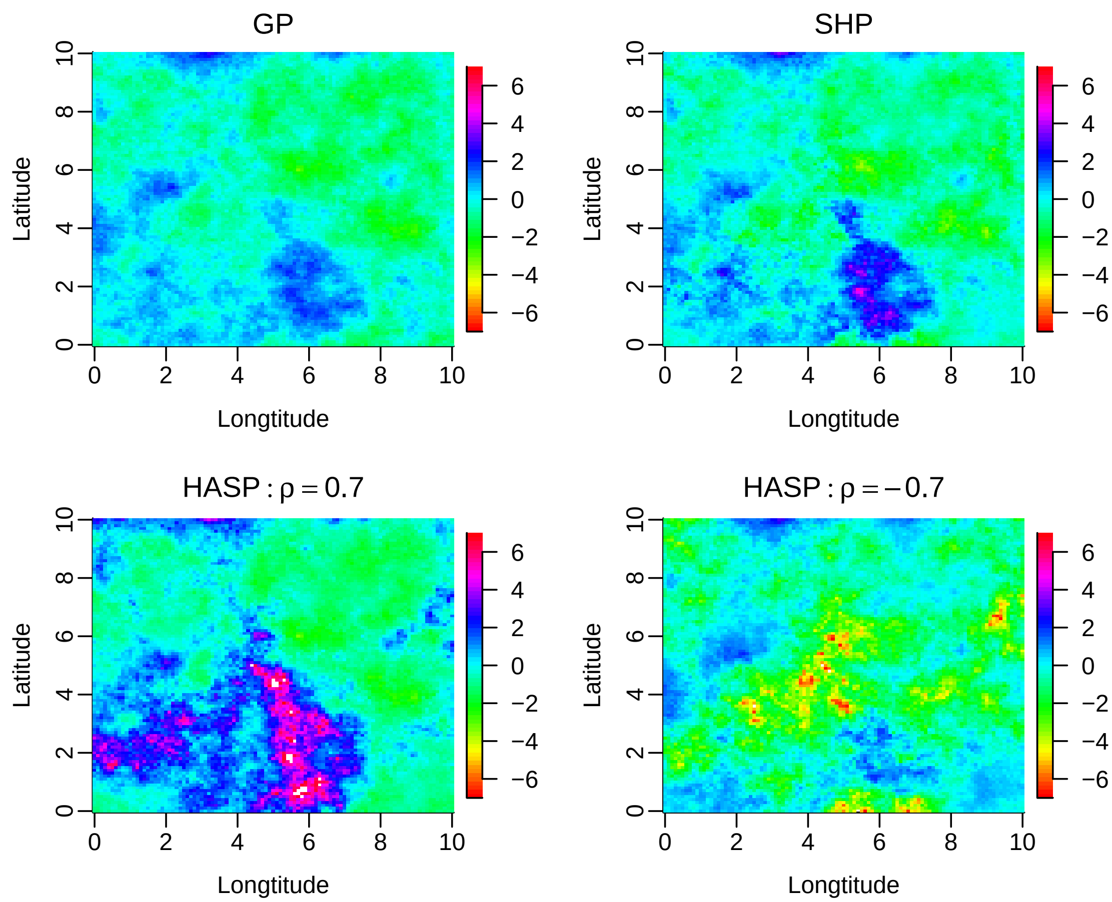

Figure S6: Five sample paths from each of the four spatial processes in $\mathbb{R}^2$: GP, GLG/SHP, HASP with positive skewness and HASP with negative skewness. An exponential correlation function is used for the GP. For the non-Gaussian processes, the same exponential correlation function is used for the marginal as well as cross correlation functions of the latent multivariate Gaussian process. Again, the sample paths for different processes bear resemblance to each other because the same seed is used for random number generation for easier comparison of the different processes. The sample paths are approximated based on a finite number of observations on a rectangular grid with an increment of 0.1 in each direction.

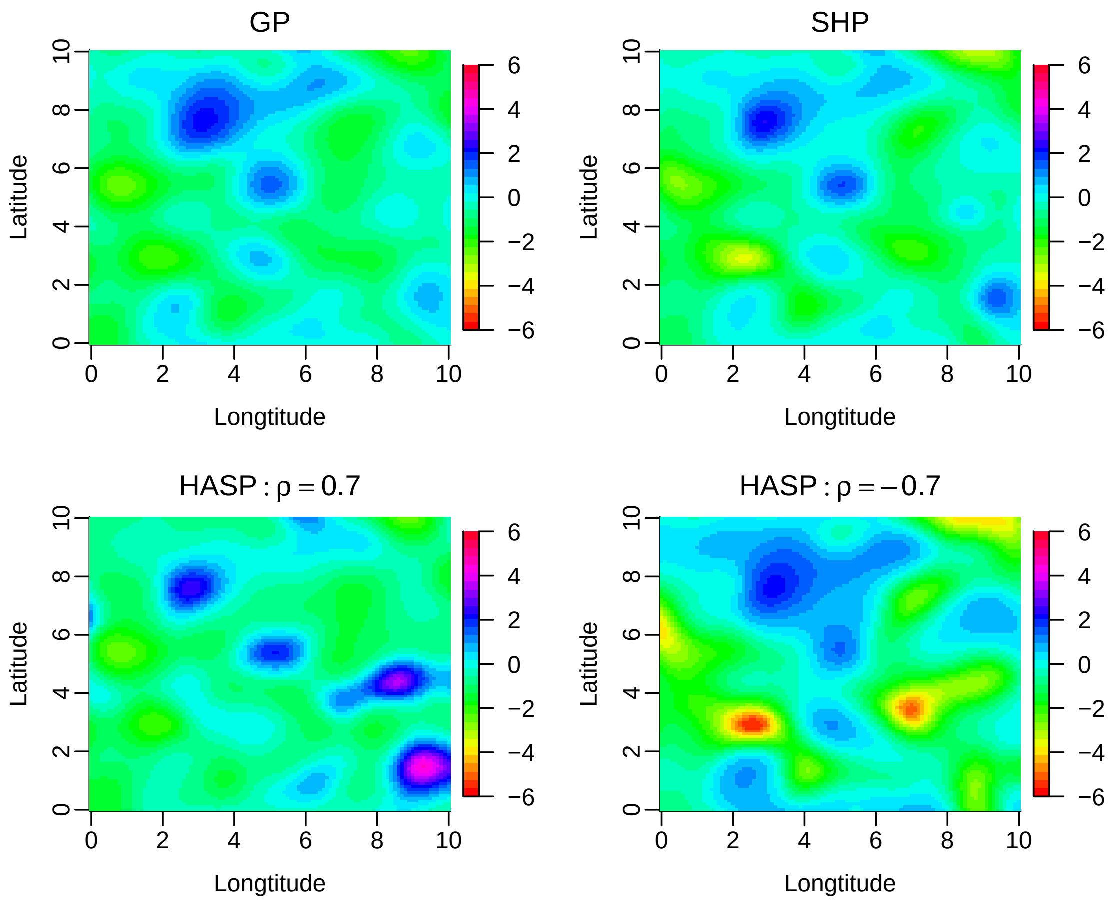

Figure S7: Five sample paths from each of the four spatial processes in $\mathbb{R}^2$: GP, GLG/SHP, HASP with positive skewness and HASP with negative skewness. A Gaussian correlation function is used for the GP. For the non-Gaussian processes, the same Gaussian correlation function is used for the marginal as well as cross correlation functions of the latent multivariate Gaussian process. The sample paths are constructed in the same way as in Figure S6.

1994; Roberts and Tweedie, 1996). The full conditional distribution of $\boldsymbol{\alpha}$ is given by

$$
\begin{aligned}
&p(\boldsymbol{\alpha}|\boldsymbol{Y},\boldsymbol{\theta}) \\
&\quad\propto |\boldsymbol{V}_\alpha|^{-1}\exp\left\{-\frac{1}{2\psi^2}[\boldsymbol{V}_\alpha^{-1}\tilde{\boldsymbol{V}}_\delta^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})-\phi\boldsymbol{\alpha}]^T \times \right.\\
&\qquad\qquad \left.\boldsymbol{\Xi}^{-1}[\boldsymbol{V}_\alpha^{-1}\tilde{\boldsymbol{V}}_\delta^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})-\phi\boldsymbol{\alpha}]\right\}\exp\left\{-\frac{1}{2\tau^2}\boldsymbol{\alpha}^T(\boldsymbol{s})\boldsymbol{P}^{-1}\boldsymbol{\alpha}\right\}.
\end{aligned}
$$

Let $\boldsymbol{Y}_\phi = \tilde{\boldsymbol{V}}_\delta^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})$ and C be a constant in $\boldsymbol{\alpha}$, then

$$
\begin{aligned}
&\log p(\boldsymbol{\alpha}|\boldsymbol{Y},\boldsymbol{\theta}) \\
&= C - \frac{1}{2}\sum_{k=1}^n \alpha(\boldsymbol{s}_k) - \frac{1}{2\psi^2}\left[(\boldsymbol{V}_\alpha^{-1}\boldsymbol{Y}_\phi)^T\boldsymbol{\Xi}^{-1}(\boldsymbol{V}_\alpha^{-1}\boldsymbol{Y}_\phi) - 2\phi\boldsymbol{\alpha}^T\boldsymbol{\Xi}^{-1}(\boldsymbol{V}_\alpha^{-1}\boldsymbol{Y}_\phi)\right] \\
&\qquad - \frac{1}{2}\boldsymbol{\alpha}^T\left[\frac{\phi^2}{\psi^2}\boldsymbol{\Xi}^{-1} + \frac{1}{\tau^2}\boldsymbol{P}^{-1}\right]\boldsymbol{\alpha} \\
&= C - \frac{1}{2}\sum_{k=1}^n \alpha(\boldsymbol{s}_k) - \frac{1}{2\psi^2}\left[(\boldsymbol{V}_\alpha^{-1}\boldsymbol{Y}_\phi - 2\phi\boldsymbol{\alpha})^T\boldsymbol{\Xi}^{-1}(\boldsymbol{V}_\alpha^{-1}\boldsymbol{Y}_\phi)\right] \\
&\qquad - \frac{1}{2}\boldsymbol{\alpha}^T\left[\frac{\phi^2}{\psi^2}\boldsymbol{\Xi}^{-1} + \frac{1}{\tau^2}\boldsymbol{P}^{-1}\right]\boldsymbol{\alpha}. \qquad \text{(S11)}
\end{aligned}
$$

We use the Metropolis-adjusted Langevin algorithm (MALA) algorithm to sequentially update the components of $\boldsymbol{\alpha}$. For $k = 1, \cdots, n$, let $l_t(\alpha(\boldsymbol{s}_k))$ denote the full conditional log likelihood function of $\alpha(\boldsymbol{s}_k)$ (up to a constant in $\alpha(\boldsymbol{s}_k)$) evaluated at the most recent draws of all other model parameters in iteration t. Then,

(a) Given the samples of $\alpha(\boldsymbol{s}_k)$ in the iteration $t-1$, $\alpha^{[t-1]}(\boldsymbol{s}_k)$, draw a sample $\alpha^*(\boldsymbol{s}_k)$ from the proposal distribution

$$N\left(\alpha^{[t-1]}(\boldsymbol{s}_k) + \frac{\varsigma_0^2}{2}\nabla l_t(\alpha^{[t-1]}(\boldsymbol{s}_k)), \varsigma_0^2\right),$$

where ς_0 is a user-defined tuning parameter.

(b) Calculate the ratio e^{Δ_t} where

$$
\begin{aligned}
\Delta_t = &\left\{l_t(\alpha^*(\boldsymbol{s}_k)) - \frac{1}{2\varsigma_0^2}\left[\alpha^{[t-1]}(\boldsymbol{s}_k) - \alpha^*(\boldsymbol{s}_k) - \frac{\varsigma_0^2}{2}\nabla l_t(\alpha^*(\boldsymbol{s}_k))\right]^2\right\} \\
&- \left\{l_t(\alpha^{[t-1]}(\boldsymbol{s}_k)) - \frac{1}{2\varsigma_0^2}\left[\alpha^*(\boldsymbol{s}_k) - \alpha^{[t-1]}(\boldsymbol{s}_k) - \frac{\varsigma_0^2}{2}\nabla l_t(\alpha^{[t-1]}(\boldsymbol{s}_k))\right]^2\right\}.
\end{aligned}
$$

(c) Set $\alpha^{[t]}(\boldsymbol{s}_k) = \alpha^*(\boldsymbol{s}_k)$ with probability $\min(1, e^{\Delta_t})$. Otherwise, set $\alpha^{[t]}(\boldsymbol{s}_k) = \alpha^{[t-1]}(\boldsymbol{s}_k)$.

A simpler form of the full conditional log likelihood function of $\alpha(\boldsymbol{s}_k),\ k = 1, \ldots, n$, can be derived from (S11), as is shown below (with subscript t omitted for simplicity). For convenience, let

$$\boldsymbol{\Omega} = \boldsymbol{\Xi}^{-1} \text{ and } \boldsymbol{\Lambda} = \frac{\phi^2}{\psi^2}\boldsymbol{\Xi}^{-1} + \frac{1}{\tau^2}\boldsymbol{P}^{-1}.$$

Now let $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Lambda}_i$ denote the i-th column vector of $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$, respectively. Since $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$ are both symmetric, $\boldsymbol{\Omega}_i^T$ and $\boldsymbol{\Lambda}_i^T$ are thus their respective i-th row vectors. In addition, let Ω_{ij} and Λ_{ij} denote the (i,j)th elements of the respective matrices. Finally, let $\boldsymbol{W}_\alpha^{-1} = (e^{\alpha(-\boldsymbol{s}_1)/2}, \cdots, e^{-\alpha(\boldsymbol{s}_n)/2})^T$ and let C_k be a scalar that does not depend on $\alpha(\boldsymbol{s}_k)$. Ignoring the terms in (S11) that do not contain $\alpha(\boldsymbol{s}_k)$, we have

$$\begin{aligned} l(\alpha(\boldsymbol{s}_k)) = C_k - \frac{1}{2}\alpha(\boldsymbol{s}_k) - \frac{1}{2\psi^2}\Bigg\{ &\left[e^{-\alpha(\boldsymbol{s}_k)/2}Y_{\phi_k} - 2\phi\alpha(\boldsymbol{s}_k)\right]\boldsymbol{\Omega}_k \cdot (\boldsymbol{W}_\alpha^{-1} \circ \boldsymbol{Y}_\phi) \\ &+ \left[e^{-\alpha(\boldsymbol{s}_k)/2}Y_{\phi_k}\right]\boldsymbol{\Omega}_k \cdot (\boldsymbol{W}_\alpha^{-1} \circ \boldsymbol{Y}_\phi - 2\phi\boldsymbol{\alpha}) \\ &- \left[e^{-\alpha(\boldsymbol{s}_k)/2}Y_{\phi_k} - 2\phi\alpha(\boldsymbol{s}_k)\right]\Omega_{kk}\left[e^{-\alpha(\boldsymbol{s}_k)/2}Y_{\phi_k}\right]\Bigg\} \\ &- \frac{1}{2}\left\{2\alpha(\boldsymbol{s}_k)\boldsymbol{\Lambda}_k \cdot \boldsymbol{\alpha} - \alpha(\boldsymbol{s}_k)^2\Lambda_{kk}\right\}. \end{aligned}$$

To calculate the gradient of $l(\alpha(\boldsymbol{s}_k))$, we can use the following general result.

Lemma S9 *Let $\boldsymbol{\omega}$ be a $p \times 1$ column vector and $\boldsymbol{M}$ be a $p \times p$ symmetric matrix whose k-th column vector is denoted as $\boldsymbol{M}_k$. Let g and h be two differentiable functions, defined as*

$$\begin{aligned} &f, g : \mathbb{R} \to \mathbb{R}; \\ &f(\boldsymbol{\omega}) = (f(\omega_1), \ldots, f(\omega_p))^T; \\ &g(\boldsymbol{\omega}) = (g(\omega_1), \ldots, g(\omega_p))^T. \end{aligned}$$

Then, for $k = 1, \ldots, p$, we have

$$\frac{\partial}{\partial \omega_k} f^T(\boldsymbol{\omega}) \boldsymbol{M} g(\boldsymbol{\omega}) = f'(\omega_k) \boldsymbol{M}_k \cdot g(\boldsymbol{\omega}) + g'(\omega_k) \boldsymbol{M}_k \cdot f(\boldsymbol{\omega}).$$

The proof is trivial, which involves expanding $f^T(\boldsymbol{\omega})\boldsymbol{M}g(\boldsymbol{\omega})$ into a quadratic form and then applying basic calculus rules. By Lemma S9 and (S11), the gradient of $l(\alpha(\boldsymbol{s}_k))$ is given by

$$\begin{aligned}
\nabla l(\alpha(\boldsymbol{s}_k)) &= \frac{\partial}{\partial \alpha(\boldsymbol{s}_k)} l(\alpha(\boldsymbol{s}_k)) \\
&= -\frac{1}{2} - \frac{1}{2\psi^2} \Bigg\{ \left[-\frac{1}{2} e^{-\alpha(\boldsymbol{s}_k)/2} Y_{\phi_k} - 2\phi \right] \boldsymbol{\Omega}_k \cdot (\boldsymbol{W}_\alpha^{-1} \circ \boldsymbol{Y}_\phi) \\
&\qquad + \left[-\frac{1}{2} e^{-\alpha(\boldsymbol{s}_k)/2} Y_{\phi_k} \right] \boldsymbol{\Omega}_k \cdot (\boldsymbol{W}_\alpha^{-1} \circ \boldsymbol{Y}_\phi - 2\phi \boldsymbol{\alpha}) \Bigg\} - \boldsymbol{\Lambda}_k \cdot \boldsymbol{\alpha}.
\end{aligned}$$

2. Update $\boldsymbol{\beta}$ and ϕ from a multivariate normal full conditional distribution. Let $\boldsymbol{X}^* = [\boldsymbol{X} \quad \tilde{\boldsymbol{W}} \circ \boldsymbol{\alpha}]$, $\boldsymbol{D} = (\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}$ and $\boldsymbol{S}_D = \psi^2 \boldsymbol{D}\tilde{\boldsymbol{V}}\boldsymbol{\Xi}\tilde{\boldsymbol{V}}\boldsymbol{D}^T$. We assume the prior distribution for $[\boldsymbol{\beta} \quad \phi]^T$ is multivariate normal with mean $\boldsymbol{m}_0$ and covariance matrix $\boldsymbol{S}_0$, $N(\boldsymbol{m}_0, \boldsymbol{S}_0)$, then the full conditional for $[\boldsymbol{\beta} \quad \phi]^T$ is

$$[\boldsymbol{\beta} \quad \phi]^T | \boldsymbol{Y}(\boldsymbol{\theta}), \boldsymbol{\alpha}, \tilde{\delta}, \kappa, \psi^2 \sim N(\boldsymbol{m}_1, \boldsymbol{S}_1),$$

where

$$\begin{aligned}
\boldsymbol{S}_1 &= (\boldsymbol{S}_0^{-1} + \boldsymbol{S}_D^{-1})^{-1} = \boldsymbol{S}_0(\boldsymbol{S}_0 + \boldsymbol{S}_D)^{-1}\boldsymbol{S}_D = \boldsymbol{S}_D(\boldsymbol{S}_0 + \boldsymbol{S}_D)^{-1}\boldsymbol{S}_0, \\
\text{and} \quad \boldsymbol{m}_1 &= (\boldsymbol{S}_0^{-1} + \boldsymbol{S}_D^{-1})^{-1}(\boldsymbol{S}_0^{-1}\boldsymbol{m}_0 + \boldsymbol{S}_D^{-1} D\boldsymbol{Y}) \\
&= \boldsymbol{S}_D(\boldsymbol{S}_0 + \boldsymbol{S}_D)^{-1}\boldsymbol{m}_0 + \boldsymbol{S}_0(\boldsymbol{S}_0 + \boldsymbol{S}_D)^{-1}\boldsymbol{D}\boldsymbol{Y}.
\end{aligned}$$

3. Update the variance parameters τ^2 and ψ^2 from their full conditional distribution. We consider independent conjugate inverse gamma priors for τ^2 and ψ^2. Since the parameters τ^2 and λ have similar effects on the covariance function of the HASP process as illustrated in Figure S2, it might be desirable that we use a highly informative prior for at least one of the two parameters.

Suppose the prior for τ^2 is $IG(a_0, b_0)$ and the prior for ψ^2 is $IG(c_0, d_0)$, then their full conditional distributions are given, respectively, by

$$IG(a_0 + 0.5n, b_0 + 0.5\boldsymbol{\alpha}^T(\boldsymbol{s})\boldsymbol{P}^{-1}\boldsymbol{\alpha}),$$

and

$$IG(c_0 + 0.5n, d_0 + 0.5\boldsymbol{Y}^{*T}(\boldsymbol{s})(\tilde{\boldsymbol{V}}\boldsymbol{\Xi}\tilde{\boldsymbol{V}})^{-1}\boldsymbol{Y}^*(\boldsymbol{s})),$$

where

$$\boldsymbol{Y}^*(\boldsymbol{s}) = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \phi\tilde{\boldsymbol{V}}\boldsymbol{\alpha}.$$

4. Update λ and κ from its full conditional distribution using a Metropolis-Hastings step. The full conditional distribution for λ is

$$\pi(\lambda|\boldsymbol{Y}, \boldsymbol{\alpha}, \phi, \psi^2, \tau^2, \boldsymbol{\beta}, \tilde{\delta}) \propto |\boldsymbol{P}|^{-1/2} \exp\left\{-\frac{1}{2\tau^2}\boldsymbol{\alpha}^T(\boldsymbol{s})\boldsymbol{P}^{-1}\boldsymbol{\alpha}\right\}\pi(\lambda),$$

while the full conditional distribution for κ is given by

$$\pi(\kappa|\boldsymbol{Y}, \boldsymbol{\alpha}, \phi, \psi^2, \tau^2, \boldsymbol{\beta}, \tilde{\delta}) \propto |\boldsymbol{\Xi}|^{-1/2} \exp\left\{-\frac{1}{2\psi^2}\boldsymbol{Y}^{*T}(\boldsymbol{s})(\tilde{\boldsymbol{V}}\boldsymbol{\Xi}\tilde{\boldsymbol{V}})^{-1}\boldsymbol{Y}^*(\boldsymbol{s})\right\}\pi(\kappa).$$

In the equations above, $\pi(\lambda)$ and $\pi(\kappa)$ denotes the probability density functions of the prior distributions of λ and κ, respectively. In practice, we can choose an exponential distribution, a gamma distribution or other distributions with positive support as the prior distributions. In our implementation, a truncated normal distribution is used as the proposal distribution in the Metropolis-Hastings algorithm.

5. Update $\tilde{\boldsymbol{\delta}}$ from its full conditional distribution using a Metropolis-Hastings step. Depending on model assumptions, $\tilde{\boldsymbol{\delta}}$ could be univariate, multivariate or even not needed in the model (when the original parameter $\boldsymbol{\delta}$ is univariate). We use a (potentially multivariate) normal distribution with mean $\tilde{\boldsymbol{m}}_0$ and covariance matrix $\tilde{\boldsymbol{S}}_0$ as the prior for

$\tilde{\boldsymbol{\delta}}$. The full conditional distribution of $\tilde{\boldsymbol{\delta}}$ is thus

$$
\begin{aligned}
&\pi(\tilde{\boldsymbol{\delta}}|\boldsymbol{Y}, \boldsymbol{\alpha}, \phi, \psi^2, \tau^2, \boldsymbol{\beta}, \kappa) \\
&\propto |\tilde{\boldsymbol{V}}|^{-1} \exp\left\{-\frac{1}{2\psi^2}\boldsymbol{Y}^{*T}(\boldsymbol{s})(\tilde{\boldsymbol{V}}\boldsymbol{\Xi}\tilde{\boldsymbol{V}})^{-1}\boldsymbol{Y}^*(\boldsymbol{s})\right\} \exp\left\{-\frac{1}{2}(\tilde{\boldsymbol{\delta}} - \tilde{\boldsymbol{m}}_0)^T\tilde{\boldsymbol{S}}_0^{-1}(\tilde{\boldsymbol{\delta}} - \tilde{\boldsymbol{m}}_0)\right\} \\
&\propto |\tilde{\boldsymbol{V}}_\delta|^{-1} \exp\left\{-\frac{1}{2\psi^2}[\tilde{\boldsymbol{V}}_\delta^{-1}\boldsymbol{V}_\alpha^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T - \phi\boldsymbol{\alpha}^T(\boldsymbol{s})] \quad \times \boldsymbol{\Xi}^{-1} \times \right. \\
&\qquad \left. [\tilde{\boldsymbol{V}}_\delta^{-1}\boldsymbol{V}_\alpha^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) - \phi\boldsymbol{\alpha}]\right\} \exp\left\{-\frac{1}{2}(\tilde{\boldsymbol{\delta}} - \tilde{\boldsymbol{m}}_0)^T\tilde{\boldsymbol{S}}_0^{-1}(\tilde{\boldsymbol{\delta}} - \tilde{\boldsymbol{m}}_0)\right\}.
\end{aligned}
$$

S7 Spatial Prediction

The objective of many spatial smoothing problems is to make predictions of the spatial process at unobserved locations. In other words, we need to calculate $E[Y(\boldsymbol{s}_0)|\boldsymbol{Y}]$ for $\boldsymbol{s}_0 \in \mathcal{D}$ as well as understand the uncertainty associated with this prediction. The commonly used kriging estimator is essentially a linear predictor based on a weighted average of the observed values, where the weights depend on the known or estimated covariance structure of the spatial process (see, e.g. Cressie, 1993). Under the Bayesian framework where we use a MCMC to sample from the posterior distribution, the spatial prediction problem becomes drawing samples from the predictive distribution $p(Y(\boldsymbol{s}_0)|\boldsymbol{Y})$. Based on these samples, we can estimate the mean $E[Y(\boldsymbol{s}_0)|\boldsymbol{Y}]$ and evaluate other properties of $p(Y(\boldsymbol{s}_0)|\boldsymbol{Y})$, such as the spread, skewness, percentiles, etc.

In the context of making spatial predictions based on the HASP model, we need to simulate from the predictive distribution

$$p(Y(\boldsymbol{s}_0)|\boldsymbol{Y}) = \int p(Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})p(\boldsymbol{\alpha}, \boldsymbol{\theta}|\boldsymbol{Y})\, d\boldsymbol{\alpha}\, d\boldsymbol{\theta}. \tag{S12}$$

We present two strategies to sample from this predictive distribution. The first approach relies on the original parameterization of the HASP model (3) and is applicable for any choice of the correlation functions. Note that $Y(\boldsymbol{s}_0)$ can be computed deterministically given a sample of $\alpha(\boldsymbol{s}_0), \epsilon(\boldsymbol{s}_0)$ and $\boldsymbol{\theta}$ from the conditional distribution $p(\alpha(\boldsymbol{s}_0), \epsilon(\boldsymbol{s}_0), \boldsymbol{\theta}|\boldsymbol{Y})$.

At the same time, note that sampling from $p(\alpha(\boldsymbol{s}_0), \epsilon(\boldsymbol{s}_0), \boldsymbol{\theta}|\boldsymbol{Y})$ is straightforward given the following result:

$$\begin{aligned} p(\alpha(\boldsymbol{s}_0), \epsilon(\boldsymbol{s}_0), \boldsymbol{\theta}|\boldsymbol{Y}) &= \int p(\alpha(\boldsymbol{s}_0), \epsilon(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta}) p(\boldsymbol{\alpha}, \boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\alpha} \\ &= \int p(\alpha(\boldsymbol{s}_0), \epsilon(\boldsymbol{s}_0)|\boldsymbol{\alpha}, \boldsymbol{\epsilon}, \boldsymbol{\theta}) p(\boldsymbol{\alpha}, \boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\alpha}, \end{aligned} \tag{S13}$$

where $p(\alpha(\boldsymbol{s}_0), \epsilon(\boldsymbol{s}_0)|\boldsymbol{\alpha}, \boldsymbol{\epsilon}, \boldsymbol{\theta})$ is a multivariate normal distribution, which can be easily obtained from the joint multivariate normal distribution $p(\alpha(\boldsymbol{s}_0), \epsilon(\boldsymbol{s}_0), \boldsymbol{\alpha}, \boldsymbol{\epsilon}|\boldsymbol{\theta})$. The second equality in (S13) holds because the sigma field generated by $(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})$ satisfies

$$\sigma\{\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta}\} = \sigma\{\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\epsilon}, \boldsymbol{\theta}\} = \sigma\{\boldsymbol{\alpha}, \boldsymbol{\epsilon}, \boldsymbol{\theta}\}.$$

Therefore, we propose the following strategy to draw samples from the predictive distribution of $Y(\boldsymbol{s}_0)|\boldsymbol{Y}$ within the MCMC algorithm for model fitting.

1. In each MCMC iteration t, record the posterior samples of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ as $\boldsymbol{\alpha}^{[t]}$ and $\boldsymbol{\theta}^{[t]}$. Calculate $\epsilon(\boldsymbol{s}_i)^{[t]}$ for $i = 1, \ldots, n$ from

$$\epsilon(\boldsymbol{s}_i)^{[t]} = \exp\left\{-\frac{\boldsymbol{H}^T(\boldsymbol{s}_i)\boldsymbol{\delta}^{[t]} + \alpha(\boldsymbol{s}_i)^{[t]}}{2}\right\}\left(Y(\boldsymbol{s}_i) - \boldsymbol{X}^T(\boldsymbol{s}_i)\boldsymbol{\beta}^{[t]}\right).$$

2. Given $\boldsymbol{\alpha}^{[t]}$, $\boldsymbol{\epsilon}^{[t]}$ and $\boldsymbol{\theta}^{[t]}$, draw a sample of $\alpha(\boldsymbol{s}_0)^{[t]}, \epsilon(\boldsymbol{s}_0)^{[t]}$ from the conditional distribution $p(\alpha(\boldsymbol{s}_0), \epsilon(\boldsymbol{s}_0), \boldsymbol{\theta}|\boldsymbol{Y})$.

3. Compute $Y(\boldsymbol{s}_0)^{[t]}$ according to

$$Y(\boldsymbol{s}_0)^{[t]} = \boldsymbol{X}^T(\boldsymbol{s}_0)\boldsymbol{\beta}^{[t]} + \exp\left\{\frac{\boldsymbol{H}^T(\boldsymbol{s}_0)\boldsymbol{\delta}^{[t]} + \alpha(\boldsymbol{s}_0)^{[t]}}{2}\right\}\epsilon(\boldsymbol{s}_0)^{[t]}.$$

   Then $Y(\boldsymbol{s}_0)^{[t]}$ is a sample from the predictive distribution $p(Y(\boldsymbol{s}_0)|\boldsymbol{Y})$.

Our second approach works specifically for the two strategies detailed in Section S4. As discussed previously, strategy I is essentially a special case of strategy II when $\boldsymbol{\Xi} = \boldsymbol{P}$.

Therefore, we will use the equation (9) to demonstrate the process to simulate from the predictive distribution. Let $\boldsymbol{\xi} = (\xi(\boldsymbol{s}_1), \ldots, \xi(\boldsymbol{s}_n))^T$. Then the equation (S13) becomes

$$\begin{aligned} p(\alpha(\boldsymbol{s}_0), \xi(\boldsymbol{s}_0), \boldsymbol{\theta}|) &= \int p(\alpha(\boldsymbol{s}_0), \xi(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta}) p(\boldsymbol{\alpha}, \boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\alpha} \\ &= \int p(\alpha(\boldsymbol{s}_0), \xi(\boldsymbol{s}_0)|\boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\theta}) p(\boldsymbol{\alpha}, \boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\alpha}, \\ &= \int p(\alpha(\boldsymbol{s}_0)|\boldsymbol{\alpha}, \boldsymbol{\theta}) p(\xi(\boldsymbol{s}_0)|\boldsymbol{\xi}, \boldsymbol{\theta}) p(\boldsymbol{\alpha}, \boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\alpha}. \end{aligned} \tag{S14}$$

As before, the second equality in (S14) holds because the sigma field generated by $(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta})$ satisfies

$$\sigma\{\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\theta}\} = \sigma\{\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\theta}\} = \sigma\{\boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\theta}\},$$

and the third equality follows from the independence of the processes $\alpha(\boldsymbol{s})$ and $\xi(\boldsymbol{s})$, $\boldsymbol{s} \in \mathcal{D}$. The Resulting procedure for sampling from $p(Y(\boldsymbol{s}_0)|\boldsymbol{Y})$ is as follows.

1. In each MCMC iteration t, record the posterior samples of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ as $\boldsymbol{\alpha}^{[t]}$ and $\boldsymbol{\theta}^{[t]}$. Calculate $\xi(\boldsymbol{s}_i)^{[t]}$ for $i = 1, \ldots, n$ from

$$\xi(\boldsymbol{s}_i)^{[t]} = \left(\psi^{[t]}\right)^{-1} \times \left[\exp\left\{-\frac{\boldsymbol{H}^T(\boldsymbol{s}_i)\tilde{\boldsymbol{\delta}}^{[t]} + \alpha(\boldsymbol{s}_i)^{[t]}}{2}\right\} \left(Y(\boldsymbol{s}_i) - \boldsymbol{X}^T(\boldsymbol{s}_i)\boldsymbol{\beta}^{[t]}\right) - \phi^{[t]}\alpha(\boldsymbol{s}_i)^{[t]}\right].$$

2. Given $\boldsymbol{\alpha}^{[t]}$, $\boldsymbol{\xi}^{[t]}$ and $\boldsymbol{\theta}^{[t]}$, draw a sample of $\alpha(\boldsymbol{s}_0)^{[t]}, \xi(\boldsymbol{s}_0)^{[t]}$ from the conditional distribution $p(\alpha(\boldsymbol{s}_0), \xi(\boldsymbol{s}_0), \boldsymbol{\theta}|\boldsymbol{Y})$. According to (S14), this can be further broken down into two steps.

   (a) Draw $\alpha(\boldsymbol{s}_0)^{[t]}$ from the normal distribution $p(\alpha(\boldsymbol{s}_0)|\boldsymbol{\alpha}^{[t]}, \boldsymbol{\theta}^{[t]})$ which is given by

$$N\left((\boldsymbol{P}_0^{[t]})^T(\boldsymbol{P}^{[t]})^{-1}\boldsymbol{\alpha}^{[t]}, (\tau^2)^{[t]}\left[\boldsymbol{P}_{00}^{[t]} - (\boldsymbol{P}_0^{[t]})^T(\boldsymbol{P}^{[t]})^{-1}\boldsymbol{P}_0^{[t]}\right]\right),$$

   where $\boldsymbol{P}_{00}^{[t]} = 1$ and $\boldsymbol{P}_0^{[t]}$ is a column vector whose i-th element is given by $\rho_\alpha(\|\boldsymbol{s}_0 - \boldsymbol{s}_i\|; \lambda^{[t]})$.

(b) Draw $\xi(\boldsymbol{s}_0)^{[t]}$ from the normal distribution $p(\xi(\boldsymbol{s}_0)|\xi(\boldsymbol{s})^{[t]}, \boldsymbol{\theta}^{[t]})$ given by

$$N\left((\boldsymbol{\Xi}_0^{[t]})^T(\boldsymbol{\Xi}^{[t]})^{-1}\boldsymbol{\alpha}^{[t]}, \left[\boldsymbol{\Xi}_{00}^{[t]} - (\boldsymbol{\Xi}_0^{[t]})^T(\boldsymbol{\Xi}^{[t]})^{-1}\boldsymbol{\Xi}_0^{[t]}\right]\right),$$

where $\boldsymbol{\Xi}_{00}^{[t]} = 1$ and $\boldsymbol{\Xi}_0^{[t]}$ is a column vector whose i-th element is given by $\rho_\alpha(\|\boldsymbol{s}_0 - \boldsymbol{s}_i\|; \kappa^{[t]})$.

3. Compute $Y(\boldsymbol{s}_0)^{[t]}$ according to

$$Y(\boldsymbol{s}_0) = \boldsymbol{X}^T(\boldsymbol{s}_0)\boldsymbol{\beta}^{[t]} + \phi^{[t]} \exp\left\{\frac{\boldsymbol{H}^T(\boldsymbol{s}_0)\tilde{\boldsymbol{\delta}}^{[t]} + \alpha(\boldsymbol{s}_0)^{[t]}}{2}\right\}\alpha(\boldsymbol{s}_0)^{[t]} + \psi^{[t]} \exp\left\{\frac{\boldsymbol{H}^T(\boldsymbol{s}_0)\tilde{\boldsymbol{\delta}}^{[t]} + \alpha(\boldsymbol{s}_0)^{[t]}}{2}\right\}\xi(\boldsymbol{s}_0)^{[t]}.$$

Then $Y(\boldsymbol{s}_0)^{[t]}$ is a sample from the predictive distribution $p(Y(\boldsymbol{s}_0)|\boldsymbol{Y})$.

References

Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107:180–193.

Askey, R. (1973). Radial characteristics functions. Technical report, University of Wisconsin-Madison, Mathematics Research Center.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York, NY.

Doob, J. L. (1942). The Brownian movement and stochastic equations. *Annals of Mathematics*, 43:351–369.

Gneiting, T. (2001). Criteria of Pólya type for radial positive definite functions. *Proceedings of the American Mathematical Society*, 129:2309–2318.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105:1167–1177.

Goulard, M. and Voltz, M. (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24:269–286.

Grenander, U. and Miller, M. I. (1994). Representations of knowledge in complex systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56:549–603.

Palacios, M. and Steel, M. (2006). Non-Gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, 101:604–618.

Pólya, G. (1949). Remarks on characteristic functions. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pages 115–123.

Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2:341–363.

Schmidt, A. M. and Gelfand, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres*, 108:1–9.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9:1135–1151.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. Springer, New York.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical review*, 36:823.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer, Berlin, 3rd edition.

Ying, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *Journal of Multivariate Analysis*, 36:280–296.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261.

Zhang, H. (2007). Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics*, 18:125–139.