Supplementary material for "Nonparametric statistical downscaling for the fusion of data of different spatiotemporal support"

C. J. Wilkie[1], C. A. Miller[1], E. M. Scott[1], R. A. O'Donnell[1], P. D. Hunter[2], E. Spyrakos[2], A. N. Tyler[2]

[1] School of Mathematics and Statistics, University of Glasgow, Glasgow, U.K.

[2] Biological and Environmental Sciences, University of Stirling, Stirling, U.K.

1 Derivation of the full conditional posterior distributions for the nonparametric statistical downscaling model

The probability density function for the **likelihood** is:

$$\begin{aligned} f(\mathbf{y}) &= \prod_{i=1}^{n} f(\mathbf{y}_i) \\ &= \frac{1}{(2\pi)^{\sum_{i=1}^{n} q_i/2} \prod_{i=1}^{n} (|\sigma_y^2 \mathbf{I}_{q_i}|^{1/2})} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{\Phi}_i \mathbf{c}_i)^{\mathrm{T}} (\sigma_y^2 \mathbf{I}_{q_i})^{-1} (\mathbf{y}_i - \mathbf{\Phi}_i \mathbf{c}_i) \right). \end{aligned}$$

The probability density functions for the **prior distributions** and **hyperprior distributions** are:

$$\begin{aligned} f(c_{ij}) &= \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{1}{2\sigma_c^2} \big(c_{ij} - (\alpha_{ij} + \beta_{ij} d_{ij})\big)^2 \right), \text{ so that} \\ f(\mathbf{c}_i) &= \prod_{j=1}^{m} f(c_{ij}) = \frac{1}{(2\pi)^{m/2} |\sigma_c^2 \mathbf{I}_m|^{1/2}} \\ &\quad \times \exp\left(-\frac{1}{2} \Big(\mathbf{c}_i - (\boldsymbol{\alpha}_i + \boldsymbol{\beta}_i \odot \mathbf{d}_i) \Big)^{\mathrm{T}} (\sigma_c^2 \mathbf{I}_m)^{-1} \Big(\mathbf{c}_i - (\boldsymbol{\alpha}_i + \boldsymbol{\beta}_i \odot \mathbf{d}_i) \Big) \right), \\ f(\mathbf{c}_j) &= \prod_{i=1}^{n} f(c_{ij}) = \frac{1}{(2\pi)^{n/2} |\sigma_c^2 \mathbf{I}_n|^{1/2}} \\ &\quad \times \exp\left(-\frac{1}{2} \Big(\mathbf{c}_j - (\boldsymbol{\alpha}_j + \boldsymbol{\beta}_j \odot \mathbf{d}_j) \Big)^{\mathrm{T}} (\sigma_c^2 \mathbf{I}_n)^{-1} \Big(\mathbf{c}_j - (\boldsymbol{\alpha}_j + \boldsymbol{\beta}_j \odot \mathbf{d}_j) \Big) \right) \text{ and} \\ f(\mathbf{c}) &= \prod_{i=1}^{n} f(\mathbf{c}_i) = \frac{1}{(2\pi)^{mn/2} |\sigma_c^2 \mathbf{I}_n|^{m/2} |\mathbf{I}_m|^{n/2}} \\ &\quad \times \exp\left(-\frac{1}{2} \mathrm{tr}\Big((\sigma_c^2)^{-1} \big(\mathbf{c} - (\boldsymbol{\alpha} + \boldsymbol{\beta} \odot \mathbf{d}) \big)^{\mathrm{T}} (\mathbf{I}_m)^{-1} \big(\mathbf{c} - (\boldsymbol{\alpha} + \boldsymbol{\beta} \odot \mathbf{d}) \big) \Big) \right); \end{aligned}$$

$$
\begin{aligned}
f(\boldsymbol{\alpha}_j) &= \frac{1}{(2\pi)^{n/2}|\sigma_\alpha^2 \exp(-\phi_\alpha \mathbf{D}_{\text{data}})|^{1/2}} \\
&\quad \times \ \exp\left(-\frac{1}{2}(\boldsymbol{\alpha}_j - \mathbf{0})^{\mathrm{T}}\big(\sigma_\alpha^2 \exp(-\phi_\alpha \mathbf{D}_{\text{data}})\big)^{-1}(\boldsymbol{\alpha}_j - \mathbf{0})\right), \text{ so that} \\
f(\boldsymbol{\alpha}) &= \prod_{j=1}^{m} f(\boldsymbol{\alpha}_j) = \frac{1}{(2\pi)^{mn/2}|\sigma_\alpha^2 \exp(-\phi_\alpha \mathbf{D}_{\text{data}})|^{m/2}|\mathbf{I}_m|^{n/2}} \\
&\quad \times \exp\left(-\frac{1}{2}\mathrm{tr}\Big(\big(\sigma_\alpha^2 \exp(-\phi_\alpha \mathbf{D}_{\text{data}})\big)^{-1}(\boldsymbol{\alpha}_j - \mathbf{0})^{\mathrm{T}}(\mathbf{I}_m)^{-1}(\boldsymbol{\alpha}_j - \mathbf{0})\Big)\right); \\
f(\boldsymbol{\beta}_j) &= \frac{1}{(2\pi)^{n/2}|\sigma_\beta^2 \exp(-\phi_\beta \mathbf{D}_{\text{data}})|^{1/2}} \\
&\quad \times \ \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_j - \mathbf{1})^{\mathrm{T}}(\sigma_\beta^2 \exp(-\phi_\beta \mathbf{D}_{\text{data}}))^{-1}(\boldsymbol{\beta}_j - \mathbf{1})\right), \text{ so that} \\
f(\boldsymbol{\beta}) &= \prod_{j=1}^{m} f(\boldsymbol{\beta}_j) = \frac{1}{(2\pi)^{mn/2}|\sigma_\beta^2 \exp(-\phi_\beta \mathbf{D}_{\text{data}})|^{m/2}|\mathbf{I}_m|^{n/2}} \\
&\quad \times \exp\left(-\frac{1}{2}\mathrm{tr}\Big(\big(\sigma_\beta^2 \exp(-\phi_\beta \mathbf{D}_{\text{data}})\big)^{-1}(\boldsymbol{\beta}_j - \mathbf{1})^{\mathrm{T}}(\mathbf{I}_m)^{-1}(\boldsymbol{\beta}_j - \mathbf{1})\Big)\right); \\
f\big((\sigma_\alpha^2)^{-1}\big) &= b_\alpha^{a_\alpha}(\sigma_\alpha^2)^{-1(a_\alpha - 1)} \exp\big(-b_\alpha(\sigma_\alpha^2)^{-1}\big)/\Gamma(a_\alpha); \\
f\big((\sigma_\beta^2)^{-1}\big) &= b_\beta^{a_\beta}(\sigma_\beta^2)^{-1(a_\beta - 1)} \exp\big(-b_\beta(\sigma_\beta^2)^{-1}\big)/\Gamma(a_\beta); \\
f\big((\sigma_y^2)^{-1}\big) &= b_y^{a_y}(\sigma_y^2)^{-1(a_y - 1)} \exp\big(-b_y(\sigma_y^2)^{-1}\big)/\Gamma(a_y); \\
f\big((\sigma_c^2)^{-1}\big) &= b_c^{a_c}(\sigma_c^2)^{-1(a_c - 1)} \exp\big(-b_c(\sigma_c^2)^{-1}\big)/\Gamma(a_c); \\
f(\mathbf{x}_i) &= \frac{1}{(2\pi)^{p_i/2}|\sigma_x^2 \mathbf{I}_{p_i}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\Psi}_i \mathbf{d}_i)^{\mathrm{T}}(\sigma_x^2 \mathbf{I}_{p_i})^{-1}(\mathbf{x}_i - \boldsymbol{\Psi}_i \mathbf{d}_i)\right), \text{ so that} \\
f(\mathbf{x}) &= \prod_{i=1}^{n} f(\mathbf{x}_i) = \frac{1}{(2\pi)^{\sum_{i=1}^n p_i/2}(\sigma_x^2)^{\sum_{i=1}^n p_i/2}\prod_{i=1}^n(|\mathbf{I}_{p_i}|^{1/2})} \\
&\quad \times \exp\left(-\frac{1}{2\sigma_x^2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\Psi}_i \mathbf{d}_i)^{\mathrm{T}}(\mathbf{I}_{p_i})^{-1}(\mathbf{x}_i - \boldsymbol{\Psi}_i \mathbf{d}_i)\right); \\
f\big((\sigma_x^2)^{-1}\big) &= b_x^{a_x}(\sigma_x^2)^{-1(a_x - 1)} \exp\big(-b_x(\sigma_x^2)^{-1}\big)/\Gamma(a_x); \text{ and} \\
f(\mathbf{d}_i) &= \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}_d|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{d}_i - \boldsymbol{\mu}_d)^{\mathrm{T}}\boldsymbol{\Sigma}_d^{-1}(\mathbf{d}_i - \boldsymbol{\mu}_d)\right).
\end{aligned}
$$

The **full conditional posterior distributions** are:

$$
\begin{aligned}
f\big((\sigma_\alpha^2)^{-1}|\cdot\big) &\propto \mathrm{Ga}\left(a_\alpha + \frac{mn}{2}, b_\alpha + \frac{1}{2}\mathrm{tr}\Big(\exp(-\phi_\alpha \mathbf{D}_{\text{data}})^{-1}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}\Big)\right); \\
f\big((\sigma_\beta^2)^{-1}|\cdot\big) &\propto \mathrm{Ga}\left(a_\beta + \frac{mn}{2}, b_\beta + \frac{1}{2}\mathrm{tr}\Big(\exp(-\phi_\beta \mathbf{D}_{\text{data}})^{-1}(\boldsymbol{\beta} - \mathbf{1})^{\mathrm{T}}(\boldsymbol{\beta} - \mathbf{1})\Big)\right); \\
f\big((\sigma_y^2)^{-1}|\cdot\big) &\propto \mathrm{Ga}\left(a_y + \sum_{i=1}^{n}\frac{q_i}{2}, b_y + \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\Phi}_i \mathbf{c}_i)^{\mathrm{T}}(\boldsymbol{y}_i - \boldsymbol{\Phi}_i \mathbf{c}_i)\right); \\
f\big((\sigma_c^2)^{-1}|\cdot\big) &\propto \mathrm{Ga}\left(a_c + \frac{mn}{2}, b_c + \frac{1}{2}\mathrm{tr}\Big(\mathbf{I}_n\big(\mathbf{c} - (\boldsymbol{\alpha} + \boldsymbol{\beta} \odot \mathbf{d})\big)^{\mathrm{T}}\mathbf{I}_m\big(\mathbf{c} - (\boldsymbol{\alpha} + \boldsymbol{\beta} \odot \mathbf{d})\big)\Big)\right);
\end{aligned}
$$

$$
\begin{aligned}
f(\boldsymbol{\alpha}_j|\cdot) &\propto \mathrm{N}(\boldsymbol{\Sigma}_{\alpha_j}\mathbf{A}_{\alpha_j}, \boldsymbol{\Sigma}_{\alpha_j}), \text{ where} \\
\boldsymbol{\Sigma}_{\alpha_j} &= \left(\left(\sigma_\alpha^2 \exp(-\phi_\alpha \mathbf{D}_{\text{data}})\right)^{-1} + (\sigma_c^2\mathbf{I}_n)^{-1}\right)^{-1} \text{ and} \\
\mathbf{A}_{\alpha_j} &= (\sigma_c^2\mathbf{I}_n)^{-1}(\mathbf{c}_j - \boldsymbol{\beta}_j \odot \mathbf{d}_j); \\
f(\boldsymbol{\beta}_j|\cdot) &\propto \mathrm{N}(\boldsymbol{\Sigma}_{\beta_j}\mathbf{A}_{\beta_j}, \boldsymbol{\Sigma}_{\beta_j}), \text{ where} \\
\boldsymbol{\Sigma}_{\beta_j} &= \left(\left(\sigma_\beta^2 \exp(-\phi_\beta \mathbf{D}_{\text{data}})\right)^{-1} + \mathbf{G}_j^{\mathrm{T}}(\sigma_c^2\mathbf{I}_n)^{-1}\mathbf{G}_j\right)^{-1} \text{ and} \\
\mathbf{A}_{\beta_j} &= \left(\sigma_\beta^2 \exp(-\phi_\beta \mathbf{D}_{\text{data}})\right)^{-1}\mathbf{1} + \mathbf{G}_j^{\mathrm{T}}(\sigma_c^2\mathbf{I}_n)^{-1}(\mathbf{c}_j - \boldsymbol{\alpha}_j); \\
f(\mathbf{c}_i|\cdot) &\propto \mathrm{N}(\boldsymbol{\Sigma}_{c_i}\mathbf{A}_{c_i}, \boldsymbol{\Sigma}_{c_i}), \text{ where} \\
\boldsymbol{\Sigma}_{c_i} &= \left(\boldsymbol{\Phi}_i^{\mathrm{T}}(\sigma_y^2\mathbf{I}_{q_i})^{-1}\boldsymbol{\Phi}_i + (\sigma_c^2\mathbf{I}_m)^{-1}\right)^{-1} \text{ and} \\
\mathbf{A}_{c_i} &= \boldsymbol{\Phi}_i^{\mathrm{T}}(\sigma_y^2\mathbf{I}_{q_i})^{-1}\mathbf{y}_i + (\sigma_c^2\mathbf{I}_m)^{-1}(\boldsymbol{\alpha}_i + \boldsymbol{\beta}_i \odot \mathbf{d}_i); \\
f\big((\sigma_x^2)^{-1}|\,\cdot\,\big) &= \mathrm{Ga}\left(a_x + \sum_{i=1}^{n}\frac{p_i}{2}, b_x + \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\Psi}_i\mathbf{d}_i)^{\mathrm{T}}(\mathbf{I}_{p_i})^{-1}(\mathbf{x}_i - \boldsymbol{\Psi}_i\mathbf{d}_i)\right); \text{ and} \\
f(\mathbf{d}_i|\cdot) &= \mathrm{N}(\boldsymbol{\Sigma}_{d_i}\mathbf{A}_{d_i}, \boldsymbol{\Sigma}_{d_i}), \text{ where} \\
\boldsymbol{\Sigma}_{d_i} &= \left(\boldsymbol{\Sigma}_d^{-1} + \boldsymbol{\Psi}_i^{\mathrm{T}}(\sigma_x^2\mathbf{I}_{p_i})^{-1}\boldsymbol{\Psi}_i + \mathbf{F}_i^{\mathrm{T}}(\sigma_c^2\mathbf{I}_m)^{-1}\mathbf{F}_i\right)^{-1} \text{ and} \\
\mathbf{A}_{d_i} &= \boldsymbol{\Sigma}_d^{-1}\boldsymbol{\mu}_d + \boldsymbol{\Psi}_i^{\mathrm{T}}(\sigma_x^2\mathbf{I}_{p_i})^{-1}\mathbf{x}_i + \mathbf{F}_i^{\mathrm{T}}(\sigma_c^2\mathbf{I}_m)^{-1}(\mathbf{c}_i - \boldsymbol{\alpha}_i).
\end{aligned}
$$

In the above equations, $\mathbf{G}_j$ and $\mathbf{F}_i$ represent diagonal matrices, with $\mathbf{d}_j$ and $\boldsymbol{\beta}_i$ as their diagonals, respectively. The indexes are $i = 1, \ldots, n$ and $j = 1, \ldots, m$, where n is the number of *in situ* sampling locations in the dataset and m is the number of basis functions fitted for each location (i.e. the chosen basis dimension). The number of remotely-sensed data available for location i is p_i and the corresponding number of *in situ* data available for location i is q_i. All other parameters are described in the main paper.

Predictions are made at new times j (for $j = 1, \ldots, \tilde{q}_i$, where $\tilde{q}_i$ is the number of times to predict at for location i) and at new locations i (for $i = 1, \ldots, \tilde{n}$, where $\tilde{n}$ is the number of locations at which to predict), by drawing from the posterior predictive distribution:

$$
\begin{aligned}
\tilde{\mathbf{y}}_i|\tilde{\mathbf{c}}_i, \sigma_y^2 &\sim \mathrm{N}_{\tilde{q}_i}(\tilde{\boldsymbol{\Phi}}_i\tilde{\mathbf{c}}_i, \sigma_y^2\mathbf{I}_{\tilde{q}_i}), \\
\tilde{c}_{ij}|\tilde{\alpha}_{ij}, \tilde{\beta}_{ij}, \tilde{d}_{ij}, \sigma_c^2 &\sim \mathrm{N}(\tilde{\alpha}_{ij} + \tilde{\beta}_{ij}\tilde{d}_{ij}, \sigma_c^2), \\
\tilde{\boldsymbol{\alpha}}_j|\boldsymbol{\alpha}_j &\sim \mathrm{N}_{\tilde{n}}\Big(\mathbf{0} + \exp(-\phi_\alpha\mathbf{D}_{\text{pred:data}})\exp(-\phi_\alpha\mathbf{D}_{\text{data}})^{-1}(\boldsymbol{\alpha}_j - \mathbf{0}), \\
&\qquad \sigma_\alpha^2\big(\exp(-\phi_\alpha\mathbf{D}_{\text{pred}}) - \exp(-\phi_\alpha\mathbf{D}_{\text{pred:data}})\exp(-\phi_\alpha\mathbf{D}_{\text{data}})^{-1}\exp(-\phi_\alpha\mathbf{D}_{\text{data:pred}})\big)\Big), \\
\tilde{\boldsymbol{\beta}}_j|\boldsymbol{\beta}_j &\sim \mathrm{N}_{\tilde{n}}\Big(\mathbf{1} + \exp(-\phi_\beta\mathbf{D}_{\text{pred:data}})\exp(-\phi_\beta\mathbf{D}_{\text{data}})^{-1}(\boldsymbol{\beta}_j - \mathbf{1}), \\
&\qquad \sigma_\beta^2\big(\exp(-\phi_\beta\mathbf{D}_{\text{pred}}) - \exp(-\phi_\beta\mathbf{D}_{\text{pred:data}})\exp(-\phi_\beta\mathbf{D}_{\text{data}})^{-1}\exp(-\phi_\beta\mathbf{D}_{\text{data:pred}})\big)\Big), \\
\tilde{\mathbf{d}}_i &\sim \mathrm{N}_m(\tilde{\boldsymbol{\Sigma}}_{d_i}\tilde{\mathbf{A}}_{d_i}, \tilde{\boldsymbol{\Sigma}}_{d_i}),
\end{aligned}
\tag{1}
$$

where:

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{d_i} &= \left(\boldsymbol{\Sigma}_d^{-1} + \tilde{\boldsymbol{\Psi}}_i^{\mathrm{T}}(\sigma_x^2\mathbf{I}_{\tilde{p}_i})^{-1}\tilde{\boldsymbol{\Psi}}_i\right)^{-1} \text{ and} \\
\tilde{\mathbf{A}}_{d_i} &= \boldsymbol{\Sigma}_d^{-1}\boldsymbol{\mu}_d + \tilde{\boldsymbol{\Psi}}_i^{\mathrm{T}}(\sigma_x^2\mathbf{I}_{\tilde{p}_i})^{-1}\tilde{\mathbf{x}}_i,
\end{aligned}
$$

where $\tilde{\mathbf{\Phi}}$ is the matrix of basis functions evaluated at the $\tilde{q}_i$ times of prediction for the *in situ* data at location i, $\tilde{\mathbf{\Psi}}$ is the matrix of basis functions evaluated at the $\tilde{p}_i$ times of data collection for the remotely-sensed data for the grid cell containing location i and $\tilde{\mathbf{x}}_i$ is the vector of remotely-sensed data for the grid cell containing the location i at which prediction is to be carried out.

Since all full conditional posterior distributions have been derived, the model is fitted using a Gibbs sampler. This is implemented in `C++`, via the `R` package `Rcpp`. Obtaining draws from $\tilde{\boldsymbol{\alpha}}_j|\boldsymbol{\alpha}_j$ and $\tilde{\boldsymbol{\beta}}_j|\boldsymbol{\beta}_j$ is potentially computationally expensive, when predictions are to be made at a large number of locations, since drawing from the multivariate Normal distributions would involve calculating the Cholesky decomposition of large matrices at each iteration of the Gibbs sampler. The covariance matrices of the distributions of both $\tilde{\boldsymbol{\alpha}}_j|\boldsymbol{\alpha}_j$ and $\tilde{\boldsymbol{\beta}}_j|\boldsymbol{\beta}_j$ are each made up of a scalar that must be updated each time (i.e. σ_α^2 and σ_β^2, respectively) multiplied by a matrix that needs only to be calculated once. Therefore, the computations can be reduced in complexity and the efficiency increased. The algorithm makes use of the fact that the Cholesky decomposition of $b\mathbf{\Sigma}$ is $(\sqrt{b}\mathbf{A})(\sqrt{b}\mathbf{A})^{\mathrm{T}} = b\mathbf{\Sigma}$, where $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{\Sigma}$ is the Cholesky decomposition of $\mathbf{\Sigma}$, so that the computation at each iteration of the sampler multiplies a matrix by a scalar, rather than the more complex Cholesky decomposition. If ϕ_α and ϕ_β were estimated within the model instead of being chosen beforehand, then this reduction in computational complexity would not be possible and the Cholesky decomposition would be required at each iteration of the Gibbs sampler.

2 Additional plots

2.1 Choice of ϕ_α and ϕ_β

Figure 1 illustrates the method for choosing the values of parameters ϕ_α and ϕ_β for the nonparametric statistical downscaling model using the Fourier basis of dimension 9. The plots show the values of model summary statistics for a range of values for each of ϕ_α and ϕ_β. Figures 1(a) and 1(b) show that root mean squared error (RMSE) and mean absolute error (MAE) are both at their lowest for $\phi_\alpha \leq 0.5$ and $\phi_\beta \leq 0.5$, while Figure 1(c) shows that the variance of the predictions is lowest for $0.01 \leq \phi_\beta \leq 0.5$. Figure 1(d) shows that the mean 95% credible interval coverage lies above the nominal 95% for all combinations of the values of ϕ_α and ϕ_β that are investigated, while Figure 1(e) shows that the mean 95% credible interval width is at its lowest for $\phi_\beta \leq 0.5$. The value of 0.1 selected for both ϕ_α and ϕ_β is therefore an appropriate choice for this dataset.

2.2 Choice of prediction locations

As stated in the main article, the prediction locations are chosen through a Delaunay triangulation of the lake, constrained by points selected along the lake boundaries. The triangulation is carried out using the `R` package `RTriangle` (Shewchuk 1996). Figure 2 shows the constraining boundary points and the inserted points, which together have a good spatial coverage of the lake.

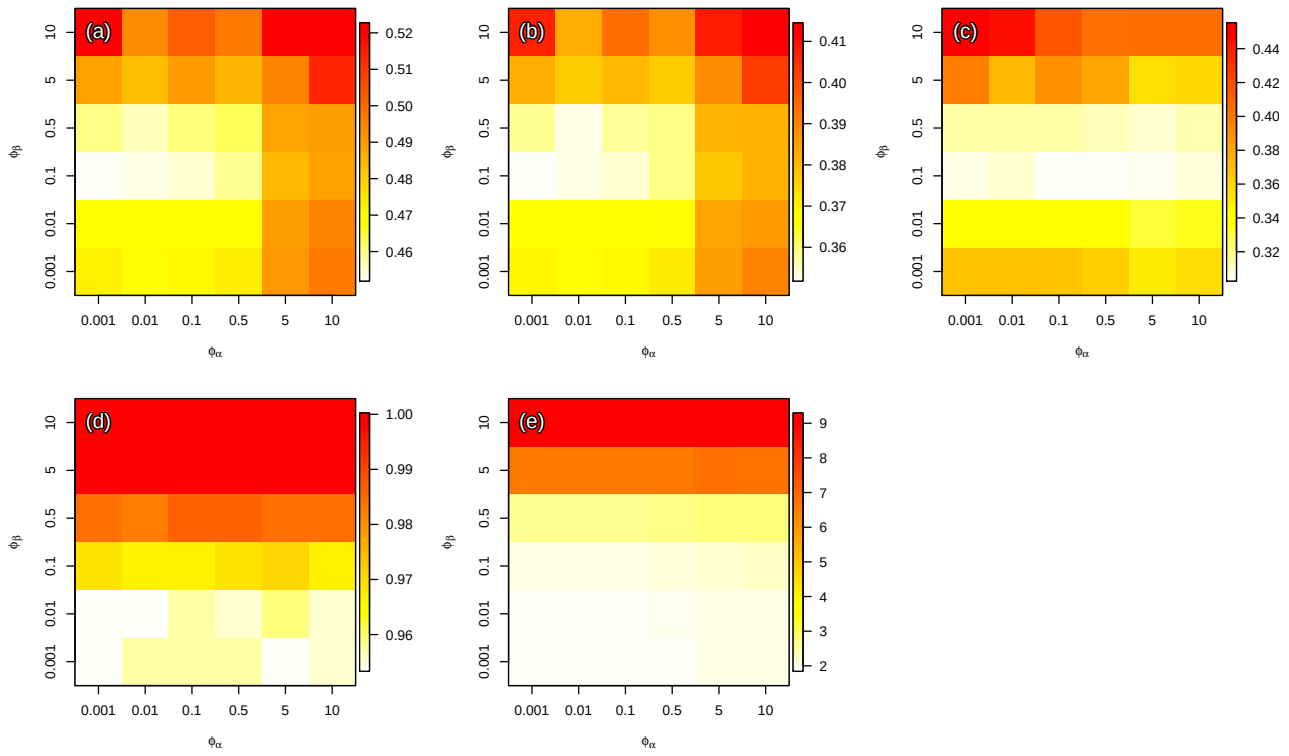

Figure 1: Plots of model summary statistics for each combination of values of ϕ_α and ϕ_β in the nonparametric statistical downscaling model, for the Fourier basis of dimension 9: (a) Root mean squared error (RMSE); (b) Mean absolute error (MAE); (c) Variance of predictions; (d) Mean 95% credible interval coverage; (e) Mean 95% credible interval width.

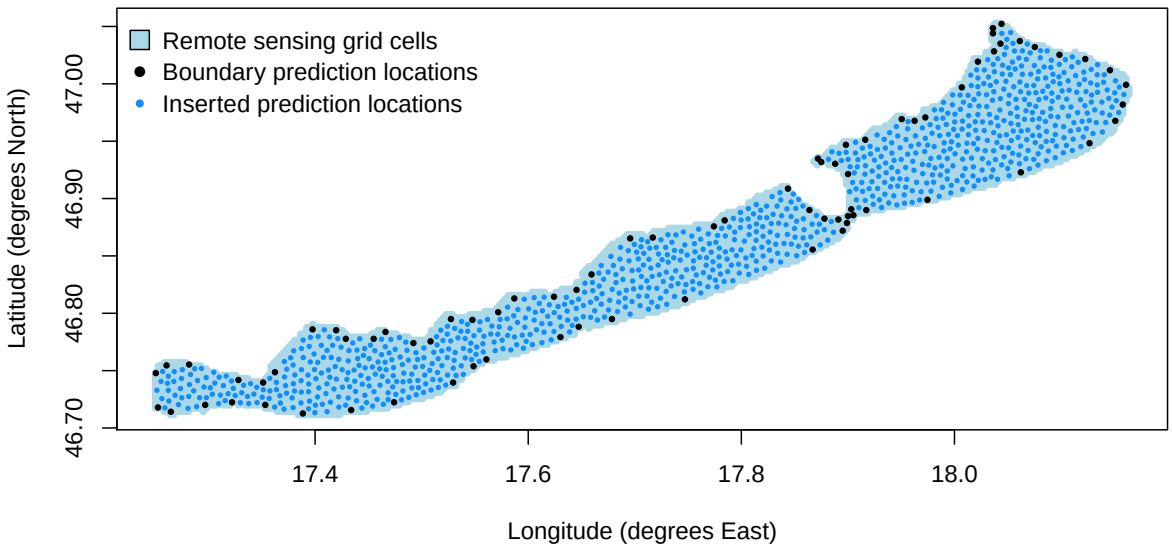

Figure 2: Plot showing the constraining boundary points and the inserted prediction locations that result from a Delaunay triangulation, overlaid upon the remote sensing grid cells.

2.3 Example trace, density and diagnostic plots for the non-parametric statistical downscaling model

Example trace and density plots are shown in Figures 3 and 4 for a selection of parameters of the nonparametric statistical downscaling model, which was fitted using a Fourier basis of dimension 9, with prediction $\tilde{y}$ made at *in situ* location 1 for 15th March 2011.

Each of the trace plots shows the "hairy caterpillar" shape, which is characteristic of MCMC chains that have converged to their posterior distributions (Gelman et al. 2014). Most of the density plots show the bell shape, which is also characteristic of chains that have converged to their posterior distributions (Gelman et al. 2014). The plots for the parameters $(\sigma_\alpha^2)^{-1}$, $(\sigma_\beta^2)^{-1}$ and $(\sigma_c^2)^{-1}$ show skewed distributions, which is expected of variance parameters. These plots provide no evidence against the assumption that the MCMC chains for these parameters have converged to their posterior distributions.

A plot of residuals versus fitted values is shown in Figure 5(a), for the nonparametric statistical downscaling model fitted using a Fourier basis of dimension 9, with predictions made at each of the locations and timepoints for which the *in situ* data are available. Figure 5(b) shows the corresponding plot of sample quantiles of the distribution of the residuals, versus the theoretical distribution of the quantiles, assuming that the residuals are Normally distributed.

Figure 6 shows plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the residuals at *in situ* location 1. The coefficients for low lags on both plots lie within, or at least nearly within, the 95% confidence intervals. These plots do not provide evidence of statistically significant positive temporal autocorrelation in the residuals for *in situ* location 1. Similar plots were produced for the residuals at each of the other *in situ* locations and all except those for one location showed similar results. As stated in the main text, the location with a large lag 1 coefficient in the ACF and PACF had only 19 *in situ* data points, which may lead to unusual patterns having a large effect on the calculated ACF and PACF.

Figure 7 shows a plot of the empirical variogram of the residuals for March 2011. The presence of spatial correlation in the residuals is assessed by comparing the empirical variogram to the boundaries of the Monte Carlo envelope. This envelope is calculated by randomly permuting the coordinates of the residuals 100 times and then calculating the variogram for each of these new datasets. The fact that the empirical variogram lies entirely within the Monte Carlo envelope provides no evidence of spatial correlation in the residuals for March 2011. Similar plots were produced for each of the 17 months for which data were available at all 9 *in situ* locations, with none indicating the presence of spatial correlation in the residuals.

References

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014), *Bayesian Data Analysis*, third edn, CRC.

Shewchuk, J. R. (1996), Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator, *in* M. C. Lin & D. Manocha, eds, 'Applied Computational Geometry: Towards Geometric Engineering', Vol. 1148, Springer, pp. 203–222.

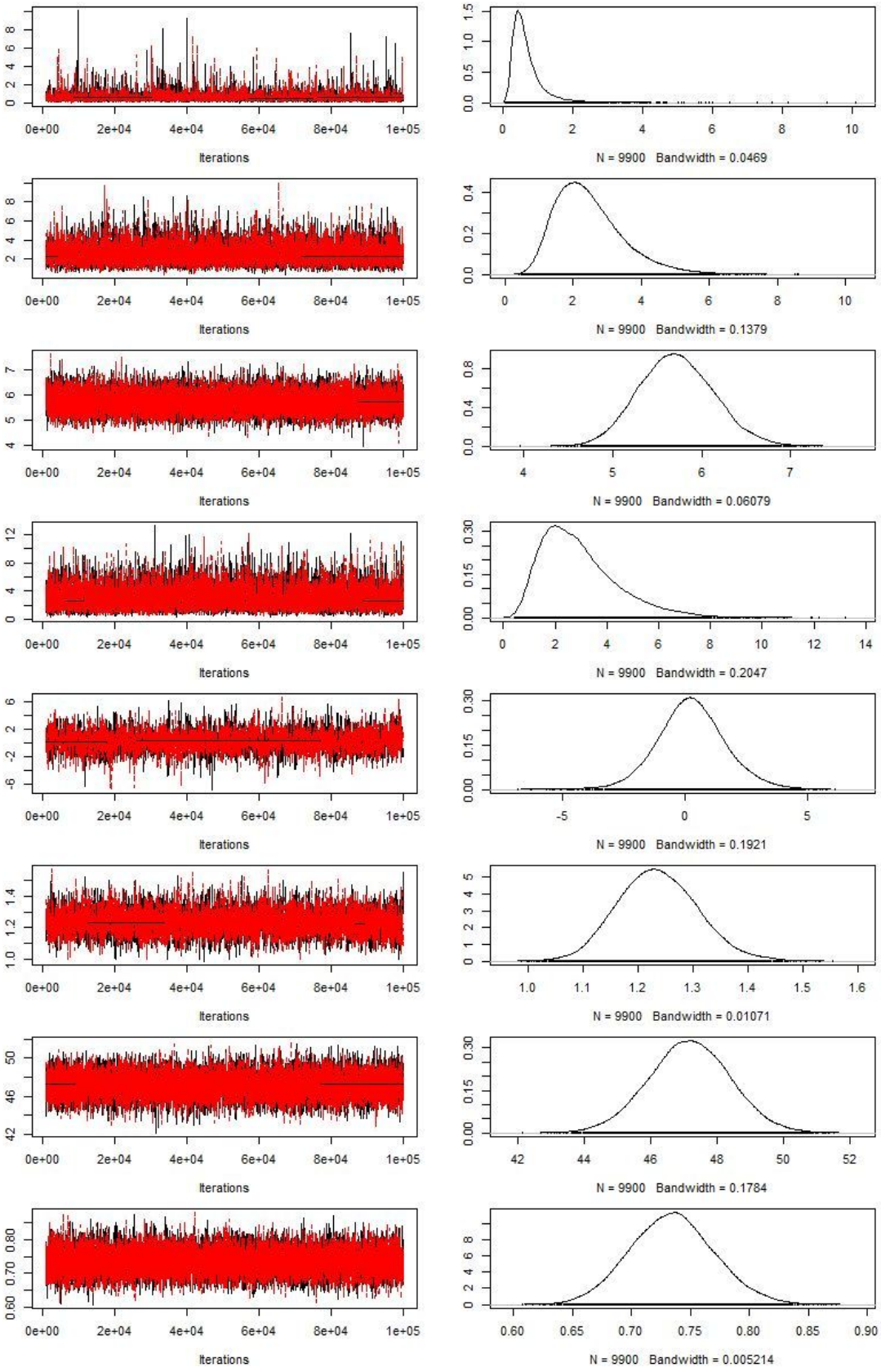

Figure 3: Trace and density plots for the parameters $(\sigma_\alpha^2)^{-1}$, $(\sigma_\beta^2)^{-1}$, $(\sigma_y^2)^{-1}$, $(\sigma_c^2)^{-1}$, $\alpha_{1,1}$, $\beta_{1,1}$, $c_{1,1}$ and $(\sigma_x^2)^{-1}$ of the nonparametric statistical downscaling model, with a Fourier basis of dimension 9.

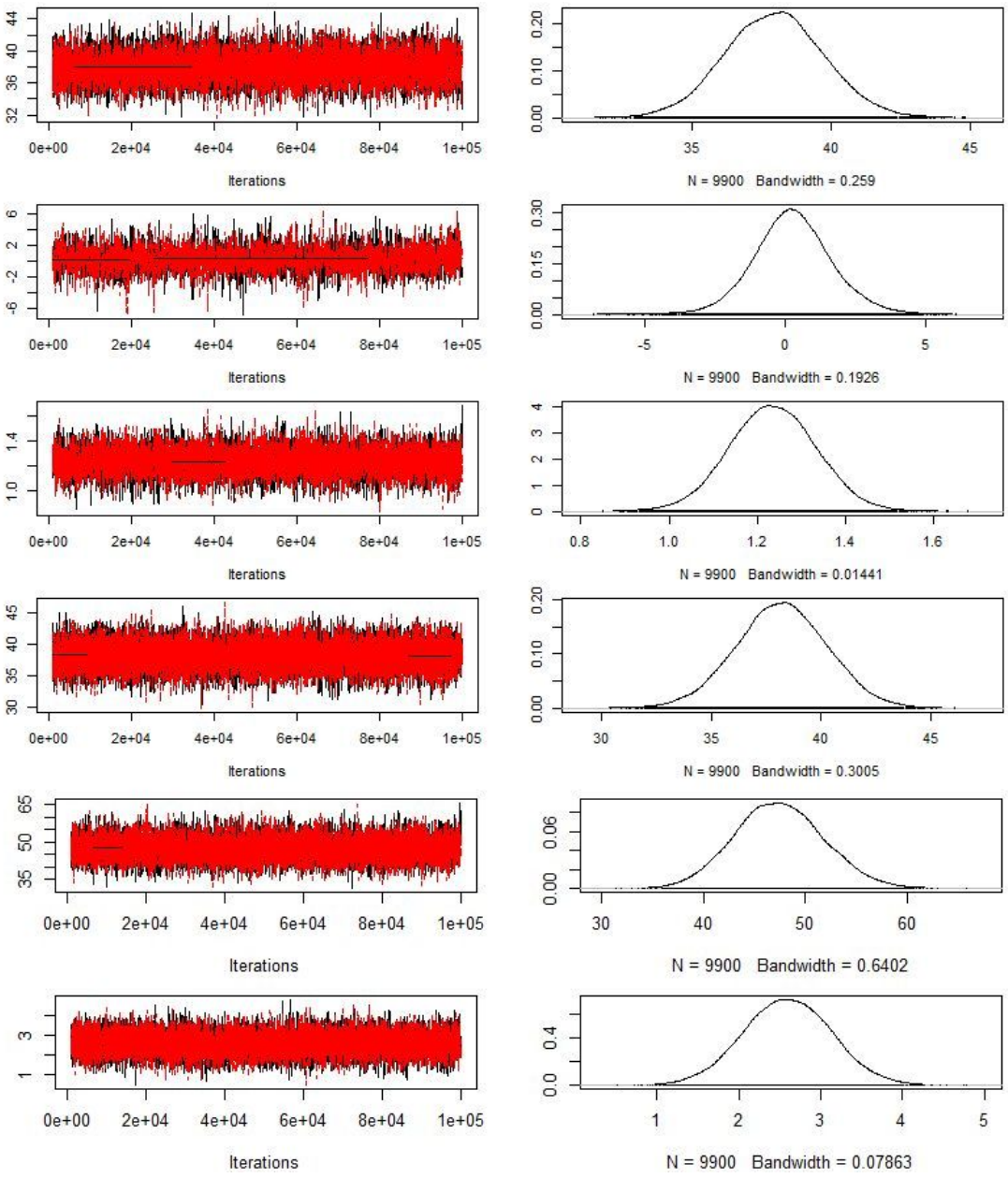

Figure 4: Trace and density plots for the parameters $d_{1,1}$, $\tilde{\alpha}_{1,1}$, $\tilde{\beta}_{1,1}$, $\tilde{d}_{1,1}$, $\tilde{c}_{1,1}$ and $\tilde{y}$ of the nonparametric statistical downscaling model, with a Fourier basis of dimension 9.

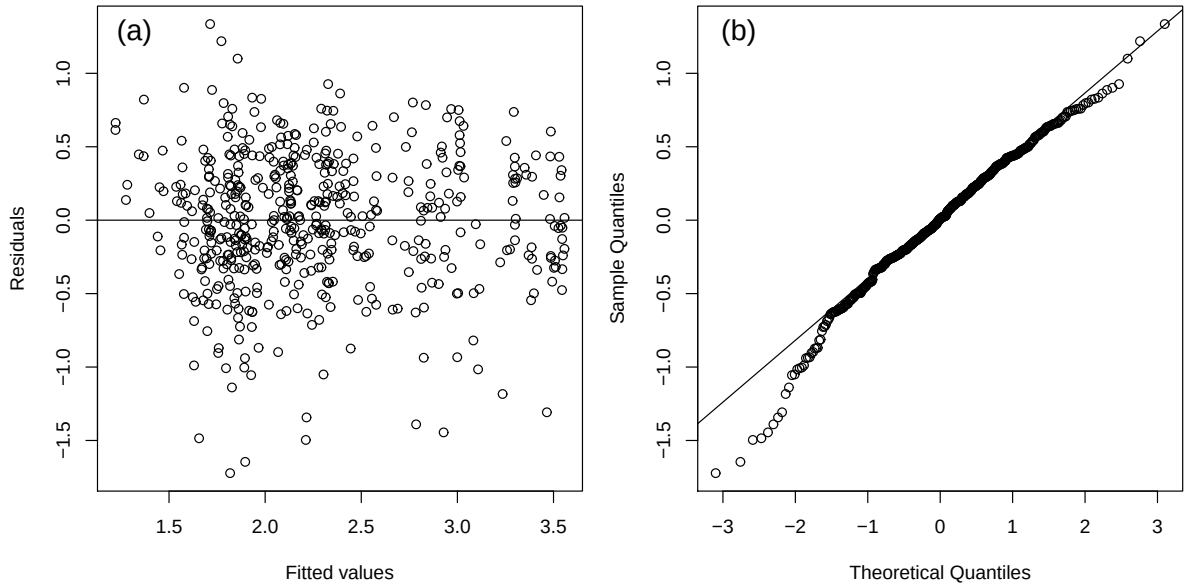

Figure 5: (a): Plot of residuals versus fitted values for the nonparametric statistical downscaling model fitted using the Fourier basis of dimension 9; (b): Plot of the sample quantiles of the distribution of the residuals versus the theoretical quantiles of the distribution of the residuals, assuming that they are Normally distributed (i.e. the Normal Q-Q plot).

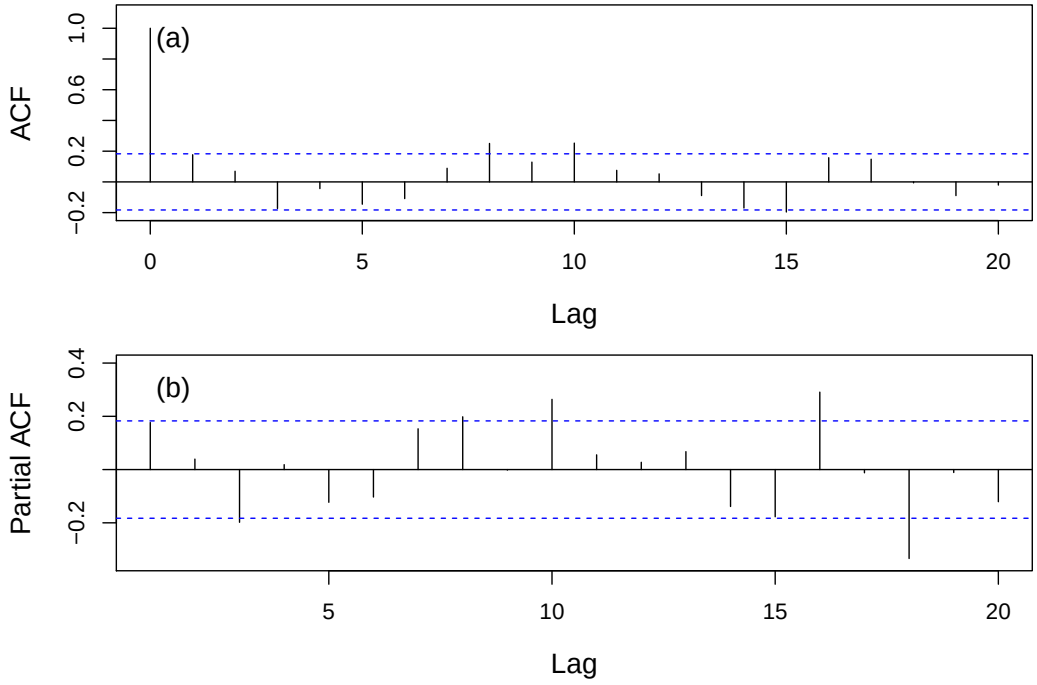

Figure 6: Plots of the (a) autocorrelation function (ACF) and (b) partial autocorrelation function (PACF) for the residuals at *in situ* location 1.

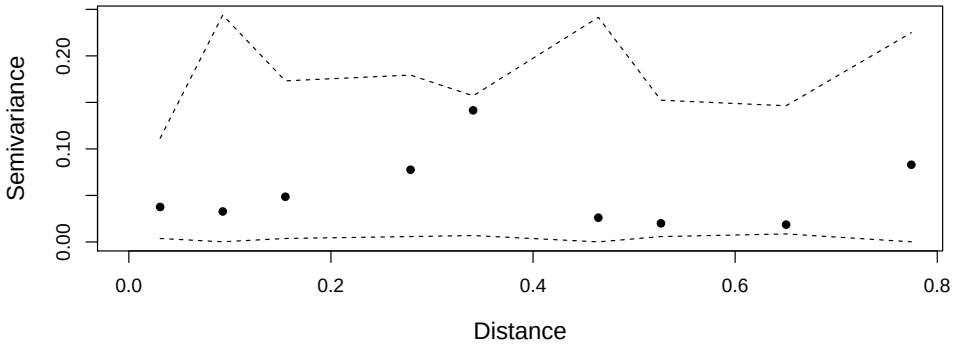

Figure 7: Plot of the empirical variogram fitted to the residuals for March 2011 (points), with the boundaries of the Monte Carlo envelope shown as dotted lines.