Web-based Supplementary Materials for Understanding Gaussian Process Fits Using an Approximate Form of the Restricted Likelihood

Maitreyee Bose, James S. Hodges, and Sudipto Banerjee

1 Web Appendix A. Spectral approximation in one dimension: technical details

1.1 Distribution of u_m from a_m and b_m

Given the assumed normal prior distributions for the a_m's and b_m's, it can be shown that the u_m's have complex normal distributions as follows.

$$\text{Assume} \quad a_m \sim N(0, \frac{1}{2M}\sigma_{\mathfrak{s}}^2\phi(\omega_m;\rho)),\ b_m \sim N(0, \frac{1}{2M}\sigma_{\mathfrak{s}}^2\phi(\omega_m;\rho))\ \ \forall m \neq 0 \text{ or } \frac{M}{2}$$

$$a_0 \sim N(0, \frac{1}{M}\sigma_{\mathfrak{s}}^2\phi(\omega_0;\rho)),\ a_{\frac{M}{2}} \sim N(0, \frac{1}{M}\sigma_{\mathfrak{s}}^2\phi(\omega_{\frac{M}{2}};\rho)),\ b_0 = b_{\frac{M}{2}} = 0.$$

Therefore, $u_m = a_m + ib_m$ has a complex normal distribution with

1. $E(u_m) = E(a_m) + iE(b_m) = 0$;
2. covariance matrix: $V(u_m) = V_{aa} + V_{bb} + i(V_{ab} - V_{ba}) = \frac{1}{M}\sigma_{\mathfrak{s}}^2\phi(\omega_m;\rho)\ \ \forall m$;
3. relation matrix: $C(u_m) = V_{aa} - V_{bb} + i(V_{ba} + V_{ab}) = 0$.

So, $E(u_m) = 0$, $E(u_m u_n) = 0$ for $m \neq n$ and $E|\ u_m\ |^2 = \frac{1}{M}\sigma_{\mathfrak{s}}^2\phi(\omega_m;\rho)$.

Note that in all the following proofs, the specific choices of the values of the s_j's and the ω_m's have a crucial role.

1.2 Proof that the spectral approximation is valid

We prove the approximate equality $\text{Cov}(g(s_j), g(s_j + d)) \simeq \sigma_{\mathfrak{s}}^2 K(d;\rho)$.

Definition of spectral density (Priestley 1981, p. 199, 211). Let $R(d)$ be the covariance function of a continuous parameter stationary process. If $R(d)$ is absolutely integrable and isotropic,

then its spectral density is defined as

$$h(\omega) = \int_{-\infty}^{\infty} \exp(-2\pi i\omega d) R(d)\ \mathrm{d}d.$$

When the process is observed only at a discrete set of integer locations, then $R(d)$ is defined only for integer values of d and the above integral has to be replaced by a discrete sum (Priestley 1981, p. 222):

$$h(\omega) = \sum_{d=-\infty}^{\infty} \exp(-2\pi i\omega d) R(d), \qquad -\frac{1}{2} \leq \omega \leq \frac{1}{2}.$$

Now, $\exp(-2\pi i\omega d)$ is a periodic function of ω with period 1, so the components in the observed discrete parameter process with frequencies $\omega - 1$, $\omega + 1$, $\omega - 2$, $\omega + 2$,.... will all appear to have frequency ω. The frequency ω is then said to be the alias of the frequencies $\omega \pm 1$, $\omega \pm 2$,..... Since every frequency outside $[-\frac{1}{2}, \frac{1}{2}]$ has an alias inside this range, the spectral density is defined only for ω in the range $[-\frac{1}{2}, \frac{1}{2}]$.

Then

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(2\pi i\omega d) h(\omega)\ \mathrm{d}\omega = R(d).$$

Proof:

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(2\pi i\omega d) h(\omega)\ \mathrm{d}\omega = \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{\varsigma=-\infty}^{\infty} \exp(-2\pi i\omega(\varsigma - d)) R(\varsigma)\ \mathrm{d}\omega$$

using the definition of spectral density

$$= \sum_{\varsigma=-\infty}^{\infty} R(\varsigma) \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(-2\pi i\omega(\varsigma - d))\ \mathrm{d}\omega$$

$$= R(d) + \sum_{\varsigma \neq d} \frac{R(\varsigma)}{2\pi i(\varsigma - d)} (\exp(\pi i(\varsigma - d)) - \exp(-\pi i(\varsigma - d)))$$

$$= R(d) + \sum_{\varsigma \neq d} \frac{R(\varsigma) \sin(\pi(\varsigma - d))}{\pi(\varsigma - d)} = R(d)$$

because sin of any integer multiple of π is 0.

Thus we have

$$\sigma_{\mathfrak{s}}^2 K(d; \rho) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(2\pi i\omega d) \sigma_{\mathfrak{s}}^2 \phi(\omega; \rho)\ \mathrm{d}\omega,$$

where d is the distance between two locations belonging to S and our proof of $\mathrm{Cov}(g(s_j), g(s_j + d)) \simeq$

$\sigma_{\mathfrak{s}}^2 K(d;\rho)$ is complete if we can show

$$\mathrm{Cov}(g(s_j), g(s_j+d)) \simeq \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(2\pi i\omega d)\sigma_{\mathfrak{s}}^2\phi(\omega;\rho)\ \mathrm{d}\omega.$$

Spectral representation theorem (Gelfand et al. 2010, p. 60). Suppose $Y_0, Y_1, ..., Y_{M-1}$ are mean zero complex random variables with $E(Y_l Y_m) = 0$ for $l \neq m = 0, 1, ..., M$ and $E|\ Y_m\ |^2 = f_m$ for each m, and suppose $\omega_0, \omega_1, ..., \omega_{M-1} \in \Re, -\frac{1}{2} \leq \omega_m \leq \frac{1}{2}\ \forall\ m$.

Consider

$$Z(s) = \sum_{m=0}^{M-1} \exp(i2\pi\omega_m s)Y_m \qquad s \in S \in \aleph.$$

Then $Z(s)$ are realizations of a weakly stationary process in $\Re$ with covariance function

$$C(d) = \sum_{m=0}^{M-1} \exp(i2\pi\omega_m d)f_m$$

where d is the distance between two locations belonging to the set of locations S.

Thus, for our approximation, by the spectral representation theorem, $\{g(s_j); s_j \in S\}$ is a weakly stationary process in $\Re$ with covariance function $\mathrm{Cov}(g(s_j), g(s_j+d))$ given by

$$C(d) = \sum_{m=0}^{M-1} \exp(i2\pi\omega_m d)E|\ u_m\ |^2$$

$$= \frac{1}{M}\sum_{m=0}^{M-1} \exp(i2\pi\omega_m d)\sigma_{\mathfrak{s}}^2\phi(\omega_m;\rho)$$

$$\simeq \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(i2\pi\omega d)\sigma_{\mathfrak{s}}^2\phi(\omega;\rho)\ \mathrm{d}\omega\ \text{ as }\ M \to \infty,$$

where the last step follows from the Riemann sum formula for the definite integral.

1.3 Orthogonality of columns of Z

For z_j the j^{th} column of Z, we need three results:

1. $z_j' z_j = 2M,\ \text{ for }\ j \in \{1, 2, ..., M-2\}, \qquad 2.\ z_j' z_j = M,\ \text{ for }\ j = M-1$
2. $z_j' z_l = 0,\ \text{ for }\ l \neq j \in \{1, 2, ..., M-1\}.$

We prove 1 below; 2 and 3 can be proved similarly.

Proof: For $j \in \{1, 3, ..., M-3\}$

$$z_j' z_j = 2\sum_{k=1}^{M}(\cos 0 + \cos(\omega_{\frac{j+1}{2}}\frac{2\pi}{M}2k)) \tag{1}$$

because $\cos(mx)\cos(nx) = \frac{1}{2}(\cos(m-n)x + \cos(m+n)x)$. Then equation (1)

$$= 2M + 2\sum_{k=1}^{M}\cos(\omega_{\frac{j+1}{2}}\frac{4\pi k}{M}) = 2M + \sum_{k=1}^{M}(e^{ck} + e^{-ck}), \text{ for } c = i\,\omega_{\frac{j+1}{2}}\frac{4\pi}{M},$$
$$= 2M + \frac{e^c(e^{Mc}-1)}{e^c-1} + \frac{e^{-Mc}(e^{Mc}-1)}{e^c-1} = 2M + \frac{(e^{Mc}-1)(e^c+e^{-Mc})}{e^c-1} = 2M$$

because $e^{Mc} = e^{i\omega_{\frac{j+1}{2}}4\pi} = \cos(\omega_{\frac{j+1}{2}}4\pi) + i\,\sin(\omega_{\frac{j+1}{2}}4\pi) = 1 + 0 = 1$.

For $j \in \{2, 4, ..., M-2\}$

$$z_j' z_j = 2\sum_{k=1}^{M}(\cos 0 - \cos(\omega_{\frac{j}{2}}\frac{2\pi}{M}2k)) \tag{2}$$

because $\sin(mx)\sin(nx) = \frac{1}{2}(\cos(m-n)x - \cos(m+n)x)$. Then equation (2)

$$= 2M - 2\sum_{k=1}^{M}\cos(\omega_{\frac{j}{2}}\frac{4\pi k}{M}) = 2M - \frac{(e^{Mc}-1)(e^c+e^{-Mc})}{e^c-1},\ c = i\,\omega_{\frac{j}{2}}\frac{4\pi}{M}, = 2M$$

because $e^{Mc} = e^{i\omega_{\frac{j}{2}}4\pi} = \cos(\omega_{\frac{j}{2}}4\pi) + i\,\sin(\omega_{\frac{j}{2}}4\pi) = 1 + 0 = 1$.

Orthogonality of X=1 and Z

We show below that the odd-numbered columns of Z sum to 0. It can be shown similarly that the even-numbered columns also sum to 0.

Proof: For $j \in \{1, 3, ..., M-3\}$, the sum of j^{th} column is

$$2\sum_{k=1}^{M}\cos(\omega_{\frac{j+1}{2}}\frac{2\pi}{M}k) = \sum_{k=1}^{M}(e^{ck} + e^{-ck}), \text{ for } c = i\,\omega_{\frac{j+1}{2}}\frac{2\pi}{M},$$
$$= \frac{e^c(e^{Mc}-1)}{e^c-1} + \frac{e^{-Mc}(e^{Mc}-1)}{e^c-1} = \frac{(e^{Mc}-1)(e^c+e^{-Mc})}{e^c-1} = 0$$

because $e^{Mc} = e^{i\omega_{\frac{j+1}{2}}2\pi} = \cos(\omega_{\frac{j+1}{2}}2\pi) + i\,\sin(\omega_{\frac{j+1}{2}}2\pi) = 1 + 0 = 1$.

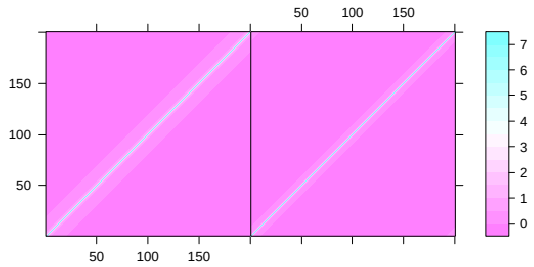
For $j = M - 1$, the sum of j^{th} column is

$$= 2\sum_{k=1}^{M} \cos(\omega_{\frac{M}{2}} \frac{2\pi}{M} k) = 2\sum_{k=1}^{M} \cos(\pi k), \text{ since } \omega_{\frac{M}{2}} = \frac{M}{2}$$
$$= \sum_{k=1}^{M} (e^{\pi k} + e^{-\pi k}), = \frac{(e^{Mc} - 1)(e^{c} + e^{-M})}{e^{c} - 1}, \text{ for } c = i\,\pi, = 0$$

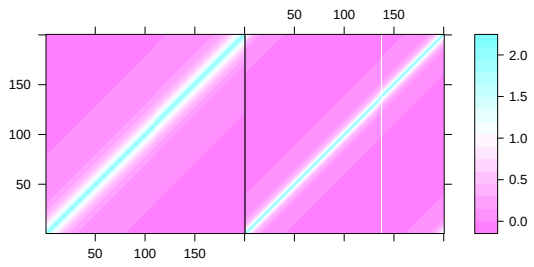
because $e^{Mc} = e^{iM\pi} = \cos(M\pi) + i\,\sin(M\pi) = 1 + 0 = 1$.

2 Web Appendix B. Covariance matrices

Figures 1 and 2 show some exact, approximate, and estimated covariance matrices for our model [GP + iid Normal error] with the spectral approximation, as discussed in the last paragraph of Section 2.3 in the main paper.

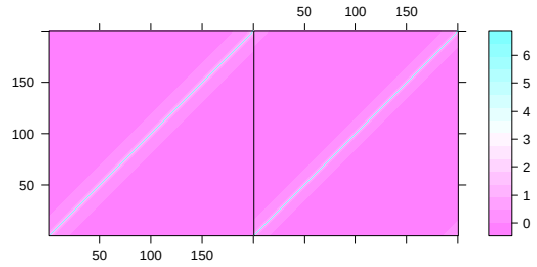


(a) Left: Exact covariance matrix of [GP + N error] with $\sigma^2_{\mathfrak{s}}$=2, σ^2_e=5 and ρ = 5 on the domain [1,2,...,199,200]. Right: Approximate covariance matrix of the same model.

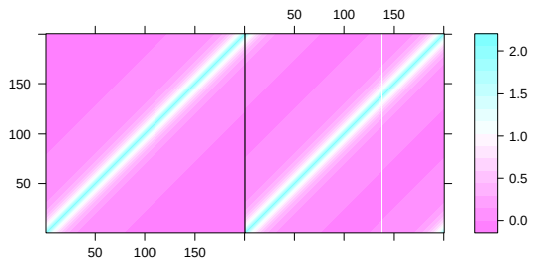


(b) Left: Exact covariance matrix of [GP + N error] with $\sigma^2_{\mathfrak{s}}$=2, σ^2_e=0.1 and $\rho = 16.67$ on the domain [1,2,...,199,200]. Right: Approximate covariance matrix of the same model.

Figure 1



(a) Left: Fitted exact covariance matrix for data with $\sigma^2_{\mathfrak{s}}$=2, σ^2_e=5 and ρ=5 using estimates from the exact restricted likelihood (RL). Right: Fitted approximate covariance using estimates from the approximate RL.



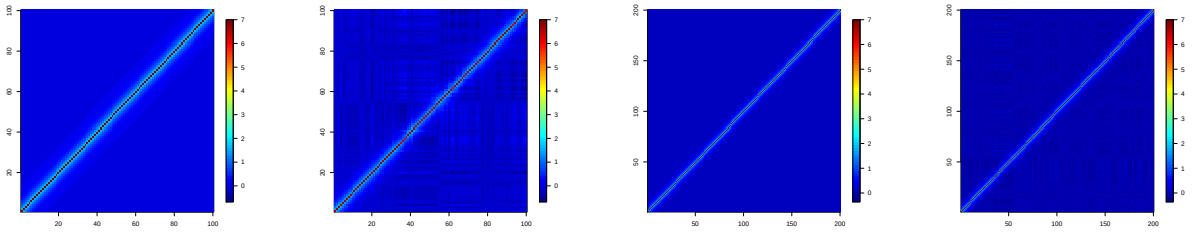
(b) Left: Fitted exact covariance matrix for data with $\sigma^2_{\mathfrak{s}}$=2, σ^2_e=0.1 and ρ=16.67 using estimates from the exact restricted likelihood (RL). Right: Fitted approximate covariance using estimates from the approximate RL.

Figure 2

2.1 Cov(y) and Cov(y^*) for M=100, and M=200

In Section 4 of the main paper, we claim that the covariance of the residuals from a regression of the data on the fixed effects in the model is reasonably well approximated by the assumed covariance structure of the data y, and implicitly argue that this approximation should be better, for a given rank of fixed-effect design matrix X, if the number of observations M is larger. Denote the residuals by y^*. Note that it is enough that Cov(y^*) resembles the covariance structure of y; it need not be the same in value as Cov(y), because the values of the parameters $\sigma^2_{\mathfrak{s}}$, σ^2_e and ρ will be different for y and for y^*. This appendix shows covariance matrices for y and for y^* (plotted using the same color scale), for the scenarios described below. Figures 3 to 6 use the exponential covariance (Matérn with ν=0.5), and X has two columns, namely a column of ones and a column of random draws from $N(0,1)$. Figures 7 to 10 also use the exponential covariance, but the second column of X is now the first column of the spectral basis matrix Z which is an extreme case because the effect of the residual transform would be concentrated on the column of Z with the largest eigenvalue. Figures 11 to 14 use the Matérn with ν=3.5 covariance, and X has two columns, namely a column of ones and a column of random draws from $N(0,1)$. Figures 15 to 18 also use the Matérn with ν=3.5 covariance, but now the second column of X is the first column of the spectral basis matrix. We consider two sets of parameter values for the two variance parameters: $\sigma^2_{\mathfrak{s}} = 2$, $\sigma^2_e = 5$, and

$\sigma^2_{\mathfrak{s}} = 2$, $\sigma^2_e = 0.1$, the object being to have two different ratios $\sigma^2_{\mathfrak{s}}/\sigma^2_e$. The two parameter values considered for the range parameter are ρ=5, and ρ=16.67.



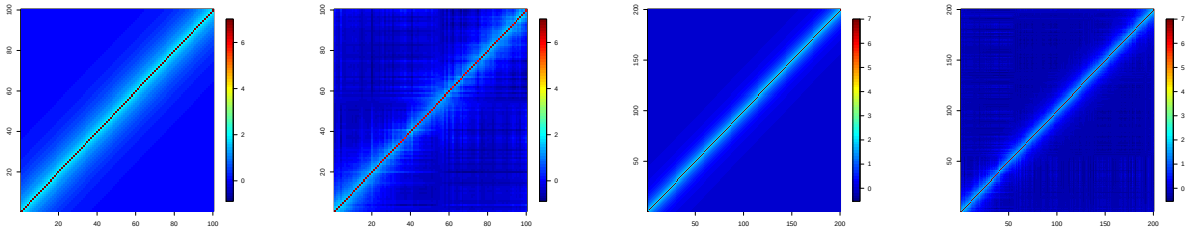
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=100.

(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=100.

(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=200.

(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=200.

Figure 3



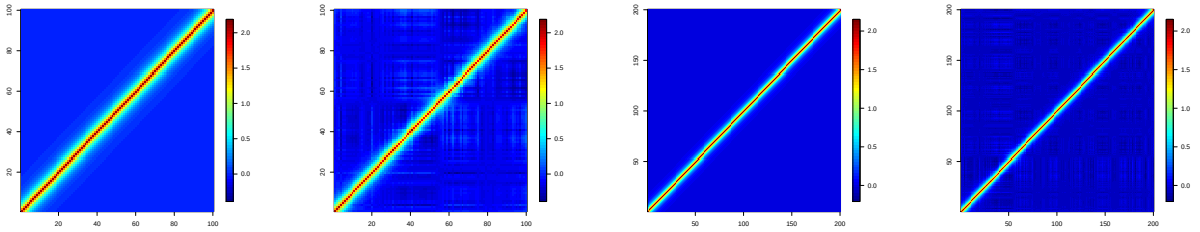
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=100.

(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=100.

(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=200.

(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=200.

Figure 4



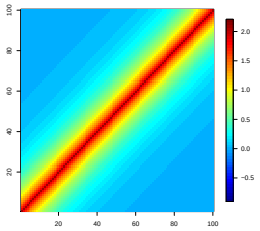
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=100.

(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=100.

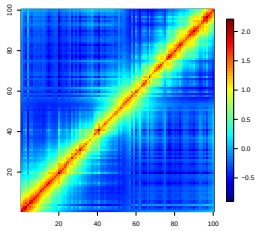
(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=200.

(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=200.

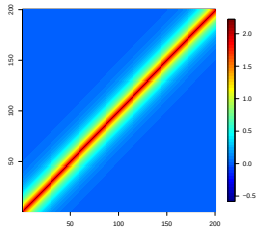
Figure 5



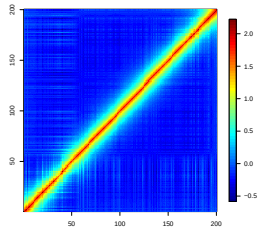
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=100.

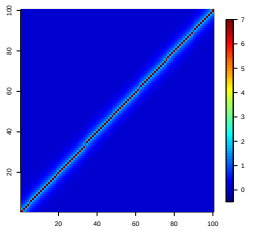


(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=200.

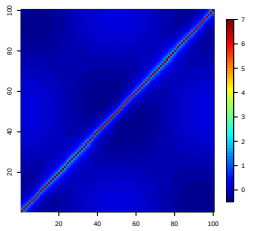


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=200.

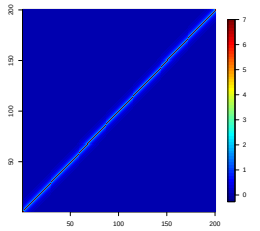
Figure 6



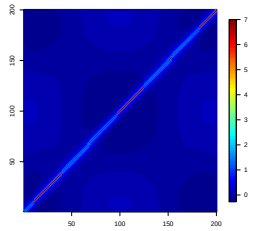
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=100.

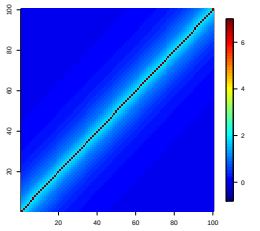


(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=200.

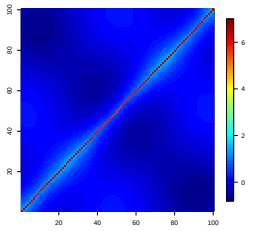


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=200.

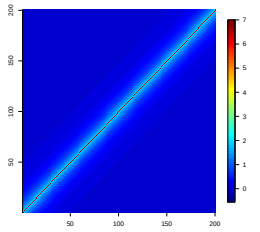
Figure 7



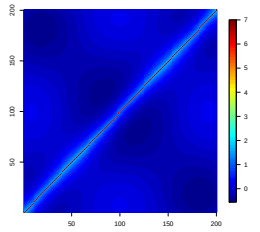
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=100.

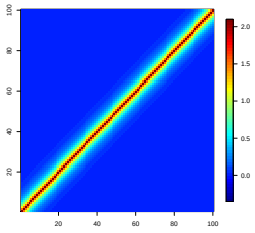


(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=200.

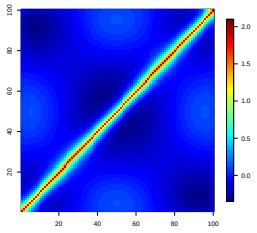


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=200.

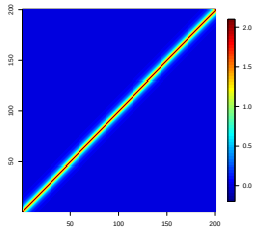
Figure 8



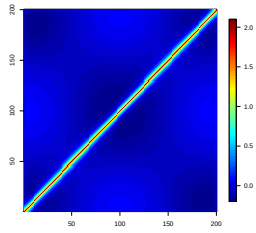
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=100.

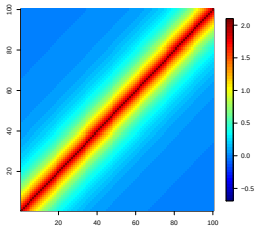


(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=200.

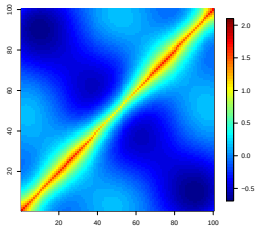


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=200.

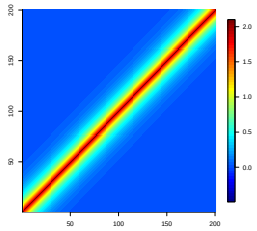
Figure 9



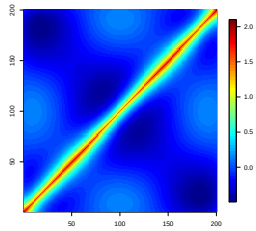
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=100.

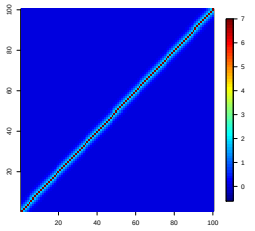


(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=200.

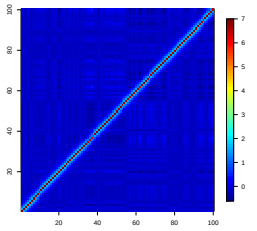


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=200.

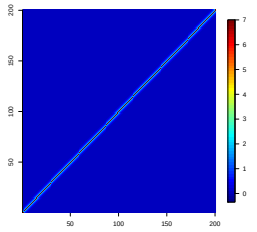
Figure 10



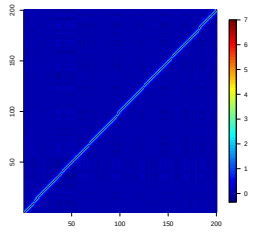
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=100.

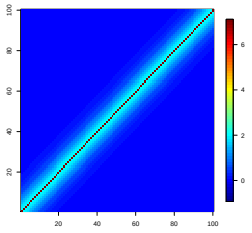


(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=200.

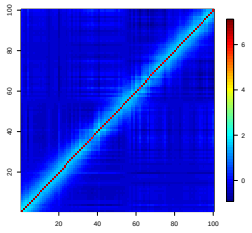


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=200.

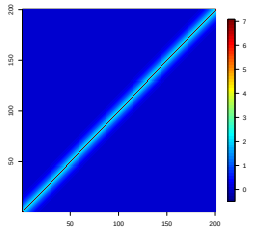
Figure 11



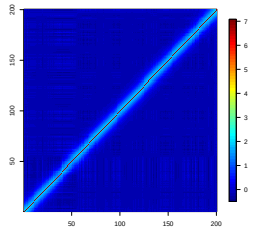
(a) Cov(y). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=5, ρ=16.67, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=5, ρ=16.67, M=100.

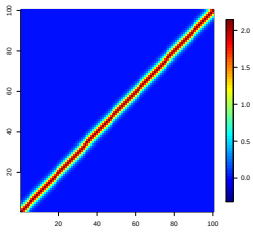


(c) Cov(y). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=5, ρ=16.67, M=200.

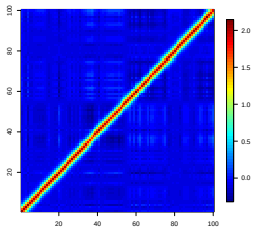


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=5, ρ=16.67, M=200.

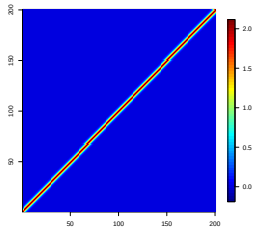
Figure 12



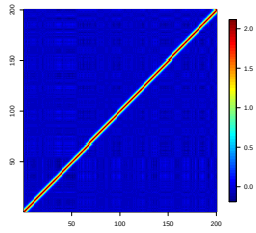
(a) Cov(y). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=0.1, ρ=5, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=0.1, ρ=5, M=100.

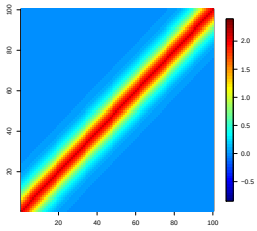


(c) Cov(y). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=0.1, ρ=5, M=200.

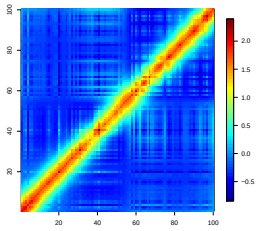


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=0.1, ρ=5, M=200.

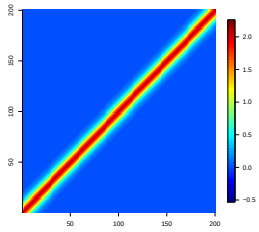
Figure 13



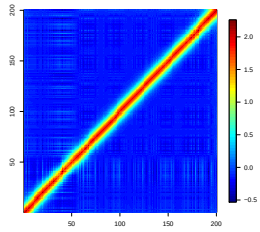
(a) Cov(y). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=0.1, ρ=16.67, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=0.1, ρ=16.67, M=100.

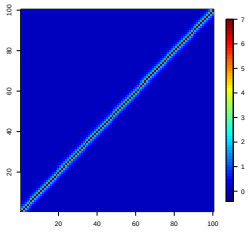


(c) Cov(y). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=0.1, ρ=16.67, M=200.

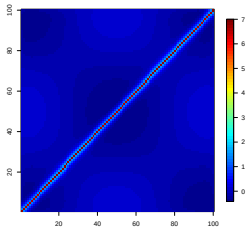


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}}$ = 2, σ^2_e=0.1, ρ=16.67, M=200.

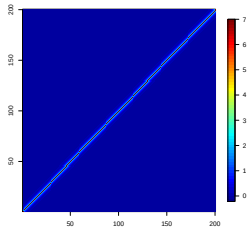
Figure 14



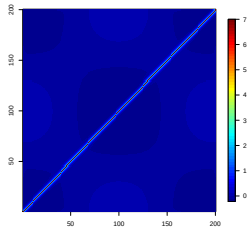
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=100.

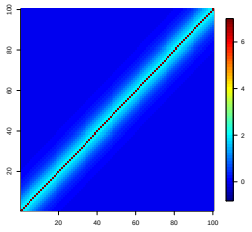


(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=200.

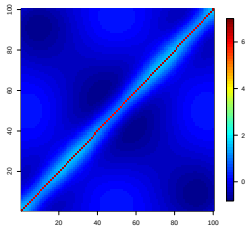


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=5, M=200.

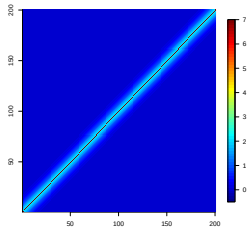
Figure 15



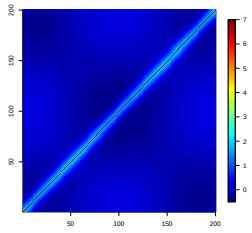
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=100.

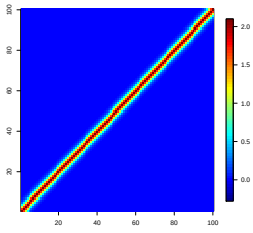


(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=200.

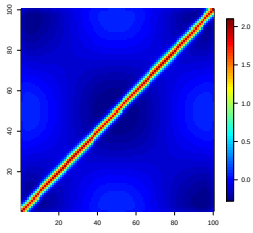


(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=5, ρ=16.67, M=200.

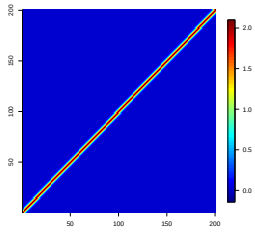
Figure 16



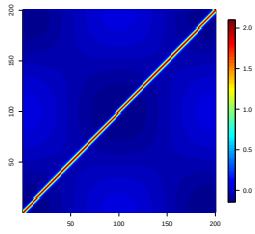
(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=100.



(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=100.

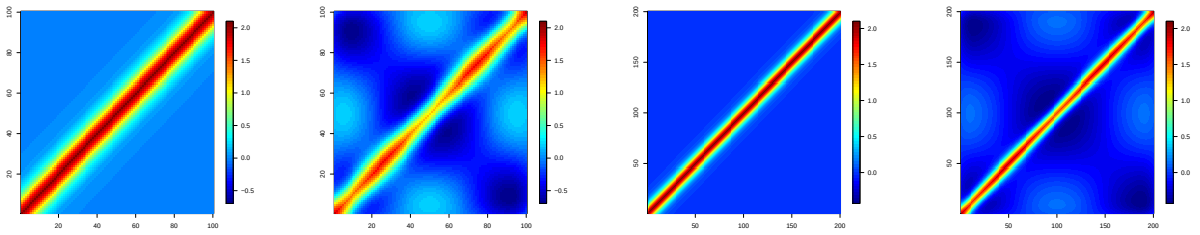


(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=200.



(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=5, M=200.

Figure 17



(a) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=100.

(b) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=100.

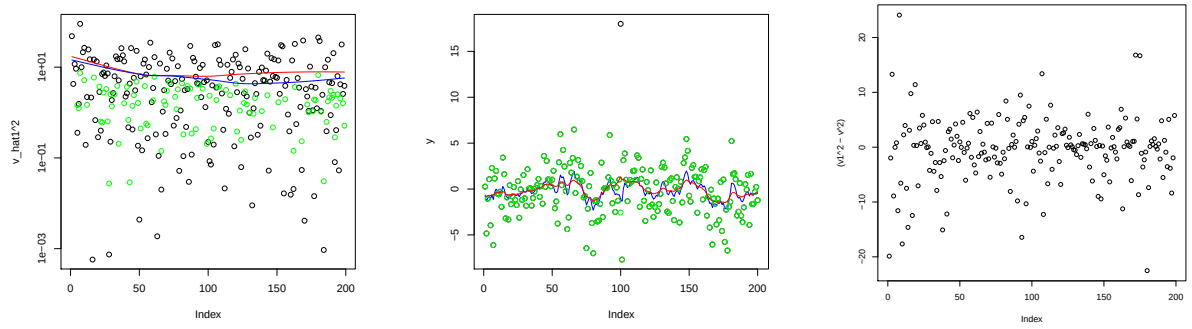
(c) Cov(y). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=200.

(d) Cov(y^*). $\sigma^2_{\mathfrak{s}} = 2$, σ^2_e=0.1, ρ=16.67, M=200.

Figure 18

3 Web Appendix C. Simulation results

This appendix supplements the simulation results reported in the main paper regarding the effects of contaminating data simulated from an intercept-only linear mixed model with GP-distributed random effect and iid Normal errors, for two contaminations: adding an outlier (Section 3.1 in the main paper; Table 1 and Figure 19a,b below) and adding a mean shift in the data series (Section 3.2 of the main paper; Tables 1 and 2 below). This appendix also presents conjectures and simulation results regarding contamination by replacing a part of the data series by data simulated from a GP with a different range parameter (ρ), which is not discussed in the main paper. Section 3.0.1 below gives these conjectures and discusses simulation results presented in Figure 19c and Tables 1 and 3 below.



(a) Green circles: log $v_j{}^2$ from simulated data; blue line: loess smooth of log $v_j{}^2$; black circles, red line: from data contaminated by outlier.

(b) Green circles: simulated data; blue line: fit; black circle: outlier added; red line: fit of data contaminated by outlier.

(c) Difference of $v_j{}^2$'s (contaminated - uncontaminated) from simulated data, contaminated by GP with ρ=16.67.

Figure 19: (a) and (b) show effects of contamination by an outlier; (c) shows effects of contamination by GP with a different range. Parameter values: $\sigma_{\mathfrak{s}}^2$=2, σ_e^2=5 and ρ=5. Estimates from exact restricted likelihood maximization.

Table 1: Estimates of GP parameters maximizing the exact and approximate restricted log likelihoods ("exact RL" and "apprx RL" respectively) for two simulated datasets. Also presented are estimates after the data is contaminated by an outlier in the middle, by a mean-shift midway, or by replacing a subseries of observations by observations from a GP with a different ρ.

	exact RL			apprx RL		
	$\sigma_{\mathfrak{s}}^2$	σ_e^2	ρ	$\sigma_{\mathfrak{s}}^2$	σ_e^2	ρ
actual values used in simulation	2	5	5	2	5	5
estimates from simulated data	1.76	4.66	4.00	1.71	4.84	8.33
from data contaminated by outlier	1.07	7.02	11.11	1.07	7.06	25.00
from data contaminated by shift	13.05	5.63	58.82	14.51	5.71	106.38
from data contaminated by another GP	1.69	4.24	3.04	1.56	4.51	7.09
actual values used in simulation	2	0.1	16.7	2	0.1	16.7
estimates from simulated data	1.95	0.11	16.7	1.94	0.15	33.33
from data contaminated by outlier	1.59	1.93	12.50	1.62	1.96	25.00
from data contaminated by shift	11.51	0.03	55.56	10.99	0.12	93.46
from data contaminated by another GP	1.60	0.09	8.55	1.56	0.18	17.86

Table 2: Average over 100 simulated datasets of estimates maximizing the exact and approximate restricted likelihoods: contamination by mean shift. Monte Carlo standard errors are in parentheses.

	exact RL			approximate RL		
	$\sigma^2_{\mathfrak{s}}$	σ^2_e	ρ	$\sigma^2_{\mathfrak{s}}$	σ^2_e	ρ
actual values	2	5	5	2	5	5
uncontaminated	2.29 (0.11)	4.75 (0.11)	6.89 (0.82)	2.39 (0.15)	4.90 (0.09)	13.51 (1.74)
contaminated	11.50 (0.36)	5.61 (0.06)	106.48 (6.06)	11.60 (0.41)	5.33 (0.07)	122.0 (6.49)
actual values	2	5	16.67	2	5	16.67
uncontaminated	2.16 (0.09)	4.89 (0.07)	26.12 (4.32)	2.17 (0.09)	4.91 (0.06)	38.11 (2.53)
contaminated	12.56 (0.59)	5.22 (0.07)	140.13 (8.31)	15.14 (0.83)	5.01 (0.05)	169.7 (11.2)
actual values	2	0.1	5	2	0.1	5
uncontaminated	1.94 (0.03)	0.099 (0.01)	4.97 (0.13)	1.97 (0.03)	0.232 (0.01)	10.24 (0.25)
contaminated	9.82 (0.17)	0.260 (0.01)	34.89 (1.29)	10.05 (0.20)	0.346 (0.01)	63.89 (2.21)
actual values	2	0.1	16.67	2	0.1	16.67
uncontaminated	2.09 (0.07)	0.093 (0.00)	18.07 (0.84)	2.01 (0.08)	0.144 (0.00)	34.84 (1.67)
contaminated	10.04 (0.29)	0.117 (0.00)	72.84 (2.96)	14.16 (0.79)	0.176 (0.00)	153.7 (9.00)
actual values	10	5	5	10	5	5
uncontaminated	10.17 (0.20)	4.99 (0.13)	5.58 (0.18)	9.98 (0.21)	5.63 (0.10)	10.87 (0.41)
contaminated	16.26 (0.49)	5.84 (0.13)	15.60 (1.32)	17.39 (0.39)	6.18 (0.11)	26.22 (1.27)
actual values	10	5	16.67	10	5	16.67
uncontaminated	10.44 (0.37)	4.80 (0.06)	17.57 (0.83)	10.54 (0.43)	5.08 (0.06)	35.60 (2.05)
contaminated	18.89 (0.84)	5.09 (0.07)	39.18 (3.05)	19.29 (0.87)	5.28 (0.06)	66.12 (4.24)
actual values	10	0.1	5	10	0.1	5
uncontaminated	9.70 (0.17)	0.21 (0.03)	5.26 (0.13)	9.65 (0.17)	0.76 (0.03)	10.34 (0.28)
contaminated	16.09 (0.35)	0.52 (0.04)	10.30 (0.34)	16.66 (0.37)	1.05 (0.04)	21.14 (0.64)
actual values	10	0.1	16.67	10	0.1	16.67
uncontaminated	9.99 (0.34)	0.11 (0.01)	17.29 (0.67)	10.02 (0.37)	0.32 (0.01)	32.70 (1.37)
contaminated	17.96 (0.77)	0.16 (0.01)	29.99 (1.32)	19.70 (1.13)	0.37 (0.01)	61.49 (5.08)

3.0.1 Contamination by changing ρ partway through the series

Suppose a draw from an intercept-only mixed linear model with GP-distributed random effect is contaminated by replacing a substantial subseries with a draw from another GP with the same $\sigma^2_{\mathfrak{s}}$ and σ^2_e but with a different range ρ. How does this contamination affect the estimates of the GP parameters? Recalling how $a_j(\rho)$ depends on ρ, we conjecture that:

1. When the contamination is by a GP with higher ρ, the v_j^2's for small j's (low frequency columns of Z) will be inflated, which will inflate $\hat{\sigma}^2_{\mathfrak{s}}$. The v_j^2's for larger j's will be comparatively unaffected, so $\hat{\sigma}^2_e$ will be comparatively unaffected. Also, the v_j^2's will decline more sharply with j, inflating $\hat{\rho}$.

2. When the contamination is by a GP with smaller ρ, the v_j^2's for large j's (high frequencies) will be inflated, leading to an inflated $\hat{\sigma}^2_e$. The v_j^2's for smaller j will be comparatively unaffected, so $\hat{\sigma}^2_{\mathfrak{s}}$ will be unaffected too. $\hat{\rho}$ will be diminished to capture the more gradual decline of the v_j^2's with j.

In the simulated data, on average contamination by a GP with higher ρ inflates $\hat{\rho}$, contamination by a GP with a smaller ρ diminishes $\hat{\rho}$, and $\hat{\sigma}^2_{\mathfrak{s}}$ and $\hat{\sigma}^2_e$ are largely unaffected, all as expected (Table 3 below). Figure 19c above shows the change in the v_j^2's for one simulated dataset, for which it is not clear what the net effect on the estimates will be (Table 1 above).

Table 3: Average over 100 simulated datasets of estimates maximizing the exact and approximate restricted likelihoods: contaminated by GP with ρ=5 and ρ=16.67 for uncontaminated GP with ρ=16.67 and ρ=5 respectively. Monte Carlo standard errors are given in parentheses.

	exact RL			approximate RL		
	$\sigma^2_{\mathfrak{s}}$	σ^2_e	ρ	$\sigma^2_{\mathfrak{s}}$	σ^2_e	ρ
actual values	2	5	5	2	5	5
uncontaminated	2.29 (0.11)	4.75 (0.11)	6.89 (0.82)	2.39 (0.15)	4.90 (0.09)	13.51 (1.74)
contaminated	2.22 (0.11)	4.79 (0.09)	14.62 (3.45)	2.28 (0.09)	4.94 (0.08)	13.42 (0.66)
actual values	2	5	16.67	2	5	16.67
uncontaminated	2.16 (0.09)	4.89 (0.07)	26.12 (4.32)	2.17 (0.09)	4.91 (0.06)	38.11 (2.53)
contaminated	2.19 (0.08)	4.90 (0.06)	18.17 (2.42)	2.28 (0.09)	4.69 (0.07)	18.65 (1.39)
actual values	2	0.1	5	2	0.1	5
uncontaminated	1.94 (0.03)	0.099 (0.01)	4.97 (0.13)	1.97 (0.03)	0.232 (0.01)	10.24 (0.25)
contaminated	1.93 (0.04)	0.116 (0.01)	6.15 (0.18)	2.03 (0.04)	0.223 (0.01)	12.99 (0.36)
actual values	2	0.1	16.67	2	0.1	16.67
uncontaminated	2.09 (0.07)	0.093 (0.00)	18.07 (0.84)	2.01 (0.08)	0.144 (0.00)	34.84 (1.67)
contaminated	2.00 (0.06)	0.110 (0.01)	11.62(0.44)	1.95 (0.06)	0.183 (0.00)	22.55 (0.92)
actual values	10	5	5	10	5	5
uncontaminated	10.17 (0.20)	4.99 (0.13)	5.58 (0.18)	9.98 (0.21)	5.63 (0.10)	10.87 (0.41)
contaminated	10.11 (0.24)	4.89 (0.12)	6.74 (0.31)	10.16 (0.23)	5.48 (0.10)	13.55 (0.49)
actual values	10	5	16.67	10	5	16.67
uncontaminated	10.44 (0.37)	4.80 (0.06)	17.57 (0.83)	10.54 (0.43)	5.08 (0.06)	35.60 (2.05)
contaminated	10.19 (0.31)	5.09 (0.09)	12.49 (0.66)	9.92 (0.32)	5.43 (0.08)	24.39 (1.03)
actual values	10	0.1	5	10	0.1	5
uncontaminated	9.70 (0.17)	0.21 (0.03)	5.26 (0.13)	9.65 (0.17)	0.76 (0.03)	10.34 (0.28)
contaminated	9.70 (0.19)	0.27 (0.03)	6.66 (0.18)	9.50 (0.20)	0.73 (0.03)	12.67 (0.38)
actual values	10	0.1	16.67	10	0.1	16.67
uncontaminated	9.99 (0.34)	0.11 (0.01)	17.29 (0.67)	10.02 (0.37)	0.32 (0.01)	32.70 (1.37)
contaminated	9.67 (0.29)	0.19 (0.02)	11.29 (0.38)	10.22 (0.29)	0.49 (0.02)	23.69 (1.01)

4 Web Appendix D. Two dimensional Gaussian process model

This section describes the spectral representation of GPs in two dimensions and develops the 2-dimensional (2-D) approximate restricted likelihood. Web Appendix E gives proofs that the approximation is valid and that the columns of Z are orthogonal to each other and to a vector of 1s (the fixed-effect design column for the intercept).

Suppose observations are made on a 2-D grid with grid points at $[\{s_{1,1}, s_{1,2}, ..., s_{1,M_1}\}\times$

$\{s_{2,1}, s_{2,2}, ..., s_{2,M_2}\}]$ and suppose the vector of observations is given by

$$y = (y(s_{1,1}, s_{2,1}), y(s_{1,1}, s_{2,2}), ..., y(s_{1,1}, s_{2,M_2}), ..., y(s_{1,M_1}, s_{2,M_2}))'. \quad (3)$$

Given data $y(s)$ at location $s \in \{(s_{1,1}, s_{2,1}), (s_{1,1}, s_{2,2}), ..., (s_{1,1}, s_{2,M_2}), ..., (s_{1,M_1}, s_{2,M_2})\}$, we want to fit the model

$$y(s) = x(s)\beta + w(s) + \epsilon(s)$$

where all variables are defined analogously to the 1-dimensional (1-D) model. Similarly, the linear mixed model to be fit to y is

$$y = X\beta + I_{M_1 M_2}\gamma + \epsilon. \quad (4)$$

The restricted likelihood has the same form as in the 1-D case, and we obtain a matrix-free approximation to it by approximating the intercept-only GP using spectral basis functions. If the model includes fixed effects X, we regress them out as in the 1-D case, assume (approximately) that the residuals follow a linear mixed model with a random effect having a GP covariance, and then approximate the GP in this model using spectral basis functions. Section 4.1 below describes the spectral approximation following Wikle (2002) and Paciorek (2007); Section 4.2 below derives the simple approximate restricted likelihood arising from applying the spectral approximation to the intercept-only GP.

4.1 The spectral approximation in 2-D

Define $(\overset{1}{\omega}_{m_1}, \overset{2}{\omega}_{m_2}) \in \{0, \frac{1}{M_1}, \frac{2}{M_1}, ..., \frac{1}{2}, -\frac{1}{2}+\frac{1}{M_1}, -\frac{1}{2}+\frac{2}{M_1}, ..., -\frac{1}{M_1}\} \times \{0, \frac{1}{M_2}, \frac{2}{M_2}, ..., \frac{1}{2}, -\frac{1}{2}+\frac{1}{M_2}, -\frac{1}{2}+\frac{2}{M_2}, ..., -\frac{1}{M_2}\}$, $m_1 = 0, 1, ..., M_1 - 1$; $m_2 = 0, 1, ..., M_2 - 1$. At observation location $(s_{1,\imath}, s_{2,j})$, $s_{1,\imath} \in \{1, 2, .., M_1\}$, $s_{2,j} \in \{1, 2, .., M_2\}$, $\imath = 1, ..., M_1$, $j = 1, ..., M_2$, the mean zero stationary isotropic GP can be approximated, using spectral basis functions, by

$$g(s_{1,\imath}, s_{2,j}) = \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} \varphi_{m_1,m_2}(2\pi s_{1,\imath}, 2\pi s_{2,j}) u_{m_1,m_2}$$

$$\overset{\text{def}}{=} \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} \exp(i2\pi(\overset{1}{\omega}_{m_1} s_{1,\imath} + \overset{2}{\omega}_{m_2} s_{2,j}))(a_{m_1,m_2} + ib_{m_1,m_2})$$

where the a_m's and b_m's have independent mean zero Gaussian prior distributions with $\mathrm{V}(a_{m_1,m_2}) = \mathrm{V}(b_{m_1,m_2}) = \frac{1}{2M_1M_2}\sigma^2_{\mathfrak{s}}\phi(\overset{1}{\omega}_{m_1}, \overset{2}{\omega}_{m_2}; \rho)$, where $\phi(\overset{1}{\omega}_{m_1}, \overset{2}{\omega}_{m_2}; \rho)$ is the spectral density function of

$K(d;\rho)$, described further below. Note that the observation locations $s_{1,\imath} \in \{1,2,..,M_1\}$, $s_{2,j} \in \{1,2,..,M_2\}$ lie on a uniformly spaced rectangular 2-D grid and the grid sizes in the two dimensions, M_1 and M_2, must be even integers.

To approximate real valued processes (Wikle 2002, Paciorek 2007), assume

$$u_{m_1,m_2} = \overline{u}_{M_1-m_1,M_2-m_2}$$

for

- $m_1 = 1,2,..,M_1/2-1,\ m_2 = 1,2,..,M_2/2$
- $m_1 = 1,2,..,M_1/2,\ m_2 = M_2/2+1, M_2/2+2,..,M_2-1.$

Finally, assume $u_{m_1,m_2} = \overline{u}_{m_1,M_2-m_2},\ \text{for } m_1 = 0,\ m_2 = 1,2,..,M_2/2-1$

and $u_{m_1,m_2} = \overline{u}_{M_1-m_1,m_2},\ \text{for } m_1 = 1,2,..,M_1/2-1,\ m_2 = 0.$

Also assume

$$b_{00} = b_{0,M_2/2} = b_{M_1/2,0} = b_{M_1/2,M_2/2} = 0.$$

Then defining $\boldsymbol{\omega} = (\overset{1}{\omega}_{m_1}, \overset{2}{\omega}_{m_2})'$ and $\boldsymbol{s} = (s_{1,\imath}, s_{2,j})'$, the approximation $g(s_{1,\imath}, s_{2,j})$ becomes

$$\begin{aligned} g(s_{1,\imath}, s_{2,j}) = 2 \sum_{(m_1,m_2)\in Q} \big(a_{m_1,m_2}\cos(2\pi\boldsymbol{\omega}'\boldsymbol{s}) - b_{m_1,m_2}\sin(2\pi\boldsymbol{\omega}'\boldsymbol{s})\big) \\ + a_{00} \\ + a_{0,M_2/2}\cos(2\pi(\overset{1}{\omega}_0 s_{1,\imath} + \overset{2}{\omega}_{M_2/2} s_{2,j})) + a_{M_1/2,0}\cos(2\pi(\overset{1}{\omega}_{M_1/2} s_{1,\imath} + \overset{2}{\omega}_0 s_{2,j})) \\ + a_{M_1/2,M_2/2}\cos(2\pi(\overset{1}{\omega}_{M_1/2} s_{1,\imath} + \overset{2}{\omega}_{M_2/2} s_{2,j})), \end{aligned} \tag{5}$$

where $Q = Q_1 \cup Q_2 \cup Q_3 \cup Q_4$, with

$Q_1 = \{(m_1,m_2) : m_1 = 1,2,...,M_1/2-1; m_2 = 1,2,...,M_2/2\},$

$Q_2 = \{(m_1,m_2) : m_1 = 1,2,...,M_1/2; m_2 = M_2/2+1, M_2/2+2,...,M_2-1\},$

$Q_3 = \{(m_1,m_2) : m_1 = 0; m_2 = 1,2,...,M_2/2-1\},$

$Q_4 = \{(m_1,m_2) : m_1 = 1,2,...,M_1/2-1; m_2 = 0\}.$

Then $(g(s_{1,1},s_{2,1}), g(s_{1,1},s_{2,2}), ..., g(s_{1,1},s_{2,M_2}), ..., g(s_{1,M_1},s_{2,M_2}))' = 1_{M_1M_2}\beta^* + Zu,$

where β^* is the intercept, Z is an $M_1M_2 \times (M_1M_2-1)$ matrix described below, $u = (a_{11}, b_{11}, a_{12}, b_{12},$

$..,a_{1,\frac{M_2}{2}}, b_{1,\frac{M_2}{2}},\ .....,\ a_{\frac{M_1}{2}-1,1}, b_{\frac{M_1}{2}-1,1}, .., a_{\frac{M_1}{2}-1,\frac{M_2}{2}}, b_{\frac{M_1}{2}-1,\frac{M_2}{2}},$

$a_{1,\frac{M_2}{2}+1}, b_{1,\frac{M_2}{2}+1}, a_{1,\frac{M_2}{2}+2}, b_{1,\frac{M_2}{2}+2}, .., a_{1,M_2-1}, b_{1,M_2-1},\ .....,$

$a_{\frac{M_1}{2},\frac{M_2}{2}+1}, b_{\frac{M_1}{2},\frac{M_2}{2}+1}, .., a_{\frac{M_1}{2},M_2-1}, b_{\frac{M_1}{2},M_2-1},$

$a_{01}, b_{01}, a_{02}, b_{02}, ..., a_{0,M_2/2-1}, b_{0,M_2/2-1},$

$a_{10}, b_{10}, a_{20}, b_{20}, ..., a_{M_1/2-1,0}, b_{M_1/2-1,0},$

$a_{0,M_2/2}, a_{M_1/2,0}, a_{M_1/2,M_2/2})'$, and

1. $\mathrm{V}(a_{m_1,m_2}) = \mathrm{V}(b_{m_1,m_2}) = \frac{1}{2M_1M_2}\sigma^2_{\mathfrak{s}}\phi(\overset{1}{\omega}_{m_1}, \overset{2}{\omega}_{m_2}; \rho)$ for $m_1 = 0, 1, .., M_1-1$, $m_2 = 0, 1, .., M_2-1$, $(m_1, m_2) \notin \{(0,0), (0, M_2/2), (M_1/2, 0), (M_1/2, M_2/2)\}$.

2. $\mathrm{V}(a_{0,M_2/2}) = \frac{1}{M_1M_2}\sigma^2_{\mathfrak{s}}\phi(\overset{1}{\omega}_0, \overset{2}{\omega}_{M_2/2}; \rho)$, $\mathrm{V}(a_{M_1/2,0}) = \frac{1}{M_1M_2}\sigma^2_{\mathfrak{s}}\phi(\overset{1}{\omega}_{M_1/2}, \overset{2}{\omega}_0; \rho)$.

3. $\mathrm{V}(a_{M_1/2,M_2/2}) = \frac{1}{M_1M_2}\sigma^2_{\mathfrak{s}}\phi(\overset{1}{\omega}_{M_1/2}, \overset{2}{\omega}_{M_2/2}; \rho)$.

Constructing G

In this construction, the covariance of the random effect vector u is a matrix G with all off-diagonal elements zero and diagonal elements given by the above variances.

The function $\phi(\boldsymbol{\omega}; \rho)$ is the spectral density, in two dimensions, of the correlation function of the GP being approximated. For example, the Matérn correlation function with smoothness ν, for Euclidean distances d, is

$$K(d; \rho, \nu) = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}d}{\rho}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}d}{\rho}\right), \tag{6}$$

and it has spectral density

$$\phi(\boldsymbol{\omega}; \rho, \nu) = \frac{\Gamma(\nu + D/2)(4\nu)^\nu}{\pi^{D/2}\Gamma(\nu)(\pi\rho)^{2\nu}}\left(\frac{4\nu}{(\pi\rho)^2} + \boldsymbol{\omega}'\boldsymbol{\omega}\right)^{-(\nu+D/2)},$$

where D is the dimension of the process (Paciorek 2007). For the exponential correlation function (Matérn with $\nu = 0.5$) in two dimensions, the spectral density is a Cauchy density

$$\phi(\boldsymbol{\omega}; \rho) = \frac{\pi\rho^2}{4}\left(1 + \frac{(\pi\rho)^2}{2}\boldsymbol{\omega}'\boldsymbol{\omega}\right)^{-3/2}. \tag{7}$$

The elements of u are ordered so that $\phi(\boldsymbol{\omega_j}; \rho)$ is a non-increasing function of j (this ordering will be the same irrespective of the value of ρ).

Constructing Z

After constructing u, the matrix Z is constructed conformably from (5) as follows.

Consider the vector of length $(M_1/2 - 1)M_2/2$

$$[2\cos(\overset{1}{\omega}_{m_1} 2\pi s_{1,i} + \overset{2}{\omega}_{m_2} 2\pi s_{2,j})]_{m_1=1,2,..,M_1/2-1; m_2=1,2,..,M_2/2},$$ where m_2 varies fastest.

Construct one such vector for each $(s_{1,\imath}, s_{2,j})$ combination; this gives us M_1M_2 such vectors. Let these vectors be $A_1, A_2, ..., A_{M_1M_2}$ respectively. Next, similarly to the above, consider the vector of length $(M_1/2-1)M_2/2$

$$[-2\sin(\overset{1}{\omega}_{m_1} 2\pi s_{1,\imath} + \overset{2}{\omega}_{m_2} 2\pi s_{2,j})]_{m_1=1,2,..,M_1/2-1; m_2=1,2,..,M_2/2}, \text{ where } m_2 \text{ varies fastest.}$$

Construct one such vector for each $(s_{1,\imath}, s_{2,j})$ combination; this gives us M_1M_2 such vectors. Let these vectors be $B_1, B_2, ..., B_{M_1M_2}$ respectively. Now construct a matrix P_1 with the r^{th} row of P_1 given by $(A_{r1}, B_{r1}, A_{r2}, B_{r2}, ..., A_{r,(M_1/2-1)M_2/2}, B_{r,(M_1/2-1)M_2/2})$, where A_{rt} denotes the t^{th} element of the vector A_r, $r = 1, 2, ..., M_1M_2$, $t = 1, 2, ..., (M_1/2-1)M_2/2$. The matrix P_1 constitutes the first $(M_1/2-1)M_2$ columns of Z.

Next, consider the vector of length $M_1/2(M_2/2-1)$

$$[2\cos(\overset{1}{\omega}_{m_1} 2\pi s_{1,\imath} + \overset{2}{\omega}_{m_2} 2\pi s_{2,j})]_{m_1=1,2,..,M_1/2; m_2=M_2/2+1, M_2/2+2,..,M_2-1},$$

where m_2 varies fastest. Construct one such vector for each $(s_{1,\imath}, s_{2,j})$ combination. Let these vectors be $C_1, C_2, ..., C_{M_1M_2}$ respectively. Next, similarly to the above, consider the vector of length $M_1/2(M_2/2-1)$

$$[-2\sin(\overset{1}{\omega}_{m_1} 2\pi s_{1,\imath} + \overset{2}{\omega}_{m_2} 2\pi s_{2,j})]_{m_1=1,2,..,M_1/2; m_2=M_2/2+1, M_2/2+2,..,M_2-1},$$

where m_2 varies fastest. Construct one such vector for each $(s_{1,\imath}, s_{2,j})$ combination; this gives us M_1M_2 such vectors. Let these vectors be $D_1, D_2, ..., D_{M_1M_2}$ respectively. Now construct a matrix P_2 with the r^{th} row of P_2 given by $(C_{r1}, D_{r1}, C_{r2}, D_{r2}, ..., C_{r,M_1/2(M_2/2-1)}, D_{r,M_1/2(M_2/2-1)})$, where C_{rt} denotes the t^{th} element of the vector C_r, $r = 1, 2, ..., M_1M_2$, $t = 1, 2, ..., M_1/2(M_2/2-1)$. The matrix P_2 constitutes the next $M_1(M_2/2-1)$ columns of Z.

Next, consider the vector

$$[2\cos(2\pi(\overset{1}{\omega}_0 s_{1,\imath} + \overset{2}{\omega}_{m_2} s_{2,j}))]_{m_2=1,2,..,M_2/2-1}.$$

Construct one such vector for every $(s_{1,\imath}, s_{2,j})$. Call them $F_1, F_2, ..., F_{M_1M_2}$ respectively. Consider the vector

$$[-2\sin(2\pi(\overset{1}{\omega}_0 s_{1,\imath} + \overset{2}{\omega}_{m_2} s_{2,j}))]_{m_2=1,2,..,M_2/2-1}.$$

Construct one such vector for every $(s_{1,\imath}, s_{2,j})$. Call them $H_1, H_2, ..., H_{M_1M_2}$ respectively. Construct a matrix P_3 with M_2-2 columns as: if r is odd the r^{th} column of P_3 is given by ($\frac{r+1}{2}^{th}$ element of

F_1, $\frac{r+1}{2}^{th}$ element of F_2,..., $\frac{r+1}{2}^{th}$ element of $F_{M_1M_2})'$; if r is even the r^{th} column of P_3 is given by ($\frac{r}{2}^{th}$ element of H_1, $\frac{r}{2}^{th}$ element of H_2,..., $\frac{r}{2}^{th}$ element of $H_{M_1M_2})'$. Then, the next M_2-2 columns of Z are P_3.

Consider the vector

$$[2\cos(2\pi(\overset{1}{\omega}_{m_1}s_{1,\imath}+\overset{2}{\omega}_0 s_{2,j}))]_{m_1=1,2,..,M_1/2-1}.$$

Construct one such vector for every $(s_{1,\imath}, s_{2,j})$. Call them $\Lambda_1, \Lambda_2, ..., \Lambda_{M_1M_2}$ respectively. Consider the vector

$$[-2\sin(2\pi(\overset{1}{\omega}_{m_1}s_{1,\imath}+\overset{2}{\omega}_0 s_{2,j}))]_{m_1=1,2,..,M_1/2-1}.$$

Construct one such vector for every $(s_{1,\imath}, s_{2,j})$. Call them $\Delta_1, \Delta_2, ..., \Delta_{M_1M_2}$ respectively. Construct a matrix P_4 with M_1-2 columns as: if r is odd the r^{th} column of P_4 is given by ($\frac{r+1}{2}^{th}$ element of Λ_1, $\frac{r+1}{2}^{th}$ element of Λ_2,..., $\frac{r+1}{2}^{th}$ element of $\Lambda_{M_1M_2})'$; if i is even the r^{th} column of P_4 is given by ($\frac{r}{2}^{th}$ element of Δ_1, $\frac{r}{2}^{th}$ element of Δ_2,..., $\frac{r}{2}^{th}$ element of $\Delta_{M_1M_2})'$. Then, the next M_1-2 columns of Z are P_4.

Finally, construct three more vectors P_5, P_6, and P_7 as follows. Construct the vector P_5

$$[\cos(\overset{1}{\omega}_0 2\pi s_{1,\imath}+\overset{2}{\omega}_{M_2/2}2\pi s_{2,j})]_{\imath=1,2,..,M_1;j=1,2,..,M_2}, \text{where } j \text{ varies fastest.}$$

Construct the vector P_6

$$[\cos(\overset{1}{\omega}_{M_1/2}2\pi s_{1,\imath}+\overset{2}{\omega}_0 2\pi s_{2,j})]_{\imath=1,2,..,M_1;j=1,2,..,M_2}, \text{where } j \text{ varies fastest.}$$

Construct the vector P_7

$$[\cos(\overset{1}{\omega}_{M_1/2}2\pi s_{1,\imath}+\overset{2}{\omega}_{M_2/2}2\pi s_{2,j})]_{\imath=1,2,..,M_1;j=1,2,..,M_2}, \text{where } j \text{ varies fastest.}$$

The last 3 columns of Z are P_5, P_6, and P_7 respectively.

Now order the columns of Z so that $\phi(\boldsymbol{\omega_j};\rho)$ is a non-increasing function of j (this ordering will be the same irrespective of the value of ρ).

4.2 The simple approximate restricted likelihood

With the definitions given above, the approximate model to be fit to the data y, if X in Section 4 above is just a column of 1's, or to be fit to the residuals y^*, if X in Section 4 also includes observed covariates, is

$$y \text{ or } y^* \approx 1_{M_1M_2}\beta^* + Zu + \epsilon^*, \tag{8}$$

with ϵ an $M_1M_2 \times 1$ vector of iid N(0,σ_e^2) errors. (If the left-hand side of equation (8) is y^*, β^* is necessarily be zero.) Z has these properties: (i)$Z'1_{M_1M_2}$=$\mathbf{0}$, and (ii)$Z'Z = M_1M_2\text{Diag}(2, 2, ..., 2,1, 1, 1)$, i.e., the columns of Z are orthogonal to each other and to the constant vector (proofs are in Web Appendix E). Premultiplying (8) by $(Z'Z)^{-0.5}Z'$, under the assumed model, the transformed data vector $v = (Z'Z)^{-0.5}Z'y$ or $(Z'Z)^{-0.5}Z'y^*$ has a Normal distribution with $\text{E}(v) = 0$ and $\text{Cov}(v) = \sigma_\mathfrak{s}^2\text{Diag}(a_j(\rho)) + \sigma_e^2 I_{M_1M_2-1}$, for $a_j(\rho) = \phi(\boldsymbol{\omega_j}; \rho)$. From the distribution of v_j, the approximate log restricted likelihood has the matrix-free form

$$\text{ALR}(\sigma_\mathfrak{s}^2, \sigma_e^2, \rho) = \text{const} - \frac{1}{2}\sum_{j=1}^{M_1M_2-1}\left(\log(\sigma_\mathfrak{s}^2 a_j(\rho) + \sigma_e^2) + v_j^2(\sigma_\mathfrak{s}^2 a_j(\rho) + \sigma_e^2)^{-1}\right). \tag{9}$$

By convention, we sort the $a_j(\rho)$ so they are in non-increasing order as j increases and the columns of Z in the same order. It is easy to show that this order does not depend on ρ. By construction, the successive column pairs of the spectral basis matrix Z decompose the data into successive frequency components, as in the 1-D case. Thus a high frequency feature, e.g., an outlier, will fall in the space spanned by the columns of Z corresponding to large j and will affect v_j^2 for larger j, which in turn will affect $\hat{\sigma}_e^2$. A low frequency feature, for example, a linear or quadratic trend in the data, will fall in the space spanned by the columns of Z corresponding to small j and thus will affect v_j^2 for smaller j, which in turn will affect $\hat{\sigma}_\mathfrak{s}^2$ and $\hat{\rho}$. A prominent low-frequency feature will inflate both $\hat{\sigma}_\mathfrak{s}^2$ and $\hat{\rho}$, which have opposite effects on the fit's smoothness, so the net effect of a low frequency feature will depend on the specifics of the data and model.

Note that to get to this point, we have made two approximations: approximating $(I_{M_1M_2} - P_X)\,\Sigma\,(I_{M_1M_2} - P_X) + \sigma_e^2(I_{M_1M_2} - P_X)$ by $\Sigma + R$, and the spectral approximation. Web Appendix F shows some examples of these covariance matrices.

5 Web Appendix E. 2-D spectral approximation: technical details

5.1 Proof that it is a valid approximation

The correlation of the approximate process is approximately equal to that of the actual GP for $M_1 \to \infty, M_2 \to \infty$.

Proof: For two dimensional frequencies $\boldsymbol{\omega}$ and two dimensional locations $\boldsymbol{s}$, the spectral density $\phi_2(\boldsymbol{\omega}; \rho)$ of the correlation function $K(\boldsymbol{s}; \rho)$ (correlation of the actual GP) is defined as

$$\phi_2(\boldsymbol{\omega}; \rho) = \int_{R^2} \exp(-2\pi i \boldsymbol{\omega}' \boldsymbol{s}) K(\boldsymbol{s}; \rho) \, \mathrm{d}\boldsymbol{s}.$$

For a process observed only a discrete uniform integer grid locations, the integral has to be replaced by a sum. Because every frequency outside $[-\frac{1}{2}, \frac{1}{2}]$ has an alias in $[-\frac{1}{2}, \frac{1}{2}]$, the spectral density is only defined for $\boldsymbol{\omega}$ in $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$.

Then

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(2\pi i \boldsymbol{\omega}' \boldsymbol{s}) \phi_2(\boldsymbol{\omega}; \rho) \, \mathrm{d}\boldsymbol{\omega} = K(\boldsymbol{s}; \rho).$$

By the spectral representation theorem, $g(\boldsymbol{s})$ follows a Normal distribution with mean 0 and correlation $K(\boldsymbol{s}; \rho) = \frac{1}{M_1 M_2} \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} \exp(i 2\pi (\overset{1}{\omega}_{m_1}, \overset{2}{\omega}_{m_2}) \boldsymbol{s}) \phi_2((\overset{1}{\omega}_{m_1}, \overset{2}{\omega}_{m_2}); \rho)$. As $M_1 \to \infty, M_2 \to \infty$, this sum is approximately

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(2\pi i \boldsymbol{\omega}' \boldsymbol{s}) \phi_2(\boldsymbol{\omega}; \rho) \, \mathrm{d}\boldsymbol{\omega}, \text{ which is equal to } K(\boldsymbol{s}; \rho).$$

5.2 Orthogonality of columns of Z, and of $X = 1$ and Z

For $s_{1,1} = 1, s_{1,2} = 2, ..., s_{1,M_1} = M_1$, and $s_{2,1} = 1, s_{2,2} = 2, ..., s_{2,M_2} = M_2$, Z satisfies the properties: (i)$Z'\mathbf{1} = \mathbf{0}$, (ii)$Z'Z = \mathrm{Diag}(2M_1M_2, 2M_1M_2, ..., 2M_1M_2, M_1M_2, M_1M_2, M_1M_2)$.

Proof of (i): We show below that the first element of $Z'\mathbf{1}$ is 0; the other elements are also 0 by similar arguments.

The first element of $Z'\mathbf{1}$ is $A_{11} + A_{21} + ... + A_{M_1M_2,1}$
$= 4 \sum_{\imath=1}^{M_1} \sum_{j=1}^{M_2} \cos(\overset{1}{\omega}_1 2\pi s_{1,\imath} + \overset{2}{\omega}_1 2\pi s_{2,j})$

$= 2\sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2}\left(e^{i(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})}+e^{-i(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})}\right)$

$= 2\sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2} e^{i(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})} + 2\sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2} e^{-i(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})}$

$= 2\sum_{k=1}^{M_1} e^{c_1 k}\sum_{k=1}^{M_2} e^{c_2 k} + 2\sum_{k=1}^{M_1} e^{-c_1 k}\sum_{k=1}^{M_2} e^{-c_2 k},\ c_1 = i\overset{1}{\omega}_1 2\pi,\ c_2 = i\overset{2}{\omega}_1 2\pi$

$= 2\Big(\frac{e^{c_1}(e^{M_1 c_1}-1)}{e^{c_1}-1}\Big)\Big(\frac{e^{c_2}(e^{M_2 c_2}-1)}{e^{c_2}-1}\Big) + 2\Big(\frac{e^{-M_1 c_1}(e^{M_1 c_1}-1)}{e^{c_1}-1}\Big)\Big(\frac{e^{-M_2 c_2}(e^{M_2 c_2}-1)}{e^{c_2}-1}\Big)$

$= 0$, since $e^{M_1 c_1} = \cos(M_1\overset{1}{\omega}_1 2\pi) + i\sin(M_1\overset{1}{\omega}_1 2\pi) = 1+0 = 1$, and $e^{M_2 c_2} = \cos(M_2\overset{2}{\omega}_1 2\pi) + i\sin(M_2\overset{2}{\omega}_1 2\pi) = 1+0 = 1$.

Proof of (ii): We show below that the $(1,1)$ element of $Z'Z$ is $2M_1M_2$, and the $(1,2)$ element of $Z'Z$ is 0; the other elements of $Z'Z$ will be $2M_1M_2$, or M_1M_2, or 0 by similar arguments.

The $(1,1)$ element of $Z'Z$ is $A_{11}^2 + A_{21}^2 + ... + A_{M_1M_2,1}^2$

$= 2\sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2}\Big(\cos(2(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j}))+1\Big)$

$= 2M_1M_2 + 2\sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2}\cos(2(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j}))$

$= 2M_1M_2 + \sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2} e^{i2(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})} + \sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2} e^{-i2(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})}$

$= 2M_1M_2 + \sum_{k=1}^{M_1} e^{c_1 k}\sum_{k=1}^{M_2} e^{c_2 k} + \sum_{k=1}^{M_1} e^{-c_1 k}\sum_{k=1}^{M_2} e^{-c_2 k},\ c_1 = 2i\overset{1}{\omega}_1 2\pi,\ c_2 = 2i\overset{2}{\omega}_1 2\pi$

$= 2M_1M_2 + \Big(\frac{e^{c_1}(e^{M_1 c_1}-1)}{e^{c_1}-1}\Big)\Big(\frac{e^{c_2}(e^{M_2 c_2}-1)}{e^{c_2}-1}\Big) + \Big(\frac{e^{-M_1 c_1}(e^{M_1 c_1}-1)}{e^{-M_1 c_1}-1}\Big)\Big(\frac{e^{-M_2 c_2}(e^{M_2 c_2}-1)}{e^{-M_2 c_2}-1}\Big)$

$= 2M_1M_2$, since $e^{M_1 c_1} = e^{M_1 i\overset{2}{\omega}_1 4\pi} = 1$, and $e^{M_2 c_2} = e^{M_2 i\overset{2}{\omega}_1 4\pi} = 1$.

The $(1,2)$ element of $Z'Z$ is $A_{11}B_{11} + A_{21}B_{21} + ... + A_{M_1M_2,1}B_{M_1M_2,1}$

$= -4\sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2}\cos(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})\sin(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})$

$= -2\sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2}\sin(2(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j}))$

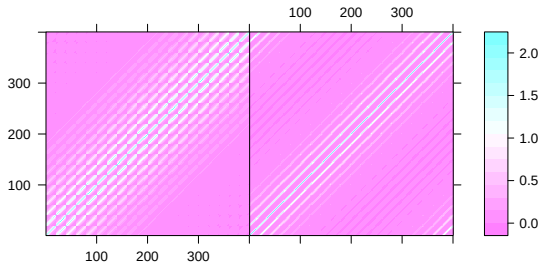
$= -\frac{1}{i}\sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2} e^{i2(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})} - \frac{1}{i}\sum_{\imath=1}^{M_1}\sum_{j=1}^{M_2} e^{-i2(\overset{1}{\omega}_1 2\pi s_{1,\imath}+\overset{2}{\omega}_1 2\pi s_{2,j})}$

$= -\frac{1}{i}\sum_{k=1}^{M_1} e^{ic_1 k}\sum_{k=1}^{M_2} e^{ic_2 k} - \frac{1}{i}\sum_{k=1}^{M_1} e^{-ic_1 k}\sum_{k=1}^{M_2} e^{-ic_2 k}$

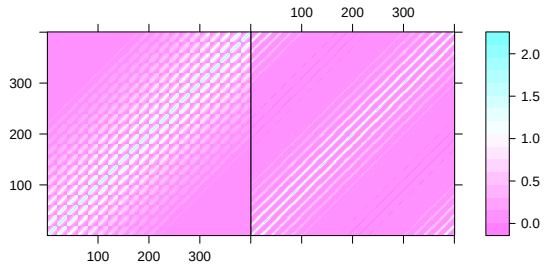
$= -\frac{1}{i}\Big(\frac{e^{c_1}(e^{M_1 c_1}-1)}{e^{c_1}-1}\Big)\Big(\frac{e^{c_2}(e^{M_2 c_2}-1)}{e^{c_2}-1}\Big) - \frac{1}{i}\Big(\frac{e^{-M_1 c_1}(e^{M_1 c_1}-1)}{e^{-M_1 c_1}-1}\Big)\Big(\frac{e^{-M_2 c_2}(e^{M_2 c_2}-1)}{e^{-M_2 c_2}-1}\Big) = 0.$

6 Web Appendix F. Covariance matrices in 2-D

The figures in this appendix show some exact, approximate, and estimated covariance matrices for our model in 2-D with the spectral approximation, referred to in the last paragraph of Web Appendix D.



(a)
Left: exact covariance matrix (including the iid errors). Right: approximate covariance matrix. $\sigma^2_{\mathfrak{s}} = 2, \sigma^2_e = 0.1, \rho = 5$.



(b) Left: Fitted exact covariance matrix using estimates from exact RL. Right: Fitted approximate covariance matrix using estimates from approximate RL.

Figure 20: 2-D GP's covariance matrices.

7 Web Appendix G. Demonstration with simulated data

This section uses a simulated dataset to show how the plot of v_j^2 vs j and added variable plots can be used to examine goodness of fit and potential covariates. We simulate data that depend on covariates, then omit those covariates from the fitted model to see how the missing covariates show themselves in the v_j^2 plot and the added variable plots in the observation and spectral domains.

Consider the data in Figure 21a. Observations were simulated at integer locations on a 20×20 regularly spaced grid from a GP with the Matérn $\nu = 0.5$ correlation function with the GP's variance and range parameters being $\sigma^2_{\mathfrak{s}} = 12$ and $\rho = 5$ and with errors distributed as iid Normal$(0, \sigma^2_e = 5)$. Then two features (covariates) were added as fixed effects to give the simulated data: The first is a north-south linear trend spanning the entire grid; the second feature is addition of the value 12 to ten randomly selected locations. Thus the first and second covariates have low and high frequency respectively.

We fit to the data an intercept-only model with the same correlation function used to simulate the data. This fit gave estimates $\hat{\sigma}^2_{\mathfrak{s}}$, $\hat{\sigma}^2_e$, and $\hat{\rho}$ equal to 21.94, 8.52, and 13.19 respectively. All three estimates are inflated: The omitted low-frequency feature inflates $\hat{\sigma}^2_{\mathfrak{s}}$ and $\hat{\rho}$ and the omitted high-frequency feature inflates $\hat{\sigma}^2_e$. Figure 21b is the v_j^2 plot with j as the plotting character. In this plot, the smooth line denotes the fit of the v_j^2, $\hat{\sigma}^2_{\mathfrak{s}} a_j(\hat{\rho}) + \hat{\sigma}^2_e$, where the estimates are obtained by maximizing the exact restricted likelihood. In Figure 21b, the v_j^2 plot has a prominent point for $j = 1$, indicating a strong north-south linear trend. The observation-domain added variable plot for the first covariate (Figure 21c), only weakly detects this missing covariate and does not have a significant slope (P = 0.18). In contrast, the spectral-domain added variable plot (Figure 21d) gives a strong signal that this covariate belongs in the model (P = 0.0005) and shows the frequencies where the covariate's signal is concentrated.

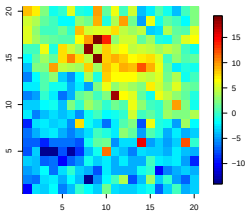
Now consider the second covariate. This missing high-frequency covariate, with no spatial pattern, is not visible in the v_j^2 plot (Figure 21b). However, both added variable plots, in the observation domain (Figure 21e) and in the spectral domain (Figure 21f), show that the covariate should be included (P $< 10^{-15}$ for both), although the former identifies the few observations (outlier locations) that drive the fit while the spectral-domain added variable plot distributes the effect of the outliers diffusely over the v_j corresponding to high frequencies.

Figure 21g is the v_j^2 plot when both covariates have been included in the model. This plot shows no sign of lack of fit; $\hat{\sigma}^2_{\mathfrak{s}}$, $\hat{\sigma}^2_e$ and $\hat{\rho}$ are now 11.98, 4.03, and 5.28 respectively, close to the values used to simulate the data.

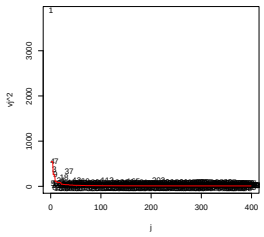
We have shown how a missing high-frequency covariate, i.e., outliers in y, can be detected from a observation-domain added variable plot or a spectral-domain added variable plot for that covariate, but is not visible in the v_j^2 plot. On the other hand, missing large-scale trends, i.e., non-stationarity in the form of a linear or quadratic trend, can be visible in the v_j^2 plot. Once detected, these latter trends may be included as covariates, and if a covariate is available that captures such trends, then an added variable plot for that covariate should have a large slope. The added variable plots in the two domains estimate the same slope for the candidate covariate but have different power for

testing the slope. For low-frequency trends, the spectral-domain added variable plot appears to have more power than the observation-domain plot.

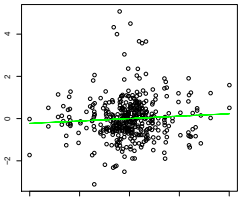
For each kind of missing covariate considered above — which were pure types used for demonstration — one kind of added variable plot shows how the signal in the data is concentrated in some y_i or v_j. In real datasets, a potential covariate may be a mix of low- and high-frequency features.



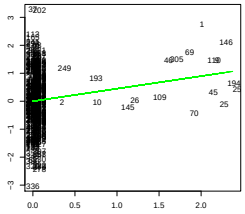
(a) Simulated data with covariates.



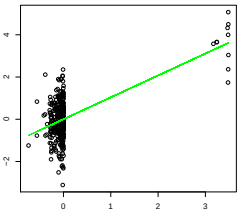
(b) v_j^2 vs j for intercept-only fit.



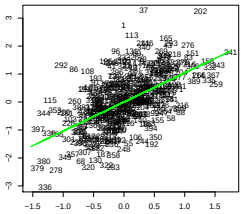
(c) Added variable plot in observation domain for low frequency covariate.



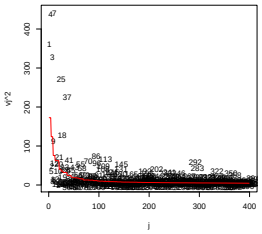
(d) Added variable plot in spectral domain for low frequency covariate.



(e) Added variable plot in observation domain for high frequency covariate.



(f) Added variable plot in spectral domain for high frequency covariate.



(g) v_j^2 vs j for fit with both covariates.

Figure 21: Plots for the analysis of the simulated data.

8 Web Appendix H. Model building in the spectral domain

The main paper's Section 7 showed the first of a sequence of model-building steps, illustrating how to use the tools. This appendix and Appendices I and J show the full sequence of steps considering both the spectral and observation domains. Again, we make no claim that this is an optimal sequence of steps; we merely intend to illustrate how the tools can be used.

Step 1: First, fit the intercept-only model. Consider the resulting plot of the v_j^2 (Figure

22a) in which the plotting symbol is j, corresponding to a particular frequency. For example, the point numbered 1 corresponds to a cubic or linear north-south trend in the data (the spectral approximation has no linear-like component), so the prominent v_1^2 indicates non-stationarity likely in the form of a north-south trend. This deviation from stationarity may be addressed by including an appropriate covariate in the model, so we draw spectral-domain added variable plots for all the covariates (Figure 22). In the added variable plots, the colors are: black for j in 1 to 100, red for j in 101 to 200, green for j in 201 to 300, blue for j in 301 to 400, pink for j in 401 to 559. Each covariate (after being pre-smoothed onto the regular grid) has been standardized by subtracting its sample mean and dividing by its sample standard deviation, so the slopes for the covariates are comparable. The p-values for the added variable plot slopes suggest that the fit can be improved by adding covariates (Table 4). The natural impulse is to include first the covariate having the largest slope in its added variable plot, in this case Elevation. However, since the v_j^2 plot for the data has v_1^2 as the most prominent point, corresponding to a north-south trend, we also want the added covariate to explain at least some of this trend, the more the better. For this purpose we calculate Cook's distance for points in the added variable plot, describing their respective influence on the regression slope; we want the point numbered 1 to have one of the largest Cook's distances. Among the potential covariates, Elevation has the smallest p-value for the added variable plot's slope, and point number 1 has the largest Cook's distance for the coefficient of Elevation, i.e., it is most highly influential in determining the slope of Elevation's added variable plot. Therefore we add Elevation as a covariate.

Slope has results similar to Elevation's, so we could have added Slope to the model at this stage. If we had, it turns out that applying the same considerations we would add Elevation in the second step, so it does not matter if we add Elevation or Slope first.

Step 2: Now fit the model with one covariate, Elevation. In Figure 23a's plot of the v_j^2, the v_j^2 for low j (low frequences) are not as striking as for the intercept-only model, but some are outstanding, suggesting deviation from stationarity; it seems the model still attributes some low-frequency data components to the GP and perhaps they can be captured using explicit covariates instead. The

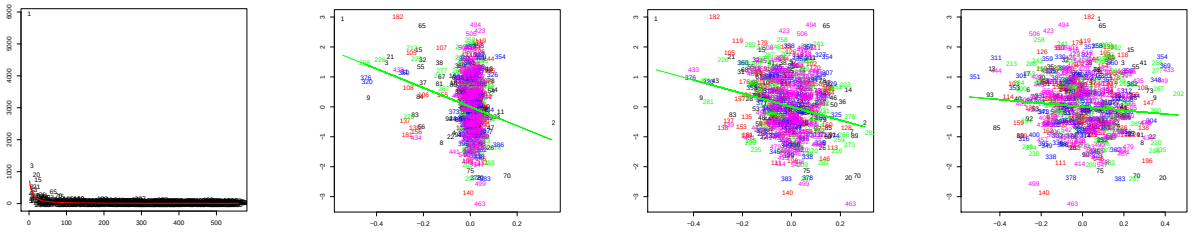
spectral-domain added variable plots (Figure 23) have large slopes for several candidate covariates (Table 5). Because v_8^2 is the largest v_j^2, we prefer to include a covariate for which the added variable plot's slope is large and which is also influenced substantially by the point $j = 8$, corresponding to a quartic (4^{th} degree polynomial) trend. The covariate SummerTC1 has the smallest p-value for its slope, but the point $j = 8$ does not have a large Cook's distance for SummerTC1. The point $j = 8$ does have a large Cook's distance for the covariate Slope. Thus, in this step we add both covariates, SummerTC1 and Slope, to the model that includes Elevation.

Step 3: Now fit the model including Elevation, Slope, and SummerTC1. The v_j^2 plot again shows low-frequency trends suggesting deviation from stationarity (Figure 24a); the point $j = 4$, corresponding to a quadratic north-south trend, is most prominent. Added variable plots for the remaining covariates (Figures 24) show that SpringTC2 has the smallest p-value but $j = 4$ is not particularly influential for its slope (Table 6). The indication of non-stationarity cannot be explained by SpringTC2, so we add a quadratic north-south trend as a covariate along with SpringTC2. This is tolerable because a quadratic tend has very low frequency; were we collaborating with subject matter experts, such a simple trend might suggest a potential covariate that could be used instead of the non-substantive quadratic.

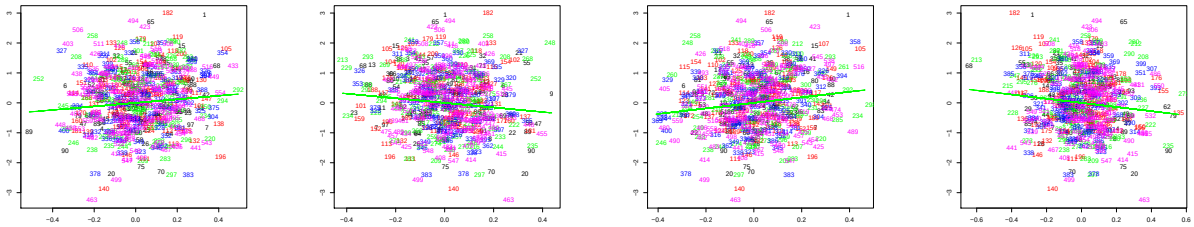
An alternative may be to add FallTC2 instead of the north-south quadratic trend. If we do that, the point $j = 4$ comes down in the result v_j^2 plot but is still prominent (Figure 25), so we add the north-south quadratic trend instead of FallTC2.

Step 4: Fit the model to the smoothed data including Elevation, Slope, SummerTC1, SpringTC2, and the north-south quadratic trend. The v_j^2 plot now has no prominent points (Figure 26a). Figure 26 shows added variable plots for the remaining covariates; based on the p-values (Table 7), we add the covariate SummerTC3.

Step 5: The best-fitting model includes Elevation, Slope, SummerTC1, SpringTC2, a north-south quadratic trend, and SummerTC3. The v_j^2 plot has no prominent points (Figure 27a) and the remaining covariates appear unable to absorb variation currently relegated to the GP or error parts of the model (Table 8, Figure 27).



(a) v_j^2 vs j. (b) Elevation. (c) Slope. (d) SpringTC2.

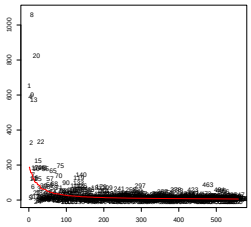


(e) SpringTC3. (f) SummerTC1. (g) SummerTC3. (h) FallTC2.

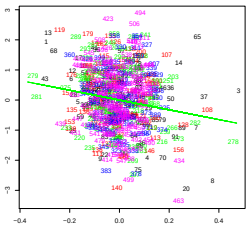
Figure 22: **Step 1** After intercept-only fit; (b) to (h) are spectral-domain added variable plots.

Candidate covariates	slope	p-value	j with top 5 Cook's dist
Elevation	-3.17	10^{-10}	1,182,9,181,434
Slope	-2.24	10^{-9}	1,182,463,70,65
SpringTC2	-0.60	0.03	20,499,354,268,369
SpringTC3	0.59	0.02	1,506,196,403,463
SummerTC1	-0.77	0.007	248,463,20,268,327
SummerTC3	0.92	0.0004	1,258,378,358,248
FallTC2	-0.69	0.004	182,463,1,212,280

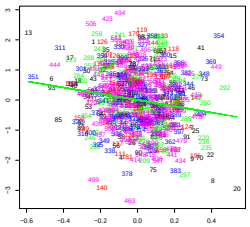
Table 4: **Step 1** Slopes of spectral-domain added variable plots iafter the intercept-only fit. The covariate chosen to be added at the next step is in bold.



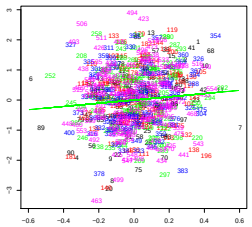
(a) v_j^2 vs j.



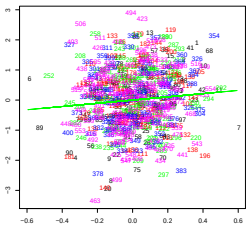
(b) Slope.



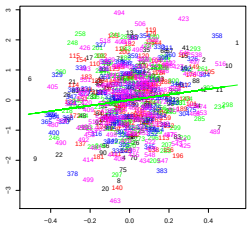
(c) SpringTC2.



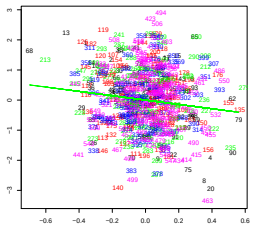
(d) SpringTC3.



(e) SummerTC1.



(f) SummerTC3.

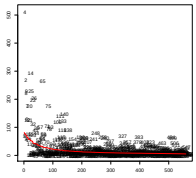


(g) FallTC2.

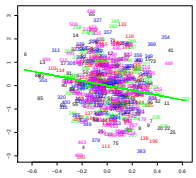
Figure 23: **Step 2** After fit with covariate Elevation; (b) to (g) are spectral-domain added variable plots.

Candidate covariates	slope	p-value	j with top 5 Cook's dist
Elevation	–	–	–
Slope	-1.63	10^{-5}	65,8,20,463,3
SpringTC2	-1.04	10^{-5}	20,354,8,13,499
SpringTC3	0.54	0.03	506,196,463,327,403
SummerTC1	-1.27	10^{-6}	20,248,463,9,327
SummerTC3	0.92	0.0003	9,1,378,378,358,489
FallTC2	-0.72	0.002	463,20,441,8,13

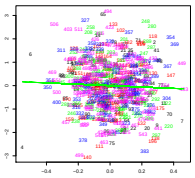
Table 5: **Step 2** Slopes of spectral-domain added variable plots after the fit with covariate Elevation. The covariates chosen to be added at the next step are in bold. The symbol "–" denotes covariates already in the model.



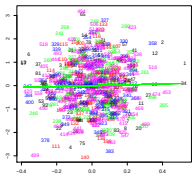
(a) v_j^2 vs j.



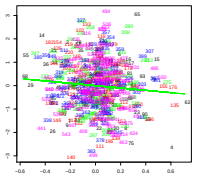
(b) SpringTC2.



(c) SpringTC3.



(d) SummerTC3.



(e) FallTC2.

Figure 24: **Step 3** After the fit with covariate Elevation, Slope, SummerTC1; (b) to (e) are spectral-domain added variable plots.

Candidate covariates	slope	p-value	pts with top 5 Cook's dist
Elevation	–	–	–
Slope	–	–	–
SpringTC2	-1.02	10^{-5}	354,41,499,25,369
SpringTC3	-0.29	0.29	4,506,354,369,280
SummerTC1	–	–	–
SummerTC3	0.18	0.57	499,2,378,111,358
FallTC2	-0.49	0.03	4,65,441,14,26

Table 6: **Step 3** Slopes of spectral-domain added variable plots after the fit with covariates Elevation, Slope, SummerTC1. The covariate chosen to be added at the next step is in bold. The symbol "–" denotes covariates already in the model.

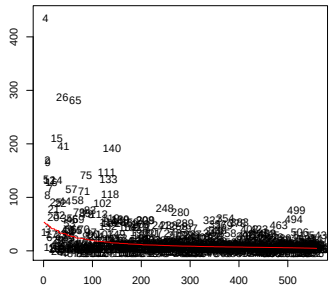
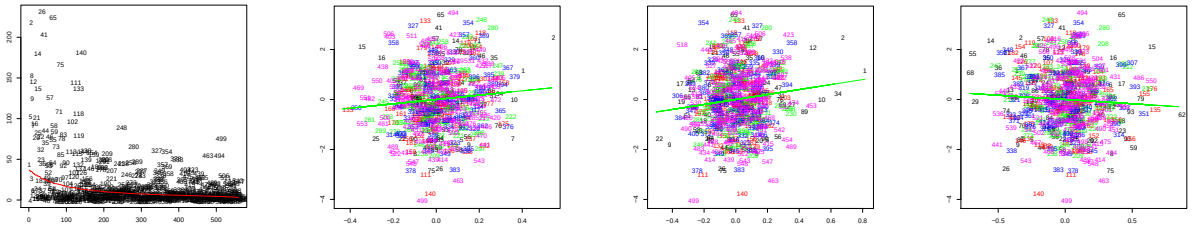


Figure 25: v_j^2 vs j for the fit with covariates Slope, Elevation, SpringTC2, SummerTC1, and FallTC2.

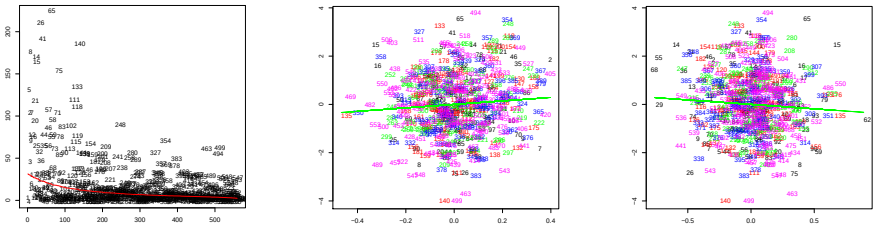


(a) v_j^2 vs j. (b) SpringTC3. (c) SummerTC3. (d) FallTC2.

Figure 26: **Step 4** After the fit with covariates Elevation, Slope, SummerTC1, SpringTC2 and north-south quadratic trend; (b) to (d) are spectral-domain added variable plots.

Candidate covariates	slope	p-value	pts with top 5 Cook's dist
Elevation	–	–	–
Slope	–	–	–
SpringTC2	–	–	–
SpringTC3	0.87	0.02	2,15,280,506,7
SummerTC1	–	–	–
SummerTC3	0.99	0.002	2,499,518,489,12
FallTC2	-0.34	0.15	65,15,26,14,55

Table 7: **Step 4** Slopes of spectral-domain added variable plots after the fit with covariates Slope, Elevation, SpringTC2, SummerTC1 and north-south quadratic trend. The covariate chosen to be added at the next step is in bold. The symbol "–" denotes covariates already in the model.



(a) v_j^2 vs j. (b) SpringTC3. (c) FallTC2.

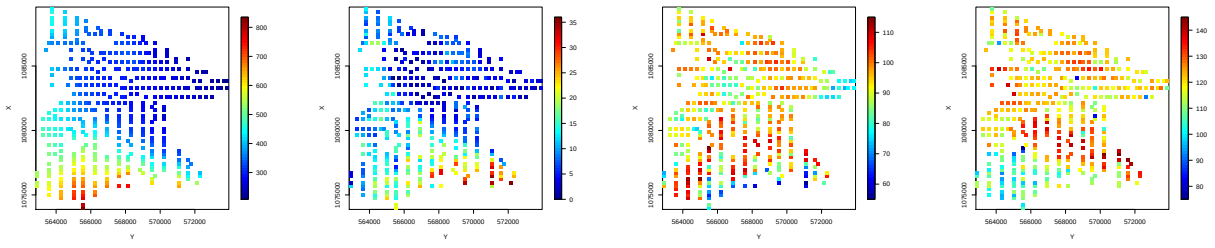
Figure 27: **Step 5** After the fit with covariates Elevation, Slope, SummerTC1, SpringTC2, north-south quadratic trend, and SummerTC3; (b) to (c) are spectral-domain added variable plots.

Candidate covariates	slope	p-value	pts with top 5 Cook's dist
Elevation	–	–	–
Slope	–	–	–
SpringTC2	–	–	–
SpringTC3	0.72	0.06	15,506,403,489,7
SummerTC1	–	–	–
SummerTC3	–	–	–
FallTC2	-0.37	0.11	15,65,26,14,55

Table 8: **Step 5** Slopes of spectral-domain added variable plots after the fit with covariates Slope, Elevation, SpringTC2, SummerTC1, north-south quadratic trend, and SummerTC3. The symbol "–" denotes covariates already in the model, in addition a north-south quadratic trend is also in the model.

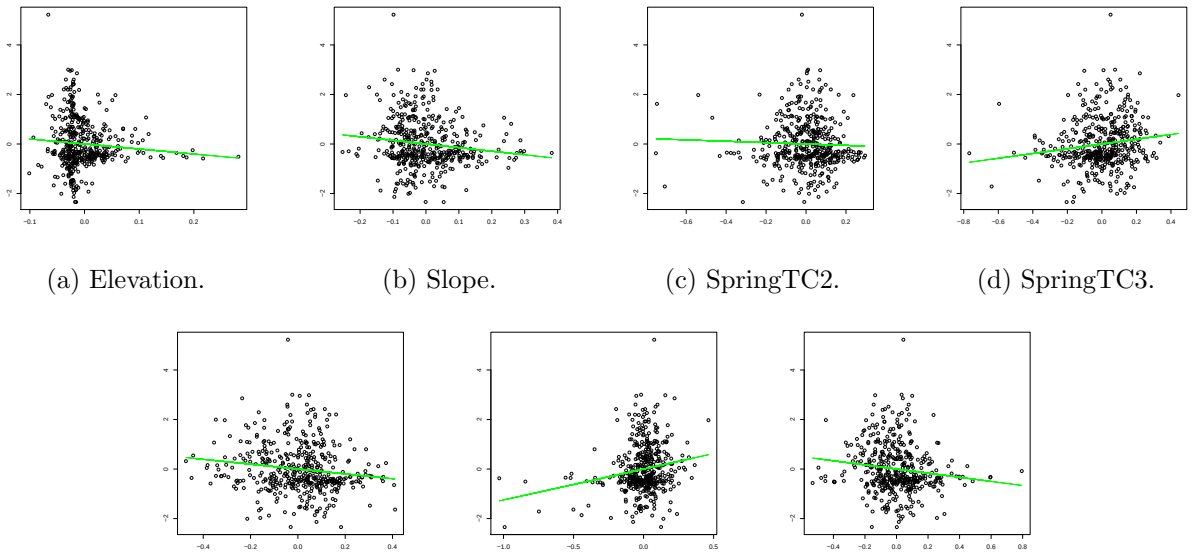
9 Web Appendix I. Observation-domain added variable plots

This Appendix shows the observation-domain added variable plots, with five steps adding covariates in the same order as they were added when considering the spectral domain in Appendix H. For each step, a figure and table below summarize the step's observation-domain added variable plots; following are comments comparing these added variable plots to those in the spectral domain. Elevation and Slope have strong low-frequency components, as shown in plots of these covariates in Figures 28a and 28b). Thus, their signal is weaker in the observation-domain added variable plots for the intercept-only model (Figure 29) than in their earlier spectral-domain added variable plots (Figure 22); compare slopes of spectral-domain added variable plots in Step 1, Table 4 with the observation-domain analogs, Table 9. Compare also the spectral-domain added variable plots for Step 2, Table 5 with the observation-domain analogs, Table 10. The p-values for Elevation and Slope are much smaller in the spectral-domain added variable plots than in the observation-domain plots; because these two covariates have relatively large power at low frequencies, they are more easily detectable in the spectral domain. SpringTC2 also has relatively large power in low frequencies so it behaves similarly: the spectral-domain added variable plot has greater power to detect its slope (compare the spectral-domain added variable plots in Step 3, Table 6 with the observation-domain analogs, Table 11). SummerTC1 has both low- and high-frequency features so it gives a rather strong signal in the added variable plots in both domains: compare the spectral-domain added variable plots in Step 2, Table 5 with the observation domain analogs, Table 10.



(a) Elevation. (b) Slope. (c) SummerTC1. (d) SpringTC2.

Figure 28: Plots of four of the covariates.



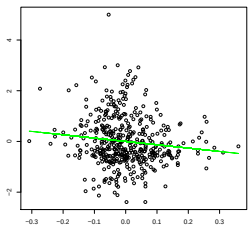
(a) Elevation. (b) Slope. (c) SpringTC2. (d) SpringTC3.

(e) SummerTC1. (f) SummerTC3. (g) FallTC2.

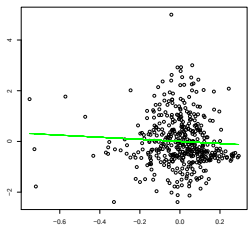
Figure 29: **Step 1** Observation-domain added variable plots after the intercept-only fit.

Candidate covariates	slope	p-value
Elevation	-2.02	0.07
Slope	-1.45	0.004
SpringTC2	-0.28	0.42
SpringTC3	0.96	0.002
SummerTC1	-0.98	0.002
SummerTC3	1.25	10^{-5}
FallTC2	-0.83	0.004

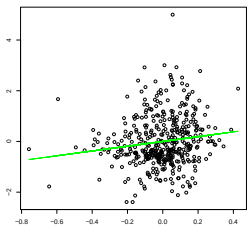
Table 9: **Step 1** Slopes of observation-domain added variable plots after the intercept-only fit. Covariate selected to be added on the basis of these plots is in bold.



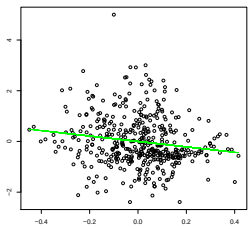
(a) Slope.



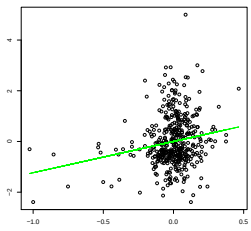
(b) SpringTC2.



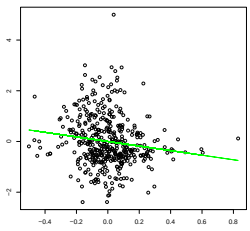
(c) SpringTC3.



(d) SummerTC1.



(e) SummerTC3.

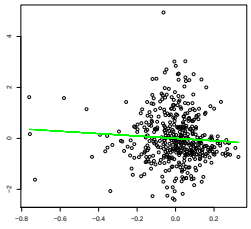


(f) FallTC2.

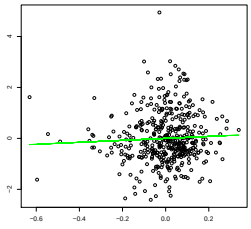
Figure 30: **Step 2** Observation-domain added variable plots after the fit with covariate Elevation.

Candidate covariates	slope	p-value
↓		
Elevation	–	–
Slope	-1.31	0.01
SpringTC2	-0.41	0.23
SpringTC3	0.94	0.002
SummerTC1	-1.07	0.0008
SummerTC3	1.23	10^{-5}
FallTC2	-0.90	0.002

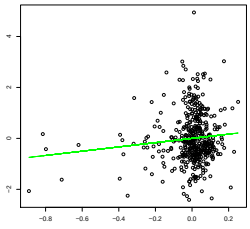
Table 10: **Step 2** Slopes of observation-domain added variable plots after the fit with Elevation. covariates selected to be added on the basis of these plots are in bold. The symbol "–" denotes covariates already in the model.



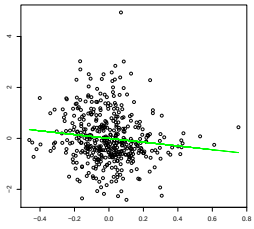
(a) SpringTC2.



(b) SpringTC3.



(c) SummerTC3.

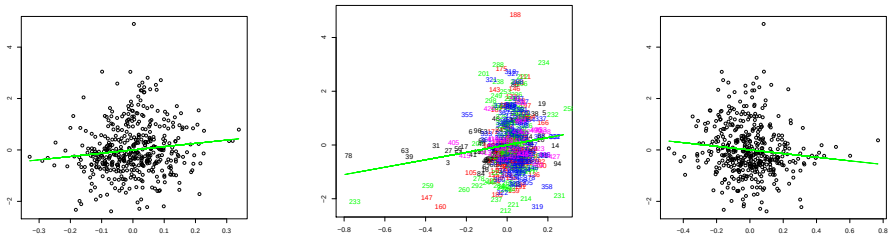


(d) FallTC2.

Figure 31: **Step 3** Observation-domain added variable plots after the fit with covariates Elevation, Slope, and SummerTC1.

Candidate covariates	slope	p-value
↓		
Elevation	–	–
Slope	–	–
SpringTC2	-0.47	0.16
SpringTC3	0.39	0.30
SummerTC1	–	–
SummerTC3	0.84	0.03
FallTC2	-0.74	0.01

Table 11: **Step 3** Slopes of observation-domain added variable plots after the fit with Elevation, Slope, and SummerTC1. covariate selected to be added on the basis of these plots is in bold. The symbol "–" denotes covariates already in the model.

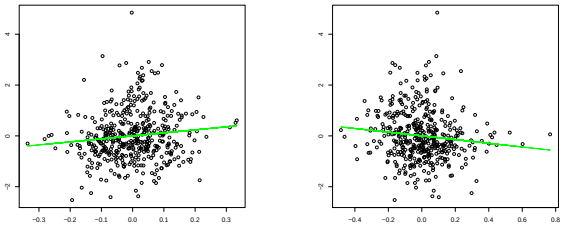


(a) SpringTC3. (b) SummerTC3. (c) FallTC2.

Figure 32: **Step 4** Observation-domain added variable plots after the fit with covariates Elevation, Slope, SummerTC1, SpringTC2, and north-south quadratic trend.

Candidate covariates	slope	p-value
↓		
Elevation	–	–
Slope	–	–
SpringTC2	–	–
SpringTC3	1.28	0.009
SummerTC1	–	–
SummerTC3	1.38	0.001
FallTC2	-0.71	0.02

Table 12: **Step 4** Slopes of observation-domain added variable plots after the fit with covariates Elevation, Slope, SummerTC1, SpringTC2, and north-south quadratic trend. covariate selected to be added on the basis of these plots is in bold. The symbol "–" denotes covariates already in the model.



(a) SpringTC3. (b) FallTC2.

Figure 33: **Step 5** Observation-domain added variable plots after the fit with covariates Elevation, Slope, SummerTC1, SpringTC2, SummerTC3, and north-south quadratic trend.

Candidate covariates	slope	p-value
↓		
Elevation	–	–
Slope	–	–
SpringTC2	–	–
SpringTC3	1.16	0.02
SummerTC1	–	–
SummerTC3	–	–
FallTC2	-0.73	0.01

Table 13: **Step 5** Slopes of observation-domain added variable plots after the fit with covariates Elevation, Slope, SummerTC1, SpringTC2, SummerTC3, and north-south quadratic trend. The symbol "–" denotes covariates already in the model.

10 Web Appendix J. Fits of the raw data using the exact restricted likelihood

The spectral-domain added variable plots used the data smoothed on a grid to select covariates. This appendix shows the improvement of the fit to the actual (not smoothed) data, when fits are made using the exact restricted likelihood. We repeat: this Appendix involves no approximations of any kind. Figure 34 shows the actual data, while Figure 35 shows the estimated fixed-effect fits at each of the five steps discussed above, on the same color scale as the observed data. Figure 36 shows the residuals obtained by subtracting the estimated fixed effects from the data at each step. Table 14 shows quantiles of absolute values of the residuals with respect to the fixed effects, which become somewhat smaller as we add covariates, with most of the change arising from adding

Elevation, Slope, and SummerTC1. Changes in the residuals are small and thus not very visible in the residual plots (Figure 36). However, it is clear from Table 14 that the fit to the data has indeed improved.

Tables 15 and 16 contain estimated coefficients of the covariates when they are added to the model and the model is re-fit at each step. The estimated coefficients are almost exactly equal to the slopes of the corresponding *observation*-domain added variable plots in Appendix I. However the p-values from ordinary Wald tests of these coefficients (taking as known the estimates of $\sigma^2_{\mathfrak{s}}$, σ^2_e, and ρ, as is typical in non-Bayesian software) are not similar to the p-values for significance of the added variable plot slopes in either the observation or spectral domain. Table 17 shows the estimates of $\sigma^2_{\mathfrak{s}}$, σ^2_e, and ρ at each step. The covariates Elevation, Slope, and Summer TC1 have the largest effect on $\hat{\sigma}^2_{\mathfrak{s}}$, reducing it by almost half, as we might expect given the relatively strong low-frequency components in these covariates. Adding Elevation to the model had the largest effect on $\hat{\rho}$, with modest reductions from adding further covariates, and adding Elevation produced most of the reduction in $\hat{\sigma}^2_e$.

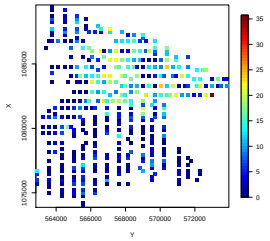
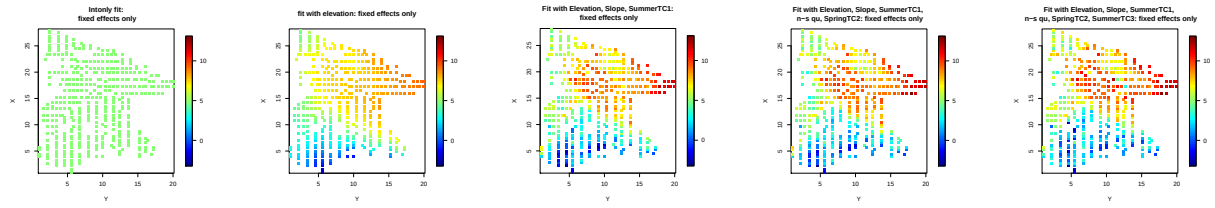


Figure 34: Observed data on the observation domain.



(a) Intercept-only fit.

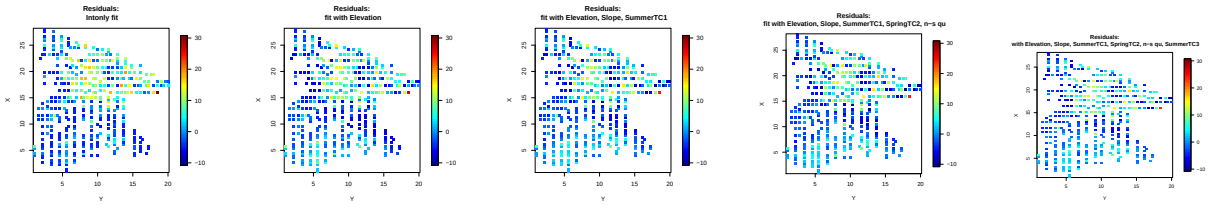
(b) Fit with Elevation.

(c) Fit with Elevation, Slope, SummerTC1.

(d) Fit with Elev, Slope, SumTC1, SprTC2, north-south quad trend.

(e) Fit with Elev, Slope, SumTC1, SprTC2, n-s quad trend, SumTC3.

Figure 35: Estimated fixed effects part of the fit for each of the five steps.



(a) Residuals from intercept-only fit.

(b) Residuals from fit with Elevation.

(c) Residuals from fit with Elevation, Slope, SummerTC1.

(d) Residuals from fit with Elev, Slope, SumTC1, SprTC2, north-south quad trend.

(e) Residuals from fit with Elev, Slope, SumTC1, SprTC2, n-s quad trend, SumTC3.

Figure 36: Residuals, i.e., data minus estimated fixed effects, for each of the five steps.

covariates in model	minimum	25%	50%	75%	maximum
Step 1: Intercept-only	0.01	2.69	4.95	5.40	30.81
Step 2: Elevation	0.00	2.10	4.20	6.70	26.79
Step 3: Elev, Slope, SumTC1	0.00	1.76	3.93	5.96	24.48
Step 4: Elev, Slope, SumTC1, SprTC2, n-s quad trend	0.03	1.71	3.77	5.91	24.09
Step 5: Elev, Slope, SumTC1, SprTC2, n-s quad, SumTC3	0.00	1.68	3.63	5.84	23.73

Table 14: Quantiles of absolute residuals, i.e., data minus estimated fixed effects.

	Step 1	**Step 1**	**Step 2**	**Step 2**
Candidate covariates	slope	p-value	slope	p-value
Elevation	-2.52	$< 10^{-12}$	–	–
Slope	-1.63	10^{-6}	-1.37	10^{-5}
SpringTC2	-0.28	0.16	-0.44	0.04
SpringTC3	0.99	10^{-5}	0.97	10^{-5}
SummerTC1	-1.03	10^{-5}	-1.09	10^{-6}
SummerTC3	1.25	10^{-7}	1.23	10^{-7}
FallTC2	-0.87	0.0001	-0.94	10^{-5}

Table 15: Coefficients for the covariates obtained by putting them into the model, and p-values from Wald tests. The symbol "–" denotes covariate already in the model.

	Step 3	**Step 3**	**Step 4**	**Step 4**	**Step 5**	**Step 5**
Candidate covariates	slope	p-value	slope	p-value	slope	p-value
Elevation	–	–	–	–	–	–
Slope	–	–	–	–	–	–
SpringTC2	-0.49	0.03	–	–	–	–
SpringTC3	0.39	0.10	1.30	0.0005	1.19	0.0008
SummerTC1	–	–	–	–	–	–
SummerTC3	0.84	0.003	1.39	10^{-6}	–	–
FallTC2	-0.76	0.0007	-0.72	0.001	-0.74	0.0005

Table 16: Coefficients for the covariates obtained by putting them into the model, and p-values from Wald tests. The symbol "–" denotes covariates already in the model.

covariates in the model	$\sigma^2_{\mathfrak{s}}$	σ^2_e	ρ
Step 1: Intercept-only	29.62	16.20	5.96
Step 2: Elev	21.96	13.82	2.85
Step 3: Elev, Slope, SumTC1	15.98	15.13	2.65
Step 4: Elev, Slope, SumTC1, SprTC2, north-south quad trend	16.04	14.37	2.28
Step 5: Elev, Slope, SumTC1, SprTC2, SumTC1, n-s quad, SumTC3	16.35	12.89	1.92

Table 17: Estimates of $\sigma^2_{\mathfrak{s}}$, σ^2_e, ρ using the exact restricted likelihood at each step.

11 Web Appendix K. Approximation Effects and Accuracy

This appendix focuses on the effects of smoothing data observed at irregular locations to a regular grid using IDW, and approximating the restricted likelihood (RL) using the spectral approximation to the GP-distributed random effect. The appendix mainly summarizes a simulation experiment examining the effect of smoothing data to the grid and the combined effect of smoothing to a grid and the spectral approximation. Subsection 11.1 discusses the intuition about IDW that motivated the simulation experiment. The spectral approximation has a large literature, which we do not attempt to summarize. Subsection 11.2 gives and interprets the simulation experiment results.

11.1 Intuition about smoothing to the grid using IDW

In the pseudo-data for a given grid point, IDW gives each actual observation a weight that is inversely proportional to the distance from the grid point to that observation's location, raised to

the power λ. A small λ gives weight to many observations and thus produces smooth pseudo-data; as λ grows, weight is concentrated on fewer observations and with very large λ, most grid points are effectively assigned the nearest observation.

The clearest intuition is about smoothing's effect on the estimated error variance, $\hat{\sigma}_e^2$: in general, $\hat{\sigma}_e^2$ will be biased downward because the weighted averaging at grid points dampens local or high-frequency variation. The power λ affects this most directly: small λ suppresses local or high-frequency variation more than large λ, so small λ should produce greater bias than large λ. As for grid density (M_1 and M_2), the downward bias of $\hat{\sigma}_e^2$ should be greatest at the extremes of very coarse and very dense grids even with large λ, though for different reasons at the two extremes. For coarse grids, with considerably fewer grid points than observations, each grid point's weighted average will be affected only by the observations closest to it; as the grid becomes coarser, the most discrepant observations are more likely to influence no grid points, so the downward bias should worsen as the grid coarsens. For dense grids, with the number of grid points approaching or greater than the number of observations, more and more individual observations will be used for more than one grid point; this repeated use will make the grid-smoothed data less variable than the actual data, so again the downward bias should worsen as the grid becomes increasingly dense. This suggests a moderate grid density is best for bias in $\hat{\sigma}_e^2$. Finally, many real datasets have measurements on non-rectangular regions, so the rectangular grid will have grid points in regions of the map that are empty of observations. Pseudo-data for such grid points are necessarily averages of observations at or very near the edge of the empty regions so they will tend to be less variable than the actual data because of repeated use of the same observations. Thus, larger empty regions should tend to have more downward bias in $\hat{\sigma}_e^2$.

Intuition about the effects of grid smoothing on the estimate of the GP's range, $\hat{\rho}$, largely follows from intuition about $\hat{\sigma}_e^2$. Smoothing to a grid should have little effect on low-frequency features of the data but will suppress power at high frequencies, so that power will tend to decline as frequency increases *faster* in the grid smoothed data than in the actual data. Using intuition about estimates based on the spectral approximation, because $\hat{\rho}$ is determined by this decline, $\hat{\rho}$ will tend to be

biased high in grid-smoothed data. As for the GP variance $\sigma^2_{\mathfrak{s}}$, in the same scheme of intuition, its estimate adjusts $\sigma^2_{\mathfrak{s}} a_j(\rho) + \sigma^2_e$ to "go through the middle of" the v^2_j for small j, and it is not clear that the aforementioned effects will induce any consistent effect on $\hat{\sigma}^2_{\mathfrak{s}}$.

11.2 Simulation experiment

In most simulation experiments done by statisticians, the purpose is to precisely estimate operating characteristics (e.g., type I error), so the experimental designs consider few conditions (simulation scenarios) and simulate many artificial datasets for each. Our purpose is different — to examine trends in estimates over many IDW settings — so our experimental design is like one we would recommend to our collaborators, with relatively few simulated datasets. (Our collaborators rarely need designs with 1000 replications per design cell.) Specifically, we simulated 20 datasets with observations at locations in the unit square; from each simulated dataset, we created 3 analysis datasets differing in the size of the empty region on the spatial map; for each analysis dataset, we considered 5 grid sizes; and for each grid size, we considered three values of the IDW smoothing parameter λ. The experimental design is thus entirely within-subject, where a "subject" is a simulated dataset (20 levels), and the factors are amount of blank space (3 levels), and estimator (16 levels, the estimates from maximizing the exact RL on the actual data plus 15 estimates for data smoothed to a grid, where 15 = 5 grid sizes $\times$ 3 λ values). Analysis of variance gave $p < 0.05$ for many effects, i.e., adequate power to detect trends of the sort described in the preceding section; below, we focus on broad trends and do not report significance tests.

Experimental settings. Twenty datasets were simulated, each with observations at 400 locations in the unit square, iid draws from a uniform distribution. The same locations were used for all 20 datasets. Simulated observations were drawn from an intercept-only linear mixed model with true $\beta = 0$ and true $\sigma^2_{\mathfrak{s}} = 12$, $\rho = 0.1$ (i.e., 1/10 of each dimension of the unit square), and $\sigma^2_e = 5$. Amount of blank space. Blank space was created by omitting observations at locations within specified regions. The levels of this factor were: no blank space; 1/8 of the unit square blank; and 1/4 of the unit square blank. The omitted regions were wedges opening upward with apex at the

point (0.5, 0.5), specifically: to omit 1/8 of the unit square, omit points having $s_2 > -2s_1 + 1.5$ and $s_2 > 2s_1 - 0.5$; to omit 1/4 of the unit square, omit points having $s_2 > -s_1 + 1$ and $s_2 > s_1$. Grid size. All grids were $M \times M$, where M was 12, 14, 16, 18, or 20. With no omitted data, a simulated dataset had 400 observations, so these grids had, respectively, 36%, 49%, 64%, 81%, and 100% as many points. Smoothing parameter λ. We considered values of 5, 10, and 100. Preliminary experiments considered $\lambda = 1$ with catastrophic results. The analyses in the main paper's Section 7 used $\lambda = 7$ for y and $\lambda = 9$ for candidate covariates by minimizing the sum over $\sigma^2_{\mathfrak{s}}$, ρ, and σ^2_e of the squared relative difference between the estimate obtained by maximizing the exact RL for the real data and the data smoothed to the grid.

For each of the 60 analysis datasets (20 simulated datasets $\times$ 3 levels of blank space), we produced estimates of $(\sigma^2_{\mathfrak{s}}, \rho, \sigma^2_e)$ by (a) maximizing the exact RL using the "real" data; then (b) maximizing the exact RL using pseudo-data smoothed to the grid; then (c) maximizing the approximate RL, based on the spectral approximation to the GP, using pseudo-data smoothed to the grid. Comparing the estimates from (b) to the exact estimates (a) isolates the effect of smoothing the data to a grid; comparing the estimates from (c) to the exact estimates (a) shows the combined effect of smoothing the data to a grid and using the spectral approximation. (We do not recommend obtaining estimates by maximizing the approximate restricted likelihood; this is simply a compact way to capture the effect of smoothing to the grid followed by the spectral approximation.) The following sections present these two comparisons; for both comparisons, preliminary analyses indicated that the logarithms of the estimates were more appropriate dependent variables for ANOVA than untransformed values.

Effect of smoothing to a grid.

Figure 37 shows the estimates obtained by maximizing the exact likelihood for the actual data ("Exact" on the horizontal axis) or the pseudo-data smoothed to the grid (labelled as "M/λ" on the horizontal axis). Figures 37 and 38 have the same vertical axes to facilitate comparison.

Broadly, biases in the estimates can be summarized as follows.

- Error variance: Estimates are biased downward, least so for moderate M and large λ though even for these settings the bias is substantial. More empty space makes the bias worse for denser grids.

- GP variance: Estimates are affected much less than error variance. The bias has no consistent direction of bias but is sensitive to both M and λ.

- GP range ρ: Estimates are generally biased high, more so for coarser grids (small M) and for more omitted data.

Regarding choices for the IDW procedure:

- Grid size: Moderate M is the best compromise: it is best for the error and GP variance, while large M is best for ρ though not by much.

- Lambda: 100 is best for error variance; for the GP variance, the best value depends on M, with moderate λ best for moderate M; and for ρ, λ has little effect.

Combined effect of smoothing to a grid and using the approximate RL.

Figure 38 shows estimates obtained by maximizing the exact RL for the actual data ("Exact" on the horizontal axis) and the approximate RL for pseudo-data smoothed to the grid (labelled "M/λ" on the horizontal axis). Figures 38 and 37 have the same vertical axes to facilitate comparison.

For data smoothed to the grid, the estimate of error variance is markedly less biased and less sensitive to the IDW settings when obtained by maximizing the approximate RL than when obtained by maximizing the exact RL. Coarser grid sizes M give modestly less bias compared to denser grids. This may depend on the true parameter values used to simulate the data; further experiments would be required to determine this. Compared to estimates obtained using the exact RL, estimates of GP variance and ρ are more biased on average using the approximate RL and also more variable, as reflected in the standard errors associated with each estimate, listed in the Figure captions. (An outlier strongly affects the average estimates of ($\sigma^2_{\mathfrak{s}}$ and ρ for $M = 20$, $\lambda = 5$). Estimates of the GP variance are generally biased high but show no particular dependence on

the IDW settings, while estimates of ρ are generally biased downward but again show no particular dependence on the IDW settings.

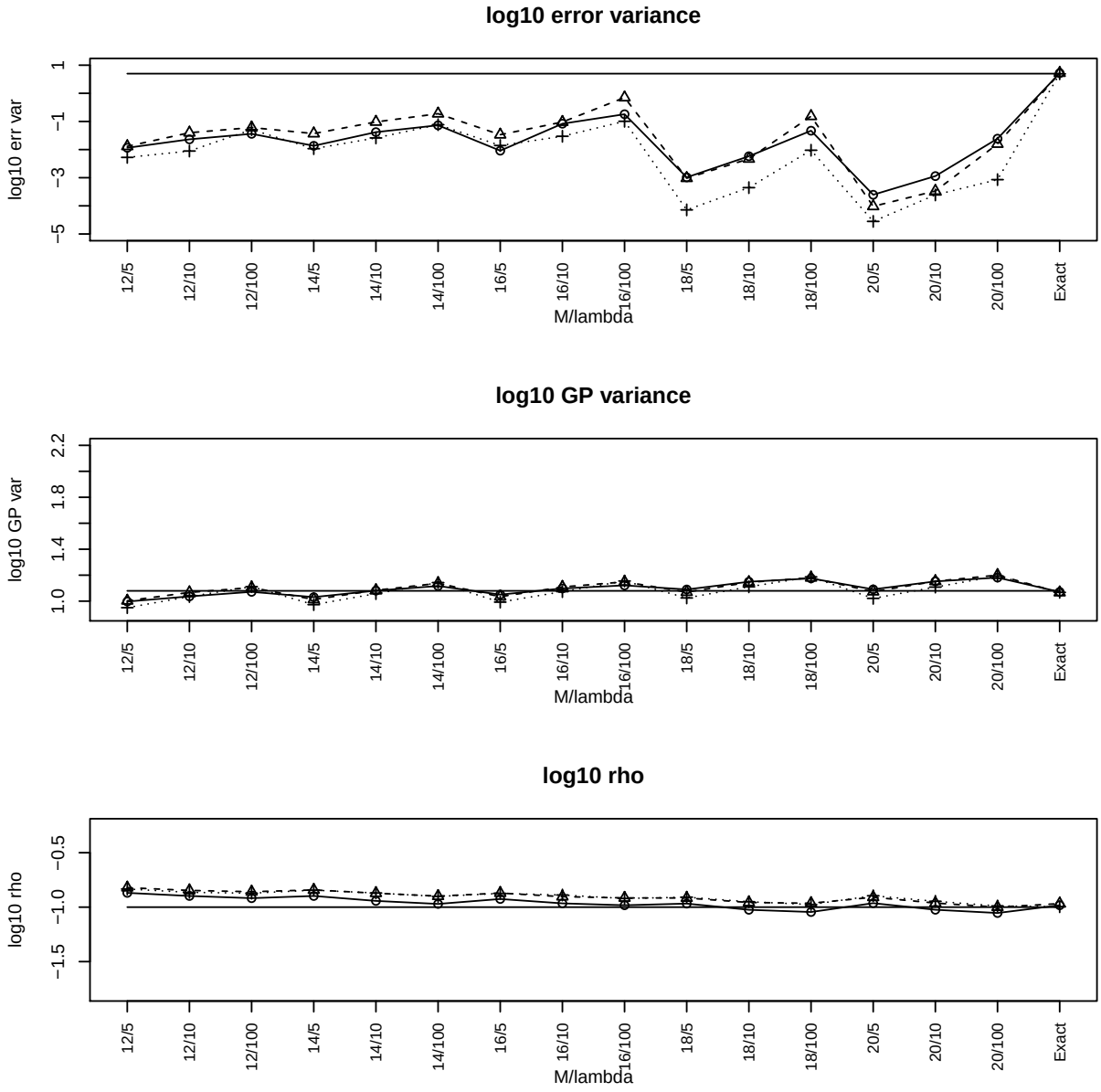


Figure 37: Estimates from maximizing the exact RL; average over 20 simulated datasets. In each panel: No empty space, solid line with circles; 1/8 empty space, dashed line with triangles; 1/4 empty space, dotted line with crosses; the solid horizontal line is the true value used to simulate the data. "Exact" is the estimate from maximizing the exact RL for the actual data. Vertical axes are the same as in Figure 38. Standard errors for individual estimates are 0.51 for log10 error variance, 0.023 for log10 GP variance, and 0.032 for log10 ρ. Standard errors for comparing pairs of estimates are similar.

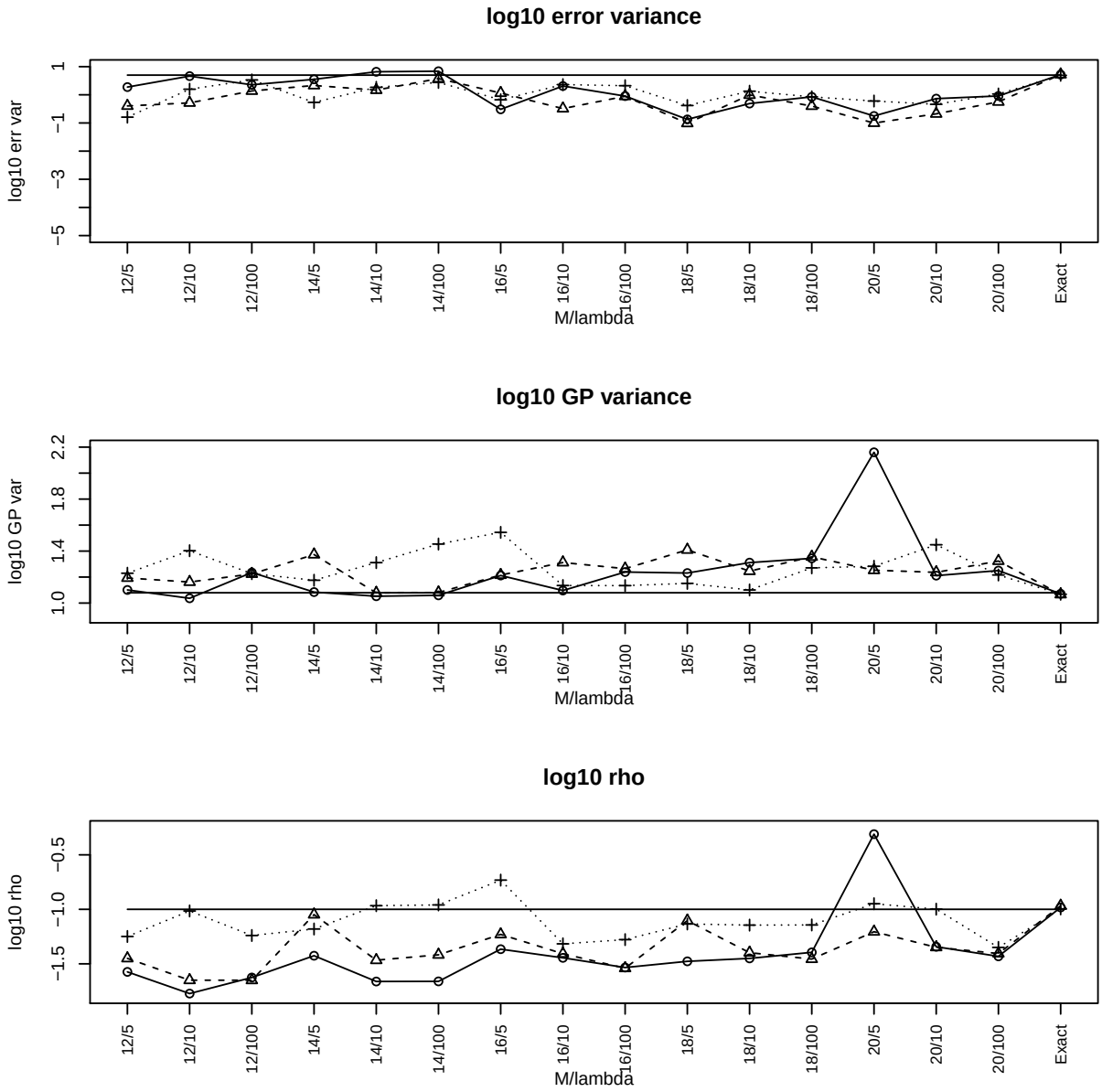


Figure 38: Estimates from maximizing the approximate RL; average over 20 simulated datasets. In each panel: No empty space, solid line with circles; 1/8 empty space, dashed line with triangles; 1/4 empty space, dotted line with crosses; the solid horizontal line is the true value used to simulate the data. "Exact" is the estimate from maximizing the exact RL for the actual data. Vertical axes are the same as in Figure 37. Standard errors for individual estimates are 0.39 for log10 error variance, 0.19 for log10 GP variance, and 0.22 for log10 ρ.

12 Web Appendix L. Spatial Confounding

This brief discussion of spatial confounding, as it relates to the spectral approximation used here, is based on Chapter 15 of Hodges (2014), especially Section 15.2.3.

A conventional non-Bayesian fit of a linear mixed model has two steps: (1) Maximize the restricted likelihood (RL) to estimate unknown parameters in the random-effect and error covariance matrices G and R, and (2) insert those estimates into G and R as if they were known to be true, and estimate fixed effect coefficients β and random effects u. (A Bayesian analysis implicitly replaces Step 1 by computing the marginal posterior of the unknowns in G and R, and replaces Step 2's plug-in estimates with the integral of the conditional posterior of (β, u) against the marginal posterior of the unknowns in G and R, which gives the marginal posterior of (β, u). If the fixed effect vector β has $\pi(\beta) \propto 1$, the following paragraph still holds; if β has a multivariate normal prior with covariance matrix Σ having a finite determinant, the following paragraph no longer holds.)

In Step 1, the RL attributes to the fixed effects *all* variation in y that lies in the column space of the fixed effect design matrix X, so that estimates of the unknowns in G and R are determined entirely by variation in y lying in the orthogonal complement to the column space of X. Thus, spatial confounding has no effect whatsoever on RL-maximizing estimates of unknowns in G and R; our approximate analysis preserves this property. In Step 2, the unknowns in G and R are set to their estimated values and the fixed effects $X\beta$ and random effects Zu then compete to explain variation in y. They compete to an extent determined by Step 1's estimates: spatial confounding can occur in Step 2 *only* if shrinkage or smoothing is not marked; this occurs if error variation (R) is not large relative to random-effect variation (G). Our approximate analysis preserves this property as well.

The foregoing, applied to the spectral approximation, may help explain spatial confounding for linear mixed models including GP-distributed random effects, the best published treatment of which, as far as we know, is Hanks et al (2015). Broadly, fixed effects with strong low-frequency components can be spatially confounded *only* if Step 1 produces an estimate of error variation that

is not large relative to random-effect variation. If this condition holds, then such fixed effects are spatially confounded by low-frequency components of the GP-distributed random effect to a degree determined by the size of their projection on the columns of Z capturing those low frequency components, and by the extent to which the latter's coefficients in u are in fact shrunk. Fixed effects with *weak* low-frequency components are generally not subject to spatial confounding: high-frequency components of the random-effect fit are shrunk a great deal and thus cannot confound fixed effects. The exception to the latter generalization occurs when error variation is estimated to be very small relative to random-effect variation, in which case little shrinkage occurs and fitting fixed effects of any kind is hazardous because the model is barely identified.

References

Fuentes, M. and Reich, B. (2010). Spectral Domain. In A. E. Gelfand, P. Diggle, P. Guttorp and M. Fuentes (eds.), *Handbook of Spatial Statistics*. CRC Press, Ch. 5.

Hanks, E.M., Schliep, E.M., Hooten, M.B., Hoeting, J.A. (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, 26:243–254.

Hodges, J. S. (2014). *Richly Parameterized Linear Models: Additive, Time Series, and Spatial Models Using Random Effects*. Chapman and Hall, CRC Press.

Paciorek, C. J. (2007). Bayesian Smoothing with Gaussian Processes using Fourier Basis Functions in the spectralGP Package. *Journal of Statistical Software*, 19(2):1–38.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

Wikle, C. (2002). Spatial Modeling of Count Data: A Case Study in Modelling Breeding Bird Survey Data on Large Spatial Domains. In A Lawson, D Denison (eds.), *Spatial Cluster Modelling*, Chapman & Hall, 199-209.