A class of multi-resolution approximations for large spatial datasets

Matthias Katzfuss and Wenlong Gong

Department of Statistics, Texas A&M University

Supplementary Material

S1 Proofs

In this section, we provide proofs for the propositions stated throughout the main manuscript. We also state and prove three lemmas that are used in the proofs of the propositions.

Proof of Proposition 1. From (2.6), we have $y_M(\cdot) = \mathbf{b}(\cdot)'\boldsymbol{\eta}$, where $\boldsymbol{\eta} \sim \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Lambda}^{-1})$ and $\mathbf{b}(\cdot)$ is a vector of deterministic functions (for given C_0, $\mathcal{Q}$, and T). Hence, it is trivial to show that $y_M(\cdot)$ is a Gaussian process with mean zero. The covariance function is derived by combining the expression for $y_M(\cdot)$ on the right-hand side of (2.4) with the equations in (2.5). $\square$

Lemma 1 (Exact predictive process). *The predictive process is exact at any* *knot location; that is, if* $x^{(m)}(\cdot)$ *is the predictive process of* $x(\cdot) \sim GP(0, C)$ *based on knots* $\mathcal{Q}_m$ *(see Definition 1), and* $\mathbf{s}_1 \in \mathcal{Q}_m$ *(or* $\mathbf{s}_2 \in \mathcal{Q}_m$*), then*

$$cov\big(x^{(m)}(\mathbf{s}_1), x^{(m)}(\mathbf{s}_2)\big) = C(\mathbf{s}_1, \mathbf{s}_2).$$

Proof of Lemma 1. By the law of total covariance, we have

$$\begin{aligned} cov\big(x^{(m)}(\mathbf{s}_1), x^{(m)}(\mathbf{s}_2)\big) &= cov\big(E(x(\mathbf{s}_1)|\mathbf{x}(\mathcal{Q}_m), E(x(\mathbf{s}_2)|\mathbf{x}(\mathcal{Q}_m)\big) \\ &= cov\big(x(\mathbf{s}_1), x(\mathbf{s}_2)\big) - E\big(cov(x(\mathbf{s}_1), x(\mathbf{s}_2)|\mathbf{x}(\mathcal{Q}_m))\big) = C(\mathbf{s}_1, \mathbf{s}_2), \end{aligned}$$

because $cov(x(\mathbf{s}_1), x(\mathbf{s}_2)|\mathbf{x}(\mathcal{Q}_m)) = 0$ if $\mathbf{s}_1 \in \mathcal{Q}_m$ (or $\mathbf{s}_2 \in \mathcal{Q}_m$). □

Proof of Proposition 2. The proof will be carried out by induction. For $l = 1$, we have $v_{m+1}(\mathbf{q}, \mathbf{s}) = \big(v_m(\mathbf{q}, \mathbf{s}) - cov(\widetilde{\tau}(\mathbf{q}), \widetilde{\tau}(\mathbf{s}))\big)T_{m+1}(\mathbf{q}, \mathbf{s}) = 0$, because using Lemma 1, we can see that $cov\big(\widetilde{\tau}(\mathbf{q}), \widetilde{\tau}(\mathbf{s})\big) = cov\big(\widetilde{\delta}^{(m)}(\mathbf{q}), \widetilde{\delta}^{(m)}(\mathbf{s})\big) = cov\big(\widetilde{\delta}(\mathbf{q}), \widetilde{\delta}(\mathbf{s})\big) = v_m(\mathbf{q}, \mathbf{s})$. For $l > 1$, assuming that $v_{m+l-1}(\mathbf{q}, \mathbf{s}) = 0$, we have

$$v_{m+l}(\mathbf{q}, \mathbf{s}) = \big(v_{m+l-1}(\mathbf{q}, \mathbf{s}) - \mathbf{b}_{m+l-1}(\mathbf{q})'\mathbf{\Lambda}_{m+l-1}^{-1}\mathbf{b}_{m+l-1}(\mathbf{s})\big) \cdot T_{m+l}(\mathbf{q}, \mathbf{s}) = 0,$$

because $\mathbf{b}_{m+l-1}(\mathbf{q}) = v_{m+l-1}(\mathbf{q}, \mathcal{Q}_{m+l-1}) = \mathbf{0}$. □

Lemma 2 (M-RA covariance at knot location $\mathbf{s}$). *If* $\mathbf{s}_1 \in \mathcal{Q}$*, then*

$$C_M(\mathbf{s}_1, \mathbf{s}_2) = \textstyle\sum_{m=0}^{M-1} v_m(\mathbf{s}_1, \mathcal{Q}_m)v_m(\mathcal{Q}_m, \mathcal{Q}_m)^{-1}v_m(\mathcal{Q}_m, \mathbf{s}_2) + v_M(\mathbf{s}_1, \mathbf{s}_2), \quad \mathbf{s}_2 \in \mathcal{D}.$$

Proof of Lemma 2. In the expression for C_M in Proposition 1, we have $v_M(\mathbf{s}_1, \mathcal{Q}_M)v_M(\mathcal{Q}_M, \mathcal{Q}_M)^{-1}v_M(\mathcal{Q}_M, \mathbf{s}_2) = v_M(\mathbf{s}_1, \mathbf{s}_2)$ for $\mathbf{s}_1 \in \mathcal{Q}$. This follows from Lemma 1 if $\mathbf{s}_1 \in \mathcal{Q}_M$, and from Proposition 2 for $\mathbf{s}_1 \in \mathcal{Q}_m$ for $m < M$ (because then both sides of the equation are zero). $\square$

LEMMA 3 (Sum of predictive processes). *For the decomposition in* (2.1), *the sum of predictive processes up to resolution m is equal in distribution to the predictive process based on the union of the knots up to resolution m, for any $m = 0, 1, \ldots, M$; that is, $\sum_{l=0}^{m} \tau_l(\cdot) \stackrel{d}{=} E(y_0(\cdot)|y_0(\cup_{l=0}^{m} \mathcal{Q}_l))$.*

Proof of Lemma 3. For $m = 1$, $\delta_1(\mathbf{s}) \perp\!\!\!\perp y_0(\mathcal{Q}_0)$, for any $\mathbf{s} \in \mathcal{D}$, because $E\big(\delta_1(\mathbf{s})y_0(\mathcal{Q}_0)\big) = E\Big(\big(y_0(\mathbf{s}) - E(y_0(\mathbf{s})|y_0(\mathcal{Q}_0))\big)y_0(\mathcal{Q}_0)\Big) = E\big(y_0(\mathcal{Q}_0)\big)E(\delta_1(\mathbf{s})) = 0$, and $y_0(\mathcal{Q}_0)$, $\delta_1(\mathbf{s})$ are jointly Gaussian. And we have $E(y_0(\cdot)|\delta_1(\mathcal{Q}_1), y_0(\mathcal{Q}_0)) = E(y_0(\cdot)|y_0(\mathcal{Q}_1), y_0(\mathcal{Q}_0))$, because for the σ-algebras

$$\sigma(\delta_1(\mathcal{Q}_1), y_0(\mathcal{Q}_0)) = \sigma\big(y_0(\mathcal{Q}_1) - E\big(y_0(\mathcal{Q}_1)|y_0(\mathcal{Q}_0)\big), y_0(\mathcal{Q}_0)\big) = \sigma\big(y_0(\mathcal{Q}_1), y_0(\mathcal{Q}_0)\big),$$

since $\sigma\big(y_0(\mathcal{Q}_1) - E\big(y_0(\mathcal{Q}_1)|y_0(\mathcal{Q}_0)\big), y_0(\mathcal{Q}_0)\big) = \sigma\big(y_0(\mathcal{Q}_1) - f\big(y_0(\mathcal{Q}_0)\big)\big), y_0(\mathcal{Q}_0)\big) \subset \sigma\big(y_0(\mathcal{Q}_1), y_0(\mathcal{Q}_0)\big)$, and the opposite also holds. Therefore,

$$\begin{aligned} E\big(\delta_1(\mathbf{s})|\delta_1(\mathcal{Q}_1)\big) &= E\big(\delta_1(\mathbf{s})|\delta_1(\mathcal{Q}_1), y_0(\mathcal{Q}_0)\big) \\ &= E\big(y_0(\mathbf{s})|\delta_1(\mathcal{Q}_1), y_0(\mathcal{Q}_0)\big) - E\big(E\big(y_0(\mathbf{s})|y_0(\mathcal{Q}_0)\big)|\delta_1(\mathcal{Q}_1), y_0(\mathcal{Q}_0)\big) \\ &= E\big(y_0(\mathbf{s})|y_0(\mathcal{Q}_1), y_0(\mathcal{Q}_0)\big) - E\big(y_0(\mathbf{s})|y_0(\mathcal{Q}_0)\big), \end{aligned}$$

And so,

$$\tau_0(\mathbf{s}) + \tau_1(\mathbf{s}) = E\big(y_0(\mathbf{s})|y_0(\mathcal{Q}_0)\big) + E\big(\delta_1(\mathbf{s})|\delta_1(\mathcal{Q}_1)\big) = E(y_0(\mathbf{s})|y_0(\mathcal{Q}_1), y_0(\mathcal{Q}_0)).$$

Then, $\delta_2(\mathbf{s}) = y_0(\mathbf{s}) - E\big(y_0(\mathbf{s})|y_0(\mathcal{Q}_0 \cup \mathcal{Q}_1)\big)$, which implies $y_0(\mathcal{Q}_0 \cup \mathcal{Q}_1) \perp\!\!\!\perp \delta_2(\mathbf{s})$. Iteratively repeat this argument to obtain $\sum_{l=0}^{m} \tau_l(\mathbf{s}) = E\big(y_0(\mathbf{s})|y_0(\cup_{l=0}^{m} \mathcal{Q}_l)\big)$. $\square$

Proof of Proposition 3. Using $T_m(\mathbf{s},\mathbf{s}) = 1$ and independence of $\widetilde{\tau}_0, \ldots, \widetilde{\tau}_M$, we have $C_M(\mathbf{s},\mathbf{s}) = var(\sum_{m=0}^{M} \widetilde{\tau}_m(\mathbf{s})) = var(\sum_{m=0}^{M} \tau_m(\mathbf{s}))$. Hence, using Lemma 3, we have $C_M(\mathbf{s},\mathbf{s}) = var\Big(E\Big(y_0(\mathbf{s})\Big|y_0(\cup_{m=0}^{M} \mathcal{Q}_m)\Big)\Big) = var(y_0(\mathbf{s})) = C_0(\mathbf{s},\mathbf{s})$, because $\mathbf{s} \in \mathcal{Q} = \cup_{m=0}^{M} \mathcal{Q}_m$. $\square$

Proof of Proposition 4. First, note that $y_0(\cdot)$ is p times (mean-square) differentiable at $\mathbf{s}$ if and only if $C_{0,\mathbf{s}}(\mathbf{h}) := C_0(\mathbf{s},\mathbf{s}+\mathbf{h})$ is $2p$ times differentiable at the origin ($2p$DO).

By Lemma 2, we have $C_{M,\mathbf{s}}(\mathbf{h}) := C_M(\mathbf{s},\mathbf{s}+\mathbf{h}) = \sum_{m=0}^{M-1} f_m(\mathbf{s},\mathbf{s}+\mathbf{h}) + v_M(\mathbf{s},\mathbf{s}+\mathbf{h})$, where $f_m(\mathbf{s}_1,\mathbf{s}_2) := \sum_{j=1}^{r_m} a_{m,j}(\mathbf{s}_1) v_m(\mathbf{q}_{m,j},\mathbf{s}_2)$, and $a_{m,j}(\mathbf{s})$ is the j-th element of the vector $\mathbf{a}_m(\mathbf{s}) = v_m(\mathcal{Q}_m,\mathcal{Q}_m)^{-1} v_m(\mathcal{Q}_m,\mathbf{s})$. We now show by induction for $m = 0, \ldots, M-1$ that

$$v_{m,\mathbf{q},\mathbf{s}}(\mathbf{h}) := v_m(\mathbf{q},\mathbf{s}+\mathbf{h}) \text{ (for any } \mathbf{q} \in \mathcal{Q}) \text{ and } f_{m,\mathbf{s}}(\mathbf{h}) := f_m(\mathbf{s},\mathbf{s}+\mathbf{h}) \text{ are at least } 2p\text{DO, and } v_{m,\mathbf{s},\mathbf{s}}(\mathbf{h}) \text{ is exactly } 2p\text{DO.} \tag{S1.1}$$

For $m = 0$, $v_{0,\mathbf{q},\mathbf{s}}(\mathbf{h}) = C_0(\mathbf{q},\mathbf{s}+\mathbf{h}) \cdot T_0(\mathbf{q},\mathbf{s}+\mathbf{h})$ is at least $2p$DO by assumption and hence so is $f_{0,\mathbf{s}}(\mathbf{h}) = \sum_{j=1}^{r_0} a_{0,j}(\mathbf{s}) v_0(\mathbf{q}_{0,j},\mathbf{s}+\mathbf{h})$. Further, $v_{0,\mathbf{s},\mathbf{s}}(\mathbf{h})$ is exactly $2p$DO. Now assume that (S1.1) holds for m. Then, using Equation 2.2, $v_{m+1,\mathbf{q},\mathbf{s}}(\mathbf{h}) = \big(v_{m,q,\mathbf{s}}(\mathbf{h}) - f_m(\mathbf{q},\mathbf{s}+\mathbf{h})\big) \cdot T_{m+1}(\mathbf{q},\mathbf{s}+\mathbf{h})$, which is at least $2p$DO, and so is $f_{m+1,\mathbf{s}}(\mathbf{h}) = \sum_{j=1}^{r_{m+1}} a_{m+1,j}(\mathbf{s}) v_{m+1,\mathbf{q}_{m,j},\mathbf{s}}(\mathbf{h})$. Also, $v_{m+1,\mathbf{s},\mathbf{s}}(\mathbf{h})$ is exactly $2p$DO. This proves (S1.1) for $m = 1, \ldots, M$.

In summary, we have $C_{M,\mathbf{s}}(\mathbf{h}) = \sum_{m=0}^{M-1} f_{m,\mathbf{s}}(\mathbf{h}) + (v_{M-1,\mathbf{s},\mathbf{s}}(\mathbf{h}) - f_{M-1,\mathbf{s}}(\mathbf{h})) \cdot T_M(\mathbf{s}, \mathbf{s}+\mathbf{h})$, where $T_{M,\mathbf{s}}(\mathbf{h}) = T_M(\mathbf{s}, \mathbf{s}+\mathbf{h})$ and $f_{m,\mathbf{s}}(\mathbf{h})$, $m = 0, \ldots, M-1$, are all at least $2p$DO, and $v_{M-1,\mathbf{s},\mathbf{s}}(\mathbf{h})$ is exactly $2p$DO.

Thus, $C_{M,\mathbf{s}}(\mathbf{h}) = C_M(\mathbf{s}, \mathbf{s}+\mathbf{h})$ is $2p$DO, and so the corresponding M-RA process $y_M(\cdot) \sim GP(0, C_M)$ is p times (mean-square) differentiable at $\mathbf{s}$. □

Proof of Proposition 5. First, note that realizations are (mean-square) continuous at $\mathbf{s} \in \mathcal{D}$, if $\lim_{\mathbf{h}\to\mathbf{0}} C_M(\mathbf{s}, \mathbf{s}+\mathbf{h}) = C_M(\mathbf{s}, \mathbf{s})$. Further, we have $\mu_M(\mathbf{s}) = E(y_M(\mathbf{s})|\mathbf{z}) = \mathbf{z}' cov(\mathbf{z})^{-1} C_M(\mathcal{S}, \mathbf{s})$. From the proof of Proposition 4, we have that $C_M(\mathbf{s}_0, \mathbf{s}+\mathbf{h}) = \sum_{m=0}^{M} \sum_{j=1}^{r_m} a_{m,j}(\mathbf{s}_0) v_m(\mathbf{q}_{m,j}, \mathbf{s}+\mathbf{h})$. It is straightforward to show using a proof by induction very similar to that for Proposition 4, that $\lim_{\mathbf{h}\to\mathbf{0}} v_m(\mathbf{q}_{m,j}, \mathbf{s}+\mathbf{h}) = v_m(\mathbf{q}_{m,j}, \mathbf{s})$ if $\lim_{\mathbf{h}\to\mathbf{0}} T_m(\mathbf{q}_{m,j}, \mathbf{s}+\mathbf{h}) = T_m(\mathbf{q}_{m,j}, \mathbf{s})$ for all m. In contrast, if $\mathbf{s}$ is on a region boundary, at least one $T_m(\mathbf{q}_{m,j}, \mathbf{s}+\mathbf{h})$ will be discontinuous as a function of $\mathbf{h}$, and so will $C_M(\mathbf{s}_0, \mathbf{s}+\mathbf{h})$ (unless $v_m(\mathbf{s}, \mathbf{s}+\mathbf{h}) = w_m(\mathbf{s}, \mathbf{s}+\mathbf{h})$ and hence the M-RA-block is exact — see Proposition 6). □

LEMMA 4 (Block-independence for exponential covariance). *Assume $y_0(\cdot) \sim GP(0, C_0)$, where C_0 is an exponential covariance function on the real line,*

$\mathcal{D} = \mathbb{R}$, and consider a domain partitioning as in (2.7) with $r_m = (J-1)J^m$ knots for $m = 0, \ldots, M-1$, which are placed such that at each resolution m a knot is located on each boundary between two subregions at resolution $m+1$. Then, for any $m = 1, \ldots, M$, if $\mathbf{s}_i \in \mathcal{D}_{i_1,\ldots,i_m}$ and $\mathbf{s}_j \in \mathcal{D}_{j_1,\ldots,j_m}$, we have $w_m(\mathbf{s}_i, \mathbf{s}_j) = 0$ (defined in (2.2)) if $(i_1, \ldots, i_m) \neq (j_1, \ldots, j_m)$.

Proof of Lemma 4. For any $m = 1, \ldots, M$, using Lemma 3, we have

$$w_m(\mathbf{s}_i, \mathbf{s}_j) = C_0(\mathbf{s}_i, \mathbf{s}_j) - C_0(\mathbf{s}_i, \mathcal{Q}^{m-1})C_0(\mathcal{Q}^{m-1}, \mathcal{Q}^{m-1})^{-1}C_0(\mathcal{Q}^{m-1}, \mathbf{s}_j),$$

where $\mathcal{Q}^{m-1} := \cup_{l=0}^{m-1} \mathcal{Q}_l$. By the law of total covariance,

$$\begin{aligned} w_m(\mathbf{s}_i, \mathbf{s}_j) &= C_0(\mathbf{s}_i, \mathbf{s}_j) - Cov\big(E\big(y_0(\mathbf{s}_i)|y_0(\mathcal{Q}^{m-1})\big), E\big(y_0(\mathbf{s}_j)|y_0(\mathcal{Q}^{m-1})\big)\big) \\ &= E\big(Cov\big(y_0(\mathbf{s}_i), y_0(\mathbf{s}_j)|y_0(\mathcal{Q}^{m-1})\big)\big). \end{aligned}$$

Because $(i_1, i_2, \ldots, i_{m-1}) \neq (j_1, j_2, \ldots, j_{m-1})$, there is a $\mathbf{q} \in \mathcal{Q}^{m-1}$ that lies between $\mathbf{s}_i$ and $\mathbf{s}_j$. As $y_0(\cdot)$ is a Markov process (e.g., Rasmussen and Williams, 2006, Ch. 6), $E\big(Cov\big(y_0(\mathbf{s}_i), y_0(\mathbf{s}_j)|y_0(\mathcal{Q}^{m-1})\big)\big) = E\big(Cov\big(y_0(\mathbf{s}_i), y_0(\mathbf{s}_j)|y_0(\mathbf{q})\big)\big) = w_m(\mathbf{s}_i, \mathbf{s}_j) = 0$. □

Proof of Proposition 6. Comparing the expression for C_M in Lemma 2 to the expression for C_0 in (2.3), it is clear that $C_M(\mathbf{s}_1, \mathbf{s}_2) = C_0(\mathbf{s}_1, \mathbf{s}_2)$ if

$$v_m(\mathbf{s}_i, \mathbf{s}_j) = w_m(\mathbf{s}_i, \mathbf{s}_j), \quad \text{for } m = 0, \ldots, M \text{ and any } \mathbf{s}_i, \mathbf{s}_j \in \mathcal{D}. \qquad \text{(S1.2)}$$

We now prove (S1.2) by induction. For $m = 0$, we have $v_0(\mathbf{s}_i, \mathbf{s}_j) =$

$C_0(\mathbf{s}_i, \mathbf{s}_j)T_0(\mathbf{s}_i, \mathbf{s}_j) = C_0(\mathbf{s}_i, \mathbf{s}_j)$, because $T_0(\mathbf{s}_i, \mathbf{s}_j) \equiv 1$ for the M-RA-block. For $m > 0$, assume that $v_{m-1}(\mathbf{s}_i, \mathbf{s}_j) = w_{m-1}(\mathbf{s}_i, \mathbf{s}_j)$. Then, we can write

$$v_m(\mathbf{s}_i, \mathbf{s}_j) = w_m(\mathbf{s}_i, \mathbf{s}_j)T_m(\mathbf{s}_i, \mathbf{s}_j). \tag{S1.3}$$

Assume that $\mathbf{s}_i \in \mathcal{D}_{i_1,\ldots,i_m}$ and $\mathbf{s}_j \in \mathcal{D}_{j_1,\ldots,j_m}$. Then, if $(i_1, \ldots, i_m) = (j_1, \ldots, j_m)$, (S1.3) holds because $T_m(\mathbf{s}_i, \mathbf{s}_j) = 1$. If $(i_1, \ldots, i_m) \neq (j_1, \ldots, j_m)$, we have $T_m(\mathbf{s}_i, \mathbf{s}_j) = 0$ but also $w_m(\mathbf{s}_i, \mathbf{s}_j) = 0$ by Lemma 4. This proves (S1.3), which proves (S1.2), which in turns proves Proposition 6. □

Proof of Proposition 7. From (3.10), we have $\mathbf{W}_{m,l}^{k+1} = (\mathbf{W}_{m,l}^k - \mathbf{X}_{m,l}^k) \circ T_{k+1}(\mathcal{Q}_m, \mathcal{Q}_l)$, where $\mathbf{X}_{m,l}^k := \mathbf{W}_{m,k}^k \mathbf{\Lambda}_k^{-1} \mathbf{W}_{l,k}^k{}'$. The (i,j)th element of this matrix is

$$(\mathbf{X}_{m,l}^k)_{i,j} = \textstyle\sum_{a,b=1}^{r_k} v_k(\mathbf{q}_{m,i}, \mathbf{q}_{k,a}) v_l(\mathbf{q}_{l,j}, \mathbf{q}_{k,b}) (\mathbf{\Lambda}_k^{-1})_{a,b}, \tag{S1.4}$$

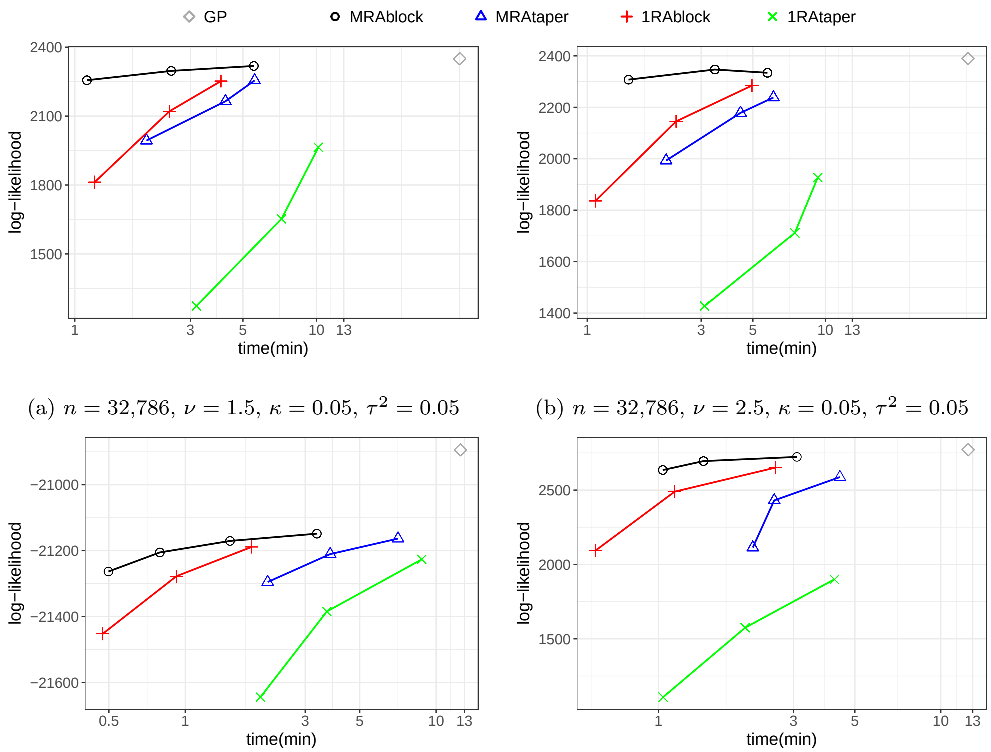
where $v_k(\mathbf{q}_{m,i}, \mathbf{q}_{k,a}) = 0$ if $\|\mathbf{q}_{m,i} - \mathbf{q}_{k,a}\| \geq d_k$, and $v_l(\mathbf{q}_{l,j}, \mathbf{q}_{k,b}) = 0$ if $\|\mathbf{q}_{l,j} - \mathbf{q}_{k,b}\| \geq d_k$. Further, we only need the (i,j)th element of $\mathbf{W}_{m,l}^{k+1}$ (and thus of $\mathbf{X}_{m,l}^k$) if $(i,j) \in \mathcal{I}_{m,l}$, because $(\mathbf{W}_{m,l}^l)_{i,j} = 0$ if $\|\mathbf{q}_{m,i} - \mathbf{q}_{l,j}\| \geq d_l$. Hence, we only need $(\mathbf{\Lambda}_k^{-1})_{a,b}$ if $\|\mathbf{q}_{m,i} - \mathbf{q}_{l,j}\| < d_l$, $\|\mathbf{q}_{m,i} - \mathbf{q}_{k,a}\| < d_k$, *and* $\|\mathbf{q}_{l,j} - \mathbf{q}_{k,b}\| < d_k$, for some $m, l \in \{k+1, \ldots, M\}$. As $d_{k+1} = d_k/J > d_{k+2} > \ldots > d_M$, this means that do not need to calculate $(\mathbf{\Lambda}_k^{-1})_{a,b}$ if $\|\mathbf{q}_{k,a} - \mathbf{q}_{k,b}\| \geq 2d_k + 2d_{k+1} = (2 + 2/J)d_k$, and so we can replace $\mathbf{\Lambda}_k^{-1}$ in $\mathbf{X}_{m,l}^k$ by $\mathbf{S}_k = \widetilde{\mathbf{\Lambda}}_k^{-1} \circ \mathbf{G}_k$.

Further, for each $(i, j) \in \mathcal{I}_{m,l}$, the time to compute (S1.4) is $\mathcal{O}(r_0^2)$, because for any $\mathbf{s} \in \mathcal{D}$, the size of the set $\{\mathbf{q} \in \mathcal{Q}_k : v_k(\mathbf{s}, \mathbf{q}) \neq 0\}$ is $\mathcal{O}(r_0)$. As $\mathcal{I}_{m,l}$ is a set of size $\mathcal{O}(r_m r_0)$, the cost of computing $\mathbf{W}_{m,l}^k$ for each m, l, k is $\mathcal{O}(r_m r_0^3)$. Thus, the total computation time for $k = 0, \ldots, l-1$, $l = 0, \ldots, m$, and $m = 0, \ldots, M$ is $\mathcal{O}(\sum_{m=0}^M \sum_{l=0}^m \sum_{k=0}^{l-1} r_m r_0^3) = \mathcal{O}(r_0^3 \sum_{m=0}^M r_m m^2) = \mathcal{O}(r_0^4 \sum_{m=0}^M J^m m^2) = \mathcal{O}(r_0^4 M^2 J^M) = \mathcal{O}(nM^2 r_0^3)$, because $n = \mathcal{O}(r_0 J^M)$ and $\sum_{m=0}^M m^2 J^m \leq 2M^2 J^M = \mathcal{O}(MJ^M)$. □

Proof of Proposition 8. We have $(\widetilde{\mathbf{\Lambda}}_{m,l})_{i,j} = 0$ if $\nexists \mathbf{s} \in \mathcal{D}$ such that $T_m(\mathbf{q}_{m,i}, \mathbf{s}) \neq 0$ and $T_l(\mathbf{q}_{l,j}, \mathbf{s}) \neq 0$, or equivalently, if $\|\mathbf{q}_{m,i} - \mathbf{q}_{l,j}\| \geq d_m + d_l$. As $d_l = d_m J^{(l-m)/d}$, the ith row $(\widetilde{\mathbf{\Lambda}}_{m,l})_{i,\cdot}$ has $\mathcal{O}(r_0 J^{(l-m)_+})$ nonzero elements, where $(x)_+ = x\mathbb{1}_{\{x \geq 0\}}$. The entire row of the matrix $\widetilde{\mathbf{\Lambda}}$ corresponding to $\mathbf{q}_{m,i}$ thus has $\mathcal{O}(r_0 \sum_{l=0}^M J^{(l-m)_+}) = \mathcal{O}(r_0(m + J^{M-m}))$ nonzero elements. As there are $\mathcal{O}(r_0 J^m)$ rows corresponding to resolution m, the total number of nonzero elements in $\widetilde{\mathbf{\Lambda}}$ is $\mathcal{O}(\sum_{m=0}^M r_0 J^m \cdot r_0(m + J^{M-m})) = \mathcal{O}(r_0^2(MJ^M + \sum_{m=0}^M mJ^m)) = \mathcal{O}(nMr_0)$, because $\sum_{m=0}^M mJ^m \leq 2MJ^M = \mathcal{O}(MJ^M)$ and $n = \mathcal{O}(r_0 J^M)$. □

S2 Additional simulation plots

We provide here additional settings for the simulation study described in Section 4 of the main document. We consider various settings for the



(a) $n = 32{,}786$, $\nu = 1.5$, $\kappa = 0.05$, $\tau^2 = 0.05$

(b) $n = 32{,}786$, $\nu = 2.5$, $\kappa = 0.05$, $\tau^2 = 0.05$

(c) $n = 32{,}786$, $\nu = 1.5$, $\kappa = 0.15$, $\tau^2 = 0.2$

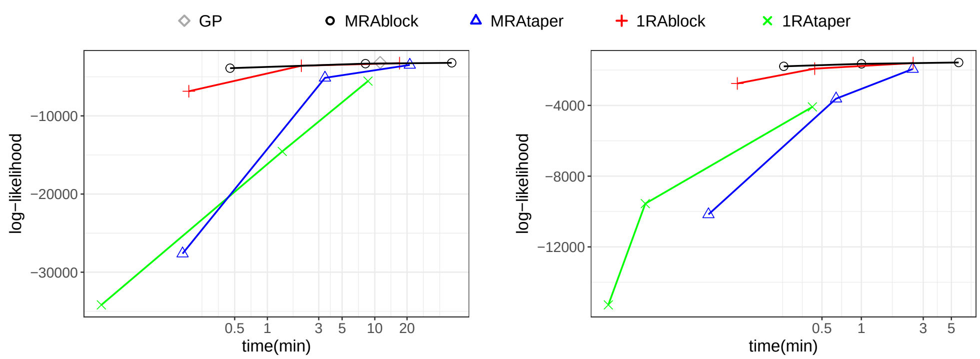
(d) $n = 40{,}960$, $\nu = 1.5$, $\kappa = 0.05$, $\tau^2 = 0.05$

Figure 1: Comparison of approximation accuracy for different sample sizes in one-dimensional space

Matérn covariance function with smoothness parameter ν, range parameter κ, and noise or nugget variance τ^2.

Bibliography

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.



(a) $n = 25{,}600$, $\nu = 1.5$, $\kappa = 0.05$, $\tau^2 = 0.05$

(b) $n = 12{,}544$, $\nu = 2.5$, $\kappa = 0.05$, $\tau^2 = 0.05$

Figure 2: Comparison of approximation accuracy for different sample sizes in two-dimensional space