Supplementary Materials for "Functional-Input Gaussian Processes with Applications to Inverse Scattering Problems"

S1. Proof of Proposition 1

For any non-zero $(\alpha_1, ..., \alpha_n)$ and $(g_1, ..., g_n)$, it follows the quadratic form:

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_j\alpha_k K(g_j, g_k) = \sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_j\alpha_k \int_\Omega\int_\Omega g_j(\mathbf{x})g_k(\mathbf{x}')\Psi(\mathbf{x},\mathbf{x}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}'$$
$$= \int_\Omega\int_\Omega \sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_j\alpha_k g_j(\mathbf{x})g_k(\mathbf{x}')\Psi(\mathbf{x},\mathbf{x}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}'$$
$$= \int_\Omega\int_\Omega \left(\sum_{j=1}^{n}\alpha_j g_j(\mathbf{x})\right)^2 \Psi(\mathbf{x},\mathbf{x}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}' \geq 0,$$

and the quadratic form is strictly greater than zero if $g_1, ..., g_n$ are linearly independent. This finishes the proof.

S2. Proof of Proposition 2

For any $a, b \in \mathbb{R}$ and $g_1, g_2 \in V$, it follows that

$$f(ag_1 + bg_2) = \sum_{j=1}^{\infty}\sqrt{\lambda_j}\langle\phi_j, ag_1 + bg_2\rangle_{L_2(\Omega)}Z_j$$
$$= \sum_{j=1}^{\infty}\sqrt{\lambda_j}\left(a\langle\phi_j, g_1\rangle_{L_2(\Omega)} + b\langle\phi_j, g_2\rangle_{L_2(\Omega)}\right)Z_j$$
$$= af(g_1) + bf(g_2),$$

which finishes the proof.

S3. Proof of Theorem 1

In order to prove the theorem, the following lemma is provided, which can be found in Proposition 10.28 of Wendland (2004).

Lemma 1. *Suppose Ψ is a symmetric and positive definite kernel on Ω. Then the integral operator $\mathcal{T}$ maps $L_2(\Omega)$ continuously into the reproducing kernel Hilbert space $\mathcal{N}_\Psi(\Omega)$. It is the adjoint of the embedding operator of the reproducing kernel Hilbert space $\mathcal{N}_\Psi(\Omega)$ into $L_2(\Omega)$, i.e., it satisfies*

$$\langle g, v\rangle_{L_2(\Omega)} = \langle g, \mathcal{T}v\rangle_{\mathcal{N}_\Psi(\Omega)}, g \in \mathcal{N}_\Psi(\Omega), v \in L_2(\Omega).$$

Now we are ready to prove Theorem 1. For any $\mathbf{u}_n = (u_1, ..., u_n)^T \in \mathbb{R}^n$, it follows that

$$\begin{aligned}
&\mathbb{E}\left(f(g) - \sum_{j=1}^n u_j f(g_j)\right)^2 = K(g,g) - 2\sum_{j=1}^n u_j K(g,g_j) + \sum_{j=1}^n\sum_{l=1}^n u_j u_l K(g_j, g_l)\\
=&\int_\Omega\int_\Omega g(\mathbf{x})g(\mathbf{x}')\Psi(\mathbf{x},\mathbf{x}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}' - 2\sum_{j=1}^n u_j \int_\Omega\int_\Omega g(\mathbf{x})g_j(\mathbf{x}')\Psi(\mathbf{x},\mathbf{x}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}'\\
&+\sum_{j=1}^n\sum_{l=1}^n u_j u_l \int_\Omega\int_\Omega g_j(\mathbf{x})g_l(\mathbf{x}')\Psi(\mathbf{x},\mathbf{x}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}'\\
=&\langle g, \mathcal{T}g\rangle_{L_2(\Omega)} - 2\sum_{j=1}^n u_j\langle g, \mathcal{T}g_j\rangle_{L_2(\Omega)} + \sum_{j=1}^n\sum_{l=1}^n u_j u_l\langle g_j, \mathcal{T}g_l\rangle_{L_2(\Omega)}\\
=&\langle \mathcal{T}g, \mathcal{T}g\rangle_{\mathcal{N}_\Psi(\Omega)} - 2\sum_{j=1}^n u_j\langle \mathcal{T}g, \mathcal{T}g_j\rangle_{\mathcal{N}_\Psi(\Omega)} + \sum_{j=1}^n\sum_{l=1}^n u_j u_l\langle \mathcal{T}g_j, \mathcal{T}g_l\rangle_{\mathcal{N}_\Psi(\Omega)}\\
=&\left\|\mathcal{T}g - \sum_{j=1}^n u_j\mathcal{T}g_j\right\|^2_{\mathcal{N}_\Psi(\Omega)}, \qquad\qquad \text{(S3.1)}
\end{aligned}$$

where the third equality is by Lemma 1.

Since $\hat{\mathbf{u}}_n := \mathbf{K}_n^{-1}\mathbf{k}_n(g)$ minimizes the MSPE, it also minimizes $\left\|\mathcal{T}g - \sum_{j=1}^n u_j\mathcal{T}g_j\right\|^2_{\mathcal{N}_\Psi(\Omega)}$, where $\mathbf{K}_n$ and $\mathbf{k}_n(g)$ are as in (2.2). By taking minimum on both sides of (S3.1), we obtain

$$\mathbb{E}\left(f(g)-\hat{f}(g)\right)^2=\min_{\mathbf{u}\in\mathbb{R}^n}\left\|\mathcal{T}g-\sum_{j=1}^n u_j\mathcal{T}g_j\right\|^2_{\mathcal{N}_\Psi(\Omega)}, \tag{S3.2}$$

which finishes the proof.

S4. Proof of Corollary 1

In order to prove this corollary, we need the following lemma, which states the asymptotic rates of the eigenvalues of K. Lemma 2 is implied by the proof of Lemma 18 of Tuo and Wang (2020). In the rest of the Supplementary Materials, we will use the following notation. For two positive sequences a_n and b_n, we write $a_n \asymp b_n$ if, for some constants $C, C' > 0$, $C \leq a_n/b_n \leq C'$. For notational simplicity, we will use $C, C', C_1, C_2, ...$ to denote the constants, of which the values can change from line to line.

Lemma 2. *Let Ψ be a Matérn kernel function defined in (3.6), and $\lambda_1 \geq \lambda_2 \geq ... > 0$ be its eigenvalues. Then, $\lambda_k \asymp k^{-2\nu/d}$.*

Proof of Corollary 1. By (3.2), we have $\mathcal{T}g(\mathbf{x})=\int_\Omega g(\mathbf{x}')\Psi(\mathbf{x},\mathbf{x}')\mathrm{d}\mathbf{x}'=\sum_{j=1}^\infty \lambda_j\langle g,\phi_j\rangle_{L_2(\Omega)}\phi_j(\mathbf{x})$. Therefore,

$$\begin{aligned}\left\|\mathcal{T}g-\sum_{j=1}^n u_j\mathcal{T}g_j\right\|^2_{\mathcal{N}_\Psi(\Omega)} &= \left\|\sum_{j=1}^\infty \lambda_j\langle g-\sum_{k=1}^n u_k g_k,\phi_j\rangle_{L_2(\Omega)}\phi_j\right\|^2_{\mathcal{N}_\Psi(\Omega)}\\ &= \sum_{j=1}^\infty \lambda_j\langle g-\sum_{k=1}^n u_k g_k,\phi_j\rangle^2_{L_2(\Omega)},\end{aligned} \tag{S4.1}$$

where the last equality holds by Theorem 10.29 of Wendland (2004). Recall that $g_k=\phi_k$ for $k=1,...,n$. Now take $u_k=\langle g,\phi_k\rangle_{L_2(\Omega)}$ for $k=1,...,n$. It follows from Theorem 1 and (S4.1) that

$$\mathbb{E}\left(f(g)-\hat{f}(g)\right)^2=\min_{\mathbf{u}\in\mathbb{R}^n}\left\|\mathcal{T}g-\sum_{j=1}^n u_j\mathcal{T}g_j\right\|_{\mathcal{N}_\Psi(\Omega)}^2$$
$$\leq\sum_{j=1}^\infty\lambda_j\langle g-\sum_{k=1}^n u_k g_k,\phi_j\rangle_{L_2(\Omega)}^2=\sum_{j=n+1}^\infty\lambda_j\langle g,\phi_j\rangle_{L_2(\Omega)}^2. \tag{S4.2}$$

This indicates that the MSPE depends on the tail behavior of the summation $\sum_{j=1}^\infty\lambda_j\langle g,\phi_j\rangle_{L_2(\Omega)}^2$. Because Ψ is a Matérn kernel function defined in (3.6), Lemma 2 implies $\lambda_j\asymp j^{-\frac{2\nu}{d}}$. Then, an explicit convergence rate can be obtained via

$$\mathbb{E}\left(f(g)-\hat{f}(g)\right)^2\leq\sum_{j=n+1}^\infty\lambda_j\langle g,\phi_j\rangle_{L_2(\Omega)}^2\leq\lambda_n\sum_{j=n+1}^\infty\langle g,\phi_j\rangle_{L_2(\Omega)}^2\leq C_1\|g\|_{L_2}^2 n^{-\frac{2\nu}{d}},$$

where we utilize $\sum_{j=n+1}^\infty\langle g,\phi_j\rangle_{L_2(\Omega)}^2\leq\sum_{j=1}^\infty\langle g,\phi_j\rangle_{L_2(\Omega)}^2=\|g\|_{L_2}^2$. This finishes the proof of (3.8).

Next, we consider the case $g\in\mathcal{N}_\Psi(\Omega)$. Theorem 10.29 of Wendland (2004) yields that

$$\|g\|_{\mathcal{N}_\Psi(\Omega)}^2=\sum_{j=1}^\infty\frac{\langle g,\phi_j\rangle_{L_2(\Omega)}^2}{\lambda_j}<\infty.$$

Then an alternative way to bound (S4.2) is by

$$\mathbb{E}\left(f(g)-\hat{f}(g)\right)^2\leq\sum_{j=n+1}^\infty\lambda_j\langle g,\phi_j\rangle_{L_2(\Omega)}^2=\sum_{j=n+1}^\infty\lambda_j^2\frac{\langle g,\phi_j\rangle_{L_2(\Omega)}^2}{\lambda_j}$$
$$\leq\lambda_n^2\sum_{j=n+1}^\infty\frac{\langle g,\phi_j\rangle_{L_2(\Omega)}^2}{\lambda_j}\leq\lambda_n^2\|g\|_{\mathcal{N}_\Psi(\Omega)}^2\leq C_2\|g\|_{\mathcal{N}_\Psi(\Omega)}^2 n^{-\frac{4\nu}{d}}. \tag{S4.3}$$

This finishes the proof of (3.9), and thus the proof of Corollary 1. □

S5. Proof of Corollary 2

The following lemma is used in the proof.

Lemma 3 (Wu and Schaback, 1993, Theorem 5.14)**.** *Let Ω be compact and convex with a positive Lebesgue measure; $\Psi(\mathbf{x}, \mathbf{x}')$ be a Matérn kernel given by (3.6) with the smoothness parameter ν. Then there exist constants c, c_0 depending only on Ω, ν and the lengthscale parameter Θ in (3.6), such that $\Psi(\mathbf{x}, \mathbf{x}) - \mathbf{r}_n(\mathbf{x})^T \mathbf{R}_n^{-1} \mathbf{r}_n(\mathbf{x}) \leq c h_{\mathbf{X}_n,\Omega}^{2\nu}$ provided that $h_{\mathbf{X}_n,\Omega} \leq c_0$, where $\mathbf{R}_n = (\Psi(\mathbf{x}_j, \mathbf{x}_k))_{jk}$ and $\mathbf{r}_n(\mathbf{x}) = (\Psi(\mathbf{x}, \mathbf{x}_1), ..., \Psi(\mathbf{x}, \mathbf{x}_n))^T$.*

Proof of Corollary 2. For any $\mathbf{u}_n = (u_1, ..., u_n)^T$, by (S4.1), we have

$$\begin{aligned}
&\left\|\mathcal{T}g - \sum_{j=1}^n u_j \mathcal{T} g_j\right\|_{\mathcal{N}_\Psi(\Omega)}^2 = \sum_{j=1}^\infty \lambda_j \langle g - \sum_{k=1}^n u_k g_k, \phi_j \rangle_{L_2(\Omega)}^2 \\
=&\sum_{j=1}^\infty \lambda_j \langle g - \sum_{k=1}^n u_k \Psi(\mathbf{x}_k, \cdot), \phi_j \rangle_{L_2(\Omega)}^2 \leq C \sum_{j=1}^\infty \langle g - \sum_{k=1}^n u_k \Psi(\mathbf{x}_k, \cdot), \phi_j \rangle_{L_2(\Omega)}^2 \\
=&C \left\| g - \sum_{k=1}^n u_k \Psi(\mathbf{x}_k, \cdot) \right\|_{L_2(\Omega)}^2 \leq C_1 \sup_{\mathbf{x}\in\Omega} \left| g(\mathbf{x}) - \sum_{k=1}^n u_k \Psi(\mathbf{x}_k, \mathbf{x}) \right|^2, \qquad \text{(S5.1)}
\end{aligned}$$

where the last equality holds because ϕ_j's are orthogonal basis in $L_2(\Omega)$. Therefore, we can take $\mathbf{u}_n = \mathbf{R}_n^{-1}\mathbf{g}_n$, where $\mathbf{g}_n = (g(\mathbf{x}_1), ..., g(\mathbf{x}_n))^T$. Then the term, $\sup_{\mathbf{x}\in\Omega} |g(\mathbf{x}) - \sum_{k=1}^n u_k \Psi(\mathbf{x}_k, \mathbf{x})|$, becomes the prediction error of the radial basis function interpolation, which is well established in the literature (Wendland, 2004).

For the completeness of this proof, we include the proof of an upper bound here. Let $v_k(\mathbf{x}) = (\mathbf{R}_n^{-1}\mathbf{r}_n(\mathbf{x}))_k$ for $k = 1, \ldots, n$. For any $\mathbf{x} \in \Omega$, the reproducing property implies that

$$
\begin{aligned}
&\left|g(\mathbf{x})-\sum_{k=1}^{n}u_k\Psi(\mathbf{x}_k,\mathbf{x})\right|^2=\left|\langle g,\Psi(\mathbf{x},\cdot)\rangle_{\mathcal{N}_\Psi(\Omega)}-\sum_{k=1}^{n}v_k(\mathbf{x})g(\mathbf{x}_k)\right|^2\\
=&\left|\langle g,\Psi(\mathbf{x},\cdot)\rangle_{\mathcal{N}_\Psi(\Omega)}-\sum_{k=1}^{n}v_k(\mathbf{x})\langle g,\Psi(\mathbf{x}_k,\cdot)\rangle_{\mathcal{N}_\Psi(\Omega)}\right|^2\\
\leq&\|g\|^2_{\mathcal{N}_\Psi(\Omega)}\left\|\Psi(\mathbf{x},\cdot)-\sum_{k=1}^{n}v_k(\mathbf{x})\Psi(\mathbf{x}_k,\cdot)\right\|^2_{\mathcal{N}_\Psi(\Omega)}\\
=&\|g\|^2_{\mathcal{N}_\Psi(\Omega)}(\Psi(\mathbf{x},\mathbf{x})-\mathbf{r}_n(\mathbf{x})^T\mathbf{R}_n^{-1}\mathbf{r}_n(\mathbf{x})),
\end{aligned}
\tag{S5.2}
$$

where the inequality is by the Cauchy-Schwarz inequality. A bound on $\Psi(\mathbf{x},\mathbf{x})-\mathbf{r}_n(\mathbf{x})^T\mathbf{R}_n^{-1}\mathbf{r}_n(\mathbf{x})$ can be obtained via Lemma 3, which gives

$$
\Psi(\mathbf{x},\mathbf{x})-\mathbf{r}_n(\mathbf{x})^T\mathbf{R}_n^{-1}\mathbf{r}_n(\mathbf{x})\leq C_2h^{2\nu}_{\mathbf{X}_n,\Omega},
\tag{S5.3}
$$

where $h_{\mathbf{X}_n,\Omega}$ is the fill distance of $\mathbf{X}_n$. Combining (3.5) with (S5.1), (S5.2) and (S5.3), we obtain that

$$
\mathbb{E}\left(f(g)-\hat{f}(g)\right)^2\leq C_3h^{2\nu}_{\mathbf{X}_n,\Omega},
$$

which finishes the proof. □

S6. Proof of Proposition 3

Without loss of generality, we assume $\gamma=1$. It suffices to show that $\mathbf{K}_n=(K(g_j,g_k))_{jk}$ is positive definite for any $g_1,...,g_n\in V$, which can be done by showing that there exist n points $\mathbf{a}_1,...,\mathbf{a}_n\in\mathbb{R}^m$ with $m<\infty$ such that $\|\mathbf{a}_j-\mathbf{a}_k\|_2=\|g_j-g_k\|_{L_2(\Omega)}$. Let $V_n=\mathrm{span}(\{g_1,...,g_n\})$, which is the linear space spanned by $g_1,...,g_n$. Clearly, V_n is a finite dimensional space. Let $\phi_1,...,\phi_m$ be an orthogonal basis of V_n with respect to $L_2(\Omega)$, with $m\leq n$, which can be found via the Gram–Schmidt process. Thus, given this basis, each g_j can be written as

$$g_j = \sum_{k=1}^{m} a_{jk}\phi_k$$

with $\mathbf{a}_j = (a_{j1}, ..., a_{jm})^T \in \mathbb{R}^m$. Then it can be verified that $\|\mathbf{a}_j - \mathbf{a}_k\|_2 = \|g_j - g_k\|_{L_2(\Omega)}$. Since $\mathbf{K}_n = (K(g_j - g_k))_{jk} = (\psi(\|g_j - g_k\|_{L_2(\Omega)}))_{jk} = (\psi(\|\mathbf{a}_j - \mathbf{a}_k\|_2))_{jk}$, which is positive definite, this finishes the proof.

S7. Proof of Theorem 2

We first provide a characterization on the function class V. Since Φ is a positive definite function, we can apply Mercer's theorem to Φ and obtain

$$\Phi(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^{\infty} \lambda_{\Phi,j}\phi_j(\mathbf{x})\phi_j(\mathbf{x}'), \quad \mathbf{x}, \mathbf{x}' \in \Omega, \tag{S7.1}$$

where $\lambda_{\Phi,1} \geq \lambda_{\Phi,2} \geq ... > 0$ are the eigenvalues and $\{\phi_k\}_{k\in\mathbb{N}}$ are the eigenfunctions, and the summation is uniformly and absolutely convergent. Because $g \in V \subset \mathcal{N}_\Phi(\Omega)$, by Theorem 10.29 of Wendland (2004), the summation

$$\|g\|^2_{\mathcal{N}_\Phi(\Omega)} = \sum_{j=1}^{\infty} \lambda^{-1}_{\Phi,j}\langle g, \phi_j\rangle^2_{L_2(\Omega)} \leq 1.$$

Definition 1. A set $\Omega \subset \mathbb{R}^d$ is said to satisfy an interior cone condition if there exists an angle $\alpha \in (0, \pi/2)$ and a radius $\mathfrak{r} > 0$ such that for every $\mathbf{x} \in \Omega$, a unit vector $\xi(\mathbf{x})$ exists such that the cone

$$C(\mathbf{x}, \xi(\mathbf{x}), \alpha, \mathfrak{r}) := \left\{\mathbf{x} + \eta\tilde{\mathbf{x}} : \tilde{\mathbf{x}} \in \mathbb{R}^d, \|\tilde{\mathbf{x}}\| = 1, \tilde{\mathbf{x}}^T\xi(\mathbf{x}) \geq \cos(\alpha), \eta \in [0, \mathfrak{r}]\right\}$$

is contained in Ω.

We need the following lemma, which is Theorem 11.8 of Wendland (2004);

also see Theorem 3.14 of Wendland (2004). Lemma 4 ensures the existence of the local polynomial reproduction.

Lemma 4 (Local polynomial reproduction)**.** *Let $l \in \mathbb{N}_0$ and $\pi_l(\mathbb{R}^d)$ be the set of d-variate polynomials with absolute degree no more than l. Suppose $\Omega \subset \mathbb{R}^d$ is bounded and satisfies an interior cone condition. Define*

$$C_1 = 2, C_2 = \frac{16(1+\sin(\alpha))^2 l^2}{3\sin^2(\alpha)}, c_0 = \frac{\mathfrak{r}}{C_2},$$

where α and $\mathfrak{r}$ are defined in Definition 1. Then for all $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \Omega$ with $h_{\mathbf{X}_n,\Omega} \leq c_0$ and every $\mathbf{x} \in \Omega$ there exist numbers $\tilde{u}_1(\mathbf{x}), \ldots, \tilde{u}_n(\mathbf{x})$ with

(1) $\sum_{j=1}^n \tilde{u}_j(\mathbf{x})p(\mathbf{x}_j) = p(\mathbf{x})$ for all $p \in \pi_l(\mathbb{R}^d)$,

(2) $\sum_{j=1}^n |\tilde{u}_j(\mathbf{x})| \leq C_1$,

(3) $\tilde{u}_j(\mathbf{x}) = 0$, if $\|\mathbf{x} - \mathbf{x}_j\| > C_2 h_{\mathbf{X}_n,\Omega}$.

The following lemma provides an upper bound of the MSPE using $\tilde{u}_j(\mathbf{x})$, which can be found from the proof of Theorem 11.9 and (11.6) of Wendland (2004).

Lemma 5. *Suppose $\Phi = r(\|\cdot\|_2) \in C^k(\mathbb{R}^d)$ is positive definite. Let Ω be a compact region satisfying an interior cone condition. Then for $h_{\mathbf{X}_n,\Omega} \leq h_0$,*

$$\Phi(\mathbf{x},\mathbf{x}) - 2\sum_{j=1}^n \tilde{u}_j \Phi(\mathbf{x},\mathbf{x}_j) + \sum_{j=1}^n \sum_{k=1}^n \tilde{u}_j \tilde{u}_k \Phi(\mathbf{x}_j,\mathbf{x}_k) \leq (1+C_1)^2 \max_{0 \leq s \leq 2C_2 h_{\mathbf{X}_n,\Omega}} |\phi(s) - p(s^2)|,$$

where $p \in \pi_{\lfloor l/2 \rfloor}(\mathbb{R})$.

Proof of Theorem 2. Define a map $h : V \to W$ between the function class V and the set $W = \{\mathbf{a} = (a_1, ..., a_n, ...)^T : \sum_{j=1}^\infty \lambda_{\Phi,j}^{-1} a_j^2 \leq 1\} \subset l_2(\mathbb{R}^\infty)$ as

$$h(g) = (\langle g, \phi_1 \rangle_{L_2(\Omega)}, ..., \langle g, \phi_n \rangle_{L_2(\Omega)}, ...)^T.$$

It can be verified that $\|g\|_{L_2(\Omega)} = \|h(g)\|_2$. Therefore, we can define a new positive definite function K_1 on W which satisfies

$$K_1(\mathbf{a}, \mathbf{a}') = \psi(\|\mathbf{a} - \mathbf{a}'\|_2) = \psi(\|g - g'\|_{L_2(\Omega)}) = K(g, g'), \quad \forall g, g' \in V, \tag{S7.2}$$

where $\mathbf{a} = h(g)$ and $\mathbf{a}' = h(g')$. Define $\mathbf{a}^{(j)} = h(g_j)$. For any $\mathbf{u}_n = (u_1, ..., u_n)^T \in \mathbb{R}^n$, it follows that

$$\begin{aligned}
&\mathbb{E}\left(f(g) - \sum_{j=1}^n u_j f(g_j)\right)^2 \\
=&K(g, g) - 2\sum_{j=1}^n u_j K(g, g_j) + \sum_{j=1}^n \sum_{k=1}^n u_j u_k K(g_j, g_k) \\
=&\psi(0) - 2\sum_{j=1}^n u_j \psi(\|\mathbf{a} - \mathbf{a}^{(j)}\|_2) + \sum_{j=1}^n \sum_{k=1}^n u_j u_l \psi(\|\mathbf{a}^{(j)} - \mathbf{a}^{(k)}\|_2).
\end{aligned} \tag{S7.3}$$

Let $\mathbf{b}_j = (a_1^{(j)}, ..., a_m^{(j)})^T$, $\mathbf{b} = (a_1, ..., a_m)^T$, $\mathbf{a}_c^{(j)} = (a_{m+1}^{(j)}, a_{m+2}^{(j)}, ...)$, and $\mathbf{a}_c = (a_{m+1}, a_{m+2}, ...)$, where m will be determined later. Then $\mathbf{a}^{(j)} = (\mathbf{b}_j^T, (\mathbf{a}_c^{(j)})^T)^T$ and $\mathbf{a} = (\mathbf{b}^T, \mathbf{a}_c^T)^T$. Applying Lemma 4 to (S7.3), we obtain that for some $\tilde{u}_j$,

$$\sum_{j=1}^n \tilde{u}_j p(\mathbf{b}_j) = p(\mathbf{b}), \text{ for all } p \in \pi_l(\mathbb{R}^m), \quad \sum_{j=1}^n |\tilde{u}_j(\mathbf{b})| \le C_1, \text{ and } \tilde{u}_j(\mathbf{b}) = 0, \text{ if } \|\mathbf{b} - \mathbf{b}_j\|_2 > C_2 h_{\mathbf{B}_n,\Omega}, \tag{S7.4}$$

when $h_{\mathbf{B}_n,\Omega} \le c_0$, where $\mathbf{B}_n = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$. Note that C_2 and c_0 depend on the interior cone condition and l. In particular, they change as the dimension of $\mathbf{b}$ and the degree l of polynomials p in (S7.4) change.

Since $\mathbf{a}, \mathbf{a}^{(j)} \in W$, it follows that $\mathbf{a} \le \sqrt{\lambda_{\Phi,j}}$ and $\mathbf{a}^{(j)} \le \sqrt{\lambda_{\Phi,j}}$. Define a set $V_2 = \bigtimes_{k=1}^m [0, \sqrt{\lambda_{\Phi,k}}]$. It can be verified that $\mathbf{b}, \mathbf{b}_j \in V_2$. Set $\alpha = \pi/6$ and $\mathfrak{r} = \sqrt{\lambda_{\Phi,m}}/2$. It can be verified that the interior cone condition is satisfied. Then $C_2 = 48l^2$ and $c_0 = \frac{\sqrt{\lambda_{\Phi,m}}}{96l^2}$.

With $\tilde{u}_j$ defined in (S7.4), by (S7.3), it follows that

$$
\begin{aligned}
&\mathbb{E}\left(f(g)-\hat{f}(g)\right)^2\\
\leq&\psi(0)-2\sum_{j=1}^{n}\tilde{u}_j\psi(\|\mathbf{a}-\mathbf{a}^{(j)}\|_2)+\sum_{j=1}^{n}\sum_{l=1}^{n}\tilde{u}_j\tilde{u}_l\psi(\|\mathbf{a}^{(j)}-\mathbf{a}^{(l)}\|_2)\\
=&\left(\psi(0)-2\sum_{j=1}^{n}\tilde{u}_j\psi(\|\mathbf{b}-\mathbf{b}_j\|_2)+\sum_{j=1}^{n}\sum_{k=1}^{n}\tilde{u}_j\tilde{u}_k\psi(\|\mathbf{b}_j-\mathbf{b}_k\|_2)\right)\\
&+\left(-2\sum_{j=1}^{n}\tilde{u}_j\left(\psi(\|\mathbf{a}-\mathbf{a}^{(j)}\|_2)-\psi(\|\mathbf{b}-\mathbf{b}_j\|_2)\right)+\sum_{j=1}^{n}\sum_{k=1}^{n}\tilde{u}_j\tilde{u}_k\left(\psi(\|\mathbf{a}^{(j)}-\mathbf{a}^{(l)}\|_2)-\psi(\|\mathbf{b}_j-\mathbf{b}_k\|_2)\right)\right)\\
:=&I_1+I_2. \qquad \text{(S7.5)}
\end{aligned}
$$

The first term can be bounded by Lemma 5, which gives

$$
I_1 \leq 9\max_{0\leq s\leq 2C_2h_{\mathbf{B}_n,\Omega}}|\psi(s)-p(s^2)| = 9\max_{0\leq s\leq\sqrt{\lambda_{\Phi,m}}}|\psi(s)-p(s^2)|, \qquad \text{(S7.6)}
$$

for some $p\in\pi_{\lfloor l/2\rfloor}(\mathbb{R})$, provided $h_{\mathbf{B}_n,\Omega}\leq c_0$. Since $\|\mathbf{b}_j-\mathbf{b}\|_2\leq\|\mathbf{a}-\mathbf{a}^{(j)}\|_2=\|g-g_j\|_{L_2(\Omega)}$, we have $h_{\mathbf{B}_n,\Omega}\leq h_{G_n,V}$ so $h_{\mathbf{B}_n,\Omega}\leq c_0$ holds.

Next, we consider bounding I_1 with a Matérn kernel function ψ. Lemma 2 implies that $\lambda_{\Phi,j}\asymp j^{-\frac{2\nu_1}{d}}$. By the expansion of modified Bessel function (Abramowitz and Stegun, 1948), ψ can be written as

$$
\psi(r)=\sum_{k=0}^{\lfloor\nu\rfloor}c_kr^{2k}+c_\psi(r),
$$

where

$$
c_\psi(r)=\begin{cases} cr^{2\nu}\log r+O(r^{2\nu}) & \nu=1,2,...\\ cr^{2\nu}+O(r^{2(\lfloor\nu\rfloor+1)}) & \text{otherwise.}\end{cases}
$$

Therefore, we can take $p(s^2) = -\sum_{k=0}^{\lfloor \nu \rfloor} c_k s^{2k}$ and obtain that

$$\max_{0\leq s\leq\sqrt{\lambda_{\Phi,m}}} |\psi(s) - p(s^2)| \leq \begin{cases} C_2\lambda_{\Phi,m}^{\nu} \log(\lambda_{\Phi,m}^{-1}) \leq C_3 m^{-\frac{2\nu\nu_1}{d}} \log m, & \nu = 2, ... \\ C_4\lambda_{\Phi,m}^{\nu} \leq C_5 m^{-\frac{2\nu\nu_1}{d}}, & \text{otherwise.} \end{cases} \tag{S7.7}$$

By (S7.6),

$$I_1 \leq 9 \max_{0\leq s\leq\sqrt{\lambda_{\Phi,m}}} |\psi(s) - p(s^2)| \leq \begin{cases} C_6 m^{-\frac{2\nu\nu_1}{d}} \log m, & \nu = 2, ... \\ C_7 m^{-\frac{2\nu\nu_1}{d}}, & \text{otherwise.} \end{cases} \tag{S7.8}$$

It remains to bound I_2. For a Matérn kernel function ψ, it can be verified that for all $s_1, s_2 \in [0, s]$,

$$|\psi(s_1) - \psi(s_2)| \leq C_8 |s_1 - s_2|^{2\tau},$$

where $\tau = \min(\nu, 1)$. Therefore, we can rewrite (S7.5) as

$$\begin{aligned} |I_2| \leq & 2C_8 \sum_{j=1}^{n} \tilde{u}_j \big| \|\mathbf{a} - \mathbf{a}^{(j)}\|_2 - \|\mathbf{b} - \mathbf{b}_j\|_2 \big|^{2\tau} + C_8 \sum_{j=1}^{n}\sum_{k=1}^{n} \tilde{u}_j \tilde{u}_k \big| \|\mathbf{a}^{(j)} - \mathbf{a}^{(k)}\|_2 - \|\mathbf{b}_j - \mathbf{b}_k\|_2 \big|^{2\tau} \\ \leq & 2C_8 \sum_{j=1}^{n} |\tilde{u}_j| \, \|\mathbf{a} - \mathbf{a}^{(j)} - (\mathbf{b} - \mathbf{b}_j)\|_2^{2\tau} + C_8 \sum_{j=1}^{n}\sum_{k=1}^{n} |\tilde{u}_j| \, |\tilde{u}_l| \, \|\mathbf{a}^{(j)} - \mathbf{a}^{(l)} - (\mathbf{b}_j - \mathbf{b}_k)\|_2^{2\tau} \\ \leq & C_8(4C_1 + 2C_1^2)\lambda_{\Phi,m+1}^{\tau} \leq C_8(4C_1 + 2C_1^2) m^{-\frac{2\tau\nu_1}{d}}. \end{aligned} \tag{S7.9}$$

Because $m^{-\frac{2\nu\nu_1}{d}} \log m \leq m^{-\frac{2\nu_1\tau}{d}} \log m$ and $\log m > 1$, combining (S7.7), (S7.9) and (S7.5) leads to

$$\mathbb{E}\left(f(g) - \hat{f}(g)\right)^2 \leq C_9 m^{-\frac{2\nu_1\tau}{d}} \log m. \tag{S7.10}$$

The last step is to compute m such that there exist n functions, $h_{G_n,V} \leq C_0 m^{-\frac{2\nu_1}{d}}$.

Since it is known that a unit ball of $\mathcal{N}_\Phi(\Omega)$ has a covering number

$$N(\delta, V, \|\cdot\|_{L_\infty}) \leq C_{10}\exp(C_{11}\delta^{-\frac{d}{\nu_1+d/2}}).$$

Thus, in order to make $h_{G_n,V} \leq C_0 m^{-\frac{2\nu_1}{d}}$, we set $\delta = C_0 m^{-\frac{2\nu_1}{d}}$. Thus, as long as $n \geq C_{12}\exp(C_{13}m^{\frac{2\nu_1}{\nu_1+d/2}})$, $h_{G_n,V} \leq C_0 m^{-\frac{2\nu_1}{d}}$ holds. This implies $m \leq C_{14}(\log n)^{\frac{\nu_1+d/2}{2\nu_1}}$. Since we require $\log m > 1$, n should satisfy $n > \exp((e/C_{14})^{\frac{2\nu_1}{\nu_1+d/2}}) =: N_0$. Plugging $m \leq C_{14}(\log n)^{\frac{\nu_1+d/2}{2\nu_1}}$ in (S7.8), (S7.9), and (S7.5), we finish the proof. □

S8. Sample path

S8.1 Linear kernel

The focus of this section is to study how the unknown parameters in the proposed linear kernel (3.1) affect the generated sample paths. We focus on the Matérn kernel function, which has the form of (3.6). Three types of unknown parameters are studied, including the d positive diagonal elements of the diagonal matrix Θ, denoted by $\boldsymbol{\theta}$, the positive scalar σ^2, and the smoothness parameter ν.

The sample paths are generated by the input functions $g(x) = \sin(\alpha x)$, where $x \in \Omega = [0, 2\pi]$ and $\alpha \in [0, 1]$. The value α indicates the frequency of the periodic function and the RKHS norm of g, i.e., $\|g\|_{\mathcal{N}_\Psi(\Omega)}$, increases monotonically with respect to α. As a result, this input function creates an analogy to the sample paths in conventional GP by studying the paths as a function of α with different parameter settings. In Figure S1, the sample paths are demonstrated with different settings of the three types of parameters. The first row illustrates the sample paths with three different settings of ν, given $\theta = 1$ and $\sigma^2 = 1$. It appears that the smoothness of the resulting sample paths is not significantly affected by the setting of ν, which typically controls the smoothness in conventional GPs. This is mainly because, unlike the conventional Matérn kernel function (Stein, 1999), the derivative of (3.1) with respect to the input

g is not directly related to the parameter ν. The middle panels in Figure S1 demonstrate the sample paths with different settings of θ, given $\nu = 2.5$ and $\sigma^2 = 1$. It shows that, as θ increases, the number of the local maxima and minima increases, which agrees with the observations in conventional GPs. Lastly, the bottom three panels show the sample paths with different settings of σ^2, given $\nu = 2.5$ and $\theta = 1$. Similar to conventional GPs, σ^2 controls the amplitude of the resulting sample paths.

S8.2 Nonlinear kernel

Based on the nonlinear kernel of (3.12) with the Matérn kernel function defined in (3.7), the sample paths are studied with respect to different settings of the γ, ν and σ^2. As shown in Figure S2, the results appear to be consistent with the observations in conventional GPs where ν controls the smoothness of the function, θ controls the number of the local maxima and minima, and σ^2 controls the amplitude of the functions.

S9. Supporting tables and figures in Sections 4 and 5

The tables and figures that present the results in Sections 4 and 5 are provided in this section.

$g(\mathbf{x})$	x_1+x_2	x_1^2	x_2^2	$1+x_1$	$1+x_2$	$1+x_1x_2$	$\sin(x_1)$	$\cos(x_1+x_2)$
$f_1(g)$	1	0.33	0.33	1.5	1.5	1.25	0.46	0.50
$f_2(g)$	1.5	0.14	0.14	3.75	3.75	2.15	0.18	0.26
$f_3(g)$	0.62	0.19	0.19	0.49	0.49	0.84	0.26	0.33

Table S1: *Training data set for the numerical study, where* $f_1(g) = \int_\Omega \int_\Omega g(\mathbf{x})\mathrm{d}x_1\mathrm{d}x_2$, $f_2(g) = \int_\Omega \int_\Omega g(\mathbf{x})^3\mathrm{d}x_1\mathrm{d}x_2$, *and* $f_3(g) = \int_\Omega \int_\Omega \sin(g(\mathbf{x})^2)\mathrm{d}x_1\mathrm{d}x_2$.

Measurements	Method	$f_1(g) = \int_\Omega \int_\Omega g$	$f_2(g) = \int_\Omega \int_\Omega g^2$	$f_3(g) = \int_\Omega \int_\Omega \sin(g)$
MSE	FIGP	$\mathbf{6.4 \times 10^{-10}}$	**0.012**	**0.016**
	FPCA	1.8×10^{-4}	0.124	0.023
	T3	0.093	1.271	0.047
Coverage (%)	FIGP	**96.33**	100	**100**
	FPCA	100	**92.33**	76.00
	T3	100	98.33	**100**
Score	FIGP	**14.899**	**2.571**	**3.458**
	FPCA	6.631	1.207	0.290
	T3	1.064	-1.364	2.047

Table S2: *Prediction results of the FIGP and basis-expansion approach for the synthetic examples (`FPCA` indicates an FPCA expansion approach and `T3` indicates the Taylor series expansion of degree 3), including MSEs, average coverage rates of the 95% prediction intervals, and the average proper scores, in which the values with better performances are boldfaced.*

References

Tuo, R. and Wang, W. (2020). Kriging prediction with isotropic Matérn correlations: robustness and experimental designs. *Journal of Machine Learning Research*, 21(187):1–38.

Wendland, H. (2004). *Scattered Data Approximation*. Cambridge University Press.

Wu, Z. and Schaback, R. (1993). Local error estimates for radial basis function interpolation of scattered data. *IMA Journal of Numerical Analysis*, 13(1):13–27.

Abramowitz, M. and Stegun, I. A. (1948). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 55. US Government Printing Office.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media..

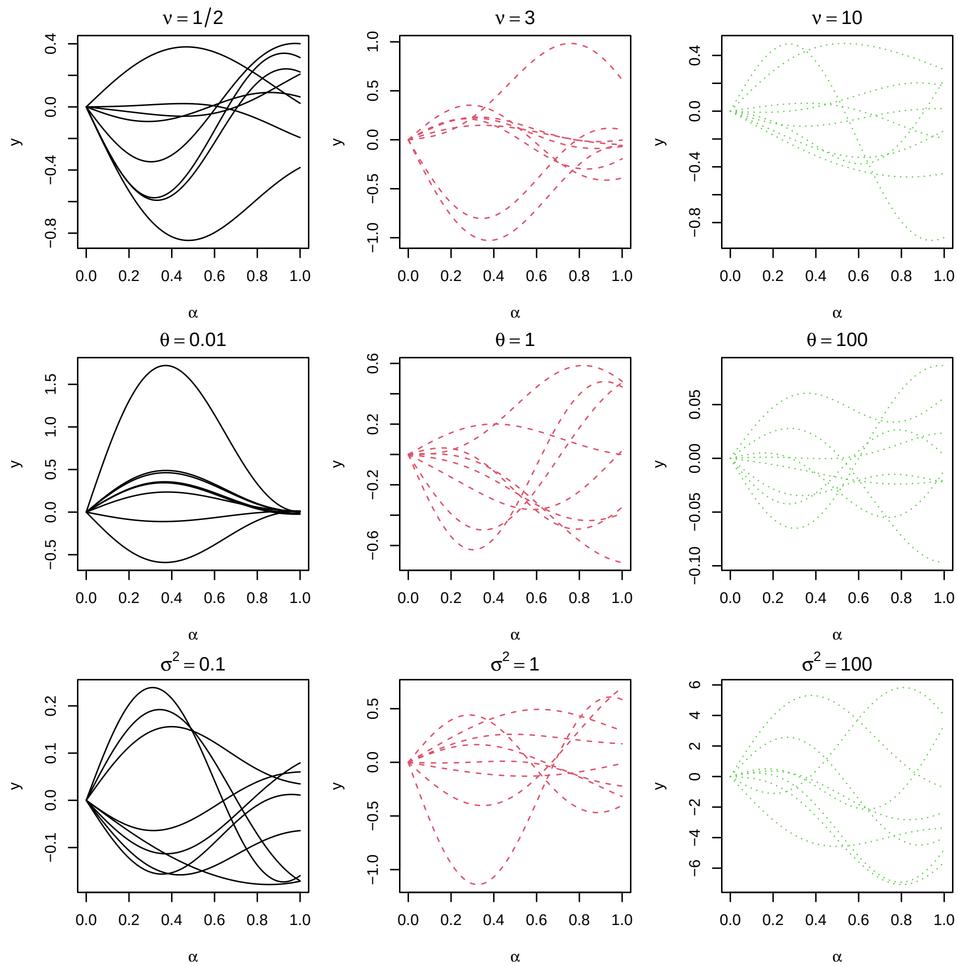

Figure S1: *Sample paths of linear kernels. Top panel shows the effect of varying the parameter* ν *with the fixed* $\theta = 1$ *and* $\sigma^2 = 1$*, middle panel shows the effect of varying the parameter* θ *with the fixed* $\nu = 2.5$ *and* $\sigma^2 = 1$*, and the bottom panel shows the effect of varying the parameter* σ^2 *with the fixed* $\nu = 2.5$ *and* $\theta = 1$.

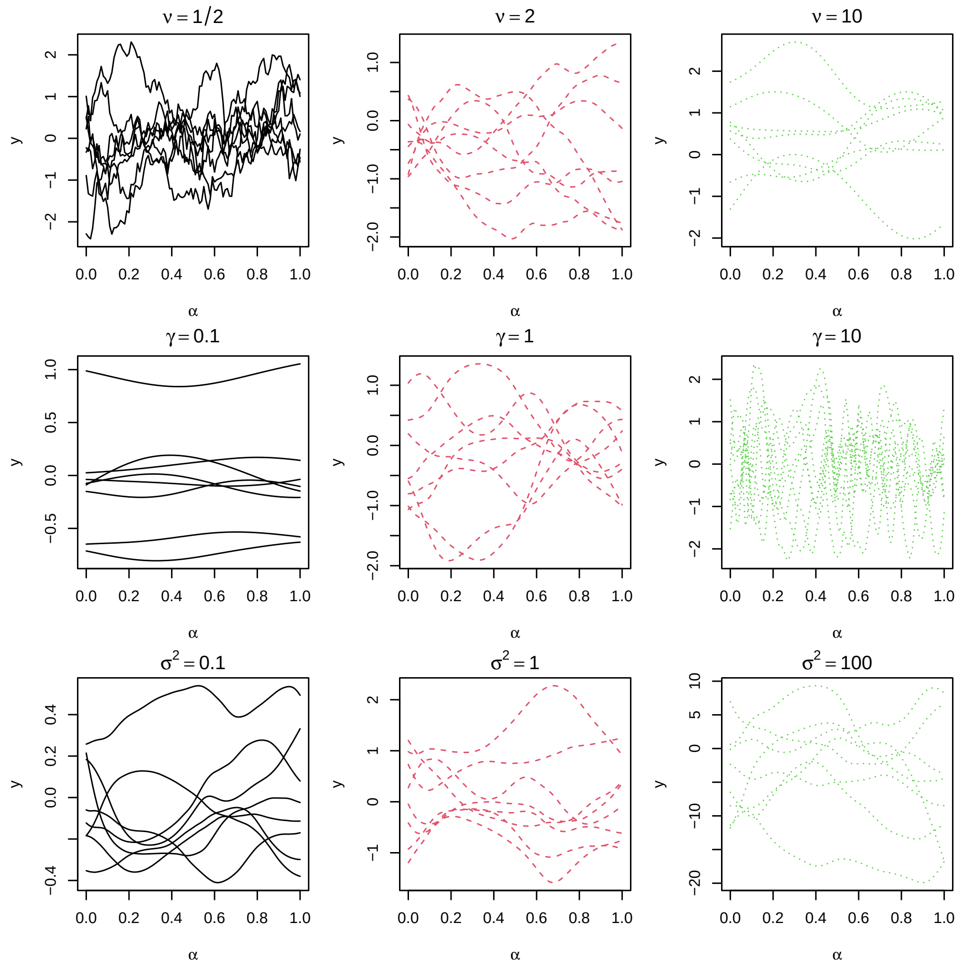

Figure S2: *Sample paths of nonlinear kernels. Top panel shows the effect of varying the parameter ν with the fixed $\gamma = 1$ and $\sigma^2 = 1$, middle panel shows the effect of varying the parameter γ with the fixed $\nu = 2.5$ and $\sigma^2 = 1$, and the bottom panel shows the effect of varying the parameter σ^2 with the fixed $\nu = 2.5$ and $\gamma = 1$.*

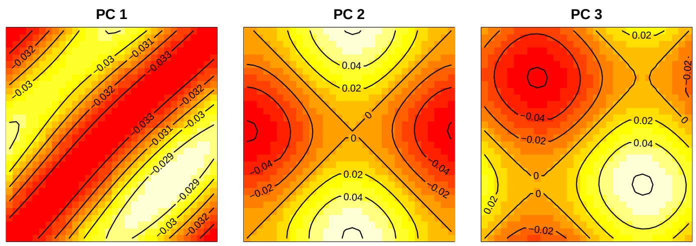

Figure S3: *Principle components, which explain more than 99.99% variations of the data.*

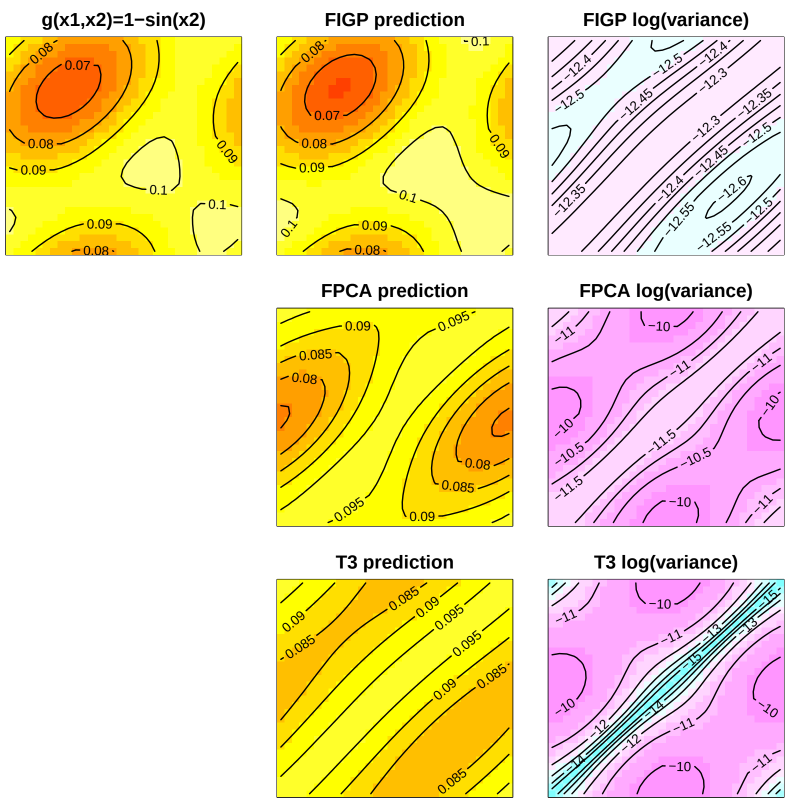

Figure S4: *Prediction on the validation function. The left panel is the true output of the functional input $g(x_1, x_2) = 1 - \sin(x_2)$, and the middle panels are the predictions of* `FIGP`*,* `FPCA`*, and* `T3`*, and the right panels are their variances in logarithm.*