| Statistica Sinica Preprint No: SS-2022-0180 | |
|---|---|
| **Title** | Functional-Input Gaussian Processes with Applications to Inverse Scattering Problems |
| **Manuscript ID** | SS-2022-0180 |
| **URL** | http://www.stat.sinica.edu.tw/statistica/ |
| **DOI** | 10.5705/ss.202022.0180 |
| **Complete List of Authors** | Chih-Li Sung,<br>Wenjia Wang,<br>Fioralba Cakoni,<br>Isaac Harris and<br>Ying Hung |
| **Corresponding Authors** | Chih-Li Sung |
| **E-mails** | sungchih@msu.edu |

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



# Functional-Input Gaussian Processes with Applications to Inverse Scattering Problems

Chih-Li Sung[1,*], Wenjia Wang[2,*], Fioralba Cakoni[3], Isaac Harris[4], Ying Hung[3]

[1]*Michigan State University*

[2]*The Hong Kong University of Science and Technology (Guangzhou)*

[3]*Rutgers, the State University of New Jersey,* [4]*Purdue University*

*Abstract:* Surrogate modeling based on Gaussian processes (GPs) is becoming increasingly popular in analysis of complex problems in science and engineering. However, despite the many studies on GP modeling, few focus on functional inputs. Motivated by an inverse scattering problem in which functional inputs representing the support and material properties of the scatterer are involved in the partial differential equations, we propose a new class of kernel functions for functional inputs of GPs. Based on the proposed GP models, we derive the asymptotic convergence properties of the resulting mean squared prediction errors, and demonstrate the finite-sample performance using numerical examples. In the application to inverse scattering, we construct a surrogate model with functional inputs, which is crucial to recovering the reflective index of an inhomogeneous isotropic scattering region of interest for a given far-field pattern.

*Key words and phrases:* Computer experiments, surrogate model, uncertainty quantification, scalar-on-function regression, functional data analysis

*These authors contributed equally to the manuscript.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



## 1. Introduction

Computer experiments that study real systems using mathematical models, such as partial differential equations, are increasingly being used to analyze complex problems in science and engineering. Such experiments typically require a great deal of time and computing resources. Therefore, based on a finite sample of computer experiments, it is crucial to build a surrogate for the actual mathematical models, which we then use for prediction, inference, and optimization. The Gaussian process (GP) model, also called kriging, is popular as a surrogate model because of its flexibility, interpolating property, and ability to perform uncertainty quantification using the predictive distribution. For additional discussions on computer experiments and surrogate modeling using GP models, see Santner et al. (2018) and Gramacy (2020).

This study is motivated by an inverse scattering problem in computer experiments that involve functional inputs and, therefore, the analysis and inference rely on a surrogate model that can take functional inputs into account. Figure 1 illustrates inverse scattering. Let the functional input $g$ represent the material properties of an inhomogeneous isotropic scattering region of interest, shown in the middle of Figure 1. For a given functional input, the far-field pattern, $u^s$, is obtained by solving partial differential equations (Cakoni et al., 2016), which is computationally intensive. Given a new far-field pattern, the goal of inverse

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



scattering is to recover the functional input using a surrogate model. To achieve this, we require a surrogate model applicable to functional inputs. In addition to inverse scattering (Cakoni et al., 2016; Kaipio et al., 2019), problems with functional inputs are frequently found in engineering applications of non-destructive testing, where measurements on the surface or exterior of an object are used to infer the interior structure. Similar problems also occur in electrical impedance tomography, where we need to recover the functional input that represents the electric conductivity from the measured current-to-voltage mapping; see, for example, Mueller and Siltanen (2020) for a discussion of the electrical impedance tomography model. Another important application is the use of computerized tomography in medical studies for interior reconstruction (Courdurier et al., 2008; Li et al., 2019).

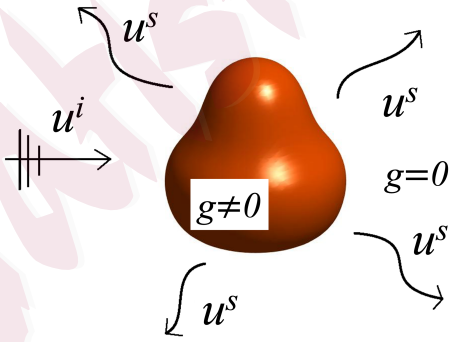


**Figure 1:** *Illustration of the inverse scattering problem.*

However, despite there being extensive studies on surrogate modeling using

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



GPs (Gramacy, 2020), few of them focus on functional inputs. To the best of our knowledge, most of the existing research on GPs with functional inputs is restricted to specific applications. For example, Nguyen and Peraire (2015) propose a functional-input GP with bilinear covariance operators, and apply it to linear partial differential equations. Morris (2012) develops a kriging model with a covariance function specifically for time-series data. Chen et al. (2021) propose a spectral-distance correlation function, and apply it to three-dimensional printing.

In functional data analysis, research that involves functional inputs is often referred to as *scalar-on-function regression* (Ramsay and Silverman, 2005; Kokoszka and Reimherr, 2017; Reiss et al., 2017). Some approaches reduce the dimension of functional inputs by using a basis-expansion approximation, and then perform a linear or nonlinear model in the reduced Euclidean space (see, e.g., Cardot et al. (1999); Ait-Saïdi et al. (2008); Yao and Müller (2010); Müller et al. (2013); McLean et al. (2014)). Other methods handle the functional inputs directly using spline approaches (see, e.g., Ferraty and Vieu (2006); Preda (2007); Baíllo and Grané (2009); Shang (2013). However, very few of these approaches incorporate GP assumptions that allow for uncertainty quantification when constructing surrogate models.

This study introduces a new class of GP surrogate models for functional

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



inputs. Recent studies on surrogate modeling apply GP to functional inputs based on a truncated basis expansion (Shi and Wang, 2008; Tan, 2019; Li and Tan, 2022). Ideas along this line are intuitive and easy to implement; however, there are three drawbacks. First, a basis expansion requires an explicit specification of the basis functions. Second, a basis expansion approximates the functional input and achieves dimension reduction by using a finite truncation of the basis functions, which can introduce additional bias into the model. Third, the curse of dimensionality makes it difficult to scale the techniques developed using a basis expansion to include high-dimensional functional inputs.

To address these problems, we propose a new GP surrogate by introducing a new class of kernel functions that are defined directly on a functional space. We show that the proposed kernels are closely connected to the idea of a basis expansion, without needing to specify individual bases, and without the loss of efficiency due to finite truncation. The procedure is general and provides a parsimonious model, especially for high-dimensional problems, in which basis-expansion approaches often require a significant number of basis functions to achieve a high quality approximation. We use simulations to compare the proposed method with those based on a basis expansion for functional inputs, and apply the proposed method to the inverse scattering problem. Our empirical results show that the proposed surrogate model outperforms those based on a

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



basis expansion in terms of prediction accuracy and uncertainty quantification.

Although the proposed surrogate models extend conventional GPs to functional inputs, the theoretical results are nontrivial extensions. These results include the convergence rates of the mean squared prediction errors (MSPEs) and the connections to the experimental design. Defining the kernels directly on a functional space reduces the model bias compared with using a basis expansion, but introduces technical challenges to the theoretical derivations. Additional scattered data approximation techniques, such as the local polynomial reproduction (Wendland, 2004), have to be applied rigorously to the study of convergence rates. These rates are further explored using the notion of fill distances, which provides a concrete connection between the performance of the proposed model and the experimental design in a functional space.

The remainder of the paper is organized as follows. In Section 2, we introduce a functional-input GP model. Then, in Section 3, we discuss a new class of kernel functions, including a linear and a nonlinear kernel, and their theoretical properties. In Section 4, we use a numerical analysis to examine the prediction accuracy of the proposed models. In Section 5, we apply the proposed framework to construct a surrogate model for an inverse scattering problem. Section 6 concludes the paper. Detailed theoretical proofs and the data and R code needed to reproduce the numerical results are provided in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



## 2. Functional-Input GP

Suppose that $V$ is a functional space consisting of functions defined on a compact and convex region $\Omega \subseteq \mathbb{R}^d$, and all functions $g \in V$ are continuous on $\Omega$, that is, $V \subset C(\Omega)$. A functional-input GP, $f : V \to \mathbb{R}$, is denoted by

$$f(g) \sim \mathcal{FIGP}(\mu, K(g, g')), \tag{2.1}$$

where $\mu$ is an unknown mean, and $K(g, g')$ is a semi-positive kernel function for $g, g' \in V$. A new class of kernel $K(g, g')$ for functional inputs is discussed in Section 3.

Given a properly defined kernel function, the estimation and prediction procedures are similar to those of the conventional GP. Assume there are $n$ realizations from the functional-input GP, where $g_1, \ldots, g_n$ are the inputs and $f(g_1), \ldots, f(g_n)$ are the outputs. We have $f(g_1), \ldots, f(g_n)$ following a multivariate normal distribution, $\mathcal{N}_n(\boldsymbol{\mu}_n, \mathbf{K}_n)$, with mean $\boldsymbol{\mu}_n = \mu \mathbf{1}_n$ and covariance $\mathbf{K}_n$, where $\mathbf{1}_n$ is a size-$n$ all-ones vector and $(\mathbf{K}_n)_{j,k} = K(g_j, g_k)$. The unknown parameters, including $\mu$ and the hyperparameters associated with the kernel function, can be estimated using likelihood-based or Bayesian approaches; for details of the estimation methods, refer to Santner et al. (2018) and Gramacy (2020).

Suppose $g \in V$ is an untried new function. By the property of the conditional

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



multivariate normal distribution, the corresponding output $f(g)$ follows a normal distribution with the mean and variance given by

$$\mathbb{E}[f(g)|\mathbf{y}_n] = \mu + \mathbf{k}_n(g)^T \mathbf{K}_n^{-1}(\mathbf{y}_n - \boldsymbol{\mu}_n) \quad \text{and} \tag{2.2}$$

$$\mathbb{V}[f(g)|\mathbf{y}_n] = K(g,g) - \mathbf{k}_n(g)^T \mathbf{K}_n^{-1}\mathbf{k}_n(g), \tag{2.3}$$

respectively, where $\mathbf{y}_n = (y_1, ..., y_n)^T$, $y_i = f(g_i)$, and $\mathbf{k}_n(g) = (K(g,g_1), ..., K(g,g_n))^T$. The conditional mean of (2.2) is used to predict $f(g)$, and the conditional variance of (2.3) can be used to quantify the prediction uncertainty.

## 3. A New Class of Kernel Functions

In this section, we introduce a new class of kernel functions for functional-input GPs. Based on the proposed models, we also derive the asymptotic convergence properties of the resulting MSPEs. Section 3.1 focuses on a linear kernel, and Section 3.2 extends the discussion to a nonlinear kernel. Practical guidance on selecting an optimal kernel is provided in Section 3.3. For notational simplicity, the mean in (2.1) is assumed to be zero in this section, but the results can be extended easily to include nonzero cases.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



### 3.1 Linear kernel for functional inputs

We first introduce a *linear kernel* for functional inputs $g_1$ and $g_2$:

$$K(g_1, g_2) = \int_\Omega \int_\Omega g_1(\mathbf{x}) g_2(\mathbf{x}') \Psi(\mathbf{x}, \mathbf{x}') \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{x}', \tag{3.1}$$

where $g_1, g_2 \in V$, and $\Psi$ is a positive-definite function defined on $\Omega \times \Omega$. The following proposition shows that this kernel function is semi-positive definite.

**Proposition 1.** *The linear kernel $K$ defined in (3.1) is semi-positive definite on $V \times V$.*

By Mercer's theorem (Rahman, 2007), we have

$$\Psi(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^{\infty} \lambda_j \phi_j(\mathbf{x}) \phi_j(\mathbf{x}'), \tag{3.2}$$

where $\mathbf{x}, \mathbf{x}' \in \Omega$, and $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ and $\{\phi_k\}_{k \in \mathbb{N}}$ are the eigenvalues and the orthonormal basis, respectively, in $L_2(\Omega)$. Given the positive-definite function $\Psi$, we can construct a GP using the Karhunen–Loève expansion:

$$f(g) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \langle \phi_j, g \rangle_{L_2(\Omega)} Z_j, \tag{3.3}$$

where $Z_j$ are independent standard normal random variables, and $\langle \phi_j, g \rangle_{L_2(\Omega)}$ is

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



the inner product of $\phi_j$ and $g$, which is $\langle \phi_j, g \rangle_{L_2(\Omega)} = \int_\Omega \phi_j(\mathbf{x}) g(\mathbf{x}) \mathrm{d}\mathbf{x}$. It can be shown that the covariance function of the constructed GP in (3.3) is $K(g_1, g_2)$, defined in (3.1), that is,

$$\begin{aligned} \mathrm{Cov}(f(g_1), f(g_2)) &= \sum_{j=1}^{\infty} \lambda_j \langle \phi_j, g_1 \rangle_{L_2(\Omega)} \langle \phi_j, g_2 \rangle_{L_2(\Omega)} \\ &= \sum_{j=1}^{\infty} \lambda_j \int_\Omega \int_\Omega g_1(\mathbf{x}) \phi_j(\mathbf{x}) g_2(\mathbf{x}') \phi_j(\mathbf{x}') \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{x}' \\ &= \int_\Omega \int_\Omega g_1(\mathbf{x}) g_2(\mathbf{x}') \Psi(\mathbf{x}, \mathbf{x}') \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{x}', \end{aligned} \tag{3.4}$$

for any $g_1, g_2 \in V$.

The proposed surrogate model is equivalent to a basis expansion as a result of the Karhunen–Loève expansion in (3.3), but the proposed method requires only a specification of the kernel function in (3.1), rather than an explicit specification of each individual basis $\phi_j$. Furthermore, we do not apply any dimension reduction or approximation to the functional input, and thus no additional bias is introduced to the surrogate. More specifically, the proposed model preserves the most information without a finite truncation of a basis expansion, because the kernel representation (3.1) is equivalent to representing each input $g$ as an element in $L_2(\Omega)$ using a basis expansion with respect to $\{\phi_j\}_{j=1}^{\infty}$. These advantages are common in kernel-based methods, such as support-vector machines (SVMs), the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



kernel principal components analysis (KPCA), and the kernel ridge regression (KRR) (Hastie et al., 2009).

**Proposition 2.** *The GP, $f(g)$, constructed as in (3.3), is linear; that is, for any $a, b \in \mathbb{R}$ and $g_1, g_2 \in V$, it follows that $f(ag_1 + bg_2) = af(g_1) + bf(g_2)$.*

The proposed kernel function has an intuitive interpretation that connects it to Bayesian modeling. In a Bayesian linear regression, the conditional mean function is assumed to be $f(\mathbf{x}) = \mathbf{x}^T\mathbf{w}$, where $\mathbf{w}$ is typically assumed to have a multivariate normal prior, that is, $\mathbf{w} \sim \mathcal{N}(0, \Sigma_{\mathbf{w}})$. Hence, for any two points, $\mathbf{x}$ and $\mathbf{x}'$, the covariance of $f(\mathbf{x})$ and $f(\mathbf{x}')$ is $\text{Cov}(f(\mathbf{x}'), f(\mathbf{x}')) = \mathbf{x}^T\Sigma_{\mathbf{w}}\mathbf{x}'$, which can be interpreted as a *weighted* inner product of $\mathbf{x}$ and $\mathbf{x}'$. The proposed model (3.3) can be viewed as an analogy to the Bayesian linear model with functional inputs, and the covariance (3.4) can be viewed as a weighted inner product of the two functions $g_1$ and $g_2$.

To understand the performance of the proposed predictor of (2.2) with the kernel function defined in (3.1), we first characterize the MSPE in the following theorem. Denote the reproducing kernel Hilbert space (RKHS) associated with a kernel $\Psi$ as $\mathcal{N}_\Psi(\Omega)$.

**Theorem 1.** *Let $\hat{f}(g) = \mathbb{E}[f(g)|\mathbf{y}_n]$ as in (2.2). For any continuous function*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



*$g \in V \subset L_2(\Omega)$, define a linear operator on $L_2(\Omega)$,*

$$\mathcal{T}g(\mathbf{x}) = \int_{\Omega} g(\mathbf{x}')\Psi(\mathbf{x}, \mathbf{x}')\mathrm{d}\mathbf{x}'.$$

*The MSPE of $\hat{f}(g)$ can be written as*

$$\mathbb{E}\left(f(g) - \hat{f}(g)\right)^2 = \min_{(u_1,\ldots,u_n)\in\mathbb{R}^n} \left\| \mathcal{T}g - \sum_{j=1}^{n} u_j \mathcal{T}g_j \right\|^2_{\mathcal{N}_\Psi(\Omega)}, \qquad (3.5)$$

*where $\|\cdot\|_{\mathcal{N}_\Psi(\Omega)}$ is the RKHS norm of $\mathcal{N}_\Psi(\Omega)$.*

By Proposition 10.28 in Wendland (2004), $\mathcal{T}g \in \mathcal{N}_\Psi(\Omega)$, and therefore the right-hand side of (3.5) exists. Theorem 1 provides a new representation of the MSPE for functional-input GPs that is an analogue to that of conventional GPs in the $L_2$ input space, and has not yet been explored in the literature. According to Theorem 1, the MSPE can be represented as the distance between $\mathcal{T}g$ and its projection on the linear space spanned by $\{\mathcal{T}g_1, ..., \mathcal{T}g_n\}$. This distance can be reduced if $g_j$ is designed so that the spanned space well approximates the space $V$. We highlight some designs of $g_j$ in the following two corollaries, which discuss the MSPE convergence rates explicitly.

In the following corollaries, the kernel function $\Psi$ is assumed to be a Matérn

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



kernel (Stein, 1999):

$$\Psi(\mathbf{x}, \mathbf{x}') = \psi(\|\Theta(\mathbf{x} - \mathbf{x}')\|_2) \quad \text{with} \tag{3.6}$$

$$\psi(r) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(2\sqrt{\nu}r)^{\nu}B_{\nu}(2\sqrt{\nu}r), \tag{3.7}$$

where $\Theta$ is a lengthscale parameter that is a $d \times d$ positive diagonal matrix, $\|\cdot\|_2$ denotes the Euclidean norm, $\sigma^2$ is a positive scalar, $B_\nu$ is the modified Bessel function of the second kind, and $\nu$ represents a smoothness parameter. We consider the Matérn kernel here because it is widely used in the computer experiment and spatial statistics literature (Santner et al., 2003; Stein, 1999). The corollaries can also be extended to a general positive kernel that has $k$ continuous derivatives, such as Wendland's compactly supported kernel (Wendland, 2004); see Wendland (2004) and Haaland and Qian (2011). Without loss of generality, we assume that $\Theta$ is an identity matrix and $\sigma^2 = 1$ for the theoretical developments in this section. These parameters, including $\Theta, \sigma^2$, and $\nu$, are discussed in greater detail in Section 4.

**Corollary 1.** *Suppose $g_j$, for $j = 1, \ldots, n$, are the first $n$ eigenfunctions of $\Psi$, that is, $g_j = \phi_j$. For $g \in V \subset L_2(\Omega)$, there exists a constant $C_1 > 0$ such that*

$$\mathbb{E}\left(f(g) - \hat{f}(g)\right)^2 \leq C_1\|g\|^2_{L_2(\Omega)}n^{-\frac{2\nu}{d}}. \tag{3.8}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



*Furthermore, if* $g \in \mathcal{N}_\Psi(\Omega)$*, then there exists a constant* $C_2 > 0$ *such that*

$$\mathbb{E}\left(f(g) - \hat{f}(g)\right)^2 \leq C_2 \|g\|^2_{\mathcal{N}_\Psi(\Omega)} n^{-\frac{4\nu}{d}}. \tag{3.9}$$

Corollary 1 represents the convergence rate analogue to that of conventional GPs (Tuo and Wang, 2020), and shows that if we design the input functions to be eigenfunctions of $\Psi$, the convergence rate of the MSPE is polynomial. If the functional space is further assumed to be the RKHS associated with the kernel $\Psi$ (i.e., $g \in \mathcal{N}_\Psi(\Omega)$), which is smaller than $L_2(\Omega)$, the convergence rate becomes faster, as shown by (3.9). This result indicates a significant difference between the proposed GP defined on a functional space and the conventional one defined on a Euclidean space. That is, the convergence results of (3.8) and (3.9) depend on the norm of the functional space in which the input $g$ lies, which is different to that of conventional GPs, which involves only the Euclidean norm.

Instead of selecting the input functions to be eigenfunctions, an alternative is to design the input functions using a set of *knots* in $\Omega$, that is, $\mathbf{X}_n \equiv \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, where $\mathbf{x}_j \in \Omega$, for $j = 1, \ldots, n$, and the convergence rate is derived in the following corollary. We first denote $h_{\mathbf{X}_n,\Omega}$ as the *fill distance* of $\mathbf{X}_n$, that is,

$$h_{\mathbf{X}_n,\Omega} := \sup_{\mathbf{x}\in\Omega} \min_{\mathbf{x}_j\in\mathbf{X}_n} \|\mathbf{x} - \mathbf{x}_j\|_2.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



Furthermore, denote $q_{\mathbf{X}_n} = \min_{1\le j\neq k\le n} \|\mathbf{x}_j - \mathbf{x}_k\|/2$, and a design $\mathbf{X}_n$ satisfying $h_{\mathbf{X}_n,\Omega}/q_{\mathbf{X}_n} \le C'$, for some constant $C'$, is called a *quasi-uniform design*.

**Corollary 2.** *(1) Suppose* $g_j(\mathbf{x}) = \Psi(\mathbf{x}, \mathbf{x}_j)$, *where* $\mathbf{x}, \mathbf{x}_j \in \Omega$, *for* $j = 1, \ldots, n$. *For* $g \in \mathcal{N}_\Psi(\Omega)$, *there exists a constant* $C_3 > 0$ *such that*

$$\mathbb{E}\left(f(g) - \hat{f}(g)\right)^2 \le C_3 \|g\|^2_{\mathcal{N}_\Psi(\Omega)} h^{2\nu}_{\mathbf{X}_n,\Omega}.$$

*(2) For a quasi-uniform design* $\mathbf{X}_n$, *there exists a positive constant* $C$ *such that* $h_{\mathbf{X}_n,\Omega} \le C n^{-1/d}$ (*Wendland, 2004; Müller, 2009*). *Therefore, there exists a constant* $C_4 > 0$ *such that*

$$\mathbb{E}\left(f(g) - \hat{f}(g)\right)^2 \le C_4 \|g\|^2_{\mathcal{N}_\Psi(\Omega)} n^{-\frac{2\nu}{d}}. \tag{3.10}$$

Compared with the results in Corollary 1, the convergence rate of the quasi-uniform designs in (3.10) is slower than the choice of eigenfunctions in (3.9). Nevertheless, despite the slower rate of convergence, designing functional inputs as $\Psi(\mathbf{x}_j, \cdot)$ with space-filling $\mathbf{x}_j$ can be relatively easier in practice than finding eigenfunctions of $\Psi$. However, if the eigenfunctions of $\Psi$ are available, then we recommend the design based on Corollary 1 (i.e., the first $n$ eigenfunctions), because the convergence rate is faster. Some kernel functions have closed-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



form expressions, such as Gaussian kernels (Zhu et al., 1997). More generally, the eigenfunctions can be approximated numerically using Nyström's method (Williams and Seeger, 2000).

The proposed linear kernel can be naturally modified to accommodate the potential nonlinearity in $f$ by enlarging the feature space using a prespecified nonlinear transformation $\mathcal{M}$ on $g$, that is, $\mathcal{M} : V \to V_1$, where $V_1$ is a function class. The resulting kernel function can be written as

$$K(g_1, g_2) = \int_\Omega \int_\Omega \mathcal{M} \circ g_1(\mathbf{x}) \mathcal{M} \circ g_2(\mathbf{x}') \Psi(\mathbf{x}, \mathbf{x}') \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{x}',$$

and the corresponding GP can be constructed as

$$f(g) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \langle \phi_j, \mathcal{M} \circ g \rangle_{L_2(\Omega)} Z_j. \tag{3.11}$$

The MSPE convergence results can be extended to (3.11). There are many possible ways to define $\mathcal{M}$ so that the feature space can be enlarged; however, this flexibility comes with a higher estimation complexity. In the next section, we propose an alternative to address the nonlinearity using a kernel function, which is computationally more efficient.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



### 3.2 Nonlinear kernel for functional inputs

In this section, we introduce a new type of kernel function for functional inputs that takes the nonlinearity into account by using a radial basis function. Let $\psi(r) : \mathbb{R}^+ \to \mathbb{R}$ be a radial basis function, the corresponding kernel of which in $\mathbb{R}^d$ is strictly positive definite for any $d \geq 1$. Note that the radial basis function of (3.7), which has a Matérn kernel, satisfies this condition. Define $K : V \times V \to \mathbb{R}$ as

$$K(g_1, g_2) = \psi(\gamma \|g_1 - g_2\|_{L_2(\Omega)}), \tag{3.12}$$

where $\|\cdot\|_{L_2(\Omega)}$ is the $L_2$-norm of a function, defined by $\|g\|_{L_2(\Omega)} = (\langle g, g \rangle_{L_2(\Omega)})^{1/2}$, and $\gamma > 0$ is a parameter that controls the decay of the kernel function with respect to the $L_2$-norm.

Although other distance metrics can be used to define the distance between two functions, such as the Fréchet distance or $L_\infty$-norm, the resulting kernel functions are not necessarily semi-positive definite, which is a required property when defining a kernel function. For example, consider an $L_\infty$-norm distance for the kernel function, that is, $K(g_1, g_2) = \psi(\gamma \|g_1 - g_2\|_{L_\infty(\Omega)})$, for any $g_1, g_2 \in L_\infty(\Omega)$, where $\psi$ has the form of (3.7), with $\nu = 2.5$ and $\sigma^2 = 1$. Given the four training functional inputs, $g_1(x_1, x_2) = x_1^2, g_2(x_1, x_2) = x_2^2, g_3(x_1, x_2) =$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



$1 + x_1, g_4(x_1, x_2) = 1 + x_2$, and $\gamma = 0.5$, the kernel matrix is

$$\mathbf{K}_n = \begin{bmatrix} 1 & 0.8286 & 0.7536 & 0.5240 \\ 0.8286 & 1 & 0.5240 & 0.7536 \\ 0.7536 & 0.5240 & 1 & 0.8286 \\ 0.5240 & 0.7536 & 0.8286 & 1 \end{bmatrix}.$$

Then, for a vector $\mathbf{a} = (1, -1, -1, 1)^T$, it follows that $\mathbf{a}^T\mathbf{K}_n\mathbf{a} = -0.2331 < 0$, which implies that the kernel function is not semi-positive. Conditions on the metric $\|\cdot\|$ that make the resulting kernel function positive definite are left to future work. In the following proposition, we show that the kernel function with $\|\cdot\|_{L_2(\Omega)}$, defined as in (3.12), is positive definite.

**Proposition 3.** *The function $K$ defined in (3.12) is positive definite on $V \times V$.*

Assume that there exists a probability measure $P$ on $V$ such that $\int_V g(t)^2 dP(g) < \infty$, for $t \in \Omega$ (Ritter, 2007). Based on the positive-definite function $K$ in (3.12), we can construct a GP using the Karhunen–Loève expansion, as follows:

$$f(g) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\varphi_j(g)Z_j, \tag{3.13}$$

where $\varphi_j$ is the orthonormal basis obtained using a generalized version of Mercer's theorem, $K(g_1, g_2) = \sum_{j=1}^{\infty} \lambda_j\varphi_j(g_1)\varphi_j(g_2)$ (Steinwart and Scovel, 2012),

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



with respect to the probability measure $P$.

The nonlinear kernel in (3.12) can be viewed as a basis expansion of the functional input based on the fact that $\|g_1 - g_2\|^2_{L_2(\Omega)} = \sum_{j=1}^{\infty} \langle \phi_j, g_1 - g_2 \rangle^2_{L_2(\Omega)}$, where $\{\phi_j\}_{j=1}^{\infty}$ are orthonormal basis functions in $L_2(\Omega)$. By a finite truncation of the basis expansion, the input $g$ can be approximated by $\{\phi_j\}_{j=1}^{M}$ in $\mathbb{R}^M$, for a positive integer $M$, and, therefore, $f(g)$ can be approximated by a GP with the correlation function $\psi(\gamma\| \cdot \|_2)$, where $\| \cdot \|_2$ is the Euclidean norm on $\mathbb{R}^M$. However, as in Section 3.1, the finite truncation introduces additional bias and requires an explicit specification of the orthonormal functions $\{\phi_j\}_{j=1}^{M}$ and $M$ in advance. Instead, the proposed method directly evaluates the correlation on a functional space using the nonlinear kernel without an approximation, requiring only that we select a proper kernel function.

Note that the $L_2$-norm in (3.12) can be replaced by any Hilbert space norm, such as the RKHS norm. Therefore, the nonlinear kernel (3.12) is flexible, and can be generalized to any target space of interest, in practice. Nevertheless, the $L_2$-norm can be approximated using numerical integration methods, such as Monte Carlo integration (Caflisch, 1998), which is computationally more efficient than, for example, the RKHS norm, which requires inverting an $N \times N$ matrix, where $N$ is the number of grid points.

Based on the proposed nonlinear kernel, the next theorem gives the conver-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



gence rate of the MSPE.

**Theorem 2.** *Suppose that $\Phi$ is a Matérn kernel function with smoothness $\nu_1$, and $\psi$ is the radial basis function of (3.7), the corresponding kernel of which is Matérn with smoothness $\nu$. Let $\tau = \min(\nu, 1)$. For any $n > N_0$, with a constant $N_0$, there exist $n$ input functions such that for any $g \in \mathcal{N}_\Phi(\Omega)$ with $\|g\|_{\mathcal{N}_\Phi(\Omega)} \leq 1$, the MSPE is bounded by*

$$\mathbb{E}\left(f(g) - \hat{f}(g)\right)^2 \leq C_5 (\log n)^{-\frac{(\nu_1 + d/2)\tau}{d}} \log\log n. \tag{3.14}$$

Based on (3.14), it appears that the convergence rate is slower than that of the conventional GP, where the inputs are defined in the Euclidean space. Although this rate can be improved, a slower rate of convergence for functional inputs is not surprising, because the input space is much larger than the Euclidean space. Note that because the RKHS generated by a Matérn kernel function with smoothness $\nu_1$ is equivalent to the (fractional) Sobolev space $H^{\nu_1 + d/2}(\Omega)$ (Wendland, 2004), the assumption of $g \in \mathcal{N}_\Phi(\Omega)$ in Theorem 2 is equivalent to $g \in H^{\nu_1 + d/2}(\Omega)$.

If $\Phi$ is a squared exponential kernel, then the corresponding RKHS is within the RKHS generated by a Matérn kernel function with any smoothness $\nu_2$. Thus, one can choose a large $\nu_2 > \nu_1$, and apply Theorem 2 to obtain the same convergence rate as in (3.14) by replacing $\nu_1$ with $\nu_2$. Therefore, we conclude

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



that the convergence rate of the RKHS generated by a squared exponential kernel is faster than that of the RKHS generated by a Matérn kernel function with a fixed $\nu_1$.

### 3.3 Selection of kernels

We have shown that the linear kernel of (3.1) results in a less flexible model, leading to a lower prediction variance, but higher bias, whereas the nonlinear kernel of (3.12) results in a more flexible model, leading to a higher variance and lower bias (Hastie et al., 2009). To find an optimal kernel function that balances the bias–variance trade-off, we use cross-validation to select the kernel by minimizing the estimated prediction error.

Although cross-validation methods are typically expensive to implement, the leave-one-out cross-validation (LOOCV) of GP models can be expressed in a closed form, which makes the computation less demanding (Zhang and Wang, 2010; Rasmussen and Williams, 2006; Currin et al., 1988). Specifically, denote $\tilde{y}_i$ as the prediction mean based on all data except the $i$th observation, and $y_i$ as the real output of the $i$th observation. For a kernel candidate $K$, which can be either the linear kernel (3.1) or the nonlinear kernel (3.12), the LOOCV error is

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{y}_i)^2=\frac{1}{n}\|\mathbf{\Lambda}_n^{-1}\mathbf{K}_n^{-1}(\mathbf{y}_n-\boldsymbol{\mu}_n)\|_2^2, \tag{3.15}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



where $\mathbf{\Lambda}_n$ is a diagonal matrix with the element $(\mathbf{\Lambda}_n)_{j,j} = (\mathbf{K}_n^{-1})_{j,j}$. Thus, we can select the optimal of the linear and nonlinear kernels by minimizing the LOOCV error.

### 3.4 Generalization to multiple functional-input variables

The linear and nonlinear kernel functions developed in Sections 3.1 and 3.2, respectively, can be extended naturally to multiple functional-input variables. For example, we have two functional-input variables, $g \in V$ and $h \in V$, and we collect $n$ inputs, $\{(g_1, h_1), \ldots, (g_n, h_n)\}$. In such cases, the linear kernel (3.1) can be rewritten as

$$K((g_1, h_1), (g_2, h_2)) = \int_\Omega \int_\Omega \left(g_1(\mathbf{x})g_2(\mathbf{x}') + h_1(\mathbf{x})h_2(\mathbf{x}')\right) \Psi(\mathbf{x}, \mathbf{x}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}',$$

and the nonlinear kernel (3.12) can be rewritten as

$$K((g_1, h_1), (g_2, h_2)) = \psi\left(\left(\gamma_1\|g_1 - g_2\|^2_{L_2(\Omega)} + \gamma_2\|h_1 - h_2\|^2_{L_2(\Omega)}\right)^{1/2}\right),$$

where $\gamma_1, \gamma_2 > 0$ are the parameters.

The nonlinear kernel can also be generalized naturally to a mixture of functional inputs and scalar inputs. That is, suppose that in addition to the two functional-input variables, $g, h \in V$, there exists a scalar input variable in the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



experiment, denoted by $z \in \Omega' \subseteq \mathbb{R}$. Then, we can define a kernel function as

$$K((g_1, h_1, z_1), (g_2, h_2, z_2)) = \psi\left(\left(\gamma_1 \|g_1 - g_2\|^2_{L_2(\Omega)} + \gamma_2 \|h_1 - h_2\|^2_{L_2(\Omega)} + \gamma_3 (z_1 - z_2)^2\right)^{1/2}\right),$$

where $\gamma_3 > 0$.

## 4. Numerical Study

In this section, we use numerical experiments to examine the performance of the proposed method. In the Supplementary Material S8, we explore the sample paths of the functional-input GP with different parameter settings.

In these numerical studies, we use the quasi-Monte Carlo integration (Morokoff and Caflisch, 1995) to numerically evaluate the integrals in the kernels. Specifically, suppose that $\Omega$ is a unit cube. Then, the linear kernel (3.1) can be approximated by

$$K(g_1, g_2) \approx \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} g_1(\mathbf{x}_i) g_2(\mathbf{x}'_j) \Psi(\mathbf{x}_i, \mathbf{x}'_j), \tag{4.1}$$

where $\{\mathbf{x}_i\}_{i=1}^N$ and $\{\mathbf{x}'_j\}_{j=1}^N$ are low-discrepancy sequences from a unit cube, for which the Sobol sequence (Sobol', 1967; Bratley and Fox, 1988) is considered

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



here. The number of points in the sequence, $N = 5,000$, is set. Similarly, the $L_2$-norm in the nonlinear kernel (3.12) can be approximated by

$$\|g_1 - g_2\|_{L_2(\Omega)} \approx \left( \frac{1}{N} \sum_{i=1}^{N} (g_1(\mathbf{x}_i) - g_2(\mathbf{x}_i))^2 \right)^{1/2}. \tag{4.2}$$

We examine the prediction performance of the proposed method using three synthetic examples, namely, a linear operator, $f_1(g) = \int_\Omega \int_\Omega g(\mathbf{x}) \mathrm{d}x_1 \mathrm{d}x_2$, and two nonlinear operators, $f_2(g) = \int_\Omega \int_\Omega g(\mathbf{x})^3 \mathrm{d}x_1 \mathrm{d}x_2$ and $f_3(g) = \int_\Omega \int_\Omega \sin(g(\mathbf{x})^2) \mathrm{d}x_1 \mathrm{d}x_2$, where $\mathbf{x} = (x_1, x_2) \in \Omega \equiv [0,1]^2$ and $g(\mathbf{x}) : [0,1]^2 \to \mathbb{R}$. We consider eight functional inputs, shown in the first row of Table S1, for each of the synthetic examples; their outputs are given in Table S1.

Three types of functional inputs are tested for the predictions: $g_9(\mathbf{x}) = \sin(\alpha_1 x_1 + \alpha_2 x_2)$, $g_{10}(\mathbf{x}) = \beta + x_1^2 + x_2^3$, and $g_{11}(\mathbf{x}) = \exp\{-\kappa x_1 x_2\}$, where $\alpha_1, \alpha_2, \beta, \kappa \in [0,1]$. Based on 100 random samples of $\alpha_1, \alpha_2, \beta$, and $\kappa$ from $[0,1]$, we evaluate the prediction performance by averaging the mean squared errors (MSEs), where MSE $= \frac{1}{3} \sum_{j=9}^{11} (f(g_j) - \hat{f}(g_j))^2$.

For the proposed method, we use a Matérn kernel function with smoothness parameter $\nu = 5/2$, which leads to a simplified form of (3.7):

$$\psi(r) = \left(1 + \sqrt{5}r + \frac{5}{3}r^2\right) \exp\left(-\sqrt{5}r\right). \tag{4.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



| Measurements | Kernel | $f_1(g)$ | $f_2(g)$ | $f_3(g)$ |
|---|---|---|---|---|
| LOOCV | linear | $\mathbf{7.9 \times 10^{-7}}$ | 1.813 | 0.454 |
| | nonlinear | $2.2 \times 10^{-6}$ | **0.227** | **0.017** |
| MSE | linear | $\mathbf{6.4 \times 10^{-10}}$ | 1.087 | 0.140 |
| | nonlinear | $3.1 \times 10^{-7}$ | **0.012** | **0.016** |

**Table 1:** *The leave-one-out cross-validation (LOOCV) errors and the mean squared errors (MSEs) for three testing functions, where* $f_1(g) = \int_\Omega \int_\Omega g$, $f_2(g) = \int_\Omega \int_\Omega g^3$, *and* $f_3(g) = \int_\Omega \int_\Omega \sin(g^2)$. *The errors corresponding to the optimal kernel are boldfaced.*

Other parameters, including $\Theta, \sigma^2$, and $\gamma$, are estimated using a maximum likelihood estimation. We use both the linear kernel (3.1) and the nonlinear kernel (3.12) for the proposed functional input GP, and report their LOOCV errors in Table 1. Following Section 3.3, the LOOCV is then used to identify the optimal kernel. By minimizing the LOOCV errors, the linear kernel is identified as the optimal choice for the linear synthetic example, $f_1(g)$, and the nonlinear kernel is identified as the optimal choice for the nonlinear synthetic examples, $f_2(g)$ and $f_3(g)$. The MSEs for the three synthetic examples are summarized in Table 1. It appears that the optimal kernels selected using the LOOCV are consistent with the selections based on minimizing the MSEs, showing that the LOOCV is a reasonable indicator of the optimal kernel when the ground truth is unknown.

The computational cost is also assessed for the two kernel choices. The numerical experiments were performed on a MacBook Pro laptop with an Apple M1 Max Chip and 32 GB of RAM. The computation for the linear kernels in each

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



of the examples takes about 9 seconds, and that for the nonlinear kernels takes less than 1 second, indicating that the linear kernel requires a greater computation than the nonlinear kernel does. This is not surprising, because the computation for linear kernels involves double integrals (see (3.1)), which require $N^2$ evaluations for the quasi-Monte Carlo integration in (4.1), whereas the nonlinear kernel (see (3.12) and (4.2)) requires only $N$ evaluations. Furthermore, the linear kernel has $d$ lengthscale parameters that need to be estimated, whereas the nonlinear kernel has only one lengthscale parameter. Nonetheless, fitting the functional-input GP model is reasonably efficient with either a linear or a nonlinear kernel, both of which take less than 10 seconds.

As a comparison, we consider the *basis-expansion* approach discussed in Section 3.2. That is, consider a functional principal component analysis (`FPCA`) with *truncated* components (Rice and Silverman, 1991; Wang et al., 2016):

$$g_i(\mathbf{x}) \approx u(\mathbf{x}) + \sum_{j=1}^{M} z_{ij}\psi_j(\mathbf{x}),$$

with the leading $M$ eigenfunctions $\{\psi_j(\mathbf{x})\}_{j=1}^{M}$ and the corresponding coefficients

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



$\{z_{ij}\}$ given by

$$\psi_j(\mathbf{x}) = \underset{\substack{\|\phi\|_2=1,\\ \langle\phi,\psi_l\rangle=0,\forall l<j}}{\operatorname{argmax}} \sum_{i=1}^{n} \left\{ \int (g_i(\mathbf{x}) - u(\mathbf{x}))\, \phi(\mathbf{x}) d\mathbf{x} \right\}^2,$$

$$z_{ij} = \int (g_i(\mathbf{x}) - u(\mathbf{x}))\, \psi_j(\mathbf{x})\, d\mathbf{x}, \tag{4.4}$$

respectively, where $n = 8$ in this example. The number of components, $M = 3$, is chosen to explain 99.46% of the variance; see Wang et al. (2016) and Mak et al. (2018). Given the training input-output pairs, $\{\mathbf{z}_i, y_i\}_{i=1}^n$, where $\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,M})$, we use a conventional GP (with a Matérn kernel) to fit the training data. The test input $\{\mathbf{z}_i\}_{i=9}^{11}$ can be obtained similarly using (4.4), and their outputs are predicted using the fitted GP.

In addition to the FPCA, we consider a Maclaurin series expansion of degree 3, which is a Taylor series expansion of a function evaluated at zero truncated to degree 3 (labeled `T3`). That is,

$$g_j(\mathbf{x}) \approx \sum_{\substack{a=0,b=0\\ a+b\leq 3}} \frac{\partial^{a+b} g_j(0,0)}{\partial x_1^a \partial x_2^b} x_1^a x_2^b.$$

The series expansion approximates the functional inputs of the examples reasonably well, with only a few nonzero coefficients. For example, the training functional input $g_1(\mathbf{x}) = x_1 + x_2$ has the coefficient one for $x_1$ and $x_2$, and zero

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



for other terms.

To evaluate the prediction performance and quantify the uncertainty, in addition to the MSEs, we consider two numerical measurements: the average coverage rate of the 95% prediction intervals, and the average proper scores. The coverage rate is the proportion of times that the interval contains the true value, and the proper score is the scoring rule of Gneiting and Raftery (2007), which is an overall measure of the accuracy of the combined prediction mean and variance predictions. Specifically, the proper score has the following form:

$$\text{proper score} = -\left(\frac{y-\mu_P}{\sigma_P}\right)^2 - \log\sigma_P^2,$$

where $y$ is the true output, $\mu_P$ is the predictive mean, and $\sigma_P^2$ is the predictive variance. A larger proper score indicates a better prediction. The results are summarized in Table S2, which shows that the proposed method, `FIGP`, outperforms the two basis-expansion approaches in terms of both predictions and uncertainty quantification. The average coverage rates of `FIGP` are close to the nominal coverage 95%, and the scores of `FIGP` are much higher than those of the two basis-expansion approaches.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



## 5. Applications to Inverse Scattering Problem

In this section, we revisit the inverse scattering problem shown in Figure 2. Let $D \subset \mathbb{R}^2$ denote an inhomogeneous isotropic scattering region of interest, and the functional input $g$, the support of which is $D$, is related to the refractive index for the region $D$ of the unknown scatterer. Given a set of finite-element simulations as the training data, the goal of inverse scattering is to recover the functional input from a given far-field pattern. To achieve this goal, an important task is to construct a surrogate model for the functional inputs.

In this study, 10 functional inputs, namely, $1, 1+x_1, 1-x_1, 1+x_1x_2, 1-x_1x_2, 1+x_2, 1+x_1^2, 1-x_1^2, 1+x_2^2$, and $1-x_2^2$, are used in the training set, and the corresponding far-field patterns are shown in Figure 2. Note that the inputs are given with explicit functional forms. In applications in which discrete realizations of the functions are available, the kernel functions can be approximated numerically using the discrete realizations in (4.1) and (4.2). As a preprocessing step, we reduce the dimension of the output images using a principal component analysis (PCA). The first three principal components, denoted by $\mathbf{u}_l \in \mathbb{R}^{1024}$, for $l = 1, 2, 3$, are shown in Figure S3, and explain more than 99.99% of the variability of the data. Therefore, given the functional input $g_i$, for $i = 1, \ldots, 10$, the output of the far-field images can be approximated by $\sum_{l=1}^{3} f_l(g_i)\mathbf{u}_l$, where $f_1(g_i), f_2(g_i), f_3(g_i)$ are the first three principal component scores.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



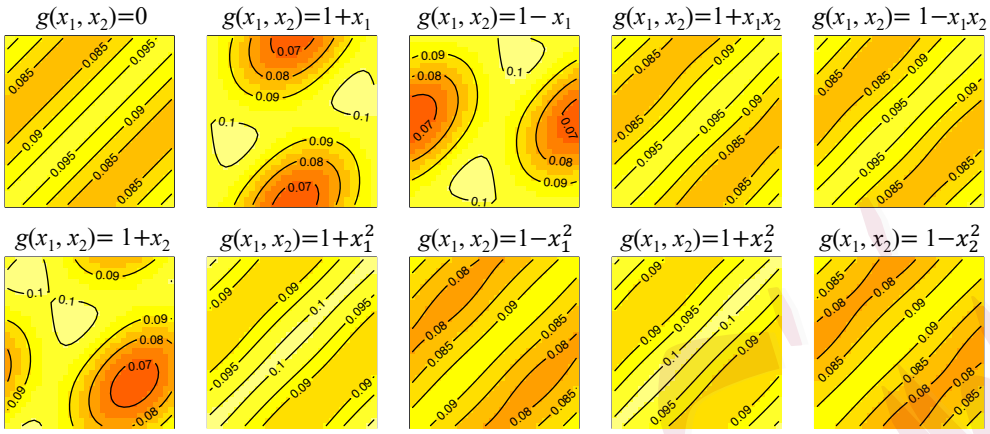


**Figure 2:** *Training data in the application of the inverse scattering problem.*

After the dimension reduction, we assume the three-dimensional outputs $f_1(g)$, $f_2(g)$, and $f_3(g)$ are mutually independent and follow the functional-input GP as the surrogate model. For any untried functional input $g \in V$, based on the results of (2.2) and (2.3), we predict the far-field pattern using the following normal distribution:

$$\mathcal{N}\left(\sum_{l=1}^{3}\mathbf{k}_l(g)^T\mathbf{K}_{n,l}^{-1}(\mathbf{f}_l-\mu_l\mathbf{1}_n)\mathbf{u}_l, \sum_{l=1}^{3}(K_l(g,g)-\mathbf{k}_l(g)^T\mathbf{K}_{n,l}^{-1}\mathbf{k}_l(g))\mathbf{u}_l\mathbf{u}_l^T\right),$$

where $\mathbf{f}_l = (f_l(g_1), \ldots, f_l(g_n))$, $\mathbf{k}_l(g) = (K_l(g, g_1), \ldots, K_l(g, g_n))^T$, $(\mathbf{K}_{n,l})_{i,j} = K_l(g_i, g_j)$, and $K_l$ is the kernel function with hyperparameters estimated based on $\mathbf{f}_l$.

We use both the linear kernel of (3.1) and the nonlinear kernel of (3.12). The optimal kernel is selected by comparing the LOOCV errors when predicting the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



far-field pattern. The LOOCV error based on the linear kernel is $3.6 \times 10^{-6}$, which is smaller than that of the nonlinear kernel, $1.2 \times 10^{-5}$. Therefore, we apply the linear kernel and examine its prediction performance for the test function, $g(\mathbf{x}) = 1 - \sin(x_2)$. Similarly to Section 4, we compare two basis-expansion approaches, `FPCA` and `T3`. The images of the true far-field patterns and their predictions, along with their variances (in logarithm), are illustrated in Figure S4. Compared with the ground truth, the predictions of `FIGP` capture the underlying structures reasonably well, with some discrepancies appearing on the lower right corner. On the other hand, the predictions of `FPCA` and `T3` both appear to deviate more from the ground truth. The MSEs and average scores are reported in Table 2, and show that the proposed method outperforms the two basis-expansion approaches in terms of prediction accuracy and uncertainty quantification.

| Measurements | `FIGP` | `FPCA` | `T3` |
|---|---|---|---|
| MSE | $\mathbf{1.10 \times 10^{-6}}$ | $1.07 \times 10^{-4}$ | $9.06 \times 10^{-5}$ |
| Score | **12.13** | 6.89 | 6.39 |

**Table 2:** *Prediction performance of the FIGP and basis-expansion approaches in the inverse scattering problem application (`FPCA` indicates an FPCA expansion approach and `T3` indicates the Taylor series expansion of degree 3), including MSEs and average proper scores, in which the values with better performance are boldfaced.*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



## 6. Conclusion

Although GP surrogates are widely used in analysis of complex systems as an alternative to direct analysis using computer experiments, the results of most existing studies do not apply to problems with functional inputs. To address this issue, we introduce two new types of kernel functions for functional-input GPs, namely, a linear and a nonlinear kernel. We also discuss the theoretical properties of the proposed surrogates, such as the convergence rate of the MSPE. The results of numerical studies and an application to surrogate modeling in an inverse scattering problem show that the proposed method exhibits high prediction accuracy.

There are extensive studies on experimental design for conventional GP surrogate models, but few on optimal designs for GPs with functional inputs. Here, we show that space-fillingness is a desirable property in terms of controlling the convergence rate of the MSPE. An interesting topic for future research is to explore the construction procedure for efficient space-filling designs with functional inputs. In addition to experimental designs, another important avenue for future work is to explore systematic approaches to efficiently identify the functional input, given an observed far-field pattern, which is the ultimate goal of inverse scattering problems. Based on the proposed GP surrogate, we aim to explore a Bayesian inverse framework that integrates computer experiments and

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



real observations. Lastly, even though the numerical studies in the Supplementary Material S8 indicate that the smoothness parameter $\nu$ in the linear kernel function does not have a significant effect on the sample paths, it is worth exploring the theoretical properties related to the choice of the parameter. This topic is left to future work.

## Supplementary Material

The online Supplementary Material includes the theoretical proofs for Propositions 1 and 3, Theorems 1 and 2, and Corollaries 1 and 2, the sample paths of the functional-input GP, the supporting tables and figures for Sections 4 and 5, and the data, and R code needed to reproduce the results in Sections 4 and 5.

## Acknowledgments

The authors gratefully acknowledge the constructive comments and suggestions of the associate editor and two anonymous reviewers. This work was supported by NSF DMS-2113407, DMS-2107891, CCF-1934924, and NSFC 12101149.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



## References

Ait-Saïdi, A., Ferraty, F., Kassa, R., and Vieu, P. (2008). Cross-validated estimations in the single-functional index model. *Statistics*, 42(6):475–494.

Baíllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response. *Journal of Multivariate Analysis*, 100(1):102–111.

Bratley, P. and Fox, B. L. (1988). Algorithm 659: Implementing Sobol's quasirandom sequence generator. *ACM Transactions on Mathematical Software (TOMS)*, 14(1):88–100.

Caflisch, R. E. (1998). Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7(1):1–49.

Cakoni, F., Colton, D., and Haddar, H. (2016). *Inverse Scattering Theory and Transmission Eigenvalues,CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: SIAM.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45(1):11–22.

Chen, J., Mak, S., Joseph, V. R., and Zhang, C. (2021). Function-on-function kriging, with applications to three-dimensional printing of aortic tissues. *Technometrics*, 63(3):384–395.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



Courdurier, M., Noo, F., Defrise, M., and Kudo, H. (2008). Solving the interior problem of computed tomography using a priori knowledge. *Inverse Problems*, 24(6):065001.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1988). A bayesian approach to the design and analysis of computer experiments. Technical report, Oak Ridge National Lab., TN (USA).

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, volume 76. Springer.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gramacy, R. B. (2020). *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. CRC Press.

Haaland, B. and Qian, P. Z. G. (2011). Accurate emulators for large-scale computer experiments. *The Annals of Statistics*, 39(6):2974–3002.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, second edition.

Kaipio, J. P., Huttunen, T., Luostari, T., Lähivaara, T., and Monk, P. B. (2019). A

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



bayesian approach to improving the born approximation for inverse scattering with high-contrast materials. *Inverse Problems*, 35(8):084001.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. Chapman and Hall/CRC.

Li, Y., Li, K., Zhang, C., Montoya, J., and Chen, G.-H. (2019). Learning to reconstruct computed tomography images directly from sinogram data under a variety of data acquisition conditions. *IEEE Transactions on Medical Imaging*, 38(10):2469–2481.

Li, Z. and Tan, M. H. Y. (2022). A Gaussian process emulator based approach for bayesian calibration of a functional input. *Technometrics*, to appear.

Mak, S., Sung, C.-L., Wang, X., Yeh, S.-T., Chang, Y.-H., Joseph, V. R., Yang, V., and Wu, C. F. J. (2018). An efficient surrogate model for emulation and physics extraction of large eddy simulations. *Journal of the American Statistical Association*, 113(524):1443–1456.

McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F., and Ruppert, D. (2014). Functional generalized additive models. *Journal of Computational and Graphical Statistics*, 23(1):249–269.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



Morokoff, W. J. and Caflisch, R. E. (1995). Quasi-monte carlo integration. *Journal of Computational Physics*, 122(2):218–230.

Morris, M. D. (2012). Gaussian surrogates for computer models with time-varying inputs and outputs. *Technometrics*, 54(1):42–50.

Mueller, J. L. and Siltanen, S. (2020). The D-bar method for electrical impedance tomography–demystified. *Inverse Problems*, 36(4):093001.

Müller, H.-G., Wu, Y., and Yao, F. (2013). Continuously additive models for nonlinear functional regression. *Biometrika*, 100(3):607–622.

Müller, S. (2009). *Komplexität und Stabilität von kernbasierten Rekonstruktionsmethoden*. PhD thesis, Niedersächsische Staats-und Universitätsbibliothek Göttingen.

Nguyen, N. C. and Peraire, J. (2015). Gaussian functional regression for linear partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 287(15 April):69–89.

Preda, C. (2007). Regression models for functional data by reproducing kernel Hilbert spaces methods. *Journal of Statistical Planning and Inference*, 137(3):829–840.

Rahman, M. (2007). *Integral Equations and Their Applications*. WIT press.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis (Second Edition)*. Springer New York, NY.

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

Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2017). Methods for scalar-on-function regression. *International Statistical Review*, 85(2):228–249.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B*, 53(1):233–243.

Ritter, K. (2007). *Average-Case Analysis of Numerical Problems*. Springer.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer Science & Business Media.

Santner, T. J., Williams, B. J., and Notz, W. I. (2018). *The Design and Analysis of Computer Experiments*. Springer New York, second edition.

Shang, H. L. (2013). Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density. *Computational Statistics & Data Analysis*, 67:185–198.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



Shi, J. Q. and Wang, B. (2008). Curve prediction and clustering with mixtures of Gaussian process functional regression models. *Statistics and Computing*, 18(3):267–283.

Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802.

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

Steinwart, I. and Scovel, C. (2012). Mercer's theorem on general domains: On the interaction between measures, kernels, and rkhss. *Constructive Approximation*, 35(3):363–417.

Tan, M. H. (2019). Gaussian process modeling of finite element models with functional inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 7(4):1133–1161.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0180



Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295.

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

Williams, C. and Seeger, M. (2000). Using the nyström method to speed up kernel machines. *Advances in Neural Information Processing Systems*, 13:682–688.

Yao, F. and Müller, H.-G. (2010). Functional quadratic regression. *Biometrika*, 97(1):49–64.

Zhang, H. and Wang, Y. (2010). Kriging and cross-validation for massive spatial data. *Environmetrics: The official journal of the International Environmetrics Society*, 21(3-4):290–304.

Zhu, H., Williams, C. K., Rohwer, R., and Morciniec, M. (1997). Gaussian regression and optimal finite dimensional linear models.