Industrial and
Systems Engineering

Augmented Gaussian Random Field: Theory And Computation

SHENG ZHANG[1], XIU YANG[2], SAMY TINDEL[1], AND GUANG LIN[1,3,4,5]

[1]Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA

[2]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA

[3]Department of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA

[4]Department of Statistics (Courtesy), Purdue University, West Lafayette, IN 47907, USA

[5]Department of Earth, Atmosphere, and Planetary Sciences (Courtesy), Purdue University, West Lafayette, IN 47907, USA

ISE Technical Report 20T-031

LEHIGH UNIVERSITY

Augmented Gaussian Random Field: Theory and Computation

Sheng Zhang[1], Xiu Yang[*2], Samy Tindel[1], and Guang Lin[†1,3,4,5]

[1]Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA
[2]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA 18015, USA
[3]Department of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA
[4]Department of Statistics (Courtesy), Purdue University, West Lafayette, IN 47907, USA
[5]Department of Earth, Atmosphere, and Planetary Sciences (Courtesy), Purdue University, West Lafayette, IN 47907, USA

September 3, 2020

Abstract We provide a rigorous theoretical foundation for incorporating data of observable and its derivatives of any order in a Gaussian-random-field-based surrogate model using tools in real analysis and probability. We demonstrate that under some conditions, the random field representing the derivatives is a Gaussian random field (GRF) given that its structure is derived from the GRF regressing the data of the observable. We propose an *augmented Gaussian random field* (AGRF) framework to unify these GRFs and calculate the prediction of this surrogate model in a similar manner as the conventional Gaussian process regression method. The advantage of our method is that it can incorporate arbitrary order derivatives and deal with missing data.
Keyword: Gaussian random field, Gaussian process regression, arbitrary order derivatives, missing data

1 Introduction

Gaussian random field (GRF) has been widely used in scientific and engineering study to construct a surrogate model (also called *response surface* or *metamodel* in different areas) of a complex system's observable based on available observations. Especially, its special case Gaussian process (GP) has become a powerful tool in applied math, statistics, machine learning, etc. [22]. Although random processes originally refer to one-dimensional random fields [1], e.g., models describing time dependent systems, the terminology GP is *interchangeable* with GRF now in most application scenarios that involve high-dimensional systems. Also, in different areas, GRF-based (or GP-based) methods have different names. For example, in geostatistics, GP regression is referred to as *Kriging*, and it has multiple variants [23, 12].

To enhance the prediction accuracy of the GRF-based surrogate model, one can incorporate all the additional information available, such as gradients, Hessian, multi-fidelity data, physical law, and empirical knowledge. For example, gradient-enhanced Kriging (GEK) uses gradients in either direct [17] or indirect [3] way to improve the accuracy; multi-fidelity Cokriging combines a small amount of high-fidelity data with a large amount of low-fidelity data from simplified or reduced models, in order to leverage low-fidelity models for speedup, while using high-fidelity data to establish accuracy and convergence guarantees [11, 9, 14, 19]; physics-informed Kriging takes advantage of well-developed simulation tools to incorporate physical laws in

*xiy518@lehigh.edu
†guanglin@purdue.edu

the resulting GRF [27, 26, 28]; GP-based numerical ordinary/partial differential equation solver intrinsically imposes the equations in the structure of the GP, and it is one of the most important tools in probabilistic scientific computing [24, 10, 2, 20, 21, 4]; inequality constraints that are not explicitly reflected in governing equations, e.g., positivity, monotonicity, can also be imposed to enhance accuracy and reduce uncertainty in the prediction [6, 15, 16, 18].

Despite the success in applying the aforementioned GRF-based methods to the construction of surrogate models for practical problems, the theory related to the accuracy, convergence, uncertainty, etc. of these methods is not well developed in general. Especially, when using the observable and its derivatives (such as gradient) jointly, e.g., to solve differential equations, one usually assumes that the random field representing the derivative is also a GRF, and its mean and covariance functions can be calculated by taking derivatives of the mean and covariance functions of the GRF representing the observable. Also, the correlation between the observable and its derivative is calculated accordingly. Intuitively, this is correct, because a linear combination of multiple Gaussian random variables is still a Gaussian random variable. However, taking limit of a linear combination is a critical step in the definition of derivative, which makes the theoretical guarantee of using derivatives in GRF-based methods less obvious. To the best of our knowledge, there is no comprehensive theory on the validity of the aforementioned assumption on the random field representing the derivative nor on its relation with the observable. Most literature uses the linearity of the derivative operator to validate this assumption, which is questionable.

In this work, we provide a rigorous theoretical framework to incorporate arbitrary order derivatives in GRF-based methods using tools in real analysis and probability. We demonstrate that under some conditions, the random field representing the observable's derivative obtained by taking derivative of the GRF representing the observable is a GRF, and its structure can be derived from the GRF representing the observable. We propose an *augmented Gaussian random field* (AGRF) framework to unify these GRFs and calculate the posterior distribution of the observable and its derivatives at new locations in a similar manner as the conventional GP regression. Our theoretical and computational framework is universal as it can incorporate derivatives of *arbitrary order*. Moreover, our approach can deal with *missing data* without difficulty, which is very useful in practice.

This paper is organized as follows. The theoretical framework is presented in Section 2, the computational framework is shown in Section 3, the numerical examples are given in Section 4, and the conclusion follows in Section 5.

2 Theoretical framework

In this section, we briefly review some fundamental definitions and an important theorem for GRF. Then we present our main theoretical framework, i.e., the AGRF, that unifies the GRFs representing the observable and its derivative. Finally, we extend it to the general case to incorporate information of derivatives of arbitrary order. For notation brevity, we consider the system's observable as a univariate function of the physical location or time. The extension to the multivariate case is straightforward. Therefore, our theoretical framework is applicable to GP methods that use the information of gradient as well as arbitrary order derivatives of the observable.

2.1 Basic concepts

In this paper, the Euclidean space $\mathbb{R}$ refers to $\mathbb{R}$ equipped with Euclidean metric and Lebesgue measure on Lebesgue-measurable sets. We begin with the definition of random fields.

Definition 2.1 *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $\mathbb{X}$ be a set. Suppose $f : \Omega \times \mathbb{X} \to \mathbb{R}$ is a mapping such that for each $x \in \mathbb{X}$, $f(\cdot, x) : \Omega \to \mathbb{R}$ is a random variable (or measurable function). Then f is called a real-valued random field on $\mathbb{X}$.*

We note that, in practical problems, $\mathbb{X}$ is typically a subset of the d-dimensional Euclidean space $\mathbb{R}^d$, i.e., $\mathbb{X} \subseteq \mathbb{R}^d$. Here, $\mathbb{X}$ is considered as a general set as in [8]. Next, we define Gaussian random fields as follows:

Definition 2.2 *Suppose f is a real-valued random field on $\mathbb{X}$ such that for every finite set of indices $x_1, \ldots, x_p \in \mathbb{X}$, $(f(\omega, x_1), \ldots, f(\omega, x_p))$ is a multivariate Gaussian random variable, or, equivalently, every linear combination of $(f(\omega, x_1), \ldots, f(\omega, x_p))$ has a univariate Gaussian (or normal) distribution. Then f is called a Gaussian random field on $\mathbb{X}$.*

Here ω is an element in the sample space Ω. A Gaussian random field is characterized by its mean and covariance functions:

Definition 2.3 *Given a Gaussian random field f on $\mathbb{X}$, its mean function is defined as the expectation:*

$$m(x) = \mathrm{E}[f(\omega, x)], \tag{1}$$

and its covariance function is defined as:

$$k(x, x') = \mathrm{E}[(f(\omega, x) - m(x))(f(\omega, x') - m(x'))]. \tag{2}$$

Here, the covariance function is also called the *kernel function*. On the other hand, the following theorem provides a perspective in the converse way:

Theorem 2.4 (Kolmogorov extension theorem) *Let $\mathbb{X}$ be any set, $m : \mathbb{X} \to \mathbb{R}$ any function, and $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ such that k is symmetric and nonnegative definite. Then there exists a Gaussian random field on $\mathbb{X}$ with mean function m and covariance function k.*

Proof 2.5 *See [8] (p. 443).*

2.2 Main results

We start our theoretical results with a limit in a Gaussian random field related to the derivative of its realization. The following theorem is a classic result with a proof in [5] (p. 84). Here, we adapt the statement into our framework and reorganize the proof.

Theorem 2.6 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ is differentiable and $k(x, x')$ is twice continuously differentiable. Then there exists a real-valued random field Df on $\mathbb{R}$ such that for each fixed $x \in \mathbb{R}$,*

$$\frac{f(\omega, x + \delta) - f(\omega, x)}{\delta} \xrightarrow[\delta \to 0]{m.s.} Df(\omega, x). \tag{3}$$

Remark 2.7 *Here "m.s." stands for mean-square convergence, i.e., convergence in L^2. Since mean-square convergence implies convergence in probability and convergence in distribution, the limit in Theorem 2.6 also holds in probability and in distribution.*

Proof 2.8 (Proof of Theorem 2.6) *Denote*

$$k_{12}(x, x') = \frac{\partial^2}{\partial x \partial x'} k(x, x'). \tag{4}$$

Let $x \in \mathbb{R}$ be fixed. For any $\delta \in \mathbb{R} - \{0\}$, define

$$\xi_\delta = \xi_\delta(\omega) = \frac{f(\omega, x + \delta) - f(\omega, x)}{\delta}. \tag{5}$$

Suppose $\delta, \tau \in \mathbb{R} - \{0\}$. Then

$$\begin{aligned}
\mathrm{E}[\xi_\delta \xi_\tau] =& \mathrm{E}\left[\frac{(f(\omega, x+\delta) - f(\omega, x))(f(\omega, x+\tau) - f(\omega, x))}{\delta\tau}\right] \\
=& \frac{1}{\delta\tau}\{\mathrm{E}[f(\omega, x+\delta)f(\omega, x+\tau)] - \mathrm{E}[f(\omega, x+\delta)f(\omega, x)] \\
&- \mathrm{E}[f(\omega, x)f(\omega, x+\tau)] + \mathrm{E}[f(\omega, x)f(\omega, x)]\} \\
=& \frac{1}{\delta\tau}\{k(x+\delta, x+\tau) + m(x+\delta)m(x+\tau) - k(x+\delta, x) - m(x+\delta)m(x) \\
&- k(x, x+\tau) - m(x)m(x+\tau) + k(x, x) + m(x)m(x)\} \\
=& \frac{1}{\delta\tau}\{k(x+\delta, x+\tau) - k(x+\delta, x) - k(x, x+\tau) + k(x, x)\} \\
&+ \frac{1}{\delta\tau}\{m(x+\delta)m(x+\tau) - m(x+\delta)m(x) - m(x)m(x+\tau) + m(x)m(x)\}.
\end{aligned} \tag{6}$$

Since $m(x)$ is differentiable and $k(x, x')$ is twice continuously differentiable,

$$\lim_{\delta,\tau\to 0} \mathrm{E}[\xi_\delta \xi_\tau] = k_{12}(x, x) + m'(x)m'(x). \tag{7}$$

Therefore,

$$\lim_{\delta,\tau\to 0} \mathrm{E}[|\xi_\delta - \xi_\tau|^2] = \lim_{\delta,\tau\to 0}\{\mathrm{E}[\xi_\delta\xi_\delta] + \mathrm{E}[\xi_\tau\xi_\tau] - 2\mathrm{E}[\xi_\delta\xi_\tau]\} = 0. \tag{8}$$

Choose a sequence $\tau_i \to 0$ $(i = 1, 2, \dots)$ such that

$$\mathrm{E}[|\xi_{\tau_{i+1}} - \xi_{\tau_i}|^2] \le \frac{1}{2^{2i}}. \tag{9}$$

Then

$$\mathrm{E}|\xi_{\tau_{i+1}} - \xi_{\tau_i}| \le \sqrt{\mathrm{E}[|\xi_{\tau_{i+1}} - \xi_{\tau_i}|^2]} \le \frac{1}{2^i}. \tag{10}$$

By monotone convergence theorem,

$$\mathrm{E}\left[\sum_{i=1}^{\infty} |\xi_{\tau_{i+1}} - \xi_{\tau_i}|\right] = \sum_{i=1}^{\infty} \mathrm{E}|\xi_{\tau_{i+1}} - \xi_{\tau_i}| \le 1 < \infty. \tag{11}$$

Thus,

$$\mathrm{P}\left(\sum_{i=1}^{\infty} |\xi_{\tau_{i+1}} - \xi_{\tau_i}| < \infty\right) = 1. \tag{12}$$

So the random variable $\eta = \xi_{\tau_1} + \sum_{i=1}^{\infty}(\xi_{\tau_{i+1}} - \xi_{\tau_i})$ is well defined, and it is a candidate to be the limit in

Eq. (3). Now, by monotone convergence theorem and Cauchy-Schwarz inequality, for any $j \geq 1$,

$$
\begin{aligned}
\mathrm{E}[|\eta - \xi_{\tau_j}|^2] =& \mathrm{E}\left[\left|\sum_{i=j}^{\infty}(\xi_{\tau_{i+1}} - \xi_{\tau_i})\right|^2\right] \leq \mathrm{E}\left[\left(\sum_{i=j}^{\infty}|\xi_{\tau_{i+1}} - \xi_{\tau_i}|\right)^2\right] \\
=& \mathrm{E}\left[\left(\sum_{i=j}^{\infty}|\xi_{\tau_{i+1}} - \xi_{\tau_i}|\right)\left(\sum_{i'=j}^{\infty}|\xi_{\tau_{i'+1}} - \xi_{\tau_{i'}}|\right)\right] \\
=& \mathrm{E}\left[\sum_{i=j}^{\infty}\sum_{i'=j}^{\infty}|\xi_{\tau_{i+1}} - \xi_{\tau_i}||\xi_{\tau_{i'+1}} - \xi_{\tau_{i'}}|\right] \\
=& \sum_{i=j}^{\infty}\sum_{i'=j}^{\infty}\mathrm{E}|\xi_{\tau_{i+1}} - \xi_{\tau_i}||\xi_{\tau_{i'+1}} - \xi_{\tau_{i'}}| \\
\leq& \sum_{i=j}^{\infty}\sum_{i'=j}^{\infty}\sqrt{\mathrm{E}[|\xi_{\tau_{i+1}} - \xi_{\tau_i}|^2]\mathrm{E}[|\xi_{\tau_{i'+1}} - \xi_{\tau_{i'}}|^2]} \\
\leq& \sum_{i=j}^{\infty}\sum_{i'=j}^{\infty}\sqrt{\frac{1}{2^{2i}}\frac{1}{2^{2i'}}} = \sum_{i=j}^{\infty}\frac{1}{2^i}\sum_{i'=j}^{\infty}\frac{1}{2^{i'}} = \left(\frac{1}{2^{j-1}}\right)^2.
\end{aligned}
\tag{13}
$$

Since

$$
\mathrm{E}[|\eta - \xi_\delta|^2] \leq \mathrm{E}[2|\eta - \xi_{\tau_j}|^2 + 2|\xi_{\tau_j} - \xi_\delta|^2] = 2\mathrm{E}[|\eta - \xi_{\tau_j}|^2] + 2\mathrm{E}[|\xi_{\tau_j} - \xi_\delta|^2], \tag{14}
$$

we have

$$
\lim_{\delta \to 0}\mathrm{E}[|\eta - \xi_\delta|^2] = 0, \tag{15}
$$

or

$$
\xi_\delta \xrightarrow[\delta \to 0]{m.s.} \eta. \tag{16}
$$

Let $Df(\omega, x) = \eta$, do the same for every x, and the proof is complete.

Remark 2.9 *For $\delta \in \mathbb{R} - \{0\}$, let*

$$
f_\delta = f_\delta(\omega, x) = \frac{f(\omega, x+\delta) - f(\omega, x)}{\delta} \tag{17}
$$

be a random field on $\mathbb{R}$. One could also consider the convergence of the family $\{f_\delta \mid \delta \in \mathbb{R} - \{0\}\}$ to Df. We have refrained from getting into this type of consideration for the sake of conciseness.

The next lemma indicates that the limit of a series of Gaussian random variables is still a Gaussian random variable. This is a classic result with various statements and proofs. Here, we give our statement and proof that fits into our framework.

Lemma 2.10 *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and ξ_δ $(\delta \in \mathbb{R})$ be a family of random variables such that ξ_δ $(\delta \neq 0)$ are Gaussian random variables with mean μ_δ and variance σ_δ^2, and*

$$
\xi_\delta \xrightarrow[\delta \to 0]{m.s.} \xi_0. \tag{18}
$$

Then ξ_0 has Gaussian distribution with mean μ_0 and variance σ_0^2, and

$$
\begin{aligned}
&\lim_{\delta \to 0}\mu_\delta = \mu_0, \\
&\lim_{\delta \to 0}\sigma_\delta^2 = \sigma_0^2.
\end{aligned}
\tag{19}
$$

Proof 2.11 *Suppose $\delta, \tau \in \mathbb{R} - \{0\}$. Since*

$$\lim_{\delta \to 0} \mathrm{E}[|\xi_\delta - \xi_0|^2] = 0 \tag{20}$$

and

$$\mathrm{E}[|\xi_\delta - \xi_\tau|^2] \leq 2\mathrm{E}[|\xi_\delta - \xi_0|^2] + 2\mathrm{E}[|\xi_\tau - \xi_0|^2], \tag{21}$$

we have

$$\lim_{\delta,\tau \to 0} \mathrm{E}[|\xi_\delta - \xi_\tau|^2] = 0. \tag{22}$$

Thus,

$$|\mu_\delta - \mu_\tau| = |\mathrm{E}[\xi_\sigma - \xi_\tau]| \leq \mathrm{E}|\xi_\sigma - \xi_\tau| \leq \sqrt{\mathrm{E}[|\xi_\sigma - \xi_\tau|^2]}, \tag{23}$$

which indicates

$$\lim_{\delta,\tau \to 0} |\mu_\delta - \mu_\tau| = 0. \tag{24}$$

Hence, there exists $\mu_0 \in \mathbb{R}$ such that

$$\lim_{\delta \to 0} \mu_\delta = \mu_0. \tag{25}$$

By (22), there exists $\Delta > 0$ such that for any δ $(0 < |\delta| \leq \Delta)$, we have

$$\mathrm{E}[|\xi_\delta - \xi_\Delta|^2] \leq 1. \tag{26}$$

So

$$\mathrm{E}[\xi_\delta^2] \leq 2\mathrm{E}[|\xi_\delta - \xi_\Delta|^2] + 2\mathrm{E}[\xi_\Delta^2] \leq 2 + 2\mathrm{E}[\xi_\Delta^2]. \tag{27}$$

Since ξ_Δ is Gaussian, $\mathrm{E}[\xi_\Delta^2]$ is finite. Now, for any δ $(0 < |\delta| \leq \Delta)$ and τ $(0 < |\tau| \leq \Delta)$, by Cauchy-Schwarz inequality,

$$\begin{aligned}
|\sigma_\delta^2 - \sigma_\tau^2| &= |(\mathrm{E}[\xi_\delta^2] - \mu_\delta^2) - (\mathrm{E}[\xi_\tau^2] - \mu_\tau^2)| \\
&\leq |\mathrm{E}[\xi_\delta^2] - \mathrm{E}[\xi_\tau^2]| + |\mu_\delta^2 - \mu_\tau^2| \\
&\leq \mathrm{E}|\xi_\delta^2 - \xi_\tau^2| + |\mu_\delta^2 - \mu_\tau^2| \\
&= \mathrm{E}|\xi_\delta - \xi_\tau||\xi_\delta + \xi_\tau| + |\mu_\delta^2 - \mu_\tau^2| \\
&\leq \sqrt{\mathrm{E}[|\xi_\delta - \xi_\tau|^2]}\sqrt{\mathrm{E}[|\xi_\delta + \xi_\tau|^2]} + |\mu_\delta^2 - \mu_\tau^2| \\
&\leq \sqrt{\mathrm{E}[|\xi_\delta - \xi_\tau|^2]}\sqrt{2\mathrm{E}[\xi_\delta^2] + 2\mathrm{E}[\xi_\tau^2]} + |\mu_\delta^2 - \mu_\tau^2| \\
&\leq \sqrt{\mathrm{E}[|\xi_\delta - \xi_\tau|^2]}\sqrt{8 + 8\mathrm{E}[\xi_\Delta^2]} + |\mu_\delta^2 - \mu_\tau^2|.
\end{aligned} \tag{28}$$

By Eqs. (22), (25), *and* (28), *we have*

$$\lim_{\delta,\tau \to 0} |\sigma_\delta^2 - \sigma_\tau^2| = 0. \tag{29}$$

Since $\sigma_\delta^2 \geq 0$ for all $\delta \in \mathbb{R} - \{0\}$, there exists $\sigma_0^2 \geq 0$ such that

$$\lim_{\delta \to 0} \sigma_\delta^2 = \sigma_0^2. \tag{30}$$

If $\sigma_0^2 = 0$, then

$$\begin{aligned}
\mathrm{E}|\xi_0 - \mu_0| &\leq \mathrm{E}|\xi_0 - \xi_\delta| + \mathrm{E}|\xi_\delta - \mu_\delta| + \mathrm{E}|\mu_\delta - \mu_0| \\
&\leq \sqrt{\mathrm{E}[|\xi_0 - \xi_\delta|^2]} + \sqrt{\mathrm{E}[|\xi_\delta - \mu_\delta|^2]} + |\mu_\delta - \mu_0| \\
&= \sqrt{\mathrm{E}[|\xi_0 - \xi_\delta|^2]} + \sqrt{\sigma_\delta^2} + |\mu_\delta - \mu_0|.
\end{aligned} \tag{31}$$

Let $\delta \to 0$ and we have

$$\mathrm{E}|\xi_0 - \mu_0| = 0, \tag{32}$$

or

$$\mathrm{P}(\xi_0 = \mu_0) = 1. \tag{33}$$

Therefore, ξ_0 *has degenerate Gaussian distribution with mean* μ_0 *and variance* 0. *If* $\sigma_0^2 > 0$, *then there exists* $\Delta' > 0$ *such that for any* δ $(0 < |\delta| \le \Delta')$, *we have*

$$\sigma_\delta^2 \ge \sigma_0^2/2 > 0. \tag{34}$$

For all δ $(|\delta| \le \Delta')$, *define*

$$\eta_\delta = \frac{\xi_\delta - \mu_\delta}{\sigma_\delta}. \tag{35}$$

Then when $0 < |\delta| \le \Delta'$, η_δ *has the standard Gaussian distribution. Now,*

$$\begin{aligned}
\mathrm{E}[|\eta_\delta - \eta_0|^2] &= \mathrm{E}\left[\left|\frac{\xi_\delta - \mu_\delta}{\sigma_\delta} - \frac{\xi_0 - \mu_0}{\sigma_0}\right|^2\right] \\
&\le 2\mathrm{E}\left[\left|\frac{\xi_\delta}{\sigma_\delta} - \frac{\xi_0}{\sigma_0}\right|^2\right] + 2\mathrm{E}\left[\left|\frac{\mu_\delta}{\sigma_\delta} - \frac{\mu_0}{\sigma_0}\right|^2\right] \\
&\le 4\mathrm{E}\left[\left|\frac{\xi_\delta}{\sigma_\delta} - \frac{\xi_\delta}{\sigma_0}\right|^2\right] + 4\mathrm{E}\left[\left|\frac{\xi_\delta}{\sigma_0} - \frac{\xi_0}{\sigma_0}\right|^2\right] + 2\left|\frac{\mu_\delta}{\sigma_\delta} - \frac{\mu_0}{\sigma_0}\right|^2 \\
&= 4\left(\frac{1}{\sigma_\delta} - \frac{1}{\sigma_0}\right)^2 \mathrm{E}[\xi_\delta^2] + \frac{4}{\sigma_0^2}\mathrm{E}[|\xi_\delta - \xi_0|^2] + 2\left|\frac{\mu_\delta}{\sigma_\delta} - \frac{\mu_0}{\sigma_0}\right|^2 \\
&= 4\frac{(\sigma_0 - \sigma_\delta)^2}{\sigma_\delta^2\sigma_0^2}(\sigma_\delta^2 + \mu_\delta^2) + \frac{4}{\sigma_0^2}\mathrm{E}[|\xi_\delta - \xi_0|^2] + 2\left|\frac{\mu_\delta}{\sigma_\delta} - \frac{\mu_0}{\sigma_0}\right|^2.
\end{aligned} \tag{36}$$

Thus,

$$\lim_{\delta\to 0} \mathrm{E}[|\eta_\delta - \eta_0|^2] = 0. \tag{37}$$

Let

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-t^2/2}dt \tag{38}$$

be the cumulative distribution function (CDF) of the standard Gaussian distribution. For any $z \in \mathbb{R}$ *and any* $\epsilon > 0$,

$$\mathrm{P}(\eta_0 \le z) \le \mathrm{P}(\eta_\delta \le z + \epsilon) + \mathrm{P}(|\eta_\delta - \eta_0| > \epsilon) \le \Phi(z+\epsilon) + \frac{\mathrm{E}[|\eta_\delta - \eta_0|^2]}{\epsilon^2}. \tag{39}$$

Let $\delta \to 0$ *and we have*

$$\mathrm{P}(\eta_0 \le z) \le \Phi(z+\epsilon). \tag{40}$$

Let $\epsilon \to 0$ *and we have*

$$\mathrm{P}(\eta_0 \le z) \le \Phi(z). \tag{41}$$

On the other hand,

$$\begin{aligned}
\mathrm{P}(\eta_0 > z) &\le \mathrm{P}(\eta_\delta > z - \epsilon) + \mathrm{P}(|\eta_\delta - \eta_0| > \epsilon) \\
&\le 1 - \Phi(z-\epsilon) + \frac{\mathrm{E}[|\eta_\delta - \eta_0|^2]}{\epsilon^2}.
\end{aligned} \tag{42}$$

Let $\delta \to 0$ *and we have*

$$\mathrm{P}(\eta_0 > z) \le 1 - \Phi(z-\epsilon). \tag{43}$$

Let $\epsilon \to 0$ *and we have*

$$\mathrm{P}(\eta_0 > z) \le 1 - \Phi(z), \tag{44}$$

or

$$\mathrm{P}(\eta_0 \leq z) \geq \Phi(z). \tag{45}$$

Combining (41) and (45), we have

$$\mathrm{P}(\eta_0 \leq z) = \Phi(z). \tag{46}$$

Thus, η_0 has standard Gaussian distribution, which implies that ξ_0 has Gaussian distribution with mean μ_0 and variance σ_0^2.

Now we define an augmented random field consisting of a Gaussian random field f and the Df associated with it. Then we prove that this augmented random field is a Gaussian random field.

Definition 2.12 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ is differentiable and $k(x, x')$ is twice continuously differentiable. The real-valued random field Df defined in Theorem 2.6 is called the derivative random field of f. Define a real-valued random field on $\mathbb{R} \times \{0, 1\}$:*

$$\tilde{f} : \Omega \times \mathbb{R} \times \{0, 1\} \to \mathbb{R} \tag{47}$$

such that

$$\begin{aligned} \tilde{f}(\omega, x, 0) &= f(\omega, x) \\ \tilde{f}(\omega, x, 1) &= Df(\omega, x). \end{aligned} \tag{48}$$

We call $\tilde{f}$ as the augmented random field of f.

Theorem 2.13 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ is differentiable and $k(x, x')$ is twice continuously differentiable. Then the augmented random field $\tilde{f}$ of f is a Gaussian random field on $\mathbb{R} \times \{0, 1\}$.*

Proof 2.14 *For any $p, q \in \mathbb{N}^{+} \cup \{0\}$ such that $p + q \geq 1$, any $x_1, \ldots, x_p, y_1, \ldots, y_q \in \mathbb{R}$, and any $c_1, \ldots, c_p, d_1, \ldots, d_q \in \mathbb{R}$, we have the linear combination:*

$$\begin{aligned} &\sum_{i=1}^{p} c_i \tilde{f}(\omega, x_i, 0) + \sum_{j=1}^{q} d_j \tilde{f}(\omega, y_j, 1) \\ =& \sum_{i=1}^{p} c_i f(\omega, x_i) + \sum_{j=1}^{q} d_j Df(\omega, y_j) \\ =& \sum_{i=1}^{p} c_i f(\omega, x_i) + \sum_{j=1}^{q} d_j \lim_{\delta_j \to 0} \frac{f(\omega, y_j + \delta_j) - f(\omega, y_j)}{\delta_j} \\ =& \lim_{\delta_1 \to 0} \cdots \lim_{\delta_q \to 0} \left\{ \sum_{i=1}^{p} c_i f(\omega, x_i) + \sum_{j=1}^{q} d_j \frac{f(\omega, y_j + \delta_j) - f(\omega, y_j)}{\delta_j} \right\}, \end{aligned} \tag{49}$$

where the limits are taken in mean-square sense. Since f is a Gaussian random field,

$$\sum_{i=1}^{p} c_i f(\omega, x_i) + \sum_{j=1}^{q} d_j \frac{f(\omega, y_j + \delta_j) - f(\omega, y_j)}{\delta_j} \tag{50}$$

has Gaussian distribution for any $\delta_1, \ldots, \delta_q \in \mathbb{R} - \{0\}$. By Lemma 2.10, the limit has Gaussian distribution. As this holds for every linear combination, $\tilde{f}$ is a Gaussian random field.

After proving the augmented Gaussian random field is well defined, we calculate its mean and covariance functions.

Theorem 2.15 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ is differentiable and $k(x, x')$ is twice continuously differentiable. Then the augmented random field $\tilde{f}$ of f has mean function:*

$$\tilde{m} : \mathbb{R} \times \{0, 1\} \to \mathbb{R} \tag{51}$$

such that

$$\begin{aligned} \tilde{m}(x, 0) &= m(x), \\ \tilde{m}(x, 1) &= \frac{d}{dx} m(x), \end{aligned} \tag{52}$$

and covariance function:

$$\tilde{k} : \mathbb{R} \times \{0, 1\} \times \mathbb{R} \times \{0, 1\} \to \mathbb{R} \tag{53}$$

such that

$$\begin{aligned} \tilde{k}(x, 0, x', 0) &= k(x, x'), \\ \tilde{k}(x, 0, x', 1) &= \frac{\partial}{\partial x'} k(x, x'), \\ \tilde{k}(x, 1, x', 0) &= \frac{\partial}{\partial x} k(x, x'), \\ \tilde{k}(x, 1, x', 1) &= \frac{\partial^2}{\partial x \partial x'} k(x, x'). \end{aligned} \tag{54}$$

Proof 2.16 *By the definition of $\tilde{f}$:*

$$\tilde{m}(x, 0) = \mathrm{E}[\tilde{f}(\omega, x, 0)] = \mathrm{E}[f(\omega, x)] = m(x). \tag{55}$$

By Theorem 2.6 and Lemma 2.10,

$$\begin{aligned} \tilde{m}(x, 1) &= \mathrm{E}[\tilde{f}(\omega, x, 1)] = \mathrm{E}[Df(\omega, x)] \\ &= \mathrm{E}\left[\lim_{\delta \to 0} \frac{f(\omega, x + \delta) - f(\omega, x)}{\delta}\right] \\ &= \lim_{\delta \to 0} \mathrm{E}\left[\frac{f(\omega, x + \delta) - f(\omega, x)}{\delta}\right] \\ &= \lim_{\delta \to 0} \frac{m(x + \delta) - m(x)}{\delta} \\ &= \frac{\mathrm{d}}{\mathrm{d}x} m(x). \end{aligned} \tag{56}$$

Similarly, by definition,

$$\begin{aligned} \tilde{k}(x, 0, x', 0) &= \mathrm{E}[(\tilde{f}(\omega, x, 0) - \tilde{m}(x, 0))(\tilde{f}(\omega, x', 0) - \tilde{m}(x', 0))] \\ &= \mathrm{E}[(f(\omega, x) - m(x))(f(\omega, x') - m(x'))] \\ &= k(x, x'). \end{aligned} \tag{57}$$

Also,

$$\begin{aligned}
&\tilde{k}(x,0,x',1)\\
=&\mathrm{E}[\tilde{f}(\omega,x,0)\tilde{f}(\omega,x',1)]-\mathrm{E}[\tilde{f}(\omega,x,0)]\mathrm{E}[\tilde{f}(\omega,x',1)]\\
=&\mathrm{E}[f(\omega,x)Df(\omega,x')]-m(x)m'(x')\\
=&\mathrm{E}\left[f(\omega,x)\left(Df(\omega,x')-\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right)\right]\\
&+\mathrm{E}\left[f(\omega,x)\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right]-m(x)m'(x')\\
=&\mathrm{E}\left[f(\omega,x)\left(Df(\omega,x')-\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right)\right]\\
&+\frac{k(x,x'+\delta)+m(x)m(x'+\delta)-k(x,x')-m(x)m(x')}{\delta}-m(x)m'(x')\\
=&\mathrm{E}\left[f(\omega,x)\left(Df(\omega,x')-\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right)\right]\\
&+\frac{k(x,x'+\delta)-k(x,x')}{\delta}+m(x)\left(\frac{m(x'+\delta)-m(x')}{\delta}-m'(x')\right).
\end{aligned}$$

Since

$$\begin{aligned}
&\left|\mathrm{E}\left[f(\omega,x)\left(Df(\omega,x')-\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right)\right]\right|\\
\leq&\mathrm{E}\left|f(\omega,x)\left(Df(\omega,x')-\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right)\right|\\
\leq&\sqrt{\mathrm{E}[|f(\omega,x)|^2]}\sqrt{\mathrm{E}\left[\left|Df(\omega,x')-\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right|^2\right]},
\end{aligned}\tag{58}$$

let $\delta\to 0$ and we have

$$\tilde{k}(x,0,x',1)=\frac{\partial}{\partial x'}k(x,x').\tag{59}$$

Similarly,

$$\tilde{k}(x,1,x',0)=\frac{\partial}{\partial x}k(x,x').\tag{60}$$

Finally,

$$\begin{aligned}
&\tilde{k}(x,1,x',1)\\
=&\mathrm{E}[Df(\omega,x)Df(\omega,x')]-m'(x)m'(x')\\
=&\mathrm{E}\left[Df(\omega,x)\left(Df(\omega,x')-\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right)\right]\\
&+\mathrm{E}\left[Df(\omega,x)\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right]-m'(x)m'(x')\\
=&\mathrm{E}\left[Df(\omega,x)\left(Df(\omega,x')-\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right)\right]\\
&+\frac{\frac{\partial}{\partial x}k(x,x'+\delta)+m'(x)m(x'+\delta)-\frac{\partial}{\partial x}k(x,x')-m'(x)m(x')}{\delta}\\
&-m'(x)m'(x')\\
=&\mathrm{E}\left[Df(\omega,x)\left(Df(\omega,x')-\frac{f(\omega,x'+\delta)-f(\omega,x')}{\delta}\right)\right]\\
&\frac{\frac{\partial}{\partial x}k(x,x'+\delta)-\frac{\partial}{\partial x}k(x,x')}{\delta}+m'(x)\left(\frac{m(x'+\delta)-m(x')}{\delta}-m'(x')\right).
\end{aligned}$$

Let $\delta \to 0$ and use the fact that $k(x, x')$ is twice continuously differentiable,

$$\tilde{k}(x, 1, x', 1) = \frac{\partial^2}{\partial x \partial x'} k(x, x'). \tag{61}$$

Corollary 2.17 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ is differentiable and $k(x, x')$ is twice continuously differentiable. Then the derivative random field Df of f is a Gaussian random field on $\mathbb{R}$ with mean function $m'(x)$ and covariance function $\partial^2 k(x, x)/\partial x \partial x'$.*

Proof 2.18 (Sketch of proof) *This is a direct conclusion from Theorem 2.6, Theorem 2.13, and Theorem 2.15.*

2.3 Extensions

The definition of the aforementioned augmented Gaussian random field can be extended to more general cases involing higher order derivatives.

Definition 2.19 *Let f be a Gaussian random field on $\mathbb{R}$ such that the derivative random field Df is a Gaussian random field on $\mathbb{R}$ with differentiable mean function and twice continuously differentiable covariance function. By Corollary 2.17, the derivative random field $D(Df)$ of Df is a Gaussian random field on $\mathbb{R}$. Define the second order derivative random field of f as $D^2 f = D(Df)$. Recursively, define the n^{th} order derivative random field of f as $D^n f = D(D^{n-1} f)$ when $D^{n-1} f$ is a Gaussian random field on $\mathbb{R}$ with differentiable mean function and twice continuously differentiable covariance function.*

Corollary 2.20 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ is n times differentiable and $k(x, x')$ is $2n$ times continuously differentiable ($n \in \mathbb{N}^+$). Then $Df, \ldots, D^n f$ are well-defined and are Gaussian random fields on $\mathbb{R}$.*

Proof 2.21 (Sketch of proof) *This corollary can be proved by applying Corollary 2.17 recursively.*

Now we can define the general augmented Gaussian random field involving higher order derivatives.

Definition 2.22 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ is n times differentiable and $k(x, x')$ is $2n$ times continuously differentiable ($n \in \mathbb{N}^+$). Define the n^{th} order augmented random field of f as:*

$$\tilde{f}^n : \Omega \times \mathbb{R} \times \{0, 1, \ldots, n\} \to \mathbb{R} \tag{62}$$

such that

$$\begin{aligned} \tilde{f}^n(\omega, x, 0) &= f(\omega, x) \\ \tilde{f}^n(\omega, x, 1) &= Df(\omega, x) \\ &\vdots \\ \tilde{f}^n(\omega, x, n) &= D^n f(\omega, x). \end{aligned} \tag{63}$$

Theorem 2.23 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ is n times differentiable and $k(x, x')$ is $2n$ times continuously differentiable ($n \in \mathbb{N}^+$). Then the n^{th} order augmented random field $\tilde{f}^n$ of f is a Gaussian random field on $\mathbb{R} \times \{0, 1, \ldots, n\}$.*

Proof 2.24 (Sketch of proof) *This can be proved in a similar way to the proof of Theorem 2.13.*

The following theorem calculates the mean and covariance functions of the n^{th} order augmented Gaussian random field.

Theorem 2.25 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ is n times differentiable and $k(x, x')$ is $2n$ times continuously differentiable ($n \in \mathbb{N}^+$). Then the n^{th} order augmented random field $\tilde{f}^n$ of f has mean function:*

$$\begin{aligned} \tilde{m}^n &: \mathbb{R} \times \{0, 1, \ldots, n\} \to \mathbb{R} \\ \tilde{m}^n(x, i) &= \frac{\mathrm{d}^i}{\mathrm{d}x^i} m(x), \end{aligned} \tag{64}$$

and covariance function:

$$\begin{aligned} \tilde{k}^n &: \mathbb{R} \times \{0, 1, \ldots, n\} \times \mathbb{R} \times \{0, 1, \ldots, n\} \to \mathbb{R} \\ \tilde{k}^n(x, i, x', j) &= \frac{\partial^{i+j}}{\partial x^i \partial x'^j} k(x, x'). \end{aligned} \tag{65}$$

Proof 2.26 (Sketch of proof) *When $i, j \in \{0, 1\}$, by Theorem 2.15, the formulas hold. Then this theorem can be proved by using induction and following a similar way to the proof of Theorem 2.15.*

Furthermore, we can extend the augmented Gaussian random field to the infinite order case, and calculate the mean and covariance functions accordingly.

Definition 2.27 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ and $k(x, x')$ are smooth. Define the infinite order augmented random field of f as:*

$$\tilde{f}^\infty : \Omega \times \mathbb{R} \times \mathbb{N} \to \mathbb{R} \tag{66}$$

such that

$$\begin{aligned} \tilde{f}^\infty(\omega, x, 0) &= f(\omega, x) \\ \tilde{f}^\infty(\omega, x, 1) &= Df(\omega, x) \\ &\vdots \\ \tilde{f}^\infty(\omega, x, n) &= D^n f(\omega, x) \\ &\vdots \end{aligned} \tag{67}$$

Theorem 2.28 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ and $k(x, x')$ are smooth. Then the infinite order augmented random field $\tilde{f}^\infty$ of f is a Gaussian random field on $\mathbb{R} \times \mathbb{N}$.*

Proof 2.29 (Sketch of proof) *This is proved in a similar way to the proof in Theorem 2.13.*

Theorem 2.30 *Let f be a Gaussian random field on $\mathbb{R}$ with mean function $m(x)$ and covariance function $k(x, x')$ such that $m(x)$ and $k(x, x')$ are smooth. Then the infinite order augmented random field $\tilde{f}^\infty$ of f has mean function:*

$$\begin{aligned} \tilde{m}^\infty &: \ \mathbb{R} \times \mathbb{N} \to \mathbb{R}, \\ \tilde{m}^\infty(x, i) &= \frac{\mathrm{d}^i}{\mathrm{d}x^i} m(x), \end{aligned} \tag{68}$$

and covariance function:

$$\begin{aligned} \tilde{k}^\infty &: \ \mathbb{R} \times \mathbb{N} \times \mathbb{R} \times \mathbb{N} \to \mathbb{R}, \\ \tilde{k}^\infty(x, i, x', j) &= \frac{\partial^{i+j}}{\partial x^i \partial x'^j} k(x, x'). \end{aligned} \tag{69}$$

Proof 2.31 (Sketch of proof) *When $i, j \in \{0, 1\}$, by Theorem 2.15, the formulas hold. Then the result is proved by induction and in a similar way to Theorem 2.15.*

3 Computational framework

In this section, we describe the computational framework for the AGRF prediction. It is based on the conventional GP regression. Specifically, in this work, we only consider the noiseless observations for demonstration purpose. Noise in the observations can be incorporated by modifying the covariance matrix as in the conventional GP regression (see [22]). Also, since we use univariate observable in the theoretical framework, we still consider this scenario here to keep consistency. Formulations for the multi-variate cases can be deduced without difficulty based on our results and the gradient-enhanced Kriging/Cokriging methods (see, e.g., [25, 13, 7]).

3.1 Posterior distribution

As in the conventional GP regression, we aim to condition the joint Gaussian prior distribution on the observations, as such to provide a posterior joint Gaussian distribution. In our framework, the observations include the collected data of the observable and its derivatives of different orders. Suppose we are given a finite collection of real-valued data pairs:

$$\begin{aligned}
\text{Observable:} \quad & (x_{0,1}, y_{0,1})\ (x_{0,2}, y_{0,2})\ \dots\ (x_{0,p_0}, y_{0,p_0}) \\
\text{First order derivative:} \quad & (x_{1,1}, y_{1,1})\ (x_{1,2}, y_{1,2})\ \dots\ (x_{1,p_1}, y_{1,p_1}) \\
\text{Second order derivative:} \quad & (x_{2,1}, y_{2,1})\ (x_{2,2}, y_{2,2})\ \dots\ (x_{2,p_2}, y_{2,p_2}) \\
\vdots \quad & \\
n^{\text{th}} \text{ order derivative:} \quad & (x_{n,1}, y_{n,1})\ (x_{n,2}, y_{n,2})\ \dots\ (x_{n,p_n}, y_{n,p_n})
\end{aligned} \tag{70}$$

with $n \geq 0$ and $p_0, p_1, p_2, \dots, p_n \geq 0$. Here $x_{i,j}$ are locations and $y_{i,j}$ are the data collected at this location. We also introduce the notation $\boldsymbol{X} = \{x_{i,j}\}$ and $\boldsymbol{y} = \{y_{i,j}\}$ where $i \in \{0, 1, \dots, n\}$ and $j \in \{1, \dots, p_i\}$. Of note, here we consider a general case, and it is not necessary that $x_{i,j} = x_{i+1,j}$. In other words, it is possible that the observable and its derivatives are sampled at different locations. We assume that a mean function and a covariance function are given for the Gaussian random field f that describes the observable:

$$m : \mathbb{R} \to \mathbb{R} \tag{71}$$

and

$$k : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \tag{72}$$

such that m is n times differentiable and k is symmetric, nonnegative definite, and $2n$ times continuously differentiable. By Theorem 2.4 and Theorem 2.23, there exists a Gaussian random field $\tilde{f}^n$ on $\mathbb{R} \times \{0, 1, \dots, n\}$ whose mean function and covariance function are given by Theorem 2.25. We use the augmented Gaussian random field $\tilde{f}^n$ to model the data such that

$$\tilde{f}^n(\omega, x_{i,j}, i) = y_{i,j} \text{ for } i \in \{0, 1, \dots, n\} \text{ and } j \in \{1, \dots, p_i\}. \tag{73}$$

The prediction of the q^{th} $(0 \leq q \leq n)$ order derivative at any $x \in \mathbb{R}$ is the posterior mean of the random variable $\tilde{f}^n(\omega, x, q)$, and the uncertainty in the prediction can be described by the confidence interval based on the posterior variance (or the standard deviation).

Since $\tilde{f}^n$ is a Gaussian random field, $\left\{\tilde{f}^n(\omega, x, q), \tilde{f}^n(\omega, x_{i,j}, i)\right\}$ satisfies the multivariate Gaussian distribution:

$$\begin{bmatrix} \tilde{f}^n(\omega, x, q) \\ \tilde{f}^n(\omega, x_{0,1}, 0) \\ \vdots \\ \tilde{f}^n(\omega, x_{n,p_n}, n) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \tilde{m}^n(x, q) \\ \tilde{m}^n(x_{0,1}, 0) \\ \vdots \\ \tilde{m}^n(x_{n,p_n}, n) \end{bmatrix}, \begin{bmatrix} \tilde{k}^n(x, q, x, q) & \tilde{k}^n(x, q, x_{0,1}, 0) & \cdots & \tilde{k}^n(x, q, x_{n,p_n}, n) \\ \tilde{k}^n(x_{0,1}, 0, x, q) & \tilde{k}^n(x_{0,1}, 0, x_{0,1}, 0) & \cdots & \tilde{k}^n(x_{0,1}, 0, x_{n,p_n}, n) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{k}^n(x_{n,p_n}, n, x, q) & \tilde{k}^n(x_{n,p_n}, n, x_{0,1}, 0) & \cdots & \tilde{k}^n(x_{n,p_n}, n, x_{n,p_n}, n) \end{bmatrix} \right). \tag{74}$$

Then the posterior distribution of $\tilde{f}^n(\omega, x, q)$ given (73) is also a Gaussian distribution:

$$\left(\tilde{f}^n(\omega, x, q) \mid \tilde{f}^n(\omega, x_{i,j}, i), i \in \{0, 1, \dots, n\}, j \in \{1, \dots, p_i\} \right) \sim \mathcal{N}(\mu, \sigma^2), \tag{75}$$

where

$$\mu = \tilde{m}^n(x, q) + \left[\tilde{k}^n(x, q, x_{0,1}, 0) \cdots \tilde{k}^n(x, q, x_{n,p_n}, n) \right] K^{-1} \left[\begin{bmatrix} y_{0,1} \\ \vdots \\ y_{n,p_n} \end{bmatrix} - \begin{bmatrix} \tilde{m}^n(x_{0,1}, 0) \\ \vdots \\ \tilde{m}^n(x_{n,p_n}, n) \end{bmatrix} \right],$$

$$\sigma^2 = \tilde{k}^n(x, q, x, q) - \left[\tilde{k}^n(x, q, x_{0,1}, 0) \cdots \tilde{k}^n(x, q, x_{n,p_n}, n) \right] K^{-1} \begin{bmatrix} \tilde{k}^n(x_{0,1}, 0, x, q) \\ \vdots \\ \tilde{k}^n(x_{n,p_n}, n, x, q) \end{bmatrix},$$

$$K = \begin{bmatrix} \tilde{k}^n(x_{0,1}, 0, x_{0,1}, 0) & \cdots & \tilde{k}^n(x_{0,1}, 0, x_{n,p_n}, n) \\ \vdots & \ddots & \vdots \\ \tilde{k}^n(x_{n,p_n}, n, x_{0,1}, 0) & \cdots & \tilde{k}^n(x_{n,p_n}, n, x_{n,p_n}, n) \end{bmatrix}.$$

Here $\tilde{m}^n$ and $\tilde{k}^n$ are calculated according to Theorem 2.25. Now, we have the posterior distribution of the q^{th} order derivative at x. The posterior mean is usually used as the prediction and the posterior standard deviation is used to describe the uncertainty (e.g., the confidence interval) in the prediction.

3.2 Hyperparameter identification

In practice, we usually assume the form of m and k, which involve hyperparameters. We denote these parameters as θ. For example, in the widely used squared exponential covariance function $k(x, x') = \sigma^2 \exp(-(x - x')^2/(2l^2))$, σ and l are hyperparameters. Similarly, the mean function m may include additional hyperparameters. For example, if m is a polynomial as in the universal Kriging, the coefficients of the polynomial are hyperparameters. Similar to the standard GP method, the AGRF method identifies the hyperparameters via maximizing the following log likelihood:

$$\begin{aligned} \log p(\boldsymbol{y}|\boldsymbol{X}, \theta) = &-\frac{p_0 + \cdots + p_n}{2} \log(2\pi) - \frac{1}{2} \log(\det(K)) \\ &- \frac{1}{2} \left[\begin{bmatrix} y_{0,1} \\ \vdots \\ y_{n,p_n} \end{bmatrix} - \begin{bmatrix} \tilde{m}^0(x_{0,1}, 0) \\ \vdots \\ \tilde{m}^n(x_{n,p_n}, n) \end{bmatrix} \right]^\top K^{-1} \left[\begin{bmatrix} y_{0,1} \\ \vdots \\ y_{n,p_n} \end{bmatrix} - \begin{bmatrix} \tilde{m}^0(x_{0,1}, 0) \\ \vdots \\ \tilde{m}^n(x_{n,p_n}, n) \end{bmatrix} \right], \end{aligned} \tag{76}$$

where

$$K = \begin{bmatrix} \tilde{k}^n(x_{0,1}, 0, x_{0,1}, 0) & \cdots & \tilde{k}^n(x_{0,1}, 0, x_{n,p_n}, n) \\ \vdots & \ddots & \vdots \\ \tilde{k}^n(x_{n,p_n}, n, x_{0,1}, 0) & \cdots & \tilde{k}^n(x_{n,p_n}, n, x_{n,p_n}, n) \end{bmatrix}. \tag{77}$$

After identifying the θ, we obtain the posterior distribution in Eq. (75).

4 Numerical examples

In this section, we present two examples to illustrate the AGRF framework. In both examples, we use the zero mean function

$$m(x) = 0, \tag{78}$$

and the squared exponential covariance function

$$k(x, x') = \sigma^2 \exp\left(\frac{-(x - x')^2}{2l^2}\right) \tag{79}$$

to construct the GRF representing the observable. The hyperparameters to identify are $\theta = (\sigma, l)$. We use the following relative L_2 error (RLE) to evaluate the accuracy of the prediction by different GRF-based methods: $\|u - \tilde{u}\|_2/\|u\|_2$, where u is the exact function and $\tilde{u}$ is the prediction by GRF-based methods.

4.1 Example 1

Consider the following function:

$$y(x) = (3x - 1)^2 \sin(14x - 4) \tag{80}$$

on $x \in [0, 1]$. In this example, we set $\theta = (\sigma, l) \in [0.1, 10] \times [0.1, 0.2]$, and we consider the following four different cases:

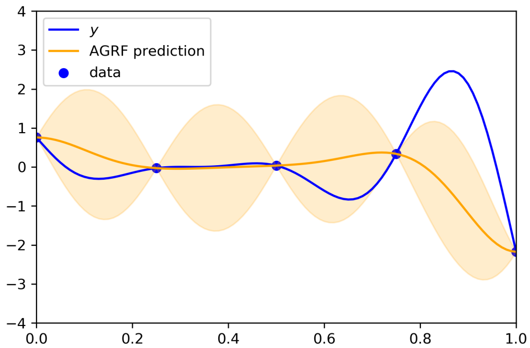
Case 1 The data include the observable at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$ only. Figure 1 shows y, y', y'' predicted by AGRF with confidence intervals. The AGRF surrogate model for y is the same as that obtained by the conventional GP regression, because the former is a generalization of the latter. However, the conventional GP regression does not provide the prediction of y' or y''.

Case 2 The data include the observable at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$ and its first order derivative at $x \in \{0.2, 0.5, 0.8\}$. Figure 2 shows that the AGRF prediction matches the exact function and its derivatives better than the results shown in Case 1. This is because the derivative information is incorporated in the model.

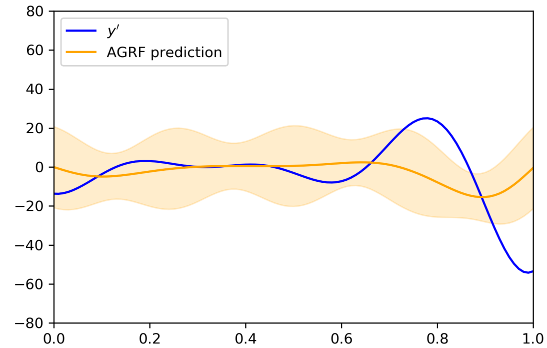
Case 3 The data include the observable at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$ and its second order derivative at $x \in \{0.1, 0.5, 0.9\}$. Figure 3 shows that, like Case 2, the AGRF prediction imitates the ups and downs that are present in the exact function but not fully reflected in the observable data. The prediction accuracy is enhanced by incorporating the second order derivative, and it is better than the prediction in Case 1.

Case 4 The data include the observable at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$, its first order derivative at $x \in \{0.2, 0.5, 0.8\}$, and its second order derivative at $x \in \{0.1, 0.5, 0.9\}$. Figure 4 demonstrates that the AGRF prediction of y almost coincides with the exact function, and the prediction of y' and y'' are also very accuracy in most regions. It is not surprising that by using all available information we can construct the most accuracy surrogate model among all four cases.

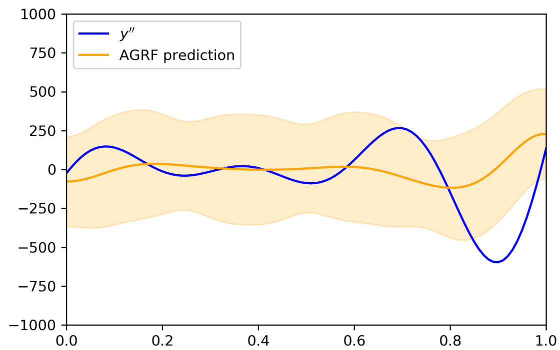
We can see that AGRF is able to integrate the observable and its derivatives of any order, regardless of the location where they are collected. As one might expect, the AGRF prediction improves when more information is available. Table 1 provides a quantitative comparison of the prediction accuracy in the four cases above, which further verifies our observations in Figures 1-4.



(a) y prediction



(b) y' prediction



(c) y'' prediction

Figure 1: [Example 1] Case 1: The data include the observable at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$. The AGRF prediction is the posterior mean and the shaded region is [posterior mean]$\pm 2\times$[posterior standard deviation].

4.2 Example 2

Consider the damped harmonic oscillator:

$$\begin{cases} F = -ky - cy' \\ F = my'' \end{cases} \tag{81}$$

where y is the displacement, y' is the velocity, y'' is the acceleration, $-ky$ is the restoring force, and $-cy'$ is the frictional force. This system can be simplified to:

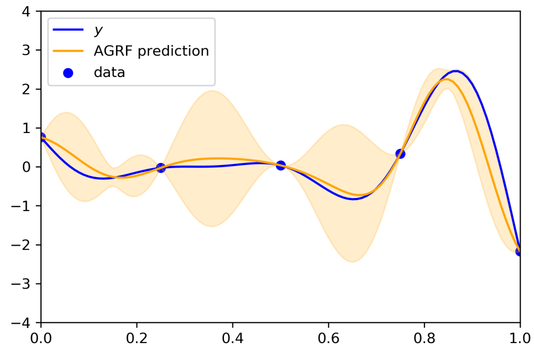
$$y'' + 2\zeta\omega_0 y' + \omega_0^2 y = 0, \tag{82}$$

where $\omega_0 = \sqrt{k/m}$ is the undamped angular frequency and $\zeta = c/\sqrt{4mk}$ is the damping ratio. When $\zeta < 1$, it has the solution:

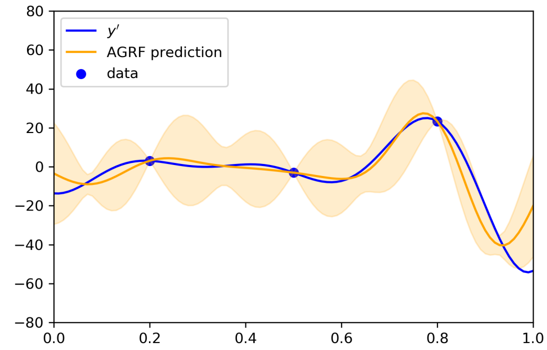
$$y(t) = A\exp(-\zeta\omega_0 t)\sin\left(\sqrt{1-\zeta^2}\omega_0 t + \phi\right), \tag{83}$$

where the amplitude A and the phase ϕ determine the behavior needed to match the initial conditions. Now, consider a specific example:

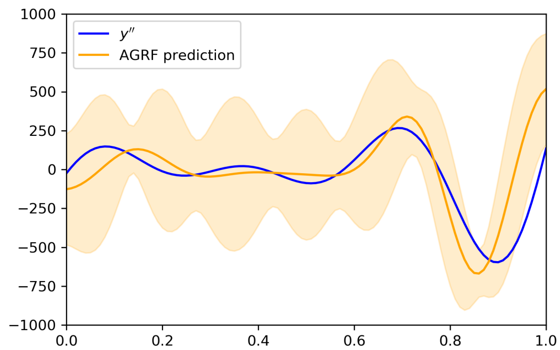
$$y(t) = \exp(-0.1 \times 20t)\sin\left(\sqrt{1-0.1^2} \times 20t\right) \tag{84}$$



(a) y prediction



(b) y' prediction



(c) y'' prediction

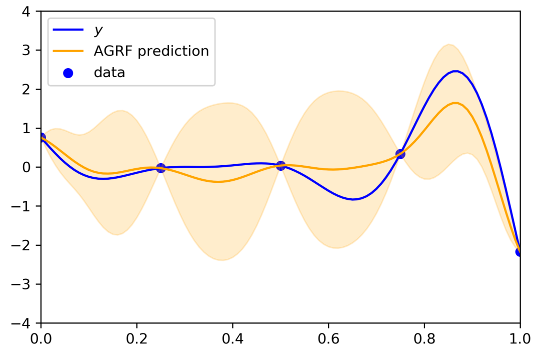
Figure 2: [Example 1] Case 2: The data include the observable at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$ and its first order derivative at $x \in \{0.2, 0.5, 0.8\}$. The AGRF prediction is the posterior mean and the shaded region is [posterior mean] $\pm 2 \times$ [posterior standard deviation].

on $t \in [0, 1]$. We use conventional GP regression (i.e., AGRF using the data of observable only), gradient-enhanced Kriging [13] (i.e., AGRF using the data of observable and its first order derivative), and AGRF (using the data of observable and all of its derivatives available) to construct surrogate models, and we set $\theta = (\sigma, l) \in [0.1, 10] \times [0.1, 0.2]$.

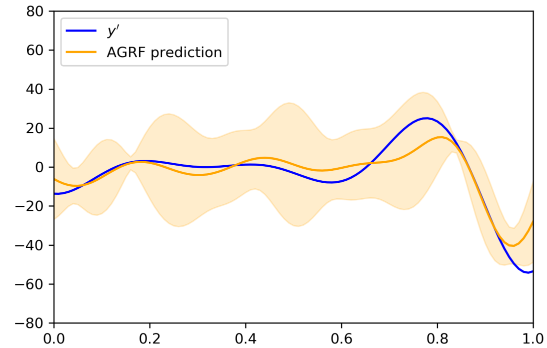
Case 1 We use the conventional GP regression. The data include the observable and its first order derivative at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$. The observable data are used to predict the displacement and the first order derivative data are used to predict the velocity, respectively. The results are shown in Figure 5. Apparently, the prediction is not accurate.

Case 2 We use GEK. The data include the observable and its first order derivative at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$. All these data are used jointly in the same random field to predict the displacement and the velocity at the same time. Figure 6 demonstrates that the prediction of y and y' is much more accurate than those in Case 1. Yet in the phase diagram plot, there is still significant discrepancy between the prediction and the exact solution when $t \in [0.25, 0.5]$.

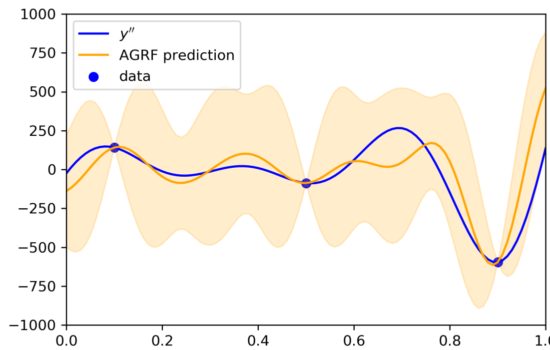
Case 3 We use AGRF. The data include the observable, its first order derivative, and its second order derivative at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$. All the data are used together in the same random field to



(a) y prediction



(b) y' prediction



(c) y'' prediction

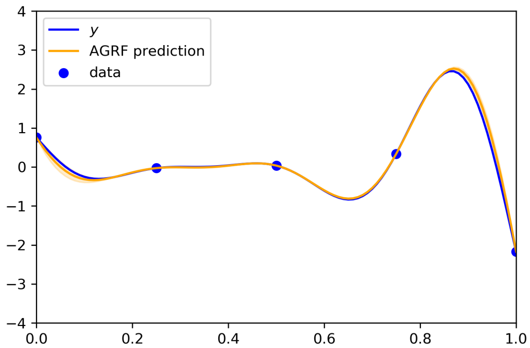
Figure 3: [Example 1] Case 3: The data include the observable at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$ and its second order derivative at $x \in \{0.1, 0.5, 0.9\}$. The AGRF prediction is the posterior mean and the shaded region is [posterior mean] $\pm 2 \times$ [posterior standard deviation].

predict the displacement and the velocity at the same time. The results in Figure 7 show much better accuracy than the corresponding ones in Figures 5 and 6, as the prediction almost coincides with the exact solution and the confidence intervals are very narrow.

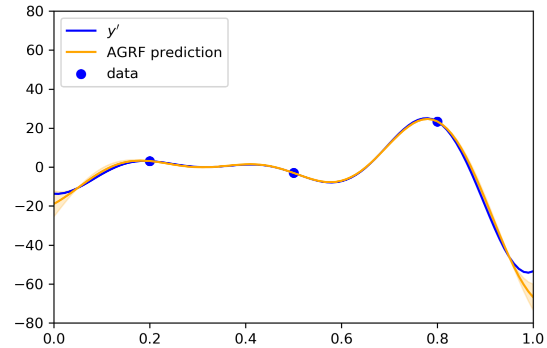
As in Example 1, we can see that the prediction is better when the observable and its derivative are used together (Case 2) than when they are used separately (Case 1). Furthermore, the inclusion of second order derivative can further enhance the prediction (Case 3). Table 2 provides a quantitative comparison of these three cases.

5 Conclusion

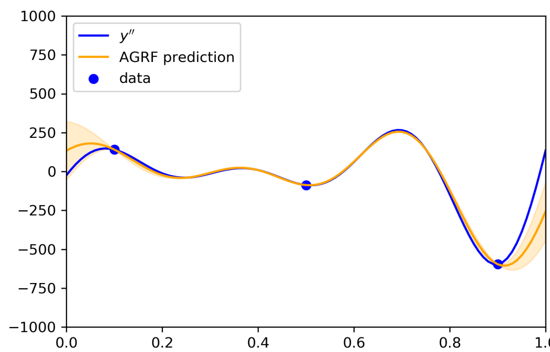
In this work, we provide a comprehensive theoretical foundation for incorporating arbitrary order derivatives in GRF-based methods. We demonstrate that under some conditions, the derivative of each realization of the GRF representing the observable can be considered as a realization of the GRF representing the corresponding derivative of the observable. Combining these GRFs, we propose the *augmented Gaussian random field* (AGRF), which is a universal framework incorporating the observable and its derivatives. Consequently, the computation of the posterior mean and variance of the observable and its derivatives



(a) y prediction



(b) y' prediction



(c) y'' prediction

Figure 4: [Example 1] Case 4: The data include the observable at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$, its first order derivative at $x \in \{0.2, 0.5, 0.8\}$, and its second order derivative at $x \in \{0.1, 0.5, 0.9\}$. The AGRF prediction is the posterior mean and the shaded region is [posterior mean] $\pm\, 2 \times$ [posterior standard deviation].

at new locations can be carried out in a similar way as in the conventional GP regression method. Our numerical results show that, in general, using more information of the system, we get better prediction in terms of the accuracy of the posterior mean and the width of the confidence interval as in the literature. The main advantage is that our universal framework provides a natural way to incorporate arbitrary order derivatives and deal with missing data, e.g., the observation of the observable or its derivative is missing at some sampling locations.

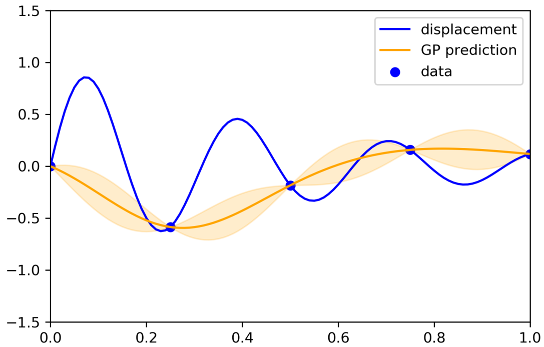
To the best of our knowledge, this is the first systematic work that proves the validity of the intuitive assumption that the derivative of a GRF is still a GRF, which is widely used in probabilistic scientific computing and GRF-based (or GP-based) regression methods. Although we use one-dimensional system for demonstration, our conclusion can be extended to multi-dimensional systems without difficulty.

Acknowledgement

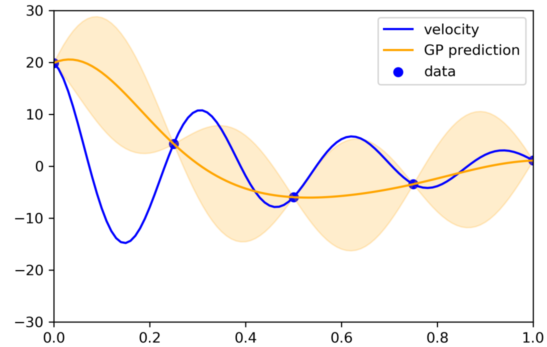
SZ was supported by National Science Foundation (NSF) Mathematical Sciences Graduate Internship (MSGI) Program sponsored by the NSF Division of Mathematical Sciences. XY was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research (ASCR) as

Table 1: [Example 1] The RLE of the AGRF prediction of y, y', y'' on $x \in [0, 1]$.

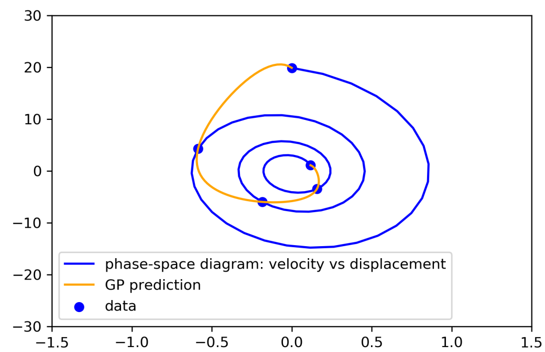
	Data	RLE of y	RLE of y'	RLE of y''
Case 1	y	1.347	0.960	1.095
Case 2	y, y'	0.356	0.412	0.805
Case 3	y, y''	0.501	0.387	0.620
Case 4	y, y', y''	0.088	0.126	0.354



(a) y prediction



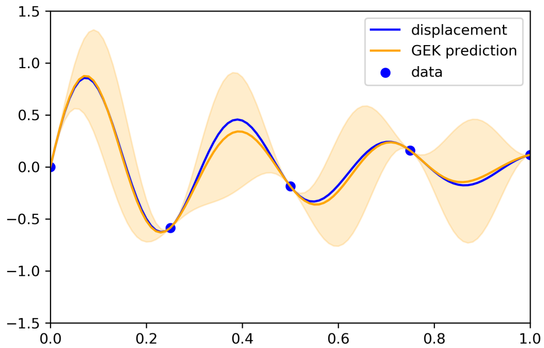
(b) y' prediction



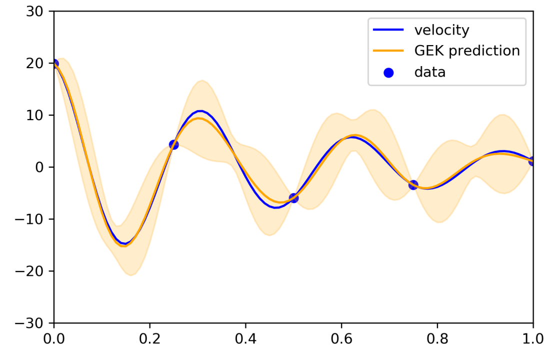
(c) Phase-space diagram prediction

Figure 5: [Example 2] Case 1: Prediction of the displacement and the velocity using conventional GP regression. The data include the observable and its first order derivative at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$. The observable data are used to predict the displacement and the first order derivative data are used to predict the velocity, respectively. The GP prediction is the posterior mean and the shaded region is [posterior mean] $\pm 2 \times$ [posterior standard deviation].

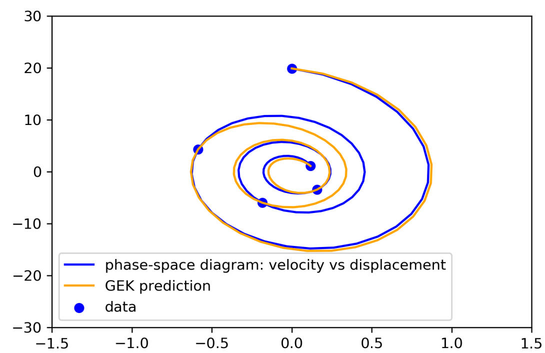
part of Multifaceted Mathematics for Rare, Extreme Events in Complex Energy and Environment Systems (MACSER). GL was supported by the National Science Foundation (DMS-1555072, DMS-1736364, CMMI-1634832 and CMMI-1560834), and Brookhaven National Laboratory Subcontract 382247.



(a) y prediction



(b) y' prediction



(c) Phase-space diagram prediction

Figure 6: [Example 2] Case 2: Prediction of the displacement and the velocity using GEK. The data include the observable and its first order derivative at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$. All the data are used jointly in the same random field to predict the displacement and the velocity at the same time. The GEK prediction is the posterior mean and the shaded region is [posterior mean] $\pm\, 2 \times$ [posterior standard deviation].

References

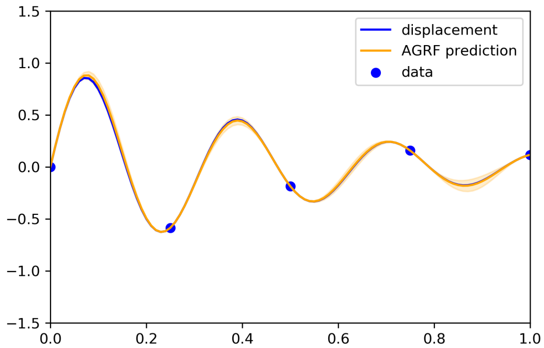
[1] Petter Abrahamsen. A review of Gaussian random fields and correlation functions, 1997.

[2] Oksana A Chkrebtii, David A Campbell, Ben Calderhead, Mark A Girolami, et al. Bayesian solution uncertainty quantification for differential equations. *Bayesian Anal.*, 11(4):1239–1267, 2016.

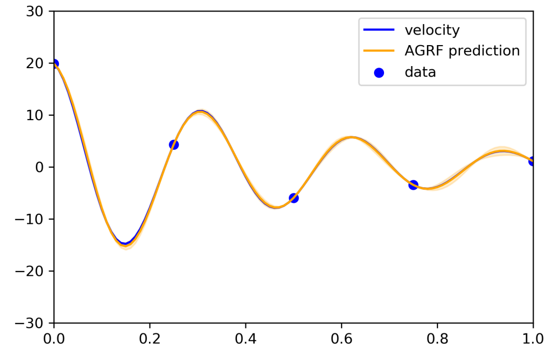
[3] H-S Chung and J Alonso. Using gradients to construct cokriging approximation models for high-dimensional design optimization problems. In *40th AIAA Aerospace Sciences Meeting & Exhibit*, page 317, 2002.

[4] Jon Cockayne, Chris J Oates, TJ Sullivan, and Mark Girolami. Bayesian probabilistic numerical methods. *SIAM Rev.*, 61(4):756–789, 2019.

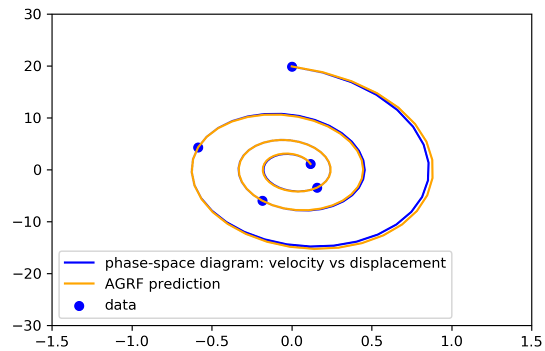
[5] Harald Cramér and M. R. Leadbetter. *Stationary and related stochastic processes: sample function properties and their applications*. Dover books on mathematics. Dover Publications, dover ed edition. OCLC: ocm56532664.



(a) y prediction



(b) y' prediction



(c) Phase-space diagram prediction

Figure 7: [Example 2] Case 3: Prediction of the displacement and the velocity using AGRF. The data include the observable, its first order derivative, and its second order derivative at $x \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$. All the data are used together in the same random field to predict the displacement and the velocity at the same time. The AGRF prediction is the posterior mean and the shaded region is [posterior mean] $\pm\, 2 \times$ [posterior standard deviation].

[6] Sébastien Da Veiga and Amandine Marrel. Gaussian process modeling with inequality constraints. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 21, pages 529–555, 2012.

[7] Yixiang Deng, Guang Lin, and Xiu Yang. Multifidelity data fusion via gradient-enhanced Gaussian process regression. *arXiv preprint arXiv:2008.01066*, 2020.

[8] R. M. Dudley. *Real analysis and probability.* Number 74 in Cambridge studies in advanced mathematics. Cambridge University Press, 2002.

[9] Alexander IJ Forrester, András Sóbester, and Andy J Keane. Multi-fidelity optimization via surrogate modelling. *Proc. R. Soc. A.*, 463(2088):3251–3269, 2007.

[10] Philipp Hennig, Michael A Osborne, and Mark Girolami. Probabilistic numerics and uncertainty in computations. *Proc. R. Soc. A.*, 471(2179):20150142, 2015.

[11] Marc C Kennedy and Anthony O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.

Table 2: [Example 2] The RLE of the prediction of the displacement and the velocity on $x \in [0, 1]$.

	Data	RLE of y	RLE of y'
Case 1	y, y' (separately)	1.334	1.551
Case 2	y, y' (together)	0.120	0.092
Case 3	y, y', y'' (together)	0.031	0.025

[12] Peter K Kitanidis. *Introduction to Geostatistics: Applications in Hydrogeology.* Cambridge University Press, 1997.

[13] Luc Laurent, Rodolphe Le Riche, Bruno Soulier, and Pierre-Alain Boucard. An overview of gradient-enhanced metamodels with applications. *Archives of Computational Methods in Engineering*, 26(1):61–106, 2019.

[14] Loic Le Gratiet and Josselin Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *Int. J. Uncertainty Quantification*, 4(5), 2014.

[15] Lizhen Lin and David B Dunson. Bayesian monotone regression using Gaussian process projection. *Biometrika*, 101(2):303–317, 2014.

[16] Andrés F López-Lopera, François Bachoc, Nicolas Durrande, and Olivier Roustant. Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA J. Uncertainty Quantification*, 6(3):1224–1255, 2018.

[17] Max D Morris, Toby J Mitchell, and Donald Ylvisaker. Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics*, 35(3):243–255, 1993.

[18] Andrew Pensoneault, Xiu Yang, and Xueyu Zhu. Nonnegativity-enforced Gaussian process regression. *arXiv preprint arXiv:2004.04632*, 2020.

[19] Paris Perdikaris, Maziar Raissi, Andreas Damianou, ND Lawrence, and George Em Karniadakis. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proc. R. Soc. A*, 473(2198):20160751, 2017.

[20] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Machine learning of linear differential equations using Gaussian processes. *J. Comput. Phys.*, 348:683–693, 2017.

[21] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Numerical Gaussian processes for time-dependent and nonlinear partial differential equations. *SIAM J. Sci. Comput.*, 40(1):A172–A198, 2018.

[22] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning.* Adaptive computation and machine learning. MIT Press, 2006. OCLC: ocm61285753.

[23] Jerome Sacks, William J Welch, Toby J Mitchell, and Henry P Wynn. Design and analysis of computer experiments. *Statistical science*, pages 409–423, 1989.

[24] Michael Schober, David K Duvenaud, and Philipp Hennig. Probabilistic ODE solvers with Runge-Kutta means. In *Adv. Neural Inf. Process. Syst.*, pages 739–747, 2014.

[25] Selvakumar Ulaganathan, Ivo Couckuyt, Francesco Ferranti, Eric Laermans, and Tom Dhaene. Performance study of multi-fidelity gradient enhanced kriging. *Struct. Multidiscipl. Optim.*, 51(5):1017–1033, 2015.

[26] Xiu Yang, David Barajas-Solano, Guzel Tartakovsky, and Alexandre M Tartakovsky. Physics-informed Cokriging: A Gaussian-process-regression-based multifidelity method for data-model convergence. *J. Comput. Phys.*, 395:410–431, 2019.

[27] Xiu Yang, Guzel Tartakovsky, and Alexandre Tartakovsky. Physics-informed Kriging: A physics-informed Gaussian process regression method for data-model convergence. *arXiv preprint arXiv:1809.03461*, 2018.

[28] Xiu Yang, Xueyu Zhu, and Jing Li. When bifidelity meets Cokriging: An efficient physics-informed multifidelity method. *SIAM J. Sci. Comput.*, 42(1):A220–A249, 2020.