

Bayesian approach to Gaussian process regression with uncertain inputs

Dongwei Ye* and Mengwu Guo^{†‡}

Department of Applied Mathematics, University of Twente

Abstract

Conventional Gaussian process regression exclusively assumes the existence of noise in the output data of model observations. In many scientific and engineering applications, however, the input locations of observational data may also be compromised with uncertainties owing to modeling assumptions, measurement errors, etc. In this work, we propose a Bayesian method that integrates the variability of input data into Gaussian process regression. Considering two types of observables – noise-corrupted outputs with fixed inputs and those with prior-distribution-defined uncertain inputs, a posterior distribution is estimated via a Bayesian framework to infer the uncertain data locations. Thereafter, such quantified uncertainties of inputs are incorporated into Gaussian process predictions by means of marginalization. The effectiveness of this new regression technique is demonstrated through several numerical examples, in which a consistently good performance of generalization is observed, while a substantial reduction in the predictive uncertainties is achieved by the Bayesian inference of uncertain inputs.

1 Introduction

Uncertainty estimation is essential for model validation in scientific computing, as computational modeling in science and engineering typically involves intrinsic system variability and/or incertitude due to a lack of knowledge. The uncertainties propagate through the modeling process and may lead to challenges in both forward and inverse problems. Bayesian statistics provides a methodology for quantifying the impact of various uncertainty sources and enables a probabilistic integration of prior knowledge into the inference process [5]. Contrary to point estimates, the Bayesian framework facilitates model predictions with probabilistic distributions and has been widely used in system design, analysis, and optimization under uncertainties.

Gaussian process (GP) regression is one of the state-of-the-art methods for supervised learning [24]. It is frequently synergized with Bayesian inference owing to its probabilistic nature [11, 1],

*Corresponding author. Email: d.ye-1@utwente.nl

[†]Email: m.guo@utwente.nl

[‡]The two authors contributed equally to this work.

and has been applied to a wide variety of tasks in computational science and engineering [7, 32, 15, 27]. The standard GP model is established based on the assumption that the input data have full certainty, while the output data are possibly corrupted by measurement noise and/or perturbed by intrinsic system stochasticity. GP regression often models such uncertainties in output data with an additive white Gaussian noise term, whose variance is calibrated as part of the hyperparameters.

However, it is often the case that the assumption of uncertainties exclusively on output data is too strong, and it hence becomes crucial to take uncertain data locations (inputs) into account. An intuitive example is the sensor placement and data measurement for the predictive monitoring of engineering assets, in which case sensors are deployed at several particular positions to examine certain mechanical behaviors of interest. Due to possible human errors, structure deformation, or interference of environment, the actual sensor locations can deviate from the desired placement, and the observational data are consequently measured from uncertain input locations. Similar circumstances also occur in mapping problems of robotic navigation [12], system identification in marine science [6], etc.

Several strategies have been proposed to embed input uncertainties into the GP regression framework, either during the training process or in the predicting phase, or both. [14] proposed that the probabilistic distribution of a test point can be integrated into model predictions via marginalization with Gaussian approximations. The premise of this method focuses on incorporating input uncertainties once a GP model is trained. On the other hand, to account uncertain data locations into training, [23] tuned GP hyperparameters by maximizing a marginal likelihood that encloses uncertain inputs, while [8] integrated input uncertainties into the GP covariance functions. The learning process may also be proceeded without any prior distributions defined of the input uncertainties, as these data locations can be inferred together with hyperparameters [22].

The existing methods tend to combine assumed input uncertainties directly into the formulations of GP regression. However, it is worth noting that the collected data can be used to probabilistically update prior assumptions on the uncertain data locations. To this end, this work proposes a new method that integrates input variability into GP regression, aided by a Bayesian inference of uncertain data locations. In such a supervised learning process, we assume the availability of two types of data: noise-corrupted outputs with fixed inputs, and those with uncertain inputs described by a prior distribution. Through the Bayes' rule, the proposed method estimates a posterior distribution of the uncertain inputs by leveraging all available information from the data, and thereafter embeds such quantified input uncertainties into GP predictions through marginalization.

2 Related work

GP modeling was first proposed by Krige for geo-statistical analysis [17] and extensively studied later to address regression problems. Many variations of GP regression models have been developed for specific contexts, such as vector-valued GPs for multi-output functions [18, 3], sparse GPs with inducing inputs to efficiently learn from large datasets [28], and deep GPs with a multi-layer network structure for an improved nonlinear expressive power [10]. The GP models have been endowed with remarkable flexibility in numerical applications [20, 21, 16, 4, 33] by their non-parametric inherent nature in probabilistic machine learning.

The studies on GP regression with uncertain input data were mainly motivated by the

demand for a robust way to quantify the propagation of uncertainties from data locations to GP predictions, while guaranteeing a good generalization performance. Most existing methods can be categorized into two types. The first type reflects the input uncertainties in GP kernel functions. [13] performed a second-order Taylor expansion to correct the covariance function with uncertain inputs, and the uncertainty in both input and output data can be learned via maximum likelihood. [8, 9] directly marginalized the distribution of data inputs in the covariance functions, which presented better generality than [13] as the high-order terms were not omitted. The second type of methods assumes that the input uncertainties result in an additional noise term in the output, which aligns with the form of heterogeneous GP [22]. This method shows the intuition that the input uncertainties have a significant influence on the area where the output changes rapidly. An alternative is to fuse the determination of data locations into hyperparameter estimation [23], for which gradient descent is used to optimize an evidence lower bound integrated with the input distribution. These available methods have been applied to diverse contexts of scientific computing, such as optimization [26, 30], active learning [34, 19], system identification [31, 2].

3 Preliminary

3.1 Gaussian process regression

A general description of the standard GP regression [24] is presented in this section. Assume that the model response $y \in \mathbb{R}$ is measured from a physical process or a computer experiment as $y = f(\mathbf{x}) + \epsilon_n$, given the corresponding input $\mathbf{x} \in \mathbb{R}^d$. Here the regression function f follows a GP prior with certain mean function, set to be zero for simplicity in this work, and a covariance (kernel) function $k_\theta(\cdot, \cdot)$ that represents the sample correlation between input locations. The covariance function is featured with hyperparameters θ , i.e., $f(\cdot) \sim \mathcal{GP}(0, k_\theta(\cdot, \cdot))$, and $\epsilon_n \sim \mathcal{N}(0, \sigma_n^2)$ is an independent white noise term. Given the observed data collection $(\mathbf{X}, \mathbf{y}) = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$, a joint normal distribution between the observed data and the output $f^*(\mathbf{x}^*)$ at a test point \mathbf{x}^* is hence defined by the GP prior as

$$\begin{bmatrix} \mathbf{y} \\ f^* \end{bmatrix} \Big| \mathbf{x}^*, \mathbf{X}, \theta, \sigma_n^2 \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} k_\theta(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & k_\theta(\mathbf{X}, \mathbf{x}^*) \\ k_\theta(\mathbf{x}^*, \mathbf{X}) & k_\theta(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix} \right). \quad (1)$$

Note that the choice of kernel should reflect certain desired properties of the regression function. The hyperparameters θ in the kernel as well as the noise variance σ_n^2 are often determined by maximizing the log-marginal likelihood:

$$\begin{aligned} (\tilde{\theta}, \tilde{\sigma}_n^2) &= \arg \max_{\theta, \sigma_n^2} \log p(\mathbf{y} | \mathbf{X}, \theta, \sigma_n^2) \\ &= \arg \max_{\theta, \sigma_n^2} \left[-\frac{1}{2} \mathbf{y}^\top (\mathbf{K}(\theta) + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}(\theta) + \sigma_n^2 \mathbf{I}| - \frac{N}{2} \log 2\pi \right], \end{aligned} \quad (2)$$

in which $\mathbf{K}(\theta) := k_\theta(\mathbf{X}, \mathbf{X})$. Conditioning on the observed data, the posterior predictive distribution of the noise-free output f^* at an unobserved point \mathbf{x}^* again follows a normal distribution:

$$f^* | \mathbf{x}^*, \mathbf{X}, \mathbf{y}, \tilde{\theta}, \tilde{\sigma}_n^2 \sim \mathcal{N}(m(\mathbf{x}^* | \mathbf{X}, \mathbf{y}), v(\mathbf{x}^* | \mathbf{X})), \quad (3)$$

where $m(\mathbf{x}^* | \mathbf{X}, \mathbf{y}) = k_{\tilde{\theta}}(\mathbf{x}^*, \mathbf{X}) [k_{\tilde{\theta}}(\mathbf{X}, \mathbf{X}) + \tilde{\sigma}_n^2 \mathbf{I}]^{-1} \mathbf{y}$ stands for the predictive mean and $v(\mathbf{x}^* | \mathbf{X}) = k_{\tilde{\theta}}(\mathbf{x}^*, \mathbf{x}^*) - k_{\tilde{\theta}}(\mathbf{x}^*, \mathbf{X}) [k_{\tilde{\theta}}(\mathbf{X}, \mathbf{X}) + \tilde{\sigma}_n^2 \mathbf{I}]^{-1} k_{\tilde{\theta}}(\mathbf{X}, \mathbf{x}^*)$ denotes the corresponding posterior variance.

3.2 Problem statement

We consider two types of available training data: $(\mathbf{X}^f, \mathbf{y}^f) = \{(\mathbf{x}_i^f, y_i^f)\}_{i=1}^{N_f}$ with fixed input locations $\mathbf{x}_i^f \in \mathbb{R}^d$ and the corresponding observations $y_i^f \in \mathbb{R}$, and $(\mathbf{X}^u, \mathbf{y}^u) = \{(\mathbf{x}_i^u, y_i^u)\}_{i=1}^{N_u}$ consisting of uncertain input locations \mathbf{x}_i^u and the corresponding observation y_i^u . Note that the extreme case with $N_f = 0$ fits in this setting and, contrarily, the problem degenerates to the standard GP when $N_u = 0$. Each uncertain input \mathbf{x}_i^u is considered as a random variable associated with an independent prior set to be a joint normal distribution,

$$\mathbf{x}_i^u \sim \mathcal{N}(\boldsymbol{\mu}_i, \text{diag}(s_{i,1}^2, s_{i,2}^2, \dots, s_{i,d}^2)), \quad i = 1, 2, \dots, N_u, \quad (4)$$

where $\boldsymbol{\mu}_i$ denotes the mean vector of the prior; the uncertain inputs are additionally assumed to have uncorrelated entries of coordinates, and the covariance matrix can thus be written as a diagonal matrix $\text{diag}(s_{i,1}^2, s_{i,2}^2, \dots, s_{i,d}^2)$ with $s_{i,j}^2$ being the variance of the j th coordinate of the i th uncertain data location. Therefore, the joint prior distribution of all the uncertain inputs is given by:

$$p(\mathbf{X}^u | \boldsymbol{\phi}) = \prod_{i=1}^{N_u} p(\mathbf{x}_i^u | \boldsymbol{\mu}_i, s_{i,1}^2, s_{i,2}^2, \dots, s_{i,d}^2). \quad (5)$$

where $\boldsymbol{\phi} = \{\boldsymbol{\mu}_i, s_{i,1}^2, s_{i,2}^2, \dots, s_{i,d}^2\}_{i=1}^{N_u}$ collects all the hyperparameters for defining the uncertain inputs' prior. In this work, we predefine $\boldsymbol{\phi}$ to let $\boldsymbol{\mu}_i$ represent an initial guess of the uncertain input location and $\pm 2s_{i,j}$ reflects the 95% confidence.

4 Method

4.1 Bayesian inference of uncertain data locations

The uncertain input locations can be inferred in a Bayesian manner from the combination of their prior distributions and the corresponding outputs, aided by the GP assumption for the regression function. To be subsequently leveraged for predictions, the posterior distribution of these uncertain inputs is directly given by the Bayes' rule as follows:

$$p(\mathbf{X}^u | \mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_n^2) \propto p(\mathbf{X}^u | \boldsymbol{\phi}) p(\mathbf{y}^u, \mathbf{y}^f | \mathbf{X}^u, \mathbf{X}^f, \boldsymbol{\theta}, \sigma_n^2), \quad (6)$$

in which the second term is a likelihood function defined by the GP, i.e.,

$$\begin{bmatrix} \mathbf{y}^f \\ \mathbf{y}^u \end{bmatrix} \Big| \mathbf{X}^u, \mathbf{X}^f, \boldsymbol{\theta}, \sigma_n^2 \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} k_{\boldsymbol{\theta}}(\mathbf{X}^f, \mathbf{X}^f) + \sigma_n^2 \mathbf{I} & k_{\boldsymbol{\theta}}(\mathbf{X}^f, \mathbf{X}^u) \\ k_{\boldsymbol{\theta}}(\mathbf{X}^u, \mathbf{X}^f) & k_{\boldsymbol{\theta}}(\mathbf{X}^u, \mathbf{X}^u) + \sigma_n^2 \mathbf{I} \end{bmatrix} \right). \quad (7)$$

This inference 6 can barely be evaluated exactly as its marginal likelihood, given as

$$p(\mathbf{y}^u, \mathbf{y}^f | \mathbf{X}^f, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_n^2) = \int p(\mathbf{X}^u | \boldsymbol{\phi}) p(\mathbf{y}^u, \mathbf{y}^f | \mathbf{X}^u, \mathbf{X}^f, \boldsymbol{\theta}, \sigma_n^2) d\mathbf{X}^u, \quad (8)$$

is typically intractable to compute in practice. Therefore, the Markov Chain Monte Carlo (MCMC) method is used to approximate the posterior distribution of uncertain data locations.

To determine the optimal values of GP hyperparameters $(\boldsymbol{\theta}, \sigma_n^2)$ in (6), we maximize an approximation of the possibly intractable marginal likelihood, in which the uncertain input locations are represented by their prior means, i.e.,

$$(\tilde{\boldsymbol{\theta}}, \tilde{\sigma}_n^2) = \arg \max_{(\boldsymbol{\theta}, \sigma_n^2)} \log p \left(\mathbf{y}^u, \mathbf{y}^f | \{\boldsymbol{\mu}_i\}_{i=1}^{N_u}, \mathbf{X}^f, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_n^2 \right), \quad (9)$$

whose specific formulation is given by (2). In this way, the GP hyperparameters can be marginalized by considering

$$\begin{aligned} & \int p(\cdot|\mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f, \phi, \boldsymbol{\theta}, \sigma_n^2) p(\boldsymbol{\theta}, \sigma_n^2 | \mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f) d\boldsymbol{\theta} d\sigma_n^2 \\ &= \int p(\cdot|\mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f, \phi, \boldsymbol{\theta}, \sigma_n^2) \delta(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}, \sigma_n^2 - \tilde{\sigma}_n^2) d\boldsymbol{\theta} d\sigma_n^2 = p(\cdot|\mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f, \phi, \tilde{\boldsymbol{\theta}}, \tilde{\sigma}_n^2), \end{aligned} \quad (10)$$

in which δ denotes the Dirac-delta function. This marginalization applies to both (6) and the Bayesian predictions in the next subsection.

4.2 Bayesian predictions for Gaussian process regression

Once the posterior description of the uncertain inputs is achieved, the predictive distribution of f^* at a test point \mathbf{x}^* can be further computed by marginalizing the posterior, i.e.,

$$\begin{aligned} & p(f^*|\mathbf{x}^*, \mathbf{X}^f, \mathbf{y}^f, \mathbf{y}^u, \phi, \tilde{\boldsymbol{\theta}}, \tilde{\sigma}_n^2) \\ &= \int p(f^*|\mathbf{x}^*, \mathbf{X}^f, \mathbf{X}^u, \mathbf{y}^f, \mathbf{y}^u, \tilde{\boldsymbol{\theta}}, \tilde{\sigma}_n^2) p(\mathbf{X}^u | \mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f, \phi, \tilde{\boldsymbol{\theta}}, \tilde{\sigma}_n^2) d\mathbf{X}^u. \end{aligned} \quad (11)$$

The MCMC-extracted samples for estimating the posterior of uncertain inputs can be reused in 11 to evaluate the integral with Monte Carlo. On the other hand, $p(f^*|\mathbf{x}^*, \mathbf{X}^f, \mathbf{X}^u, \mathbf{y}^f, \mathbf{y}^u, \tilde{\boldsymbol{\theta}}, \tilde{\sigma}_n^2)$ corresponds to the predictive distribution (3) of a GP model conditioning on the combination of both fixed data pairs and those with sampled locations of uncertain inputs. The first and second moments of the marginal predictive distribution 11 are derived as follows:

$$\mathbb{E}[f^*|\mathbf{x}^*, \mathbf{X}^f, \mathbf{y}^f, \mathbf{y}^u] = \mathbb{E}_{\mathbf{X}^u | \mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f} [m(\mathbf{x}^* | \mathbf{X}^f \cup \mathbf{X}^u, \mathbf{y}^f \cup \mathbf{y}^u)], \quad (12)$$

and similarly

$$\begin{aligned} & \mathbb{V}ar[f^*|\mathbf{x}^*, \mathbf{X}^f, \mathbf{y}^f, \mathbf{y}^u] \\ &= \mathbb{E}_{\mathbf{X}^u | \mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f} [v(\mathbf{x}^* | \mathbf{X}^f \cup \mathbf{X}^u)] + \mathbb{V}ar_{\mathbf{X}^u | \mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f} [m(\mathbf{x}^* | \mathbf{X}^f \cup \mathbf{X}^u, \mathbf{y}^f \cup \mathbf{y}^u)], \end{aligned} \quad (13)$$

where $m(\cdot)$ and $v(\cdot)$ were defined for the mean and variance of a GP prediction, $\mathbb{E}_{\mathbf{X}^u | \mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f}$ and $\mathbb{V}ar_{\mathbf{X}^u | \mathbf{X}^f, \mathbf{y}^u, \mathbf{y}^f}$ indicate the expectation and variance values over the posterior distribution of \mathbf{X}^u , respectively.

5 Experiments

The setting of numerical experiments and their results are provided in this section. We use the mean squared error (MSE) to measure estimation errors of the uncertain input locations, and the mean squared prediction error (MSPE) over $N_{\text{test}} = 100$ test points for GP predictions. These two error metrics are defined as follows,

$$\text{MSE} = \frac{1}{N_u} \mathbb{E}_{\mathbf{X}_u} [\|\mathbf{X}_u - \hat{\mathbf{X}}_u\|^2], \quad \text{MSPE} = \frac{1}{N_{\text{test}}} \mathbb{E}_{\mathbf{X}_u} \mathbb{E}_{f^* | \mathbf{X}_u} [\|f^*(\mathbf{X}^*) - \hat{f}(\mathbf{X}^*)\|^2], \quad (14)$$

in which $\hat{\mathbf{X}}_u$ collects the actual locations of uncertain data inputs, \mathbf{X}^* denotes the test points, \hat{f} is the ground truth function for regression, and \mathbf{X}_u may follow its prior *or* posterior distribution.

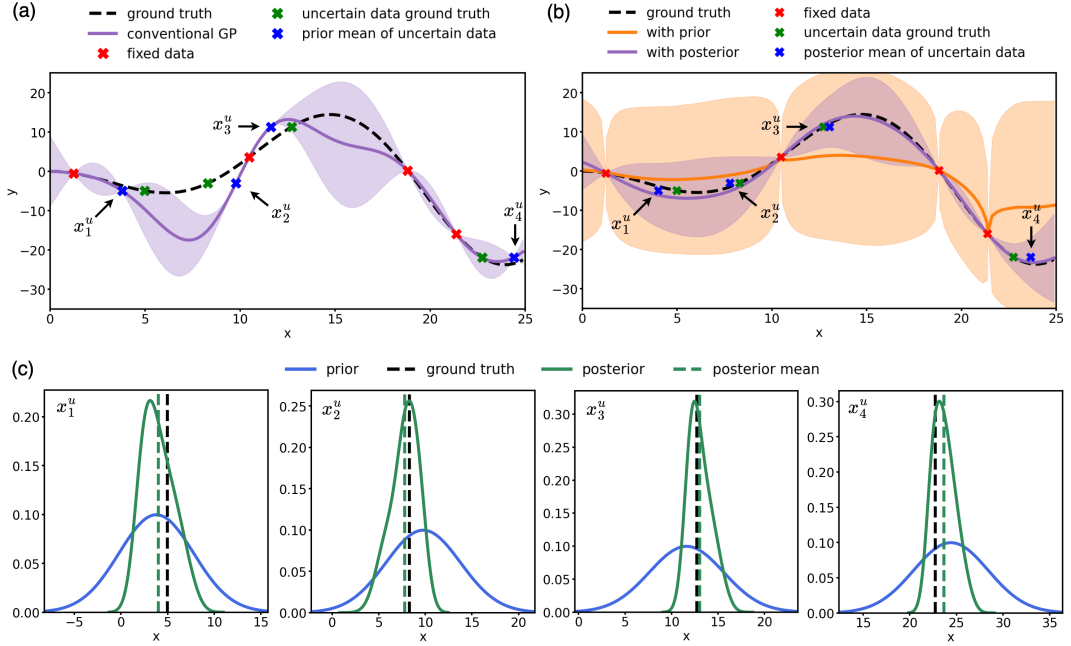


Figure 1: (a) Data points with fixed and uncertain input locations generated from latent function $y = -x \sin(x/3)$, and the result of standard GP fitting; x_1^u to x_4^u denote the four uncertain input locations. (b) A comparison of GP predictions marginalized over the prior and posterior distributions of the uncertain input locations. (c) Prior and posterior marginal distributions of each uncertain input location. Note that color bands in (a) and (b) represent the $\pm 2\sigma$ level of GP predictions in this figure. The predictions are depicted based on 100 test points distributed evenly over the domain of interest.

To demonstrate the details of the proposed method, an example with eight data points is presented first. The training data are generated from a one-dimensional latent function $y = -x \sin(x/3)$. We assume that four of the data points have fixed locations, while the other four have uncertain inputs x_i^u ($i = 1, 2, 3, 4$), for which a prior distribution is defined. The prior means of these uncertain data locations are computed by appending a perturbation to given locations, i.e., $\mu_i = \hat{x}_i^u + \epsilon$, where \hat{x}_i^u denotes a given actual location on the latent function and $\epsilon \sim \mathcal{U}(0, 2)$ is a randomly generated term to realize a perturbation. The original locations of these eight data points are sampled by the quasi-Monte Carlo method using a Sobol sequence [29]. The data points together with the latent function of ground truth are shown in Figure 1(a). We are interested in the regression over $[0, 8\pi]$ with an independent prior distribution $x_i^u \sim \mathcal{N}(\mu_i, s^2)$ for each uncertain input location. Assuming that we do not have much confidence in the priors, a relatively large variance $s^2 = 4$ is presumed. The measurement noise is omitted in this example for demonstration purpose.

Figure 1(a) depicts the predictive distribution of a conventional GP regression, which is trained with fixed data inputs and prior means of uncertain inputs, paired with their corresponding observations. It is evident that this GP regression fails to incorporate the input uncertainties, and the mean and variance of the GP predictive function are misled by the perturbed mean locations

of uncertain inputs in this interpolation task.

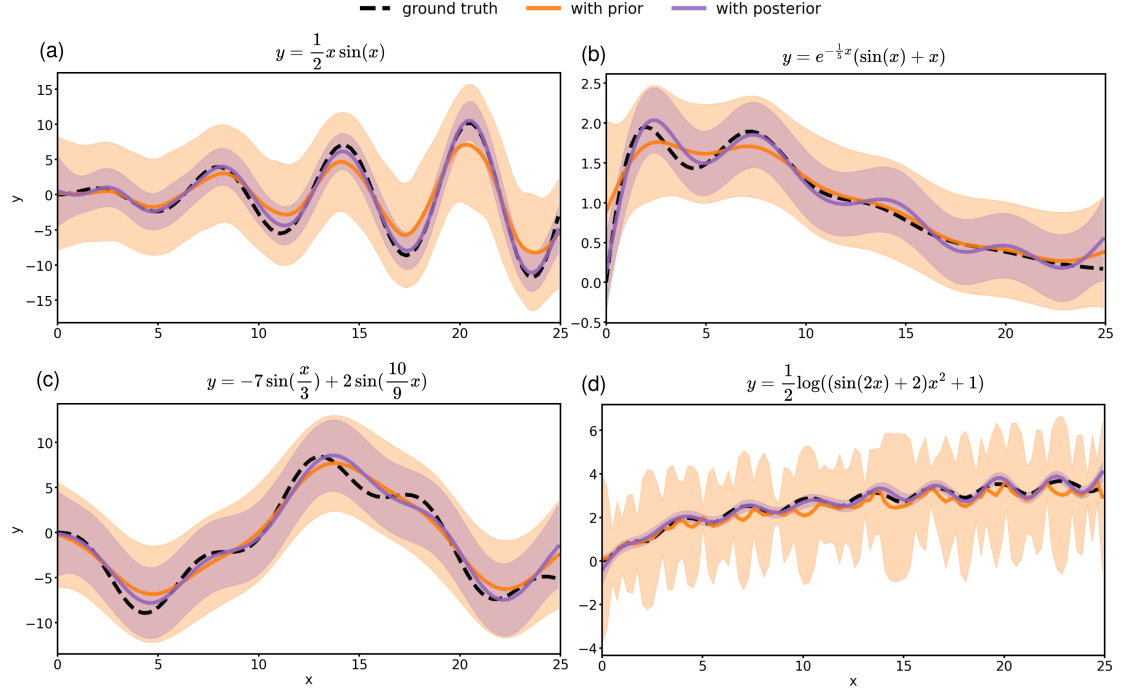


Figure 2: A comparison of GP predictions marginalized over the prior and posterior distributions of the uncertain input locations in four examples. Color bands represent the $\pm 2\sigma$ level of GP predictions in this figure. There are 30 training data points with fixed inputs and another 30 with uncertain inputs, and the predictions are based on 100 test points distributed evenly over $[0, 8\pi]$.

The results of the proposed Bayesian approach to GP regression is presented in Figures 1(b) and (c). A posterior estimation of uncertain data locations is performed first through Bayesian inference. As the marginal likelihood (evidence) is computationally intractable, an MCMC sampling with the Metropolis algorithm is employed to directly fetch samples for the posterior distribution of uncertain inputs. The probability density functions of both the prior (in blue) and the posterior (in green) distributions are shown in Figure 1(c), in which the posterior distributions are approximated by the technique of kernel density estimation [25]. It is clear that the update from prior to posterior has made the mean estimates of the uncertain inputs closer to the actual locations, and that the prior variance has meanwhile been reduced by leveraging all existing data. It is worth noting that, in practice, the ground truth function is generally unknown, and thus it is typically impossible to claim if the inference has indeed improved our knowledge in uncertain input locations. Nevertheless, the Bayesian inference of these locations exploits all the available information from training data to reduce/update their uncertainties. Subsequently, the samples that MCMC fetches from the posterior are further used to construct the predictions for new (unseen) locations over the entire domain $[0, 8\pi]$. To show a reduction of uncertainty in GP regression as a result of Bayesian inference for uncertain inputs, we present in Figure 1(b) the GP predictions marginalized over the prior (in orange) and posterior (in purple) distributions of uncertain inputs. The former is affected by the significant uncertainties assumed in the prior and

Numerical examples	Uncertain input locations			GP predictions		
	prior vs ground truth	posterior vs ground truth	relative reduction	prior vs ground truth	posterior vs ground truth	relative reduction
(a) $y = \frac{1}{2}x \sin(x)$	1.10	0.45	59.1%	16.42	2.33	85.8%
(b) $y = e^{-\frac{1}{5}x}(\sin(x) + x)$	1.10	0.84	23.6%	0.12	0.05	58.3%
(c) $y = -7 \sin(\frac{x}{3}) + 2 \sin(\frac{10}{9}x)$	1.10	0.89	19.1%	8.45	4.92	41.8%
(d) $y = \frac{1}{2} \log((\sin(2x) + 2)x^2 + 1)$	1.10	0.32	70.9%	1.56	0.05	96.8%

Table 1: A comparison of mean squared error (MSE) for the uncertain inputs between prior and posterior distributions against ground truth locations, and that of mean squared prediction error (MSPE) between prior and posterior GP predictions against ground truth functions. The relative reduction refers to the percentage of corresponding error reduction from prior to posterior.

results in high variance in the GP prediction. The mean function also deviates from the ground truth. On the contrary, the predictive mean (in the latter) by the proposed method captures the latent function well, and the corresponding predictive variance shows a good representation of the uncertain inputs, i.e., a relatively wide uncertainty band is observed around uncertain locations (x_1^u to x_4^u), while the variance pinches to a small value at fixed points. The MSPE at 100 test locations, respectively for the prior and posterior predictions by GP, are 143.2 and 11.9.

Results for another four functions are presented in Figure 2 and Table 1. More data are utilized in these examples – 30 each for fixed and uncertain inputs, and output measurements are assumed to be corrupted by noise. Such existence of output noise plays a role of regularization, and the value of σ_n^2 can be fine-tuned to prevent over- and under-fitting in GP regression. Similarly, predictions with and without an update of input location inference are shown for comparison. In Figure 2(a), both predictive mean functions have captured the pattern in the latent function; however, the one marginalized with the prior slightly underestimates at each wave peak and trough, while the other captures these values better. On the other hand, the $\pm 2\sigma$ level of the GP predictions, represented by the width of color bands, has been significantly reduced because of the update of uncertain input locations by Bayesian inference. In the cases (b) and (c), substantial improvement in the GP predictive distribution has been observed according to the comparison of MSPE in Table 1. Moreover, the advantage of our proposed method in Figure 2(d) is more evident. The result with prior information shows a poor predictive mean estimation, while our method captures the latent function well, and the narrow uncertainty band shows a high confidence in the prediction. Among the four examples, in general, the proposed method with Bayesian inference for uncertain data locations consistently achieves a meaningful reduction in the predictive uncertainties and offers an improved predictive mean estimation for the GP regression.

6 Conclusion

In this work, a Bayesian method for Gaussian process regression with uncertain data locations is discussed and demonstrated. Through a Bayesian inference, all available information from the data is leveraged to update the knowledge on uncertain input locations. By marginalizing a posterior distribution for the input uncertainties, the predictive distribution for new test points can be achieved via Gaussian process regression. Numerical experiments show that, in comparison with the results without Bayesian updating of data locations, the proposed method presents a

consistently improved performance in generalization and predictive uncertainty reduction.

Acknowledgement

The authors acknowledge the financial support from Sectorplan Bèta (NL) under the focus area *Mathematics of Computational Science*.

References

- [1] A. M. Alaa and M. van der Schaar. Bayesian inference of individualized treatment effects using multi-task Gaussian processes. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, 2017.
- [2] H. Bijl, T. B. Schön, J.-W. van Wingerden, and M. Verhaegen. System identification through online sparse Gaussian process regression with input noise. *IFAC Journal of Systems and Control*, 2:1–11, 2017.
- [3] I. Bilonis and N. Zabarás. Multi-output local Gaussian process regression: Applications to uncertainty quantification. *Journal of Computational Physics*, 231(17):5718–5746, 2012.
- [4] N. Botteghi, M. Guo, and C. Brune. Deep kernel learning of dynamical models from high-dimensional noisy data. *Scientific Reports*, 12(1):21530, Dec 2022.
- [5] G. E. Box and G. C. Tiao. *Bayesian inference in statistical analysis*. John Wiley & Sons, 2011.
- [6] M. C. Burkhart, Y. Heo, and V. M. Zavala. Measurement and verification of building systems under uncertain data: A Gaussian process modeling approach. *Energy and Buildings*, 75:189–198, 2014.
- [7] E. T. Y. Chang, M. Strong, and R. H. Clayton. Bayesian sensitivity analysis of a cardiac cell model using a Gaussian process emulator. *PLOS ONE*, 10(6):1–20, 06 2015.
- [8] P. Dallaire, C. Besse, and B. Chaib-draa. Learning Gaussian process models from uncertain data. In C. S. Leung, M. Lee, and J. H. Chan, editors, *Neural Information Processing*, pages 433–440, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
- [9] P. Dallaire, C. Besse, and B. Chaib-draa. An approximate inference with Gaussian process to latent functions from uncertain data. *Neurocomputing*, 74(11):1945–1955, 2011.
- [10] A. Damianou and N. D. Lawrence. Deep Gaussian processes. In C. M. Carvalho and P. Ravikumar, editors, *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, volume 31 of *Proceedings of Machine Learning Research*, pages 207–215, Scottsdale, Arizona, USA, 29 Apr–01 May 2013. PMLR.
- [11] R. Frigola, F. Lindsten, T. B. Schön, and C. E. Rasmussen. Bayesian inference and learning in Gaussian process state-space models with particle MCMC. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26, 2013.

- [12] M. Ghaffari Jadidi, J. V. Miro, and G. Dissanayake. Warped Gaussian processes occupancy mapping with uncertain inputs. *IEEE Robotics and Automation Letters*, 2(2):680–687, 2017.
- [13] A. Girard. Approximate methods for propagation of uncertainty with Gaussian process models. *PQDT - Global*, page 168, 2004.
- [14] A. Girard, C. Rasmussen, J. Quiñonero-Candela, and R. Murray-Smith. Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002.
- [15] M. Gulian, M. Raissi, P. Perdikaris, and G. Karniadakis. Machine learning of space-fractional differential equations. *SIAM Journal on Scientific Computing*, 41(4):A2485–A2509, 2019.
- [16] M. Guo and J. S. Hesthaven. Reduced order modeling for nonlinear structural analysis using Gaussian process regression. *Computer Methods in Applied Mechanics and Engineering*, 341:807–826, 2018.
- [17] D. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139, 1951.
- [18] H. Liu, J. Cai, and Y.-S. Ong. Remarks on multi-output Gaussian process regression. *Knowledge-Based Systems*, 144:102–121, 2018.
- [19] H. Liu, Y.-S. Ong, X. Shen, and J. Cai. When Gaussian process meets big data: A review of scalable GPs. *IEEE Transactions on Neural Networks and Learning Systems*, 31(11):4405–4423, 2020.
- [20] A. Marrel, B. Iooss, B. Laurent, and O. Roustant. Calculations of Sobol indices for the Gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3):742–751, 2009.
- [21] A. Marrel, B. Iooss, F. Van Dorpe, and E. Volkova. An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics & Data Analysis*, 52(10):4731–4744, 2008.
- [22] A. Mchutchon and C. Rasmussen. Gaussian process training with input noise. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24, 2011.
- [23] J. Quiñonero-Candela and S. T. Roweis. Data imputation and robust training with Gaussian processes. NIPS, 2003.
- [24] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [25] M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9(2):65–78, 1982.
- [26] K. M. Ryan, J. Kristensen, Y. Ling, S. Ghosh, I. Asher, and L. Wang. A Gaussian process modeling approach for fast robust design with uncertain inputs. volume Volume 7A: Structures and Dynamics of *Turbo Expo: Power for Land, Sea, and Air*. American Society of Mechanical Engineers, 06 2018.

- [27] F. Sahli Costabal, P. Perdikaris, E. Kuhl, and D. E. Hurtado. Multi-fidelity classification using Gaussian processes: Accelerating the prediction of large-scale computational models. *Computer Methods in Applied Mechanics and Engineering*, 357:112602, 2019.
- [28] E. Snelson and Z. Ghahramani. Sparse Gaussian processes using pseudo-inputs. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems*, volume 18. MIT Press, 2006.
- [29] I. M. Sobol and Y. L. Levitan. A pseudo-random number generator for personal computers. *Computers and Mathematics with Applications*, 37(4-5):33–40, 1999.
- [30] H. Wang, J. Yuan, and S. H. Ng. Gaussian process based optimization algorithms with input uncertainty. *IJSE Transactions*, 52(4):377–393, 2020.
- [31] Y. Xue, Y. Liu, C. Ji, G. Xue, and S. Huang. System identification of ship dynamic model based on Gaussian process regression with input noise. *Ocean Engineering*, 216:107862, 2020.
- [32] D. Ye, A. Nikishova, L. Veen, P. Zun, and A. G. Hoekstra. Non-intrusive and semi-intrusive uncertainty quantification of a multiscale in-stent restenosis model. *Reliability Engineering & System Safety*, 214:107734, 2021.
- [33] D. Ye, P. Zun, V. Krzhizhanovskaya, and A. G. Hoekstra. Uncertainty quantification of a three-dimensional in-stent restenosis model with surrogate modelling. *Journal of The Royal Society Interface*, 19(187):20210864, 2022.
- [34] X. Yue, Y. Wen, J. H. Hunt, and J. Shi. Active learning for Gaussian process considering uncertainties with application to shape control of composite fuselage. *IEEE Transactions on Automation Science and Engineering*, 18(1):36–46, 2021.