

Uncertainty Quantification in Probabilistic Machine Learning Models: Theory, Methods, and Insights

Marzieh Ajirak,* Anand Ravishankar,** Petar M. Djurić**

**Weill Cornell Medicine, Cornell University, New York, NY*

*†Department of Electrical and Computer Engineering,
Stony Brook University, Stony Brook, NY*

maa4083@med.cornell.edu, {anand.ravishankar, petar.djuric}@stonybrook.edu

Abstract—Uncertainty Quantification (UQ) is essential in probabilistic machine learning models, particularly for assessing the reliability of predictions. In this paper, we present a systematic framework for estimating both epistemic and aleatoric uncertainty in probabilistic models. We focus on Gaussian Process Latent Variable Models and employ scalable Random Fourier Features-based Gaussian Processes to approximate predictive distributions efficiently. We derive a theoretical formulation for UQ, propose a Monte Carlo sampling-based estimation method, and conduct experiments to evaluate the impact of uncertainty estimation. Our results provide insights into the sources of predictive uncertainty and illustrate the effectiveness of our approach in quantifying the confidence in the predictions.

Index Terms—Uncertainty quantification, aleatoric, epistemic, Gaussian process latent variable models, predictive distributions.

I. INTRODUCTION

Uncertainty Quantification (UQ) plays a fundamental role in probabilistic machine learning (ML). ML models are increasingly used in applications that require making decisions and where reliable predictions are essential. This is especially important in fields that demand reliable choices under uncertainty, such as medical diagnosis [1] autonomous systems [2], and scientific modeling in general [3]. However, most traditional ML models provide only point predictions without assessing their confidence. This, in turn, makes it difficult to quantify the reliability of the results [5].

To address this limitation, probabilistic ML models provide a foundational framework for representing uncertainty [4]. By explicitly modeling uncertainty [10], these models not only improve reliability in high-risk applications, but also promote better generalization by distinguishing between known and unknown regions in the input space. Moreover, uncertainty estimates increase model interpretability by clarifying the sources of uncertainty and revealing potential limitations and biases, which emphasizes the role of UQ in characterizing and managing the uncertainty inherent in models and their predictions while providing valuable insights into reliability and confidence calibration. Given the broad impact of UQ, improving uncertainty estimation remains a pressing challenge.

In probabilistic ML models, uncertainty can be categorized into two main types:

- 1) Epistemic uncertainty (\mathcal{E}): Also known as model uncertainty, this type of uncertainty arises due to a lack of knowledge about the model, which, in turn, results from insufficient training data. It reflects the model's confidence in its parameters and structure.
- 2) Aleatoric uncertainty (\mathcal{A}): This uncertainty, often called data uncertainty, originates from intrinsic noise in the data-generating process and represents variability that remains irreducible, even with unlimited data.

The total predictive uncertainty in a model is given by

$$\mathcal{T} = \mathcal{E} + \mathcal{A}, \quad (1)$$

where \mathcal{T} represents the total uncertainty in the model's prediction.

In this paper, we focus on developing a principled framework for estimating both epistemic and aleatoric uncertainty in probabilistic ML models. Since probabilistic ML models comprise a broad range of approaches, we focus on making the problem more tractable by studying the UQ of predictions made by Gaussian process latent variable models (GPLVMs) [9]. Specifically, we explore how uncertainty can be efficiently quantified using scalable approximations, such as Random Fourier Features (RFF)-based Gaussian processes (GPs), which we use to construct GPLVMs.

Our contributions include:

- A systematic formulation of epistemic and aleatoric uncertainty estimation.
- A scalable approach to UQ using RFF-based GPs.
- Methodology for computing epistemic and aleatoric uncertainties, along with insights gained from experimental results.

The paper is organized as follows. First, we provide background on probabilistic ML models in Section II, followed by a brief overview of GPLVMs in Section III. In Section IV, we present the theoretical foundation for evaluating epistemic and aleatoric uncertainties, while in Section V we provide details of the methods used for these evaluations. Section VI discusses experimental results and insights gained from various experiments. Finally, Section VII concludes with final remarks.

This work was supported by NSF under Award 2212506.

II. BACKGROUND

A probabilistic ML model learns a predictive distribution over outputs \mathbf{y}_n given an input \mathbf{x}_n . The general formulation takes the following form:

$$p(\mathbf{y}_n|\mathbf{x}_n, \mathcal{D}) = \int p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}, \quad (2)$$

where $\mathcal{D} = \{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^N$ is the training dataset, $p(\boldsymbol{\theta}|\mathcal{D})$ is the posterior distribution over the model parameters, and $p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta})$ represents the likelihood function.

In many cases, ML models approximate this likelihood function using a deterministic mapping from inputs to outputs. In such models, a function $f(\mathbf{x}_n, \boldsymbol{\theta})$ produces a single output estimate,

$$\hat{\mathbf{y}}_n = f(\mathbf{x}_n, \hat{\boldsymbol{\theta}}), \quad (3)$$

where $\hat{\boldsymbol{\theta}}$ is some estimate of $\boldsymbol{\theta}$ obtained from the training data \mathcal{D} . However, probabilistic models extend this by modeling the entire predictive distribution rather than just a point estimate as given by (2).

We now describe how this predictive distribution quantifies the uncertainty of the prediction through its total covariance. Let the predictive distribution be given as in (2). Then, we can express the predictive mean as follows:

$$\mathbb{E}_{p(\mathbf{y}_n|\mathbf{x}_n, \mathcal{D})}[\mathbf{y}_n] = \mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})}[\mathbb{E}_{p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta})}[\mathbf{y}_n]]. \quad (4)$$

To find the total predictive covariance, we apply the law of total covariance and obtain

$$\begin{aligned} \text{Cov}_{p(\mathbf{y}_n|\mathbf{x}_n, \mathcal{D})}[\mathbf{y}_n] &= \text{Cov}_{p(\boldsymbol{\theta}|\mathcal{D})}[\mathbb{E}_{p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta})}[\mathbf{y}_n]] \\ &\quad + \mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})}[\text{Cov}_{p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta})}[\mathbf{y}_n]]. \end{aligned} \quad (5)$$

This decomposition separates epistemic uncertainty (first term) from aleatoric uncertainty (second term), i.e.,

$$\mathcal{E} = \text{Cov}_{p(\boldsymbol{\theta}|\mathcal{D})}[\mathbb{E}_{p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta})}[\mathbf{y}_n]], \quad (6)$$

$$\mathcal{A} = \mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})}[\text{Cov}_{p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta})}[\mathbf{y}_n]]. \quad (7)$$

In words, the epistemic uncertainty measures how much the expected prediction varies due to uncertainty in the parameters $\boldsymbol{\theta}$. It is obtained by first computing the expectation over the predictive distribution given the parameters and then computing the covariance of this expectation across different model parameters sampled from the posterior of the parameters. The aleatoric uncertainty is the expected predictive covariance due to data noise under the model's posterior.

III. GPLVMs AND INFERENCE WITH GPLVMs

We now briefly introduce GPLVMs with a focus on the uncertainties in their predictions. GPLVMs are powerful tools for modeling high-dimensional data by learning low-dimensional latent representations [9]. They establish a connection between a low-dimensional latent space and a high-dimensional observed space through GPs. Specifically, GPLVMs model data $\mathbf{Y} \in \mathbb{R}^{N \times d_y}$ in a high-dimensional space \mathcal{Y} , where $\mathcal{Y} = \mathbb{R}^{d_y}$, by assuming that these data are generated from latent variables \mathbf{X} in a lower-dimensional space, \mathcal{X} , with $\mathbf{X} \in \mathbb{R}^{N \times d_x}$ and

$\mathcal{X} = \mathbb{R}^{d_x}$, where $d_x \ll d_y$. Here, N represents the number of paired observations between \mathcal{X} and \mathcal{Y} .

More specifically, each observed vector $\mathbf{y}_n \in \mathbb{R}^{d_y}$, $n = 1, 2, \dots, N$ is generated from a corresponding latent vector $\mathbf{x}_n \in \mathbb{R}^{d_x}$, where the mapping $\mathbf{x}_n \mapsto \mathbf{y}_n$ is modeled using d_y GPs [7]. In other words, for a given latent variable \mathbf{x}_n , the observed variable \mathbf{y}_n is drawn according to

$$\mathbf{y}_n \sim \mathcal{N}(\mathbf{f}, \boldsymbol{\Sigma}_y), \quad (8)$$

where $\boldsymbol{\Sigma}_y = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_{d_y}^2\}$, and \mathbf{f} is a vector of d_y functions, i.e., $\mathbf{f}(\mathbf{x}_n) = [f_1(\mathbf{x}_n), f_2(\mathbf{x}_n), \dots, f_{d_y}(\mathbf{x}_n)]^\top$, each generated from a GP prior,

$$f_d(\cdot) \sim \mathcal{GP}(0, \kappa_d), \quad d = 1, 2, \dots, d_y. \quad (9)$$

Figure 1 presents a visual representation of a GPLVM.

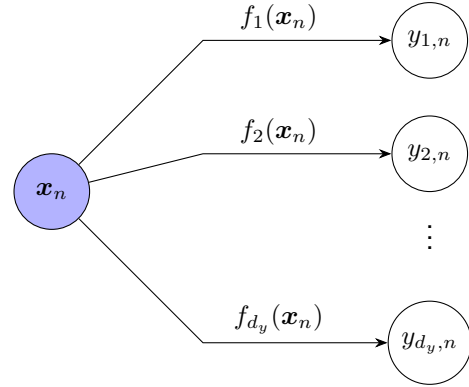


Fig. 1: A GPLVM with its basic elements, the latent vector \mathbf{x}_n and d_y GPs, one for each observed output.

Now we provide a brief overview of inference with GPLVMs. Given the observed (training) data $\mathbf{Y} \in \mathbb{R}^{N \times d_y}$, the main objective during training is to learn how to draw samples from the posterior distribution,

$$p(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y)p(\mathbf{X})p(\boldsymbol{\theta})p(\boldsymbol{\Sigma}_y), \quad (10)$$

where $\boldsymbol{\theta}$ represents the model parameters, $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y)$ is the likelihood function, where we treat \mathbf{Y} as observed variables and view $\mathbf{X}, \boldsymbol{\theta}$, and $\boldsymbol{\Sigma}_y$ as the unknowns to be inferred, and $p(\mathbf{X})$, $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\Sigma}_y)$ are the respective priors of \mathbf{X} , $\boldsymbol{\theta}$, and $\boldsymbol{\Sigma}_y$. Samples from the posterior can be drawn using methods such as Hamiltonian Monte Carlo (HMC) or other Markov Chain Monte Carlo (MCMC) techniques.

For testing, let $\mathbf{y}_* = [\mathbf{y}_{o,*}^\top, \mathbf{y}_{u,*}^\top]^\top$, where $\mathbf{y}_{o,*}$ and $\mathbf{y}_{u,*}$ represent the observed and unobserved parts of \mathbf{y}_* . The goal is now to determine the predictive distribution of $\mathbf{y}_{u,*}$ given $\mathbf{y}_{o,*}$, $p(\mathbf{y}_{u,*}|\mathbf{y}_{o,*}, \mathbf{Y})$, using the learned posterior of the GPLVM in (10). We achieve this in three steps:

- 1) First we infer the posterior distribution of the latent variable \mathbf{x}_* that corresponds to the partially observed test point $\mathbf{y}_{o,*}$. This posterior is given by

$$\begin{aligned} p(\mathbf{x}_*|\mathbf{y}_{o,*}, \mathbf{Y}) &\propto p(\mathbf{x}_*) \int p(\mathbf{y}_{o,*}|\mathbf{x}_*, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y) \\ &\quad \times p(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y|\mathbf{Y})d\mathbf{X}d\boldsymbol{\theta}d\boldsymbol{\Sigma}_y, \end{aligned} \quad (11)$$

where $p(\mathbf{y}_{o,*}|\mathbf{x}_*, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y)$ is a Gaussian distribution, and $p(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y|\mathbf{Y})$ is given by (10).

- 2) Next we find the predictive distribution of the unobserved components $y_{d,*}$, $d \in \mathcal{S}_u$, where \mathcal{S}_u represents the set of indices corresponding to the unobserved outputs. This is done using the GPs associated with the respective components of $\mathbf{y}_{u,*}$, which leads to the predictive distribution

$$p(y_{d,*}|\mathbf{y}_{o,*}, \mathbf{Y}) \propto \int p(y_{d,*}|\mathbf{x}_*, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y) p(\mathbf{x}_*|\mathbf{y}_{o,*}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y) \times p(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y|\mathbf{Y}) d\mathbf{X} d\boldsymbol{\theta} d\mathbf{x}_* d\boldsymbol{\Sigma}_y, \quad (12)$$

where $p(\mathbf{x}_*|\mathbf{y}_{o,*}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y)$ was determined in step 1, and $p(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y|\mathbf{Y})$ is defined by (10).

- 3) From (12), we can draw samples of each $y_{d,*}$ and use them to approximate its predictive distribution. Often, this predictive distribution is approximated by a Gaussian factorized across dimensions, i.e.,

$$p(\mathbf{y}_{u,*}|\mathbf{y}_{o,*}, \mathbf{Y}) = \prod_{d \in \mathcal{S}_u} \mathcal{N}(y_{d,*}|\mu_{d,*}, \sigma_{d,*}^2), \quad (13)$$

where $\mu_{d,*}$ is the posterior GP mean for dimension d , and $\sigma_{d,*}^2$ is the corresponding posterior variance. The variance $\sigma_{d,*}^2$ quantifies the uncertainty in each predicted component of $\mathbf{y}_{u,*}$ and it includes both aleatoric and epistemic uncertainty, where the epistemic uncertainty arises from the uncertainty of the latent variables and the model parameters.

In order to expand our work on UQ to as wide a set of problems as possible, we explore the use of Random Fourier Features (RFF)-based GPs [6]. These GPs approximate an adopted kernel function by a finite-dimensional RFF mapping by

$$\kappa(\mathbf{x}, \mathbf{x}') \approx \phi(\mathbf{x})^\top \phi(\mathbf{x}'), \quad (14)$$

where $\phi(\mathbf{x}) \in \mathbb{R}^{J \times 1}$ is given by

$$\phi(\mathbf{x}) = \sqrt{\frac{2}{J}} \left[\cos(\boldsymbol{\omega}_1^\top \mathbf{x}), \sin(\boldsymbol{\omega}_1^\top \mathbf{x}), \dots, \sin(\boldsymbol{\omega}_{J/2}^\top \mathbf{x}) \right]^\top, \quad (15)$$

where $\boldsymbol{\omega}_j$ are random frequencies, which according to Bochner's theorem [8], are drawn from the power spectral density of the chosen kernel. The functions of interest shown in Fig. 1 are then modeled by

$$f_d(\mathbf{x}_n) = \phi(\mathbf{x}_n)^\top \boldsymbol{\theta}_d. \quad (16)$$

Thus, after training the RFF-based GPs, the log-likelihood used to optimize \mathbf{x}_* becomes

$$\mathcal{L} = \sum_{d \in \mathcal{S}_o} \log \mathcal{N}(y_{o,d}|\phi(\mathbf{x}_*)^\top \hat{\boldsymbol{\theta}}_d, \hat{\sigma}_{o,d}^2) + \log p(\mathbf{x}_*), \quad (17)$$

where $\hat{\boldsymbol{\theta}}_d$ and $\hat{\sigma}_{o,d}^2$ represent estimated values from the training data.

IV. UNCERTAINTY QUANTIFICATION: THE THEORY

Given the above description of GPLVMs, we first formulate the problem. We are given a training dataset $\mathcal{D} = \{\mathbf{y}_n\}_{n=1}^N$, which is used to construct a GPLVM. The construction involves estimating the latent variables \mathbf{X} , the parameters of the GP models, $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_d]$, and $\boldsymbol{\Sigma}_y$. The estimates of \mathbf{X} , $\boldsymbol{\Theta}$, and $\boldsymbol{\Sigma}_y$ can be samples from $p(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_y|\mathbf{Y})$, which we denote by $\mathbf{X}^{(m)}$, $\boldsymbol{\Theta}^{(m)}$, and $\boldsymbol{\Sigma}_y^{(m)}$, $m = 1, \dots, M$. After training, a test observation \mathbf{y}_* is provided, where some elements are missing. The missing variables are predicted using Gaussian predictive distributions constructed by the GPs that correspond to these variables. *Our goal is to quantify the epistemic and aleatoric uncertainties associated with these predictions.*

As already pointed out, we use the law of total variance and decompose the total predictive variance of $y_{d,*}$ by

$$\text{Var}[y_{d,*}] = \underbrace{\mathbb{E}_{p(\mathbf{x}_*|\mathbf{y}_{o,*}, \mathbf{Y})} [\text{Var}_{p(y_{d,*}|\mathbf{x}_*, \boldsymbol{\theta}_d)} [y_{d,*}]]}_{\text{Aleatoric Uncertainty}} + \underbrace{\text{Var}_{p(\mathbf{x}_*|\mathbf{y}_{o,*}, \mathbf{Y})} [\mathbb{E}_{p(y_{d,*}|\mathbf{x}_*, \boldsymbol{\theta}_d)} [y_{d,*}]]}_{\text{Epistemic Uncertainty}}. \quad (18)$$

We focus first on the epistemic uncertainty when making predictions of $y_{d,*}$. This uncertainty comes from three main sources

- 1) Uncertainty in training latent variables \mathbf{X} : Recall that \mathbf{X} was inferred from data and that it follows a posterior distribution $p(\mathbf{X}|\mathbf{Y})$, which affects both $\boldsymbol{\theta}_d$ and \mathbf{x}_* .
- 2) Uncertainty in test latent variable \mathbf{x}_* : The posterior $p(\mathbf{x}_*|\mathbf{y}_{o,*}, \mathbf{Y})$ also introduces variability in predictions.
- 3) Uncertainty in function parameters $\boldsymbol{\theta}_d$: Since $\boldsymbol{\theta}_d$ is also inferred from data, its posterior $p(\boldsymbol{\theta}_d|\mathbf{X}, \mathbf{Y})$ also contributes to epistemic uncertainty.

The predicted value of $y_{d,*}$ has a mean that is linear in $\boldsymbol{\theta}_d$, i.e.,

$$\mathbb{E}_{p(y_{d,*}|\mathbf{x}_*, \boldsymbol{\theta}_d)} [y_{d,*}] = \phi(\mathbf{x}_*)^\top \mathbb{E}[\boldsymbol{\theta}_d|\mathbf{X}, \mathbf{Y}]. \quad (19)$$

Thus, the epistemic uncertainty simplifies to

$$\sigma_{\text{epistemic}, d,*}^2 = \text{Var}_{p(\mathbf{x}_*|\mathbf{y}_{o,*}, \mathbf{Y})} [\phi(\mathbf{x}_*)^\top \mathbb{E}[\boldsymbol{\theta}_d|\mathbf{X}, \mathbf{Y}]]. \quad (20)$$

Now we use the law of total variance, and we decompose the epistemic uncertainty into contributions from the training latent variables \mathbf{X} and test latent variable \mathbf{x}_* ,

We incorporate the uncertainty in \mathbf{X} by applying the law of total variance again. We write

$$\begin{aligned} \sigma_{\text{epistemic}, d,*}^2 &= \mathbb{E}_{p(\mathbf{X}|\mathbf{Y})} \left[\text{Var}_{p(\mathbf{x}_*|\mathbf{y}_{o,*}, \mathbf{X}, \mathbf{Y})} [\phi(\mathbf{x}_*)^\top \mathbb{E}[\boldsymbol{\theta}_d|\mathbf{X}, \mathbf{Y}]] \right] \\ &\quad + \text{Var}_{p(\mathbf{X}|\mathbf{Y})} \left[\mathbb{E}_{p(\mathbf{x}_*|\mathbf{y}_{o,*}, \mathbf{X}, \mathbf{Y})} [\phi(\mathbf{x}_*)^\top \mathbb{E}[\boldsymbol{\theta}_d|\mathbf{X}, \mathbf{Y}]] \right], \end{aligned} \quad (21)$$

where the first term accounts for uncertainty due to test input variability, reflecting how variations in \mathbf{x}_* variations influence the prediction, and the second term accounts for uncertainty

from inferred latent variables and represents the impact of different inferred values of \mathbf{X} .

We recall that

$$\mathbb{E}_{p(y_{d,*}|\mathbf{x}_*,\boldsymbol{\theta}_d)}[y_{d,*}] = \boldsymbol{\phi}(\mathbf{x}_*)^\top \boldsymbol{\theta}_d, \quad (22)$$

and write

$$\begin{aligned} \text{Var}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X},\mathbf{Y})}[\boldsymbol{\phi}(\mathbf{x}_*)^\top \mathbb{E}[\boldsymbol{\theta}_d|\mathbf{X},\mathbf{Y}]] \\ = \mathbb{E}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X},\mathbf{Y})}[\boldsymbol{\phi}(\mathbf{x}_*)^\top \text{Cov}[\boldsymbol{\theta}_d|\mathbf{X},\mathbf{Y}]\boldsymbol{\phi}(\mathbf{x}_*)] \\ + \text{Var}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X},\mathbf{Y})}[\boldsymbol{\phi}(\mathbf{x}_*)^\top \mathbb{E}[\boldsymbol{\theta}_d|\mathbf{X},\mathbf{Y}]]. \end{aligned} \quad (23)$$

Next, we take the expectation over $p(\mathbf{X}|\mathbf{Y})$ and obtain

$$\begin{aligned} \sigma_{\text{epistemic},d,*}^2 = \mathbb{E}_{p(\mathbf{X}|\mathbf{Y})} \left[\boldsymbol{\mu}_\phi^\top \text{Cov}[\boldsymbol{\theta}_d|\mathbf{X},\mathbf{Y}]\boldsymbol{\mu}_\phi \right. \\ \left. + \text{Tr}(\text{Cov}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X},\mathbf{Y})}[\boldsymbol{\phi}(\mathbf{x}_*)]\text{Cov}[\boldsymbol{\theta}_d|\mathbf{X},\mathbf{Y}]) \right] \\ + \text{Tr}(\text{Cov}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{Y})}[\mathbb{E}_{p(y_{d,*}|\mathbf{x}_*,\boldsymbol{\theta}_d)}[y_{d,*}]]), \end{aligned} \quad (24)$$

where

$$\boldsymbol{\mu}_\phi = \mathbb{E}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X},\mathbf{Y})}[\boldsymbol{\phi}(\mathbf{x}_*)]. \quad (25)$$

We observe that

- the uncertainty of the function parameters is quantified by $p(\boldsymbol{\theta}_d|\mathbf{X},\mathbf{Y})$, more specifically by $\text{Cov}[\boldsymbol{\theta}_d|\mathbf{X},\mathbf{Y}]$. We note it affects the predictions through both expectation and variance terms,
- the uncertainty due to \mathbf{x}_* is quantified by $p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X},\mathbf{Y})$ and is tracked via $\text{Var}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X},\mathbf{Y})}[\cdot]$, and
- the effect of the uncertainty on the predictions due to \mathbf{X} and quantified by $p(\mathbf{X}|\mathbf{Y})$ appears in $\text{Var}_{p(\mathbf{X}|\mathbf{Y})}[\cdot]$.

Now we turn our attention to the aleatoric uncertainty. In our GPLVM framework, we assume that

$$y_{d,*} \sim \mathcal{N}(\boldsymbol{\phi}(\mathbf{x}_*)^\top \boldsymbol{\theta}_d, \sigma_d^2), \quad (26)$$

and, thus,

$$\text{Var}_{p(y_{d,*}|\mathbf{x}_*,\boldsymbol{\theta}_d)}[y_{d,*}] = \sigma_d^2. \quad (27)$$

We note that since \mathbf{x}_* is random, σ_d^2 is also random. We take the expectation over $p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{Y})$ and write

$$\sigma_{\text{aleatoric},d,*}^2 = \mathbb{E}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{Y})}[\sigma_d^2]. \quad (28)$$

From training, the noise variance σ_d^2 is inferred from the posterior

$$p(\sigma_d^2|\mathbf{Y}) = \int p(\sigma_d^2|\mathbf{X},\mathbf{Y})p(\mathbf{X}|\mathbf{Y})d\mathbf{X}. \quad (29)$$

Thus, the full expression for aleatoric uncertainty is

$$\sigma_{\text{aleatoric},d,*}^2 = \mathbb{E}_{p(\mathbf{X}|\mathbf{Y})}\mathbb{E}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X},\mathbf{Y})}[\sigma_d^2]. \quad (30)$$

V. UNCERTAINTY QUANTIFICATION: THE METHODS

In this section we discuss the actual computation of the epistemic and aleatoric uncertainties. We propose an approximation based on a Monte Carlo sampling method. We compute the epistemic uncertainty defined by (24) using the following approach:

- Draw M samples $\mathbf{X}^{(m)} \sim p(\mathbf{X}|\mathbf{Y})$.
- For each $\mathbf{X}^{(m)}$, draw L samples

$$\mathbf{x}_*^{(m,\ell)} \sim p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X}^{(m)},\mathbf{Y}).$$

The Monte Carlo estimate is then

$$\begin{aligned} \sigma_{\text{epistemic},d,*}^2 \approx \frac{1}{M} \sum_{m=1}^M \left[\boldsymbol{\mu}_\phi^\top \text{Cov}[\boldsymbol{\theta}_d|\mathbf{X}^{(m)},\mathbf{Y}]\boldsymbol{\mu}_\phi \right. \\ \left. + \text{Tr}(\text{Cov}_{p(\mathbf{x}_*|\mathbf{y}_{o,*},\mathbf{X}^{(m)},\mathbf{Y})}[\boldsymbol{\phi}(\mathbf{x}_*)]\text{Cov}[\boldsymbol{\theta}_d|\mathbf{X}^{(m)},\mathbf{Y}]) \right] \\ + \frac{1}{M} \sum_{m=1}^M \text{Tr} \left(\frac{1}{L} \sum_{\ell=1}^L (\boldsymbol{\phi}(\mathbf{x}_*^{(m,\ell)})^\top \mathbb{E}[\boldsymbol{\theta}_d|\mathbf{X}^{(m)},\mathbf{Y}])^2 \right). \end{aligned} \quad (31)$$

The aleatoric uncertainty from (30) is approximated by

$$\sigma_{\text{aleatoric},d,*}^2 \approx \frac{1}{M} \sum_{m=1}^M \left[\frac{1}{L} \sum_{\ell=1}^L \sigma_d^{2(m)} \right]. \quad (32)$$

VI. UNCERTAINTY QUANTIFICATION: THE INSIGHTS FROM EXPERIMENTS

In this section, we describe our experiments and present insights gained from them. We generated $N = 1,000$ four-dimensional observations \mathbf{y}_n , with 800 used for training and the remaining 200 reserved for testing. Each observed vector \mathbf{y}_n was obtained by first drawing a latent variable $\mathbf{x}_n \in \mathbb{R}^2$ according to $\mathbf{x}_n \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, and then applying four different functions to obtain $y_{d,n}$ for $d = 1, 2, 3, 4$. More specifically, the observations were generated as follows:

- 1) linear function (baseline):

$$y_{1,n} = \mathbf{w}_1^\top \mathbf{x}_n + \epsilon_{1,n}, \quad \epsilon_{1,n} \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad (33)$$

- 2) nonlinear squared function:

$$y_{2,n} = (\mathbf{w}_2^\top \mathbf{x}_n)^2 + \epsilon_{2,n}, \quad \epsilon_{2,n} \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad (34)$$

- 3) periodic function:

$$y_{3,n} = \sin(\mathbf{w}_3^\top \mathbf{x}_n) + \epsilon_{3,n}, \quad \epsilon_{3,n} \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad (35)$$

- 4) discontinuous step function:

$$y_{4,n} = \begin{cases} 1 + \epsilon_{4,n}, & \text{if } \mathbf{w}_4^\top \mathbf{x}_n > 0, \\ -1 + \epsilon_{4,n}, & \text{otherwise,} \end{cases}, \quad \epsilon_{4,n} \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad (36)$$

where the variance of the noise in all the outputs was $\sigma_\epsilon^2 = 1$, and the weights $\mathbf{w}_d \in \mathbb{R}^2$ were drawn independently from $\mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{I}_2)$, with $\sigma_w^2 = 1$.

During training, we drew $M = 100$ samples from $p(\mathbf{X}|\mathbf{Y})$ using the variational distribution of \mathbf{X} . For each test sample,

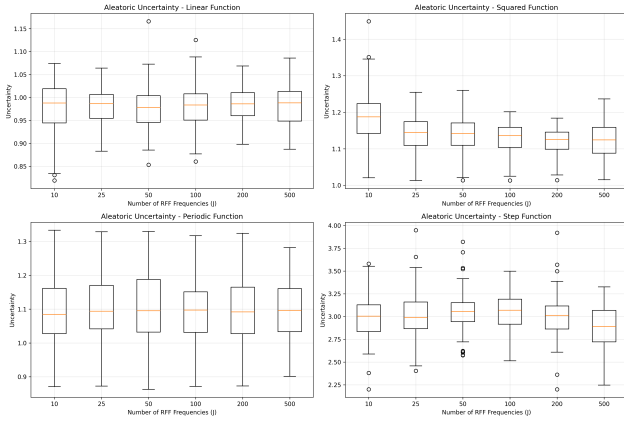


Fig. 2: The estimated aleatoric uncertainties in each of the four predicted outputs of \mathbf{y}_* for various values of J . It was assumed that one of the outputs was missing while the remaining ones were observed.

$\mathbf{y}_{o,*}$ and $\mathbf{X}^{(m)}$, we generated $L = 100$ samples of \mathbf{x}_* according to $\mathbf{x}_*^{(m,\ell)} \sim p(\mathbf{x}_* | \mathbf{y}_{o,*}, \mathbf{X}^{(m)}, \mathbf{Y})$. We used the generated training samples $\mathbf{X}^{(m)}$ to compute the epistemic and aleatoric uncertainties defined by (31) and (32), respectively. We conducted experiments where the number of drawn frequencies took values from the set $J \in \{10, 25, 50, 100, 200, 500\}$. All the frequencies were obtained by sampling from the radial-basis-function kernel.

During evaluation, we computed the aleatoric and epistemic uncertainties as described in the previous section. In each trial, we assumed that one of the output variables, $y_{d,*}$, was missing and had to be predicted using the remaining variables. For example, to predict $y_{1,*}$, we used the observed values $y_{2,*}$, $y_{3,*}$, and $y_{4,*}$ to estimate the latent variable \mathbf{x}_* , which was then used to infer $y_{1,*}$. The boxplots in Figs. 2 and 3 summarize average aleatoric and epistemic variances computed on 200 test samples per trial, following training. Each boxplot is constructed from 50 such averages obtained across independent simulations. The circles denote outlier trials. The plots illustrate the consistency of uncertainty estimates and the variability across test instances.

We observe that the aleatoric uncertainty of $y_{1,*}$, which corresponds to the linear function, was estimated accurately and did not vary across values of J . In contrast, the aleatoric uncertainty for $y_{4,*}$, which corresponds to the step function, was significantly overestimated due to the limited ability of GPs to model discontinuities. The epistemic uncertainty reached its highest value for $y_{4,*}$, which reveals the model's lack of confidence in regions with abrupt changes, and its lowest value for $y_{3,*}$, the periodic function output, which is smooth and more suitable for GP models.

VII. CONCLUSIONS

In this paper, we presented a systematic framework for UQ in probabilistic ML models with focus on GPLVMs. We provided a theoretical formulation for decomposing predic-

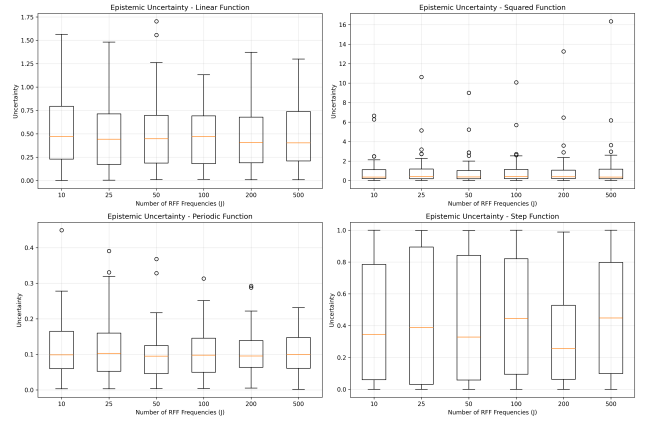


Fig. 3: The estimated epistemic uncertainties in each of the four predicted outputs of \mathbf{y}_* for various values of J . It was assumed that one of the outputs was missing while the remaining ones were observed.

tive uncertainty into epistemic and aleatoric components and introduced a Monte Carlo sampling-based approach for their estimation. Using RFF-based GPs, we demonstrated a scalable and efficient method to approximate predictive distributions. Our experimental results provided insight into the impact of uncertainty estimation on model reliability. The computation of uncertainties relied on Monte Carlo simulations, which in turn introduced additional sources of uncertainty. This was not accounted for in this paper and is left for future work.

REFERENCES

- [1] M. Abdar, A. Khosravi, S. M. S. Islam, U. R. Acharya, and A. V. Vasilakos. The need for quantification of uncertainty in artificial intelligence for clinical data analysis: increasing the level of trust in the decision-making process. *IEEE Systems, Man, and Cybernetics Magazine*, 8(3):28–40, 2022.
- [2] S. Cheng, C. Quilodr  n-Casas, S. Ouala, A. Farchi, C. Liu, P. Tandoe, R. Fablet, D. Lucor, B. Iooss, J. Brajard, et al. Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. *IEEE/CAA Journal of Automatica Sinica*, 10(6):1361–1387, 2023.
- [3] R. Ghanem, D. Higdon, H. Owahdi, et al. *Handbook of Uncertainty Quantification*, volume 6. Springer New York, 2017.
- [4] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT press, 2012.
- [5] T. Papamarkou, M. Skoularidou, K. Palla, L. Aitchison, J. Arbel, D. Dunson, M. Filippone, V. Fortuin, P. Hennig, J. M. Hern  ndez-Lobato, et al. Position: Bayesian deep learning is needed in the age of large-scale AI. In *Forty-first International Conference on Machine Learning*, 2024.
- [6] A. Rahimi and B. Recht. Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20, 2007.
- [7] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*, volume 2. MIT Press, 2006.
- [8] W. Rudin. *Fourier Analysis on Groups*. Courier Dover Publications, 2017.
- [9] M. Titsias and N. D. Lawrence. Bayesian Gaussian process latent variable model. In *Proceedings of the Thirtieth International Conference on Artificial Intelligence and Statistics*, pages 844–851. JMLR Workshop and Conference Proceedings, 2010.
- [10] T. Wang, Y. Wang, J. Zhou, B. Peng, X. Song, C. Zhang, X. Sun, Q. Niu, J. Liu, S. Chen, et al. From aleatoric to epistemic: Exploring uncertainty quantification techniques in artificial intelligence. *arXiv preprint arXiv:2501.03282*, 2025.