MODELLISTICA E CALCOLO SCIENTIFICO

MOX

MODELING AND SCIENTIFIC COMPUTING

MOX-Report No. 56/2025

Improvements on uncertainty quantification with variational autoencoders

Tonini, A.; Bui-Thanh, T.; Regazzoni, F.; Dede', L; Quarteroni, A.

IMPROVEMENTS ON UNCERTAINTY QUANTIFICATION WITH VARIATIONAL AUTOENCODERS

ANDREA TONINI[1,*], TAN BUI-THANH[2,3], FRANCESCO REGAZZONI[1], LUCA DEDE'[1] AND ALFIO QUARTERONI[4,5]

ABSTRACT. Inverse problems aim to determine model parameters of a mathematical problem from given observational data. Neural networks can provide an efficient tool to solve these problems. In the context of Bayesian inverse problems, Uncertainty Quantification Variational AutoEncoders (UQ-VAE), a class of neural networks, approximate the posterior distribution mean and covariance of model parameters. This allows for both the estimation of the parameters and their uncertainty in relation to the observational data. In this work, we propose a novel loss function for training UQ-VAEs, which includes, among other modifications, the removal of a sample mean term from an already existing one. This modification improves the accuracy of UQ-VAEs, as the original theoretical result relies on the convergence of the sample mean to the expected value (a condition that, in high dimensional parameter spaces, requires a prohibitively large number of samples due to the curse of dimensionality). Avoiding the computation of the sample mean significantly reduces the training time in high dimensional parameter spaces compared to previous literature results. Under this new formulation, we establish a new theoretical result for the approximation of the posterior mean and covariance for general mathematical problems. We validate the effectiveness of UQ-VAEs through three benchmark numerical tests: a Poisson inverse problem, a non affine inverse problem and a 0D cardiocirculatory model, under the two clinical scenarios of systemic hypertension and ventricular septal defect. For the latter case, we perform forward uncertainty quantification.

Keywords: Bayesian inverse problem, deep learning, partial differential equations, uncertainty quantification, variational autoencoders.
AMS Subject Classification: 68T07, 62C10

1. INTRODUCTION

Forward problems describe mathematical models that map a set of parameters to a corresponding set of observational data. On the other hand, inverse problems aim to determine the model parameters from (noisy) observational data by minimizing a loss function that quantifies the mismatch between the observational data and the model outputs. Deterministic inverse problems seek to identify the single values of the parameters.[24] However, even when the forward map from parameters to outputs exists, the inverse map may not, leading to the ill-posedness of deterministic inverse problems. To address this issue,

[1]MOX, Dipartimento di Matematica, Politecnico di Milano, Milan, Italy.

[2]Department of Aerospace Engineering and Engineering Mechanics, the University of Texas at Austin, Austin, Texas, USA..

[3]The Oden Institute for Computational Engineering and Sciences, the University of Texas at Austin, Austin, Texas, USA..

[4](Professor Emeritus) MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci, Milan, 20133, Italy.

[5](Professor Emeritus) Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

*Corresponding author. E-mail: andrea.tonini@polimi.it.

Tikhonov regularization is often applied to the loss function by adding a term that penalizes the norm of the parameters.[39]

Bayesian inverse problems aim to determine the posterior distribution of the parameters, that is their probability distribution given an observational data.[37] The solution to a Bayesian inverse problem accounts for uncertainty in the parameters estimate that depends on the noise in the observational data.

Neural Networks (NNs) have gained increasing attention for solving inverse problems in recent years.[5,31,32,50] They offer a significant reduction in the computational cost of solving inverse problems that would otherwise require numerous iterations of an optimization method.[4,6,23] A single iteration of the chosen optimization method requires solving the forward problem, which can make inverse problems prohibitively expensive when the forward one is computationally demanding.

Tikhonov Networks (TNets) have been shown to effectively solve deterministic inverse problems.[25] Their loss function consists of two terms: one incorporating the prior knowledge of the parameters and another enforcing a physics-aware constraint. Data randomization has been theoretically and numerically demonstrated to implicitly regularize the weights and biases of the NN. Tikhonov autoencoders (TAEN) further enhance the use of data randomization, enabling the training of the NN generating new data by the randomization of a single sample.[45] InVAErt networks address the ill-posedness nature of deterministic inverse problems by extending the observational data space with latent variables, enabling the identification of different parameters generating the same observational data.[40,41] NETworks Tikhonov (NETT) have been applied to inverse problems in imaging, leveraging a data-driven regularization approach.[21] Additionally, convolutional Neural Networks (CNN) have been widely explored for image-based inverse problems.[8,17,22]

In the context of Bayesian inverse problems, Uncertainty Quantification Variational AutoEncoders (UQ-VAE) have been developed to estimate the mean and covariance of the posterior distribution of parameters. In Ref. 14, the authors established the theoretical foundations of UQ-VAEs, deriving a loss function based on a family of Jensen-Shannon divergences.[26] This approach comes from generative modeling that employs Jensen-Shannon divergences to compute the posterior parameter distribution and generate new data samples.[9,38] In Ref. 42, a novel loss function was proposed to improve the estimation accuracy of UQ-VAEs, while enhancing their generalization capabilities. However, this approach required an increased computational cost for evaluating the training loss function, leading to prolonged training times.

In this work, we build upon the results of Ref. 42 proposing a novel loss function by removing a sample mean term from the original loss function, among other modifications. This approach improves accuracy and reduces the training time compared to previous literature results. The time reduction advantage becomes more pronounced as the dimensionality of the parameter space increases. We obtain a theoretical result analogous to the original one for affine forward problems and extend it to general forward problems under additional assumptions, thereby strengthening the theoretical background of UQ-VAEs. We validate our approach on a Poisson problem and compare it in terms of training time and accuracy with the previous one on a non linear problem. Furthermore, we test our approach on two test cases involving 0D cardiocirculatory models,[1,10,12,28,43] specifically modeling systemic hypertension[27] and ventricular septal defect.[7,36] For the latter case, we perform forward uncertainty quantification to further assess the performances of the new approach.

The paper is structured as follows: in section 2, we introduce the new loss function and UQ-VAEs; in section 3, we present the numerical tests on the Poisson problem and the 0D cardiocirculatory model; in section 4, we draw our conclusions.

2. BAYESIAN INVERSE PROBLEMS

We present the general framework of Bayesian inverse problems and UQ-VAEs. Additionally, we propose a novel loss function for training the UQ-VAEs addressing previous limitations.[42] The new loss function reduces the computational cost of training the UQ-VAE while extending the theoretical result of the previous approach.

Bayesian inverse problems are used to estimate the parameters of a mathematical model along with their uncertainties, given noisy observational data. Let U and Y be random variables representing the parameters and the noisy observational data, respectively. We assume the existence of a map $\mathcal{F}$ from the parameters to the observational data in absence of noise. Let E denote the noise random variable. The mathematical model under consideration is:

$$Y = \mathcal{F}(U) + E. \tag{1}$$

We focus on the case where U and Y are finite dimensional real random variables. Specifically, each parameter is a vector $\mathbf{u} \in \mathbb{R}^D$ and each observational data is a vector $\mathbf{y} \in \mathbb{R}^O$, with $D, O \in \mathbb{N}$. Thus, the map is defined as $\mathcal{F} : \mathbb{R}^D \to \mathbb{R}^O$. The solution of a Bayesian inverse problem is the posterior distribution of the parameters $p_{U|Y}(\mathbf{u}|\mathbf{y})$, which is the distribution of the parameters conditioned to the given noisy observational data.

Using the Bayes' theorem,[2] it is possible to retrieve an implicit formula for the posterior distribution that, in most of the cases, is not possible to compute explicitly as it requires the evaluation of an integral. Generally, one looks for the mean and covariance, namely the first two moments, of the posterior distribution to characterize the expected parameter and its uncertainty subjected to the noisy observational data. A standard approach to solve Bayesian inverse problems is by computing the Maximum A Posteriori estimate (MAP) parameter $\mathbf{u}_{MAP} \in \mathbb{R}^D$, that is the most likely parameter to observe given a noisy observational data $\mathbf{y} \in \mathbb{R}^O$, and the Laplace approximation of the covariance matrix $\Gamma_{Lap} \in \mathbb{R}^{D \times D}$.[11] By the Bayes' theorem, the posterior distribution can be expressed as:

$$p_{U|Y}(\mathbf{u}|\mathbf{y}) = \frac{p_{U,Y}(\mathbf{u}, \mathbf{y})}{p_Y(\mathbf{y})} = \frac{p_{Y|U}(\mathbf{y}|\mathbf{u}) \, p_U(\mathbf{u})}{p_Y(\mathbf{y})} = \frac{p_{Y|U}(\mathbf{y}|\mathbf{u}) \, p_U(\mathbf{u})}{\int_{\mathbb{R}^D} p_{Y|U}(\mathbf{y}|\mathbf{u}) p_U(\mathbf{u}) \mathrm{d}\mathbf{u}}, \tag{2}$$

where $p_{U,Y}(\mathbf{u}, \mathbf{y})$, $p_Y(\mathbf{y})$, $p_U(\mathbf{u})$ and $p_{Y|U}(\mathbf{y}|\mathbf{u})$ are the probability density functions (pdf) of the joint probability distribution of U and Y, the marginal pdfs of Y and U, and the conditioned probability of Y with respect to U (likelihood), respectively. We assume that E and U are independent, so the likelihood function is:

$$p_{Y|U}(\mathbf{y}|\mathbf{u}) = p_E(\mathbf{y} - \mathcal{F}(\mathbf{u})), \tag{3}$$

where p_E is the pdf of the noise random variable E. Furthermore, we assume that $U \sim \mathcal{N}(\boldsymbol{\mu}_{pr}, \Gamma_{pr})$ and $E \sim \mathcal{N}(\boldsymbol{\mu}_E, \Gamma_E)$, where $\boldsymbol{\mu}_{pr} \in \mathbb{R}^D$ and $\boldsymbol{\mu}_E \in \mathbb{R}^O$ are the means of the parameters and noise random variables, respectively, and $\Gamma_{pr} \in \mathbb{R}^{D \times D}$ and $\Gamma_E \in \mathbb{R}^{O \times O}$ are the corresponding covariance matrices. In this work, we restrict our analysis to symmetric positive definite (SPD) covariance matrices, as well as their inverses, to ensure the well definition of the loss function used in training the variational autoencoders. Using

(3) and the normality assumptions on U and E, the joint pdf $p_{U,Y}(\mathbf{u}, \mathbf{y})$ can be rewritten as:

$$p_{U,Y}(\mathbf{u}, \mathbf{y}) = p_E(\mathbf{y} - \mathcal{F}(\mathbf{u})) \, p_U(\mathbf{u}) = \frac{1}{(2\pi)^{\frac{O+D}{2}} |\Gamma_E|^{\frac{1}{2}} |\Gamma_{pr}|^{\frac{1}{2}}}$$
$$\exp\left(-\frac{1}{2} \left(\|\mathbf{y} - \mathcal{F}(\mathbf{u}) - \boldsymbol{\mu}_E\|^2_{\Gamma_E^{-1}} + \|\mathbf{u} - \boldsymbol{\mu}_{pr}\|^2_{\Gamma_{pr}^{-1}} \right) \right),$$

where $|\cdot|$ denotes the matrix determinant and $\|\cdot\|_{\Gamma^{-1}}$ represents the norm induced by an SPD matrix Γ^{-1}. Maximizing the posterior pdf $p_{U|Y}(\mathbf{u}|\mathbf{y})$ with respect to $\mathbf{u}$ yields the MAP estimate:

$$\mathbf{u}_{MAP} := \underset{\mathbf{u} \in \mathbb{R}^D}{\arg\max} \, p_{U|Y}(\mathbf{u}|\mathbf{y})$$

(4)
$$= \underset{\mathbf{u} \in \mathbb{R}^D}{\arg\min} \left(\|\mathbf{y} - \mathcal{F}(\mathbf{u}) - \boldsymbol{\mu}_E\|^2_{\Gamma_E^{-1}} + \|\mathbf{u} - \boldsymbol{\mu}_{pr}\|^2_{\Gamma_{pr}^{-1}} \right).$$

The first term of the optimization problem derives from the likelihood function and quantifies the mismatch between the observational data and model predictions. The second term encodes prior knowledge about the parameters and serves as a regularization term. Solving this optimization problem typically requires gradient-based methods, which can be computationally expensive, particularly when evaluating the map $\mathcal{F}$ is costly.

To quantify uncertainty in the parameter estimate due to noise, the Laplace approximation Γ_{Lap} of the covariance matrix can be used:[11]

(5)
$$\Gamma_{Lap} = \left(J_{\mathcal{F}}(\mathbf{u}_{MAP})^{\mathrm{T}} \Gamma_E^{-1} J_{\mathcal{F}}(\mathbf{u}_{MAP}) + \Gamma_{pr}^{-1} \right)^{-1},$$

where $J_{\mathcal{F}}(\mathbf{u}_{MAP}) \in \mathbb{R}^{O \times D}$ is the Jacobian of $\mathcal{F}$ evaluated at $\mathbf{u}_{MAP}$. Computing Γ_{Lap} requires evaluating $J_{\mathcal{F}}(\mathbf{u}_{MAP})$, which is significantly more expensive than a single evaluation of $\mathcal{F}$.

When evaluating $\mathcal{F}$ is expensive (such as when $\mathcal{F}$ represents the solution operator of a complex partial differential equation), computing $\mathbf{u}_{MAP}$ and Γ_{Lap} becomes infeasible. UQ-VAEs,[14,42] a class of NNs, offer a way to mitigate this computational burden. Although the most advanced UQ-VAE formulation[42] presents strong generalization capabilities, even with small datasets, it has two main drawbacks when $\mathcal{F}$ is not affine: an inaccurate estimate of the posterior covariance matrix and a high computational cost for training. Even if the latter is compensated by the efficient solution of Bayesian inverse problems, reducing the training cost is a priority. In this work, we propose a modification to the loss function used for training UQ-VAEs.[42] The new loss function extends the previous theoretical result while significantly reducing computational costs. This reduction enables also faster testing of the NN architectures, indirectly improving the accuracy of the posterior mean and covariance matrix estimates.

Employing variational inference, we approximate the posterior distribution $p_{U|Y}(\mathbf{u}|\mathbf{y})$ using a Gaussian distribution:

(6)
$$q_\phi(\mathbf{u}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_{post}(\mathbf{y}, \phi), \Gamma_{post}(\mathbf{y}, \phi)),$$

where $\boldsymbol{\mu}_{post}(\mathbf{y}, \phi) \in \mathbb{R}^D$, $\Gamma_{post}(\mathbf{y}, \phi) \in \mathbb{R}^{D \times D}$ and ϕ denotes a set of hyper-parameters (in section 2.2, ϕ is the set of weights and biases of the UQ-VAE). For conciseness, we omit the explicit dependence of $\boldsymbol{\mu}_{post}$ and Γ_{post} on $\mathbf{y}$ and ϕ.

2.1. **Loss function.** We describe the loss function and the theoretical result presented in Ref. 42, modifying the former to reduce its computational cost.

Given $\alpha \in (0, 1)$ and an observational data $\mathbf{y} \in \mathbb{R}^O$, the loss function proposed in the previous work is:

$$
\begin{aligned}
L_\alpha(\boldsymbol{\mu}, C) = (1 - \alpha) \left(\|\boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}}\|^2_{\Gamma^{-1}} + \mathrm{tr}\left(\Gamma^{-1}\Gamma_{\mathrm{pr}}\right) \right) + \\
\alpha \left(\|\boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}}\|^2_{\Gamma_{\mathrm{pr}}^{-1}} + \mathrm{tr}\left(\Gamma_{\mathrm{pr}}^{-1}\Gamma\right) \right) + \\
\alpha \mathbb{E}_{\mathcal{N}(\boldsymbol{\mu},\Gamma)} \left[\|\mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathcal{F}(\mathbf{u})\|^2_{\Gamma_{\mathrm{E}}^{-1}} \right],
\end{aligned}
\tag{7}
$$

where $\mathbb{E}_{\mathcal{N}(\boldsymbol{\mu},\Gamma)}[\cdot]$ is the expected value with respect to the probability distribution $\mathcal{N}(\boldsymbol{\mu}, \Gamma)$, $\boldsymbol{\mu} \in \mathbb{R}^D$, $\Gamma \in \mathbb{R}^{D \times D}$ is SPD and C is its Cholesky factor (a lower triangular matrix with positive diagonal entries). This loss function is an upper bound for a family of Jensen-Shannon divergences (JSD)[26] summed to a Kullback-Leibler divergence, which measure the difference between the posterior distribution $p_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y})$ and the distribution $\mathcal{N}(\boldsymbol{\mu}, \Gamma)$. The first two lines come from the Kullback-Leibler divergences between the prior and the distribution $\mathcal{N}(\boldsymbol{\mu}, \Gamma)$ and the third line represents the expected squared error on the observational data.

When $\mathcal{F}$ is affine (i.e., $\mathcal{F}(\mathbf{u}) = F\mathbf{u} + \mathbf{f}$, where $F \in \mathbb{R}^{O \times D}, \mathbf{f} \in \mathbb{R}^O$), (4) and (5) can be expressed as:

$$
\mathbf{u}_{\mathrm{MAP}} = \Gamma_{\mathrm{Lap}} \left(F^{\mathrm{T}}\Gamma_{\mathrm{E}}^{-1} (\mathbf{y} - \mathbf{f} - \boldsymbol{\mu}_{\mathrm{E}}) + \Gamma_{\mathrm{pr}}^{-1}\boldsymbol{\mu}_{\mathrm{pr}} \right),
\tag{8}
$$

$$
\Gamma_{\mathrm{Lap}} = \left(F^{\mathrm{T}}\Gamma_{\mathrm{E}}^{-1}F + \Gamma_{\mathrm{pr}}^{-1} \right)^{-1}.
\tag{9}
$$

Moreover, the expected value in (7) is computed exactly:

$$
\begin{aligned}
\mathbb{E}_{\mathcal{N}(\boldsymbol{\mu},\Gamma)} \left[\|\mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathcal{F}(\mathbf{u})\|^2_{\Gamma_{\mathrm{E}}^{-1}} \right] = \\
\|\mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathbb{E}_{\mathcal{N}(\boldsymbol{\mu},\Gamma)}[\mathcal{F}(\mathbf{u})]\|^2_{\Gamma_{\mathrm{E}}^{-1}} + \mathrm{tr}\left(\Gamma_{\mathrm{E}}^{-1}\mathrm{Cov}_{\mathcal{N}(\boldsymbol{\mu},\Gamma)}[\mathcal{F}(\mathbf{u})]\right) = \\
\|\mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - F\boldsymbol{\mu} - \mathbf{f}\|^2_{\Gamma_{\mathrm{E}}^{-1}} + \mathrm{tr}\left(\Gamma_{\mathrm{E}}^{-1}F\Gamma F^{\mathrm{T}}\right),
\end{aligned}
\tag{10}
$$

where $\mathrm{Cov}_{\mathcal{N}(\boldsymbol{\mu},\Gamma)}[\cdot]$ is the covariance with respect to the probability distribution $\mathcal{N}(\boldsymbol{\mu}, \Gamma)$. Minimizing this loss function yields the following result.[42]

Theorem 1. *Assume that* $\mathcal{F}(\mathbf{u}) = F\mathbf{u} + \mathbf{f}$, *where* $F \in \mathbb{R}^{O \times D}, \mathbf{f} \in \mathbb{R}^O$, *and* $\mathbf{y} \in \mathbb{R}^O$. *Suppose that the prior and noise models are Gaussian and independent,* $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$ *and* $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{E}}, \Gamma_{\mathrm{E}})$, *respectively. Then, the posterior distribution* $p_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y})$ *is* $\mathcal{N}(\mathbf{u}_{\mathrm{MAP}}, \Gamma_{\mathrm{Lap}})$, *where* $(\mathbf{u}_{\mathrm{MAP}}, \Gamma_{\mathrm{Lap}})$ *are given by* (8) *and* (9).

Let $\alpha \in (0, 1)$ *and* $(\hat{\boldsymbol{\mu}}, \hat{C})$ *(with* $\hat{\Gamma} = \hat{C}\hat{C}^{\mathrm{T}}$) *be the stationary points of the loss function*

$$
\begin{aligned}
L_\alpha(\boldsymbol{\mu}, C) = (1 - \alpha) \left(\|\boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}}\|^2_{\Gamma^{-1}} + \mathrm{tr}\left(\Gamma^{-1}\Gamma_{\mathrm{pr}}\right) \right) + \\
\alpha \left(\|\boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}}\|^2_{\Gamma_{\mathrm{pr}}^{-1}} + \mathrm{tr}\left(\Gamma_{\mathrm{pr}}^{-1}\Gamma\right) \right) + \\
\alpha \left(\|\mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - F\boldsymbol{\mu} - \mathbf{f}\|^2_{\Gamma_{\mathrm{E}}^{-1}} + \mathrm{tr}\left(\Gamma_{\mathrm{E}}^{-1}F\Gamma F^{\mathrm{T}}\right) \right).
\end{aligned}
\tag{11}
$$

Let

$$
\boldsymbol{\mu}_{\mathrm{post}} = \frac{1 - \alpha}{\alpha} \Gamma_{\mathrm{Lap}}\hat{\Gamma}^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{\mathrm{pr}}) + \hat{\boldsymbol{\mu}},
\tag{12}
$$

$$
\Gamma_{\mathrm{post}} = \hat{\Gamma}A^{-1}\hat{\Gamma},
\tag{13}
$$

with

$$A = \frac{1 - \alpha}{\alpha} \left((\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{\mathrm{pr}})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{\mathrm{pr}})^{\mathrm{T}} + \Gamma_{\mathrm{pr}} \right).$$

Then $(\boldsymbol{\mu}_{\mathrm{post}}, \Gamma_{\mathrm{post}}) = (\mathbf{u}_{\mathrm{MAP}}, \Gamma_{\mathrm{Lap}})$.

Remark. In practice, we minimize the loss function (11) to find $(\hat{\boldsymbol{\mu}}, \hat{\Gamma})$ and we compute $(\boldsymbol{\mu}_{\mathrm{post}}, \Gamma_{\mathrm{post}})$ using (12) and (13).

Remark. The matrix A is SPD since it is a positive scalar multiple of the sum of two symmetric matrices that are positive semidefinite and positive definite. Consequently, $\hat{\Gamma} A^{-1} \hat{\Gamma}$ is also SPD, as Γ_{post} and Γ_{Lap} are.

When the map $\mathcal{F}$ is not affine, $\mathbb{E}_{\mathcal{N}(\boldsymbol{\mu}, \Gamma)} \left[\| \mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathcal{F}(\mathbf{u}) \|_{\Gamma_{\mathrm{E}}^{-1}}^2 \right]$ cannot be computed in a closed form. To approximate this expectation, the central limit theorem is used, leading to the following loss function:

$$(1 - \alpha) \left(\| \boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}} \|_{\Gamma^{-1}}^2 + \mathrm{tr} \left(\Gamma^{-1} \Gamma_{\mathrm{pr}} \right) \right) +$$

$$\alpha \left(\| \boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}} \|_{\Gamma_{\mathrm{pr}}^{-1}}^2 + \mathrm{tr} \left(\Gamma_{\mathrm{pr}}^{-1} \Gamma \right) \right) +$$

(14) $$\alpha \frac{1}{K} \sum_{k=1}^{K} \| \mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathcal{F}(\mathbf{u}_{\mathrm{draw}}^{\mathrm{k}}) \|_{\Gamma_{\mathrm{E}}^{-1}}^2,$$

where $K \in \mathbb{N}$ and, for $k = 1, \ldots, K$, $\mathbf{u}_{\mathrm{draw}}^{\mathrm{k}} = \boldsymbol{\mu} + C \boldsymbol{\varepsilon}^{\mathrm{k}}$, with $\boldsymbol{\varepsilon}^{\mathrm{k}} \in \mathbb{R}^{\mathrm{D}}$ and $\boldsymbol{\varepsilon}^{\mathrm{k}} \sim \mathcal{N}(\mathbf{0}, I)$. Therefore, at each iteration of the optimization algorithm, $\mathbf{u}_{\mathrm{draw}}^{\mathrm{k}}$ are drawn from $\mathcal{N}(\boldsymbol{\mu}, \Gamma)$.

Approximating the expectation in (7) is typically expensive, even when Quasi-Monte Carlo methods are used,[15] as a large number of samples K is required for convergence. Given the Cholesky factorizations of Γ_{pr} and Γ_{E} the computational cost of evaluating (14) is $O(D^3 + KD^2 + K\mathrm{c}(\mathcal{F}) + KO^2)$, where $\mathrm{c}(\mathcal{F})$ is the cost of an evaluation of $\mathcal{F}$. $O(D^3)$ is the cost of solving D linear systems with matrix Γ_{post} or Γ_{pr}. $O(KD^2)$, $O(K\mathrm{c}(\mathcal{F}))$ and $O(KO^2)$ are the costs of computing K samples $\mathbf{u}_{\mathrm{draw}}^{\mathrm{k}}$, of K evaluations of $\mathcal{F}$ and of solving K linear systems with matrix Γ_{E}, respectively. The computational costs of the other terms are lower than the previous ones.

We propose a novel loss function that significantly reduces the previous computational cost:

$$L_\theta(\boldsymbol{\mu}, C) = \theta^2 \mathrm{tr} \left(\Gamma^{-1} \Gamma_{\mathrm{pr}} \right) + \| \boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}} \|_{\Gamma_{\mathrm{pr}}^{-1}}^2 + \theta^2 \mathrm{tr} \left(\Gamma_{\mathrm{pr}}^{-1} \Gamma \right) +$$

$$\| \mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathcal{F}(\boldsymbol{\mu}) \|_{\Gamma_{\mathrm{E}}^{-1}}^2 +$$

(15) $$\mathrm{tr} \left(\Gamma_{\mathrm{E}}^{-1} \left(\tilde{\mathcal{F}}(\theta C + \boldsymbol{\mu} \mathbf{1}_{\mathrm{D}}^{\mathrm{T}}) - \mathcal{F}(\boldsymbol{\mu}) \mathbf{1}_{\mathrm{D}}^{\mathrm{T}} \right) \left(\tilde{\mathcal{F}}(\theta C + \boldsymbol{\mu} \mathbf{1}_{\mathrm{D}}^{\mathrm{T}}) - \mathcal{F}(\boldsymbol{\mu}) \mathbf{1}_{\mathrm{D}}^{\mathrm{T}} \right)^{\mathrm{T}} \right),$$

where $\tilde{\mathcal{F}}(A) = [\mathcal{F}(A_{:,1}) \quad \ldots \quad \mathcal{F}(A_{:,\mathrm{D}})] \in \mathbb{R}^{\mathrm{O} \times \mathrm{D}}$, for $A \in \mathbb{R}^{\mathrm{D} \times \mathrm{D}}$, is the map $\mathcal{F}$ applied column wise to A. The vector $\mathbf{1}_{\mathrm{D}} \in \mathbb{R}^{\mathrm{D}}$ consists of ones and the optimal value for $\theta \in \mathbb{R} \setminus \{0\}$ is analyzed in section 3.1.

The new loss function fixes $\alpha = 1/2$ in the previous one (7) and removes the term $\| \boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}} \|_{\Gamma^{-1}}^2$, since it vanishes for $\boldsymbol{\mu} = \boldsymbol{\mu}_{\mathrm{pr}}$ as $\| \boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}} \|_{\Gamma_{\mathrm{pr}}^{-1}}^2$ does. The coefficient θ is a (small) perturbation needed to obtain the following theoretical result. The coefficient θ^2 balances the derivatives of L_θ with respect to C. The expected value of (7) is substituted with the second and third lines of (15) which are more efficient to compute and mimic the expected value in the affine case (10).

With this novel loss function, we are able to establish the following new theorem.

Theorem 2. *Assume that $\mathcal{F}$ is a $C^1(\mathbb{R}^D, \mathbb{R}^O)$ function, $\mathbf{y} \in \mathbb{R}^O$ and that the prior and noise models are Gaussian and independent, $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$ and $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{E}}, \Gamma_{\mathrm{E}})$, respectively. Let $(\hat{\boldsymbol{\mu}}(\theta), \hat{C}(\theta))$ be the stationary points of $L_\theta(\boldsymbol{\mu}, C)$. Suppose that $\underline{\hat{\boldsymbol{\mu}}} = \lim_{\theta \to 0} \hat{\boldsymbol{\mu}}(\theta)$ and $\underline{\hat{C}} = \lim_{\theta \to 0} \hat{C}(\theta)$ exist and are bounded in norm, $\underline{\hat{C}}$ is invertible and the minimization problem (4) has only one stationary point $\mathbf{u}_{\mathrm{MAP}}$.*
Let $\underline{\hat{\Gamma}} = \underline{\hat{C}} \underline{\hat{C}}^{\mathrm{T}}$ and define:

$$\boldsymbol{\mu}_{\mathrm{post}} = \underline{\hat{\boldsymbol{\mu}}},$$
$$\Gamma_{\mathrm{post}} = \underline{\hat{\Gamma}} \Gamma_{\mathrm{pr}}^{-1} \underline{\hat{\Gamma}}.$$

Then $(\boldsymbol{\mu}_{\mathrm{post}}, \Gamma_{\mathrm{post}}) = (\mathbf{u}_{\mathrm{MAP}}, \Gamma_{\mathrm{Lap}})$, where $(\mathbf{u}_{\mathrm{MAP}}, \Gamma_{\mathrm{Lap}})$ are given by (4) and (5).

Proof. We compute the gradients of $L_\theta(\boldsymbol{\mu}, C)$:

$$\frac{\partial L_\theta}{\partial \boldsymbol{\mu}} = 2\Gamma_{\mathrm{pr}}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_{\mathrm{pr}}) - 2J_{\mathcal{F}}^{\mathrm{T}}(\boldsymbol{\mu})\Gamma_{\mathrm{E}}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathcal{F}(\boldsymbol{\mu})) +$$

$$2\sum_{k=1}^{D} \left(J_{\mathcal{F}}(\theta C_{:,k} + \boldsymbol{\mu}) - J_{\mathcal{F}}(\boldsymbol{\mu})\right)^{\mathrm{T}} \Gamma_{\mathrm{E}}^{-1}\left(\mathcal{F}(\theta C_{:,k} + \boldsymbol{\mu}) - \mathcal{F}(\boldsymbol{\mu})\right),$$

$$\frac{\partial L_\theta}{\partial C} = -2\theta^2 \Gamma^{-1} \Gamma_{\mathrm{pr}} C^{-\mathrm{T}} + 2\theta^2 \Gamma_{\mathrm{pr}}^{-1} C +$$
$$2\theta \left[J_{\mathcal{F}}^{\mathrm{T}}(\theta C_{:,1} + \boldsymbol{\mu}) \Gamma_{\mathrm{E}}^{-1}\left(\mathcal{F}(\theta C_{:,1} + \boldsymbol{\mu}) - \mathcal{F}(\boldsymbol{\mu})\right) \quad \ldots \right].$$

Setting the gradients to vanish and omitting the dependence of $(\hat{\boldsymbol{\mu}}(\theta), \hat{C}(\theta))$ on θ, we get:

$$\Gamma_{\mathrm{pr}}^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{\mathrm{pr}}) - J_{\mathcal{F}}^{\mathrm{T}}(\hat{\boldsymbol{\mu}})\Gamma_{\mathrm{E}}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathcal{F}(\hat{\boldsymbol{\mu}})) +$$

(16) $$\sum_{k=1}^{D} \left(J_{\mathcal{F}}(\theta \hat{C}_{:,k} + \hat{\boldsymbol{\mu}}) - J_{\mathcal{F}}(\hat{\boldsymbol{\mu}})\right)^{\mathrm{T}} \Gamma_{\mathrm{E}}^{-1}\left(\mathcal{F}(\theta \hat{C}_{:,k} + \hat{\boldsymbol{\mu}}) - \mathcal{F}(\hat{\boldsymbol{\mu}})\right) = \mathbf{0},$$

$$-\hat{\Gamma}^{-1}\Gamma_{\mathrm{pr}}\hat{\Gamma}^{-1} + \Gamma_{\mathrm{pr}}^{-1} +$$

(17) $$\frac{1}{\theta}\left[J_{\mathcal{F}}^{\mathrm{T}}(\theta \hat{C}_{:,1} + \hat{\boldsymbol{\mu}})\Gamma_{\mathrm{E}}^{-1}\left(\mathcal{F}(\theta \hat{C}_{:,1} + \hat{\boldsymbol{\mu}}) - \mathcal{F}(\hat{\boldsymbol{\mu}})\right) \quad \ldots \right] \hat{C}^{-1} = 0.$$

As $\mathcal{F} \in C^1(\mathbb{R}^D, \mathbb{R}^O)$ and $\underline{\hat{C}}$ is bounded in norm, we can expand the terms $\mathcal{F}(\theta \hat{C}_{:,k} + \hat{\boldsymbol{\mu}})$ and $J_{\mathcal{F}}(\theta \hat{C}_{:,k} + \hat{\boldsymbol{\mu}})$ for $\theta \to 0$ as:

$$\mathcal{F}(\theta \hat{C}_{:,k} + \hat{\boldsymbol{\mu}}) = \mathcal{F}(\hat{\boldsymbol{\mu}}) + \theta J_{\mathcal{F}}(\hat{\boldsymbol{\mu}})\hat{C}_{:,k} + o(\theta\|\hat{C}_{:,k}\|_2) = \mathcal{F}(\hat{\boldsymbol{\mu}}) + \theta J_{\mathcal{F}}(\hat{\boldsymbol{\mu}})\hat{C}_{:,k} + o(\theta),$$
$$J_{\mathcal{F}}(\theta \hat{C}_{:,k} + \hat{\boldsymbol{\mu}}) = J_{\mathcal{F}}(\hat{\boldsymbol{\mu}}) + o(\|\hat{C}_{:,k}\|_2) = J_{\mathcal{F}}(\hat{\boldsymbol{\mu}}) + o(1),$$

where $o(\cdot)$ indicates the small-o notation. The boundedness of $\underline{\hat{C}}$ is needed to remove $\|\hat{C}_{:,k}\|_2$ from $o(\cdot)$. By computing the limit for $\theta \to 0$ of equation (16), using $\mathcal{F} \in C^1(\mathbb{R}^D, \mathbb{R}^O)$ and that $\hat{\boldsymbol{\mu}}, \hat{C}$ exist and are bounded, we obtain:

$$\Gamma_{\mathrm{pr}}^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{\mathrm{pr}}) - J_{\mathcal{F}}^{\mathrm{T}}(\hat{\boldsymbol{\mu}})\Gamma_{\mathrm{E}}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathcal{F}(\hat{\boldsymbol{\mu}})) + o(\theta) \to$$
$$\Gamma_{\mathrm{pr}}^{-1}\left(\underline{\hat{\boldsymbol{\mu}}} - \boldsymbol{\mu}_{\mathrm{pr}}\right) - J_{\mathcal{F}}^{\mathrm{T}}\left(\underline{\hat{\boldsymbol{\mu}}}\right)\Gamma_{\mathrm{E}}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_{\mathrm{E}} - \mathcal{F}\left(\underline{\hat{\boldsymbol{\mu}}}\right)\right) = \mathbf{0},$$

By the above equation, we observe that $\boldsymbol{\mu}_{\mathrm{post}} = \underline{\hat{\boldsymbol{\mu}}}$ satisfies the same equation of the stationary point $\mathbf{u}_{\mathrm{MAP}}$ of equation (4), therefore, by the uniqueness of the stationary point of (4), $\boldsymbol{\mu}_{\mathrm{post}}$ is equal to $\mathbf{u}_{\mathrm{MAP}}$.

Since $\hat{\underline{C}}$ is invertible, $\hat{\underline{\Gamma}}^{-1}$ exists. We compute the limit of equation (17) for $\theta \to 0$:

$$- \hat{\Gamma}^{-1} \Gamma_{\mathrm{pr}} \hat{\Gamma}^{-1} + \Gamma_{\mathrm{pr}}^{-1} + \frac{1}{\theta} \left(J_{\mathcal{F}}^{\mathrm{T}}(\hat{\boldsymbol{\mu}}) + o(1) \right) \Gamma_{\mathrm{E}}^{-1} \left(\theta J_{\mathcal{F}}(\hat{\boldsymbol{\mu}}) \hat{C} + o(\theta) \right) \hat{C}^{-1} \to$$

$$(18) \quad - \hat{\underline{\Gamma}}^{-1} \Gamma_{\mathrm{pr}} \hat{\underline{\Gamma}}^{-1} + \Gamma_{\mathrm{pr}}^{-1} + J_{\mathcal{F}}^{\mathrm{T}} \left(\hat{\underline{\boldsymbol{\mu}}} \right) \Gamma_{\mathrm{E}}^{-1} J_{\mathcal{F}} \left(\hat{\underline{\boldsymbol{\mu}}} \right) = 0.$$

Since $\hat{\underline{\boldsymbol{\mu}}}$ is equal to $\mathbf{u}_{\mathrm{MAP}}$, we get:

$$\Gamma_{\mathrm{pr}}^{-1} + J_{\mathcal{F}}^{\mathrm{T}} \left(\hat{\underline{\boldsymbol{\mu}}} \right) \Gamma_{\mathrm{E}}^{-1} J_{\mathcal{F}}^{\mathrm{T}} \left(\hat{\underline{\boldsymbol{\mu}}} \right) = \Gamma_{\mathrm{Lap}}^{-1}.$$

Therefore, from equation (18), Γ_{post} is equal to Γ_{Lap}. $\qquad\square$

Remark. The above theorem only holds for $\theta \to 0$. In practice, θ can be chosen arbitrarily small without incurring additional computational cost, as long as machine precision does not hinder the evaluation of the loss.

Remark. The $\lim_{\theta \to 0} L_\theta(\boldsymbol{\mu}, C)$ corresponds to the loss function of minimization problem (4), which makes the result $\boldsymbol{\mu}_{\mathrm{post}} = \mathbf{u}_{\mathrm{MAP}}$ intuitive. However, it must be shown that the minimization of $L_\theta(\boldsymbol{\mu}, C)$ with respect to $\boldsymbol{\mu}$ and the limit as $\theta \to 0$ commute, as we have done. Moreover, although taking $\theta \to 0$ recovers the minimization problem for $\mathbf{u}_{\mathrm{MAP}}$, the θ dependent terms still enable the estimation of the posterior covariance matrix.

Remark. When $\mathcal{F}(\mathbf{u}) = F\mathbf{u} + \mathbf{f}$ is affine, the result of Theorem 2 holds $\forall \theta \in \mathbb{R} \setminus \{0\}$ without any hypotheses on $\hat{\boldsymbol{\mu}}$, $\hat{C}$ and the minimization problem (4). It is sufficient to substitute in (16) and (17) the expression of $\mathcal{F}$ and $J_{\mathcal{F}}(\mathbf{u}) = F$.

Remark. The above theorem does not assume a specific posterior distribution therefore the mean and covariance estimates hold in general. We choose to approximate the posterior distribution using a multivariate Gaussian distribution, as shown in (6).

Remark. The assumption of a unique stationary point for (4) is restrictive and generally holds only when the prior dominates the loss function.

The computational cost of evaluating (15) is $O(D^3 + Dc(\mathcal{F}) + DO^2)$, which follows from the same cost analysis as for (14). Thanks to the curse of dimensionality, the number of samples K to accurately estimate an expected value is generally much higher than the sample space dimension D. Consequently, evaluating (15) is significantly less expensive than evaluating (14).

2.2. Uncertainty quantification variational autoencoders.
An autoencoder consists in the composition of two maps called encoder and decoder. The encoder maps the inputs of the autoencoder to a latent space, while the decoder maps the outputs of the encoder to the original space. Autoencoders typically approximate the identity function, meaning that their inputs and outputs should match. In our method, the encoder is a NN $\varphi : \mathbb{R}^O \to \mathbb{R}^{D + \frac{D(D+1)}{2}}$, while the decoder corresponds to the map $\mathcal{F}$ or an approximation of it. The latent space of the UQ-VAE represents the mean $\boldsymbol{\mu}$ and the Cholesky factor C of the multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \Gamma)$ (Figure 1).

The following Corollary, proven in Ref. 42 and easy to extend to the new loss function (15), demonstrates the convergence of the outputs of a single-layer linear encoder to the posterior mean and covariance in the case of $\mathcal{F}$ affine.

Corollary 3. *Consider a single-layer linear encoder. Let $L_\theta(\phi) = L_\theta(\boldsymbol{\mu}(\phi), \Gamma(\phi))$, where ϕ is the set of weights and biases of the NN. Under the assumptions of Theorem 2 and $\mathcal{F}$*
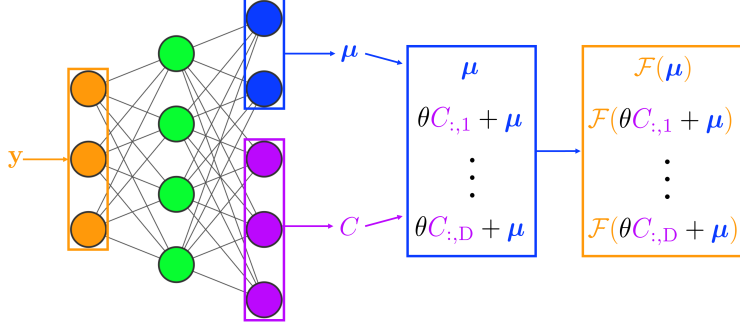
FIGURE 1. Structure of the UQ-VAE when $\mathcal{F}$ is evaluated.

affine, we define:

$$\boldsymbol{\mu} = W_{\boldsymbol{\mu}}\mathbf{y} + \mathbf{b}_{\boldsymbol{\mu}}, \tag{19}$$

$$C = \text{vec}_{\text{L}}^{-1}(\mathbf{l}) + \text{diag}(\boldsymbol{\sigma}), \tag{20}$$

$$\boldsymbol{\sigma} = \exp\left(W_{\boldsymbol{\sigma}}\mathbf{y} + \mathbf{b}_{\boldsymbol{\sigma}}\right), \tag{21}$$

$$\mathbf{l} = W_{\mathbf{l}}\mathbf{y} + \mathbf{b}_{\mathbf{l}}, \tag{22}$$

where $W_{\boldsymbol{\mu}} \in \mathbb{R}^{\text{D}\times\text{O}}$ and $W_{\boldsymbol{\sigma}} \in \mathbb{R}^{\text{D}\times\text{O}}, W_{\mathbf{l}} \in \mathbb{R}^{\frac{\text{D}(\text{D}-1)}{2}\times\text{O}}$ are the encoder weight matrices, $\mathbf{b}_{\boldsymbol{\mu}} \in \mathbb{R}^{\text{D}}, \mathbf{b}_{\boldsymbol{\sigma}} \in \mathbb{R}^{\text{D}}$ and $\mathbf{b}_{\mathbf{l}} \in \mathbb{R}^{\frac{\text{D}(\text{D}-1)}{2}}$ are the encoder biases. The functions $\exp$, vec_{L} and diag denote the elementwise exponential function, the vectorization of a strictly lower triangular matrix and the construction of a diagonal matrix from its vector diagonal, respectively.

The stationary points $\hat{\phi}$ of $L_{\theta}(\phi)$ are such that $(\boldsymbol{\mu}_{\text{post}}(\hat{\phi}), \Gamma_{\text{post}}(\hat{\phi})) = (\mathbf{u}_{\text{MAP}}, \Gamma_{\text{Lap}})$.

For deeper encoders with non linear activation functions, $(\boldsymbol{\mu}, C)$ is computed as in (19), (20), (21) and (22), using the encoder's outputs in place of $W_{\boldsymbol{\mu}}\mathbf{y} + \mathbf{b}_{\boldsymbol{\mu}}$, $W_{\boldsymbol{\sigma}}\mathbf{y} + \mathbf{b}_{\boldsymbol{\sigma}}$ and $W_{\mathbf{l}}\mathbf{y} + \mathbf{b}_{\mathbf{l}}$.

In many applications, evaluating $\mathcal{F}$ is computationally expensive or $\mathcal{F}$ can be unknown. In such cases, we approximate the map $\mathcal{F}$ using a second NN $\psi : \mathbb{R}^{\text{D}} \to \mathbb{R}^{\text{O}}$ (Figure 2). To account for the approximation error introduced by the NN ψ, we use an error random variable $E_{\text{dec}} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{dec}}, \Gamma_{\text{dec}})$ in the model (1):

$$\mathcal{F}(U) = \psi(U) + E_{\text{dec}} \Rightarrow Y = \psi(U) + E_{\text{dec}} + E.$$

We assume that E_{dec} and E are independent random variables, so that their sum $E_{\text{dec}} + E$ is distributed as $\mathcal{N}(\boldsymbol{\mu}_{\text{dec}} + \boldsymbol{\mu}_{\text{E}}, \Gamma_{\text{dec}} + \Gamma_{\text{E}})$. Defining $\boldsymbol{\mu}_{\tilde{\text{E}}} = \boldsymbol{\mu}_{\text{dec}} + \boldsymbol{\mu}_{\text{E}}$ and $\Gamma_{\tilde{\text{E}}} = \Gamma_{\text{dec}} + \Gamma_{\text{E}}$, the corresponding loss function becomes:

$$\theta^2 \text{tr}\left(\Gamma^{-1}\Gamma_{\text{pr}}\right) + \|\boldsymbol{\mu} - \boldsymbol{\mu}_{\text{pr}}\|^2_{\Gamma_{\text{pr}}^{-1}} + \theta^2 \text{tr}\left(\Gamma_{\text{pr}}^{-1}\Gamma\right) +$$

$$\|\mathbf{y} - \mu_{\tilde{\text{E}}} - \psi(\boldsymbol{\mu})\|^2_{\Gamma_{\tilde{\text{E}}}^{-1}} +$$

$$\text{tr}\left(\Gamma_{\tilde{\text{E}}}^{-1}\left(\tilde{\psi}(\theta C + \boldsymbol{\mu}\mathbf{1}_{\text{D}}^{\text{T}}) - \psi(\boldsymbol{\mu})\mathbf{1}_{\text{D}}^{\text{T}}\right)\left(\tilde{\psi}(\theta C + \boldsymbol{\mu}\mathbf{1}_{\text{D}}^{\text{T}}) - \psi(\boldsymbol{\mu})\mathbf{1}_{\text{D}}^{\text{T}}\right)^{\text{T}}\right), \tag{23}$$

where $\tilde{\psi}(A) = [\psi(A_{:,1}) \quad \ldots \quad \psi(A_{:,\text{D}})] \in \mathbb{R}^{\text{O}\times\text{D}}$, for $A \in \mathbb{R}^{\text{D}\times\text{D}}$.

When the UQ-VAE consists of two NNs, its training is performed in two stages: at first the decoder NN is trained and after the encoder one.[45] Simultaneous training of both NNs
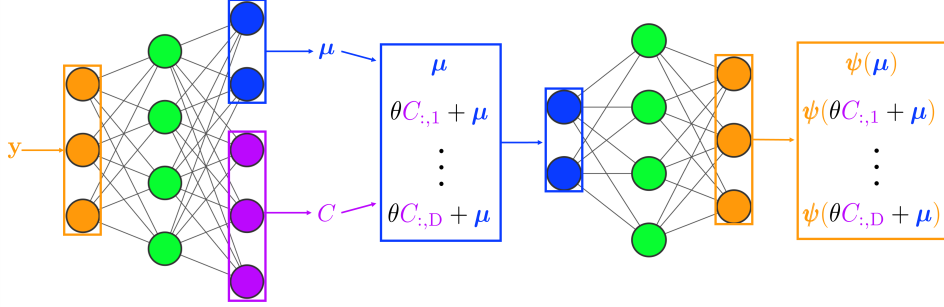
FIGURE 2. Structure of the UQ-VAE using a NN to approximate $\mathcal{F}$.

may lead to unpredictable results, as our analysis of the minima of the loss function (15) assumes a fixed decoder during optimization.

Remark. In practice, we always use fully connected feed forward NNs as encoders and decoders. However, the proposed approach is compatible with other NN architectures.

Remark. We train the decoder NN using the squared error between its outputs and the noiseless observational data.

Remark. After training the decoder NN, we estimate $\boldsymbol{\mu}_{\text{dec}}$ and Γ_{dec} by computing the sample mean and covariance of $\mathcal{F}(U) - \psi(U)$ on a test set of approximately $29,000$ samples.

Remark. When a dataset $\{\mathbf{y}^{(\text{m})}\}_{\text{m}=1}^{\text{M}}$, with $M \in \mathbb{N}$, is available, we train the encoder by minimizing the average of the loss functions (15) or (23) over the dataset.

2.2.1. *Weights and biases initialization and dataset normalization.* The initialization of the weights and biases of the encoder plays a crucial role in its training. A good initial guess is obtained by setting $\boldsymbol{\mu} \sim \boldsymbol{\mu}_{\text{pr}}$ and $C \sim C_{\text{pr}}$ (with C_{pr} the Cholesky factor of Γ_{pr}), as these values minimize the term $\theta^2 \text{tr}\left(\Gamma^{-1}\Gamma_{\text{pr}}\right) + \|\boldsymbol{\mu} - \boldsymbol{\mu}_{\text{pr}}\|_{\Gamma_{\text{pr}}^{-1}}^2 + \theta^2 \text{tr}\left(\Gamma_{\text{pr}}^{-1}\Gamma\right)$ in the loss function (15). The encoder weights are initialized according to the Xavier uniform initializer.[13] Then, all the weights in the last layer are scaled by 10^{-4}. This ensures that, up to a small discrepancy, the output of the encoder is given by the biases of the last layer. All encoder biases are set to zero, except for those of the last layer which are fixed to $\boldsymbol{\mu}_{\text{pr}}$ and $\mathbf{c}_{\text{pr}}$, where $\mathbf{c}_{\text{pr}}$ is given by applying the inverse of the right hand side transformation of (20) to $C_{\text{pr}} \in \mathbb{R}^{\text{D}\times\text{D}}$.

Since the loss function (15) depends on the mean and covariance of both parameters and noise, normalizing the dataset (to enhance the UQ-VAE generalization capability) also affects these quantities.

We consider a dataset $\{(\mathbf{u}^{(\text{m})}, \mathbf{y}^{(\text{m})})\}_{\text{m}=1}^{\text{M}}$, where $M \in \mathbb{N}$, $\mathbf{y}^{(\text{m})} = \mathcal{F}(\mathbf{u}^{(\text{m})}) + \varepsilon^{(\text{m})}$, with $\mathbf{u}^{(\text{m})}$ and $\varepsilon^{(\text{m})}$ are sampled from $\mathcal{N}(\boldsymbol{\mu}_{\text{pr}}, \Gamma_{\text{pr}})$ and $\mathcal{N}(\boldsymbol{\mu}_{\text{E}}, \Gamma_{\text{E}})$, respectively. To normalize the data within a given range, e.g., $[0, 1]$, we introduce vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{\text{D}}$ and $\mathbf{c}, \mathbf{d} \in \mathbb{R}^{\text{O}}$ such that:

$$\bar{\mathbf{u}}^{(\text{m})} = \mathbf{u}^{(\text{m})} \odot \mathbf{a} + \mathbf{b} \in [0, 1]^{\text{D}},$$

$$\bar{\mathbf{y}}^{(\text{m})} = \mathbf{y}^{(\text{m})} \odot \mathbf{c} + \mathbf{d} \in [0, 1]^{\text{O}}.$$

As shown in Section 2.4.1 of Ref. 42, the normalized random variables $\bar{Y}$, $\bar{U}$ and $\bar{E}$ satisfy a model analogous to (1) in the normalized scale, with $\bar{U} \sim N(\bar{\boldsymbol{\mu}}_{\text{pr}}, \bar{\Gamma}_{\text{pr}})$ and

$\bar{E} \sim N(\bar{\boldsymbol{\mu}}_{\mathrm{E}}, \bar{\Gamma}_{\mathrm{E}})$, where:

$$\bar{\boldsymbol{\mu}}_{\mathrm{pr}} = \boldsymbol{\mu}_{\mathrm{pr}} \odot \mathbf{a} + \mathbf{b},$$
$$\bar{\Gamma}_{\mathrm{pr}} = \Gamma_{\mathrm{pr}} \odot (\mathbf{a}\mathbf{a}^{\mathrm{T}}),$$
$$\bar{\boldsymbol{\mu}}_{\mathrm{E}} = \boldsymbol{\mu}_{\mathrm{E}} \odot \mathbf{c},$$
$$\bar{\Gamma}_{\mathrm{E}} = \Gamma_{\mathrm{E}} \odot (\mathbf{c}\mathbf{c}^{\mathrm{T}}).$$

The same normalization is applied to observational data both with and without noise. Once $(\bar{\boldsymbol{\mu}}_{\mathrm{post}}, \bar{\Gamma}_{\mathrm{post}})$ in the normalized scale is computed, we revert to the original scale using the transformation:

$$\boldsymbol{\mu}_{\mathrm{post}} = (\bar{\boldsymbol{\mu}}_{\mathrm{post}} - \mathbf{b}) \oslash \mathbf{a},$$
$$\Gamma_{\mathrm{post}} = \bar{\Gamma}_{\mathrm{post}} \odot ((\mathbf{1}_{\mathrm{D}} \oslash \mathbf{a})(\mathbf{1}_{\mathrm{D}} \oslash \mathbf{a})^{\mathrm{T}}).$$

3. NUMERICAL RESULTS

We carry out three benchmark numerical tests to show the capabilities of UQ-VAEs: in section 3.1, we solve Bayesian inverse problems for the Poisson equation and study the impact of varying the number of neurons and layers of both the decoder and the encoder, as well as modifying θ and the robustness with respect to the noise level η, the number of observations O, the prior distribution $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$; in section 3.2, we face Bayesian inverse problems involving an exponential map $\mathcal{F}$, investigate the effect of varying the dimensions of the parameter and data space, D and O, and compare the performance, both in terms of computational time and accuracy, of UQ-VAEs trained using the loss function L_α (7) or L_θ (15); in section 3.3, we address Bayesian inverse problems for a 0D cardiocirculatory model in the context of systemic hypertension (section 3.3.1) and ventricular septal defects (section 3.3.2). The activation functions of the NNs are *ReLU* functions and the data are normalized to the range $[0, 1]$. Additionally, we always use the *Adam* optimizer.[18] We tuned the number of neurons and layers of each NN to minimize the validation loss functions. The generated datasets consist of M samples, which are split in $\lfloor 9M/10 \rfloor$ training samples and $M - \lfloor 9M/10 \rfloor$ validation samples, where $\lfloor \cdot \rfloor$ denotes the lower integer part.

In the following sections, the true posterior mean and covariance matrix are approximated using Sobol' sequences,[33] a Quasi-Monte Carlo method that accelerates the convergence of the sample mean to the expected value. Using (2) and (3), we obtain:

$$(24) \qquad \mathbb{E}_{\mathrm{P}_{\mathrm{U|Y}}(\mathbf{u|y})}[\mathbf{u}] = \frac{\mathbb{E}_{\mathrm{P}_{\mathrm{U}}(\mathbf{u})}[\mathbf{u}\, p_{\mathrm{E}}(\mathbf{y} - \mathcal{F}(\mathbf{u}))]}{\mathbb{E}_{\mathrm{P}_{\mathrm{U}}(\mathbf{u})}[p_{\mathrm{E}}(\mathbf{y} - \mathcal{F}(\mathbf{u}))]},$$

$$\mathrm{Cov}_{\mathrm{P}_{\mathrm{U|Y}}[\mathbf{u|y}]}(\mathbf{u}) = \frac{\mathbb{E}_{\mathrm{P}_{\mathrm{U}}(\mathbf{u})}[\mathbf{u}\mathbf{u}^{\mathrm{T}}\, p_{\mathrm{E}}(\mathbf{y} - \mathcal{F}(\mathbf{u}))]}{\mathbb{E}_{\mathrm{P}_{\mathrm{U}}(\mathbf{u})}[p_{\mathrm{E}}(\mathbf{y} - \mathcal{F}(\mathbf{u}))]} -$$

$$(25) \qquad \frac{\mathbb{E}_{\mathrm{P}_{\mathrm{U}}(\mathbf{u})}[\mathbf{u}\, p_{\mathrm{E}}(\mathbf{y} - \mathcal{F}(\mathbf{u}))]\mathbb{E}_{\mathrm{P}_{\mathrm{U}}(\mathbf{u})}[\mathbf{u}\, p_{\mathrm{E}}(\mathbf{y} - \mathcal{F}(\mathbf{u}))]^{\mathrm{T}}}{\mathbb{E}_{\mathrm{P}_{\mathrm{U}}(\mathbf{u})}[p_{\mathrm{E}}(\mathbf{y} - \mathcal{F}(\mathbf{u}))]^2}.$$

This procedure is computationally expensive and can only be performed for simple problems.

3.1. **Poisson equation.** We consider the Poisson equation on the unit square $\Omega = (0, 1)^2$:

$$(26) \qquad \begin{cases} -\nabla(e^{\mathrm{u}(\mathbf{x})}\nabla y(\mathbf{x})) = \frac{17}{4}\pi^2 \sin(2\pi x_1) \sin(2\pi x_2) & \mathbf{x} \in \Omega, \\ y(\mathbf{x}) = 0 & \mathbf{x} \in \partial\Omega, \end{cases}$$

where $u(\mathbf{x})$ is the (log) diffusion parameter. We approximate the solution $y(\mathbf{x})$ of the Poisson problem by the linear finite element method on a triangular mesh of $D = 17^2 = 289$ degrees of freedom.

We define a Gaussian random field on the whole domain Ω, with mean and covariance functions $\mu_{\mathrm{pr}}(\mathbf{x})$ and $\mathcal{C}_{\mathrm{pr}}(\mathbf{x}, \mathbf{z})$, for $\mathbf{x}, \mathbf{z} \in (0, 1)$. The parameter U is given by the evaluation of this Gaussian random field at the degrees of freedom of the mesh, obtaining $U \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$. We define μ_{pr} to be zero in Ω and the covariance function $\mathcal{C}_{\mathrm{pr}}$ to ensure the well-posedness of infinite dimensional Bayesian inverse problems.[3,37] We define the differential operator $\mathcal{A}$:

$$\mathcal{A}u = -\gamma \Delta u + \delta u \quad \text{in } \Omega,$$

where $\gamma, \delta > 0$. The operator $\mathcal{A}^{-2}(\mathbf{x}, \mathbf{z}) = G_2(\mathbf{x}, \mathbf{z})$, for $(\mathbf{x}, \mathbf{z}) \in \Omega$, is defined as the solution operator of:

$$\begin{cases} \mathcal{A}G_1(\mathbf{x}, \mathbf{z}) = \delta_{\mathbf{x}} & \mathbf{z} \in \Omega, \\ \nabla G_1(\mathbf{x}, \mathbf{z}) \cdot \mathbf{n}(\mathbf{z}) + \beta G_1(\mathbf{x}, \mathbf{z}) = 0 & \mathbf{z} \in \partial\Omega, \\ \mathcal{A}G_2(\mathbf{x}, \mathbf{z}) = G_1(\mathbf{x}, \mathbf{z}) & \mathbf{z} \in \Omega, \\ \nabla G_2(\mathbf{x}, \mathbf{z}) \cdot \mathbf{n}(\mathbf{z}) + \beta G_2(\mathbf{x}, \mathbf{z}) = 0 & \mathbf{z} \in \partial\Omega, \end{cases}$$

where, for $\mathbf{x} \in \Omega$, $\delta_{\mathbf{x}}$ is the Dirac delta centered in $\mathbf{x}$, $\mathbf{n}$ is the unit exterior normal to $\partial\Omega$ and β is chosen to reduce the boundary artifacts.[8] We set $\gamma = 0.1$ and $\delta = 0.5$. The covariance function is given by $\mathcal{C}_{\mathrm{pr}}(\mathbf{x}, \mathbf{z}) = \mathcal{A}^{-\frac{3}{2}}(\mathbf{x}, \mathbf{z})$. We compute $\mathcal{C}_{\mathrm{pr}}$ using the same mesh and finite element method as for solving the Poisson problem, via the Python library for inverse problems *hIPPYlib*.[46–48] Finally, we define $(\boldsymbol{\mu}_{\mathrm{pr}})_i = \mu_{\mathrm{pr}}(\mathbf{x}_i)$ and $(\Gamma_{\mathrm{pr}})_{i,j} = \mathcal{C}_{\mathrm{pr}}(\mathbf{x}_i, \mathbf{x}_j)$ for $i, j = 1, \cdots, D$, where $\mathbf{x}_i$ are the coordinates of the degrees of freedom of the mesh. The observational data consist of noisy evaluations of the solution $y(\mathbf{x})$ of (26) at $O = 20$ random points in the domain Ω.

We generate a dataset $\{(\mathbf{u}^{(m)}, \mathbf{y}^{(m)})\}_{m=1}^{M}$, where $M \in \mathbb{N}$ and $\mathbf{u}^{(m)}$ is sampled from $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$. Let $\tilde{\mathbf{y}}^{(m)}$ be the noiseless observations of the solution of (26) with parameter $\mathbf{u}^{(m)}$. Then, $\mathbf{y}^{(m)} = \tilde{\mathbf{y}}^{(m)} + \boldsymbol{\varepsilon}^{(m)}$, for $m = 1, \cdots, M$, where $\boldsymbol{\varepsilon}^{(m)}$ is sampled from $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{E}}, \Gamma_{\mathrm{E}})$ with $\boldsymbol{\mu}_{\mathrm{E}} = \mathbf{0}$ and Γ_{E} is a diagonal matrix with $(\Gamma_{\mathrm{E}})_{i,i} = (\eta \max_{j,m} |\tilde{y}_j^{(m)}|)^2$, for $i = 1, \cdots, O$ and $\eta \in \mathbb{R}$. We fix $\eta = 0.01$.

To normalize the data to $[0, 1]$, we apply a uniform normalization to the parameters and observational data across the entire mesh. We define:

$$a = (\max_{i,m} u_i^{(m)} - \min_{i,m} u_i^{(m)})^{-1},$$

$$b = -a \min_{i,m} u_i^{(m)},$$

$$c = (\max_{i,m} y_i^{(m)} - \min_{i,m} y_i^{(m)})^{-1},$$

$$d = -c \min_{i,m} y_i^{(m)}$$

and set $\bar{u}(\mathbf{x}) = u(\mathbf{x})a + b$ and $\bar{y}(\mathbf{x}) = y(\mathbf{x})c + d$ in the continuous framework. Referring to section 2.2.1, in the discrete setting, $\mathbf{a} = a\mathbf{1}_D$, $\mathbf{b} = b\mathbf{1}_D$, $\mathbf{c} = c\mathbf{1}_O$ and $\mathbf{d} = d\mathbf{1}_O$. We can derive the Poisson problem in the normalized scale:

$$\begin{cases} -\nabla(e^{\frac{\bar{u}(\mathbf{x})}{a}} \nabla \bar{y}(\mathbf{x})) = c e^{\frac{b}{a}} \frac{17}{4} \pi^2 \sin(2\pi x_1) \sin(2\pi x_2) & \mathbf{x} \in \Omega, \\ \bar{y}(\mathbf{x}) = d & \mathbf{x} \in \partial\Omega. \end{cases}$$

Decoder						
Neurons \ Layers	1	2	3	4	5	6
50	$5.8 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$	$2.2 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$
100	$7.4 \cdot 10^{-4}$	$3.1 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$	$1.8 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$	$\mathbf{1.2 \cdot 10^{-4}}$
150	$6.4 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$
200	$6.4 \cdot 10^{-4}$	$3.0 \cdot 10^{-4}$	$2.4 \cdot 10^{-4}$	$1.8 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$
250	$5.9 \cdot 10^{-4}$	$3.4 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$	$1.8 \cdot 10^{-4}$

Encoder				
Neurons \ Layers	1	2	3	4
500	26.8	24.0	25.2	25.7
750	25.6	24.2	24.6	25.5
1000	24.5	**23.5**	24.9	25.6
1250	24.5	24.0	25.3	26.5

TABLE 1. Final validation loss functions for the decoder and the encoder for the Poisson problem.

We perform hyperparameter tuning to select the optimal number of layers and neurons per layer for both the decoder and the encoder neural networks. We first determine the best architecture for the decoder network and then use this decoder to train the encoder with parameter $\theta = 10^{-4}$. According to the hyperparameter tuning results (Table 1), we select a decoder with 6 hidden layers and 100 neurons per layer, trained on a dataset of $M = 4096$ samples. The chosen encoder has 2 hidden layers and 1000 neurons per layer and is trained on a subset of 100 samples of the decoder dataset. Although the dimension of the encoder dataset is small with respect to the decoder one, we show in the following that it still generalizes to unseen data.

We generate 2^{19} samples from the distribution $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$ to estimate (24) and (25).

We analyze the UQ-VAE performances varying θ in the loss function (23). For the posterior means $\mathbb{E}_{\mathrm{p}_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]$ and variances $\mathrm{Var}_{\mathrm{p}_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]$ (the diagonal of (25)), with $m \in \{1, \dots, M\}$, we compute the relative error with respect to the infinity norms at each degree of freedom:

$$(27) \qquad \mathbf{e}_{\mathbb{E}_{\mathrm{p}_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]} = \frac{|\boldsymbol{\mu}_{\mathrm{post}}^{(\mathrm{m})} - \mathbb{E}_{\mathrm{p}_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]|}{\|\mathbb{E}_{\mathrm{p}_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]\|_{\infty}},$$

$$\mathbf{e}_{\mathrm{Var}_{\mathrm{p}_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]} = \frac{|\mathrm{diag}\left(\Gamma_{\mathrm{post}}^{(\mathrm{m})}\right) - \mathrm{Var}_{\mathrm{p}_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}(\mathbf{u})|}{\|\mathrm{Var}_{\mathrm{p}_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]\|_{\infty}},$$

where $(\boldsymbol{\mu}_{\mathrm{post}}^{(\mathrm{m})}, \Gamma_{\mathrm{post}}^{(\mathrm{m})})$ are the UQ-VAE mean and covariance estimates for the m-th sample, $|\cdot|$ is applied element-wise and $\|\cdot\|_{\infty}$ is the vector infinite norm. The maximum over the mesh of these errors is the relative error on the m-th sample in the infinity norm.

The average relative errors on the posterior mean for a test set of 100 samples share similar trends and magnitudes for all values of θ (Figure 3). The maximum average relative error on the posterior variance is more than 50% for $\theta = 10^{-5} \vee 10^{-6}$, due to the small gradient of L_{θ} (23) with respect to C which scales with θ^2. For the other values of θ the errors on the variance are similar. We choose to use in all the tests that follow (also in section 3.2 and section 3.3) $\theta = 10^{-4}$, because Theorem 2 holds for $\theta \to 0$.

For $\theta = 10^{-4}$, the average relative error on the posterior mean and the posterior variance are less than 11%. Even if there are differences in $\mathbb{E}_{\mathrm{p}_{\mathrm{U|Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]$ and $\boldsymbol{\mu}_{\mathrm{post}}^{(\mathrm{m})}$, they estimate the true diffusion coefficients $\mathbf{u}^{(\mathrm{m})}$ with the same accuracy (Figure 4), showing that the
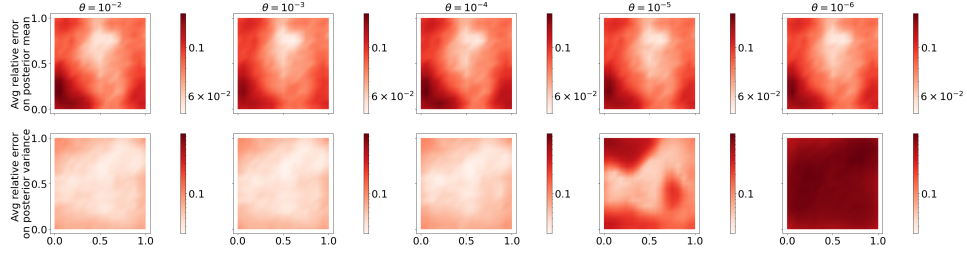
FIGURE 3. Average relative errors (in logarithmic scale) of the UQ-VAE mean $\boldsymbol{\mu}_{\text{post}}^{(\text{m})}$ and variance $\Gamma_{\text{post}}^{(\text{m})}$ on the posterior mean $\mathbb{E}_{\text{P}_{\text{U}|\text{Y}}(\mathbf{u}|\mathbf{y}^{(\text{m})})}[\mathbf{u}]$ and on the posterior variance $\text{Var}_{\text{P}_{\text{U}|\text{Y}}(\mathbf{u}|\mathbf{y}^{(\text{m})})}[\mathbf{u}]$ in dependence of θ for a test set of 100 samples for the Poisson problem.
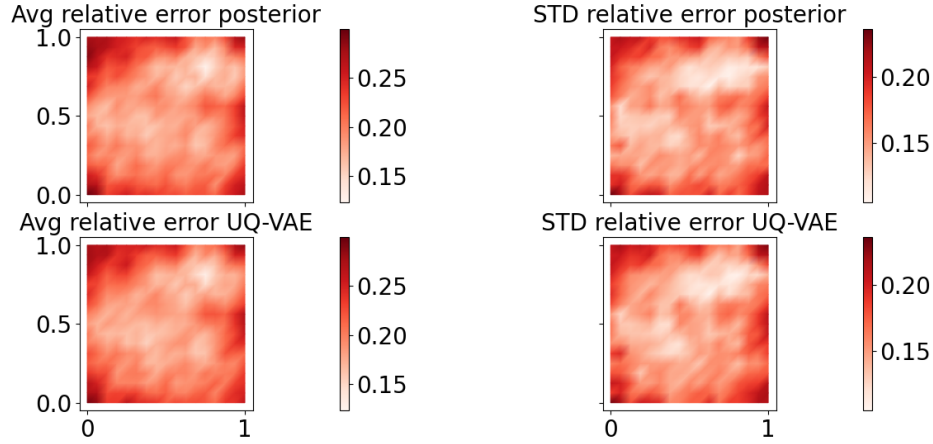

FIGURE 4. Averages and standard deviations of the relative errors of the posterior mean $\mathbb{E}_{\text{P}_{\text{U}|\text{Y}}(\mathbf{u}|\mathbf{y}^{(\text{m})})}[\mathbf{u}]$ and $\boldsymbol{\mu}_{\text{post}}^{(\text{m})}$ on the true diffusion coefficients $\mathbf{u}^{(\text{m})}$ for a test set of 100 samples for the Poisson problem.

	Offline phase	Online phase (100 samples)
UQ-VAE	$5.5h$	$10.74s$
Sobol' sequences	$0.6h$	$8880s$

TABLE 2. Comparison between the time required to solve Bayesian inverse problems with UQ-VAE or approximating (24) and (25) with Sobol' sequences.

UQ-VAE approach is effective in estimating parameters in a noisy setting. The relative errors on the true diffusion coefficients are computed by substituting $\mathbb{E}_{\text{P}_{\text{U}|\text{Y}}(\mathbf{u}|\mathbf{y}^{(\text{m})})}[\mathbf{u}]$ with $\mathbf{u}^{(\text{m})}$ in (27).

The offline computational time for the UQ-VAE (generating the dataset and training the decoder and the encoder) is $5.5h$, whereas the one for generating 2^{19} samples to estimate (24) and (25) is $0.6h$ (Table 2). The UQ-VAE approach is advantageous on the online
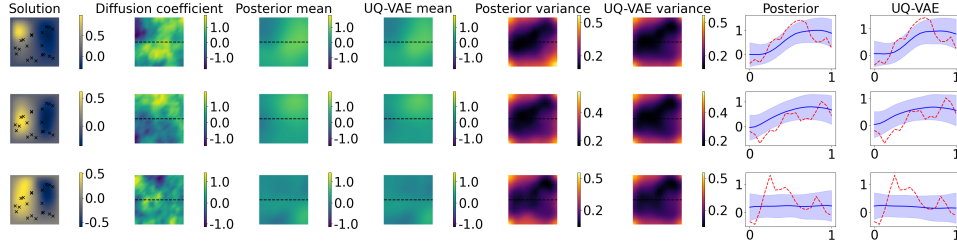
FIGURE 5. True solutions $y^{(\mathrm{m})}(\mathbf{x})$ with marked observation points, true diffusion coefficients $\mathbf{u}^{(\mathrm{m})}$, posterior and UQ-VAE mean ($\mathbb{E}_{\mathrm{P}_{\mathrm{U}|\mathrm{Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]$ and $\boldsymbol{\mu}_{\mathrm{post}}^{(\mathrm{m})}$) and variance ($\mathrm{Var}_{\mathrm{P}_{\mathrm{U}|\mathrm{Y}}(\mathbf{u}|\mathbf{y}^{(\mathrm{m})})}[\mathbf{u}]$ and $\Gamma_{\mathrm{post}}^{(\mathrm{m})}$) estimates for three samples. The last two columns are the posterior or UQ-VAE means (continue blue lines) and true diffusion coefficients (dashed red lines) evaluated at $y = 0.5$. The width of the shaded areas is equal to two times the standard deviation.

phase, when the Bayesian inverse problems are solved, with an average elapsed time of $0.1s$, compared to the $88.8s$ required to estimate (24) and (25).

We analyze, for three test samples, the posterior and UQ-VAE estimates for the mean and variance (Figure 5). The three diffusion coefficients are highly irregular and, when the data are not informative enough, the estimates are poor, like in the third case. For the first two cases, both the posterior and UQ-VAE mean are capable of capturing the trend and the magnitude of the true diffusion coefficient. The two estimates are smoother than the true diffusion coefficient. The UQ-VAE variance is comparable to the posterior one. The estimated mean and variance together are able to catch the irregularity of the true diffusion coefficient, both for the posterior and the UQ-VAE estimates.

We test the UQ-VAE approach on different values for the noise level η and the number of observation points O (Figure 6). For all the values of η and O the posterior and UQ-VAE mean estimates are in agreement. For $\eta = 0.05$ the UQ-VAE variance estimates slightly differ from the posterior ones, whereas for $\eta = 0.01$ the two variances are similar. This also happens for $\eta = 0.005$ with $O = 5 \vee 20$. In all such cases, the variance is smaller near the observational data. For low values of η and large number of observations O, the posterior distribution is less regular and the variance estimates get worse, even with 2^{19} Quasi-Monte Carlo samples, making the comparison with the UQ-VAE estimates unfair (as can be seen for $\eta = 0.005$ and $O = 40$). Indeed, as η decreases or O increases, the determinant of the noise covariance matrix Γ_{E} decreases. The reduction of the regularity of Γ_{E} corresponds to a lower regularity for $p_{\mathrm{U}|\mathrm{Y}}(\mathbf{u}|\mathbf{y})$ (2), due to its dependence on $\|\cdot\|_{\Gamma_{\mathrm{E}}^{-1}}$ and $|\Gamma_{\mathrm{E}}|^{-1/2}$. Therefore, (24) and (25) require an elevated number of samples to be estimated accurately, resulting in a high (and unaffordable) computational cost.

Finally, we test the UQ-VAE approach for different exponents ξ of the covariance function $\mathcal{C}_{\mathrm{pr}}(\mathbf{x}, \mathbf{z}) = \mathcal{A}^{-\xi}$ (Figure 7). We consider three different values of ξ in $(1, 2)$ and use the same neural network architecture for all cases. For $\xi = 3/2$, we obtain the results already analyzed in Figures 3 and 4. For $\xi = 7/4$, the posterior variance is estimated with similar accuracy as in the $\xi = 3/2$, while the errors in the estimates of the true parameters and posterior means are slightly higher. This increase in error is due to hyperparameters tuning being performed exclusively for the case $\xi = 3/2$. Nonetheless, the errors remain
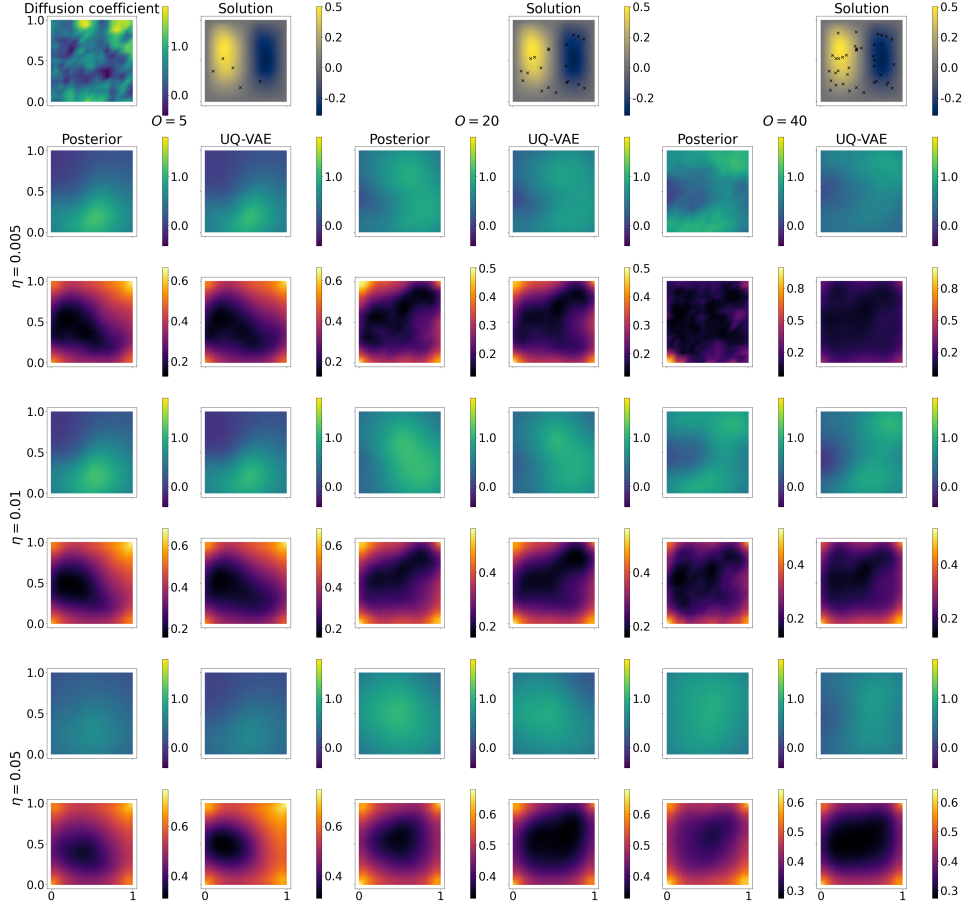
FIGURE 6. Posterior and UQ-VAE estimates for a test sample, with noise level η varying across rows and number of observation points O varying across columns. The first row shows the true diffusion coefficient and the corresponding solution with marked observation points. Estimated means are shown in the second, fourth and sixth rows, while the third, fifth, seventh rows display the corresponding estimated variances.

of the same order of magnitude. In the case $\xi = 5/4$, the average error on the true parameters is comparable to the other two cases, whereas the errors on the posterior means and variances are higher. As for a low noise level η and a high number of observations O, smaller values of ξ reduce the regularity of the prior distribution, which in turn decreases the regularity of the posterior distribution. As a result, (24) and (25) would require a large (and computationally infeasible) number of samples to be estimated accurately.

3.2. **Exponential Bayesian inverse problem.** We test the UQ-VAE approach on a non linear Bayesian inverse problem with forward map $\mathcal{F}(\mathbf{u}) = e^{\mathbf{u}}$, where the exponentiation is applied element wise. We compare the training time and accuracy of the UQ-VAE trained with loss functions L_α (7) and L_θ (15). To compute L_α, we use a Sobol' sequence with $K = 2^{12}$ samples. In this test, we compare the UQ-VAE estimates with the maximum
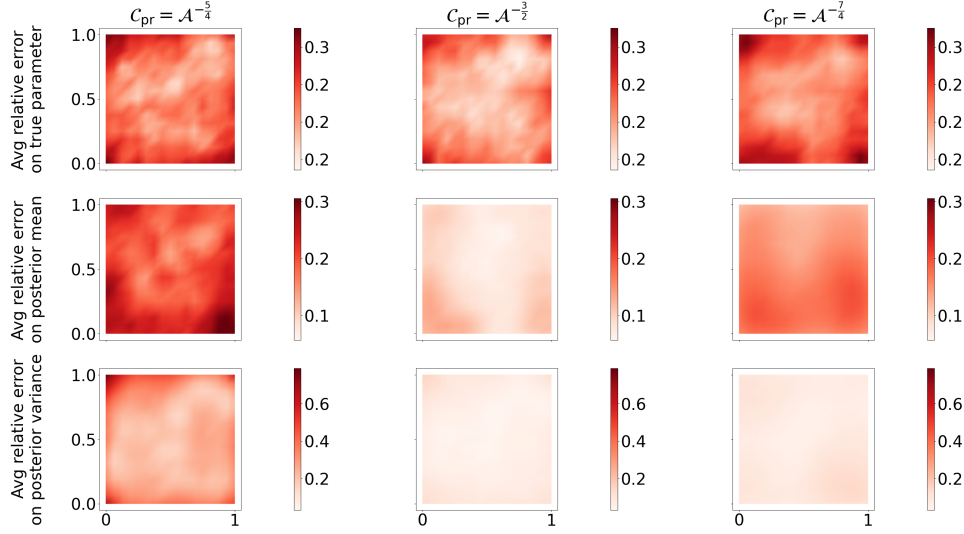
FIGURE 7. Average relative errors of UQ-VAE estimates on the true parameters and posterior means and variances on 100 test samples.
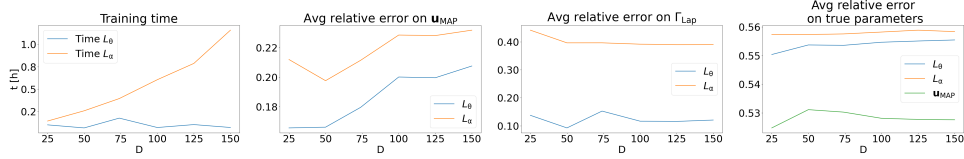

FIGURE 8. Performance comparison in terms of computational time and accuracy of UQ-VAEs (tested on 100 samples) trained with the loss function L_α and L_θ on the non linear Bayesian inverse problem with forward map $\mathcal{F}(\mathbf{u}) = e^{\mathbf{u}}$.

a posteriori parameter $\mathbf{u}_{\mathrm{MAP}}$ (4) and the Laplace approximation of the covariance matrix Γ_{Lap} (5) because the true posterior mean (24) and covariance (25) are expensive to estimate with an exponential map $\mathcal{F}$.

We set the dimensions of the parameter and observational data space equal ($D = O$) and vary them from 25 and 150. We set $(\mu_{\mathrm{pr}})_{\mathrm{i}} = -3$, for $i = 1, \ldots, D$, and $\Gamma_{\mathrm{pr}} = 4I$, where $I \in \mathbb{R}^{\mathrm{D} \times \mathrm{D}}$ is the identity matrix.

We generate a dataset $\{(\mathbf{u}^{(\mathrm{m})}, \mathbf{y}^{(\mathrm{m})})\}_{\mathrm{m}=1}^{\mathrm{M}}$, where $M \in \mathbb{N}$ and $\mathbf{u}^{(\mathrm{m})}$ is sampled from $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$. The noiseless observations are denoted by $\tilde{\mathbf{y}}^{(\mathrm{m})} = \mathcal{F}(\mathbf{u}^{(\mathrm{m})})$. The noisy observations are then defined as $\mathbf{y}^{(\mathrm{m})} = \tilde{\mathbf{y}}^{(\mathrm{m})} + \boldsymbol{\varepsilon}^{(\mathrm{m})}$, for $m = 1, \cdots, M$, where $\boldsymbol{\varepsilon}^{(\mathrm{m})}$ is sampled from $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{E}}, \Gamma_{\mathrm{E}})$ with $\boldsymbol{\mu}_{\mathrm{E}} = \mathbf{0}$ and Γ_{E} diagonal has entries $(\Gamma_{\mathrm{E}})_{\mathrm{i},\mathrm{i}} = (\eta \max_{m} |\tilde{y}_{\mathrm{i}}^{(\mathrm{m})}|)^2$, for $i = 1, \cdots, O$ and $\eta \in \mathbb{R}$. We fix $\eta = 0.01$.

We use the true map $\mathcal{F}$ as the decoder. The encoder consists of 3 hidden layers with 1000 neurons each and is trained on a dataset of 100 samples.

While the training time of the UQ-VAE using L_α loss increases with D and O, the training time with L_θ loss is nearly constant (Figure 8). This is because the computational cost of the training with L_θ is mainly driven by the network dimension. In contrast, with
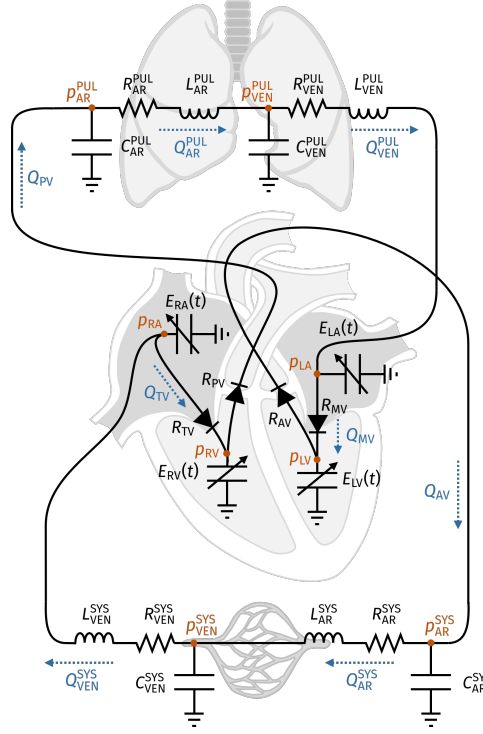
FIGURE 9. 0D cardiocirculatory model. We depict pressures and flow
rates in red and blue, respectively, and parameters in black.

L_α, the cost of sampling significantly contributes to the training time, even for small values
of D. The average errors (on a dataset of 100 samples) in the estimation of $\mathbf{u}_{\mathrm{MAP}}$, Γ_{Lap}
and true parameters follow a similar trend for both loss functions. Nonetheless, training
with L_θ yields lower errors across all metrics. The errors in estimating the true parameters
with both loss functions are slightly higher than those of $\mathbf{u}_{\mathrm{MAP}}$, but still comparable.

3.3. **0D cardiocirculatory model.** 0D (or lumped-parameter) cardiocirculatory models
represent the human cardiovascular system as an electrical circuit: the current corresponds
to the blood flow through vessels and valves, the electric potential represents the blood
pressure, the electric resistance plays the role of the resistance to blood flow, the capaci-
tance represents the vessel compliance and the inductance corresponds to the blood inertia.
In this context, Bayesian inverse problems are used to personalize the model for individual
patients and provide insights into their clinical condition, while accounting for the uncer-
tainty in the estimated parameters due to measurement errors (and poor observability of
the parameters themselves).

A 0D cardiocirculatory model partitions the cardiovascular system into distinct com-
partments (e.g. right atrium, systemic arteries/veins). The 0D cardiocirculatory model
used here is the one proposed in Ref. 28 (Figure 9). The model consists in the four cardiac
chambers, the systemic and pulmonary circulation, split into arterial and venous compart-
ments.
The 0D cardiocirculatory model depends on the heart rate (HR), which determines the
heartbeat period $T_{\mathrm{HB}} = 60/\mathrm{HR}$, and on the parameters $\breve{\mathbf{u}}$ listed in Table 3. A subset $\mathbf{u}$ of

	Parameter	Unit	Reference value	Description
	EA_{LA}	mmHg/mL	0.2273	Left atrial active elastance
	EB_{LA}	mmHg/mL	0.209	Left atrial passive elastance
	$V_{\mathrm{U,LA}}$	mL	2.0	Left atrial unstressed volume
	EA_{LV}	mmHg/mL	3.0391	Left ventricular active elastance
	EB_{LV}	mmHg/mL	0.10	Left ventricular passive elastance
	$V_{\mathrm{U,LV}}$	mL	16.0	Left ventricular unstressed volume
	EA_{RA}	mmHg/mL	0.0429	Right atrial active elastance
	EB_{RA}	mmHg/mL	0.0636	Right atrial passive elastance
	$V_{\mathrm{U,RA}}$	mL	2.0	Right atrial unstressed volume
	EA_{RV}	mmHg/mL	0.6683	Right ventricular active elastance
	EB_{RV}	mmHg/mL	0.07	Right ventricular passive elastance
	$V_{\mathrm{U,RV}}$	mL	16.0	Right ventricular unstressed volume
	$R_{\min}$	mmHg· s/mL	0.0075	Minimal valve resistance
GSA	$R_{\max}$	mmHg· s/mL	75006.2	Maximal valve resistance
	$R_{\mathrm{AR}}^{\mathrm{SYS}}$	mmHg· s/mL	0.588	Systemic arterial resistance
	$C_{\mathrm{AR}}^{\mathrm{SYS}}$	mL/mmHg	0.96	Systemic arterial compliance
	$L_{\mathrm{AR}}^{\mathrm{SYS}}$	mmHg· s^2/mL	0.005	Systemic arterial inertia
	$R_{\mathrm{VEN}}^{\mathrm{SYS}}$	mmHg· s/mL	0.352	Systemic venous resistance
	$C_{\mathrm{VEN}}^{\mathrm{SYS}}$	mL/mmHg	60.0	Systemic venous compliance
	$L_{\mathrm{VEN}}^{\mathrm{SYS}}$	mmHg· s^2/mL	0.0005	Systemic venous inertia
	$R_{\mathrm{AR}}^{\mathrm{PUL}}$	mmHg· s/mL	0.104	Pulmonary arterial resistance
	$C_{\mathrm{AR}}^{\mathrm{PUL}}$	mL/mmHg	5.0	Pulmonary arterial compliance
	$L_{\mathrm{AR}}^{\mathrm{PUL}}$	mmHg· s^2/mL	0.0005	Pulmonary arterial inertia
	$R_{\mathrm{VEN}}^{\mathrm{PUL}}$	mmHg· s/mL	0.0105	Pulmonary venous resistance
	$C_{\mathrm{VEN}}^{\mathrm{PUL}}$	mL/mmHg	16.0	Pulmonary venous compliance
	$L_{\mathrm{VEN}}^{\mathrm{PUL}}$	mmHg· s^2/mL	0.0005	Pulmonary venous inertia
	HR	s	75	Heart rate
	tC_{LA}	s	$0.79T_{\mathrm{HB}}$	Time of left atrial contraction
	TC_{LA}	s	$0.11T_{\mathrm{HB}}$	Duration of left atrial contraction
	tR_{LA}	s	$tC_{\mathrm{LA}} + TC_{\mathrm{LA}}$	Time of left atrial relaxation
	TR_{LA}	s	$0.8T_{\mathrm{HB}}$	Duration of left atrial relaxation
	tC_{LV}	s	0.0	Time of left ventricular contraction
	TC_{LV}	s	$0.35T_{\mathrm{HB}}$	Duration of left ventricular contraction
No GSA	tR_{LV}	s	$tC_{\mathrm{LV}} + TC_{\mathrm{LV}}$	Time of left ventricular relaxation
	TR_{LV}	s	$0.4T T_{\mathrm{HB}}$	Duration of left ventricular relaxation
	tC_{RA}	s	$0.8T_{\mathrm{HB}}$	Time of right atrial contraction
	TC_{RA}	s	$0.1T_{\mathrm{HB}}$	Duration of right atrial contraction
	tR_{RA}	s	$tC_{\mathrm{RA}} + TC_{\mathrm{RA}}$	Time of left atrial relaxation
	TR_{RA}	s	$0.7T_{\mathrm{HB}}$	Duration of left atrial relaxation
	tC_{RV}	s	0.0	Time of right ventricular contraction
	TC_{RV}	s	$0.3T_{\mathrm{HB}}$	Duration of left ventricular contraction
	tR_{RV}	s	$tC_{\mathrm{RV}} + TC_{\mathrm{RV}}$	Time of right ventricular relaxation
	TR_{RV}	s	$0.4T_{\mathrm{HB}}$	Duration of left ventricular relaxation

TABLE 3. List of parameters and their reference values for an ideal healthy individual

these parameters is used to define the model (1), depending on the specific application. The hemodynamics of the cardiovascular system are described by the following dynamical

system:

$$(28) \quad \begin{cases} \dot{V}_{\mathrm{LA}}(t) = Q_{\mathrm{VEN}}^{\mathrm{PUL}}(t) - Q_{\mathrm{MV}}(t), \\ \dot{V}_{\mathrm{LV}}(t) = Q_{\mathrm{MV}}(t) - Q_{\mathrm{AV}}(t), \\ C_{\mathrm{AR}}^{\mathrm{SYS}} \dot{p}_{\mathrm{AR}}^{\mathrm{SYS}}(t) = Q_{\mathrm{AV}}(t) - Q_{\mathrm{AR}}^{\mathrm{SYS}}(t), \\ L_{\mathrm{AR}}^{\mathrm{SYS}} \dot{Q}_{\mathrm{AR}}^{\mathrm{SYS}}(t) = -R_{\mathrm{AR}}^{\mathrm{SYS}} Q_{\mathrm{AR}}^{\mathrm{SYS}}(t) + p_{\mathrm{AR}}^{\mathrm{SYS}}(t) - p_{\mathrm{VEN}}^{\mathrm{SYS}}(t), \\ C_{\mathrm{VEN}}^{\mathrm{SYS}} \dot{p}_{\mathrm{VEN}}^{\mathrm{SYS}}(t) = Q_{\mathrm{AR}}^{\mathrm{SYS}}(t) - Q_{\mathrm{VEN}}^{\mathrm{SYS}}(t), \\ L_{\mathrm{VEN}}^{\mathrm{SYS}} \dot{Q}_{\mathrm{VEN}}^{\mathrm{SYS}}(t) = -R_{\mathrm{VEN}}^{\mathrm{SYS}} Q_{\mathrm{VEN}}^{\mathrm{SYS}}(t) + p_{\mathrm{VEN}}^{\mathrm{SYS}}(t) - p_{\mathrm{RA}}(t), \\ \dot{V}_{\mathrm{RA}}(t) = Q_{\mathrm{VEN}}^{\mathrm{SYS}}(t) - Q_{\mathrm{TV}}(t), \\ \dot{V}_{\mathrm{RV}}(t) = Q_{\mathrm{TV}}(t) - Q_{\mathrm{PV}}(t), \\ C_{\mathrm{AR}}^{\mathrm{PUL}} \dot{p}_{\mathrm{AR}}^{\mathrm{PUL}}(t) = Q_{\mathrm{PV}}(t) - Q_{\mathrm{AR}}^{\mathrm{PUL}}(t), \\ L_{\mathrm{AR}}^{\mathrm{PUL}} \dot{Q}_{\mathrm{AR}}^{\mathrm{PUL}}(t) = -R_{\mathrm{AR}}^{\mathrm{PUL}} Q_{\mathrm{AR}}^{\mathrm{PUL}}(t) + p_{\mathrm{AR}}^{\mathrm{PUL}}(t) - p_{\mathrm{VEN}}^{\mathrm{PUL}}(t), \\ C_{\mathrm{VEN}}^{\mathrm{PUL}} \dot{p}_{\mathrm{VEN}}^{\mathrm{PUL}}(t) = Q_{\mathrm{AR}}^{\mathrm{PUL}}(t) - Q_{\mathrm{VEN}}^{\mathrm{PUL}}(t), \\ L_{\mathrm{VEN}}^{\mathrm{PUL}} \dot{Q}_{\mathrm{VEN}}^{\mathrm{PUL}}(t) = -R_{\mathrm{VEN}}^{\mathrm{PUL}} Q_{\mathrm{VEN}}^{\mathrm{PUL}}(t) + p_{\mathrm{VEN}}^{\mathrm{PUL}}(t) - p_{\mathrm{LA}}(t), \end{cases}$$

coupled with suitable initial conditions $\mathbf{x}_0$ and with $t \in (0, T]$. The model unknowns are the volumes of the left atrium (V_{LA}) and ventricle (V_{LV}) and of the right atrium (V_{RA}) and ventricle (V_{RV}), the circulatory pressures of the systemic arteries ($p_{\mathrm{AR}}^{\mathrm{SYS}}$) and veins ($p_{\mathrm{VEN}}^{\mathrm{SYS}}$) and of the pulmonary arteries ($p_{\mathrm{AR}}^{\mathrm{PUL}}$) and veins ($p_{\mathrm{VEN}}^{\mathrm{PUL}}$), the circulatory fluxes of the systemic arteries ($Q_{\mathrm{AR}}^{\mathrm{SYS}}$) and veins ($Q_{\mathrm{VEN}}^{\mathrm{SYS}}$) and of the pulmonary arteries ($Q_{\mathrm{AR}}^{\mathrm{PUL}}$) and veins ($Q_{\mathrm{VEN}}^{\mathrm{PUL}}$).

The fluxes through the valves (tricuspid, pulmonary, mitral and aortic) depend on the pressure jump between the upstream and downstream compartments and the resistance through the valves is given by the non linear function R_{valve} as follows:

$$Q_{\mathrm{TV}}(t) = Q_{\mathrm{valve}}(p_{\mathrm{RA}}(t) - p_{\mathrm{RV}}(t)), \quad Q_{\mathrm{MV}}(t) = Q_{\mathrm{valve}}(p_{\mathrm{LA}}(t) - p_{\mathrm{LV}}(t)),$$
$$Q_{\mathrm{PV}}(t) = Q_{\mathrm{valve}}(p_{\mathrm{RV}}(t) - p_{\mathrm{AR}}^{\mathrm{PUL}}(t)), \quad Q_{\mathrm{AV}}(t) = Q_{\mathrm{valve}}(p_{\mathrm{LV}}(t) - p_{\mathrm{AR}}^{\mathrm{SYS}}(t)),$$

$$Q_{\mathrm{valve}}(\Delta p) = \frac{\Delta p}{R_{\mathrm{valve}}(\Delta p)}, \quad R_{\mathrm{valve}}(\Delta p) = \sqrt{R_{\max} R_{\min}} \left(\frac{R_{\max}}{R_{\min}} \right)^{\frac{\arctan(-100\pi\Delta p)}{\pi}}.$$

Each cardiac chamber is described as a pressure generator and depends on the periodic function $e_{\mathrm{c}}(t)$ that models the periodicity of each heartbeat.

$$p_{\mathrm{c}}(t) = E_{\mathrm{c}}(t)(V_{\mathrm{c}}(t) - V_{\mathrm{U,c}}),$$
$$E_{\mathrm{c}}(t) = EB_{\mathrm{c}} + EA_{\mathrm{c}} e_{\mathrm{c}}(t),$$

$$e_{\mathrm{c}}(t) = \begin{cases} \frac{1}{2} \left[1 - \cos\left(\frac{\pi}{TC_{\mathrm{c}}} \mathrm{mod}\,(t - tC_{\mathrm{c}}, T_{\mathrm{HB}}) \right) \right] & \text{if } \mathrm{mod}\,(t - tC_{\mathrm{c}}, T_{\mathrm{HB}}) < TC_{\mathrm{c}}, \\ \frac{1}{2} \left[1 + \cos\left(\frac{\pi}{TR_{\mathrm{c}}} \mathrm{mod}\,(t - tR_{\mathrm{c}}, T_{\mathrm{HB}}) \right) \right] & \text{if } \mathrm{mod}\,(t - tR_{\mathrm{c}}, T_{\mathrm{HB}}) < TR_{\mathrm{c}}, \\ 0 & \text{otherwise}, \end{cases}$$

for $c \in \{LA, LV, RA, RV\}$.

We solve the discrete dynamical system for 25 heartbeats ($T = 25 T_{\mathrm{HB}}$) to approach its periodic state and focus on the last heartbeat. We compute the model outputs $\check{\mathbf{y}}$ as functions

Model output	Unit	Range	Model value	Description
$LA_{\text{I}-\text{Vmax}}$	mL/m^2	$[16,34]^{20}$	33.1	Indexed maximal left atrial volume
LA_{Pmax}	mmHg	$[6,20]^{16}$	12.0	Maximal left atrial pressure
LA_{Pmin}	mmHg	$[-2,9]^{16}$	8.1	Minimal left atrial pressure
LA_{Pmean}	mmHg	$[4,12]^{16}$	10.9	Mean left atrial pressure
LV_{SV}	mL	$[30,80]^{20}$	67.6	Left ventricular stroke volume
CI	L/min/m^2	$[2.8,4.2]^{16}$	2.8	Cardiac index
$LV_{\text{I}-\text{EDV}}$	mL/m^2	$[50,90]^{16}$	66.0	Indexed left ventricular end diastolic volume
LV_{ESV}	mL	$[18,52]^{20}$	50.6	Left ventricular end systolic volume
LV_{EF}	%	$[53,73]^{20}$	57.2	Left ventricular ejection fraction
LV_{Pmax}	mmHg	$[90,140]^{16}$	113.2	Maximal left ventricular pressure
LV_{Pmin}	mmHg	$[4,12]^{16}$	5.6	Minimal left ventricular pressure
RA_{Pmax}	mmHg	$[2,14]^{16}$	8.4	Maximal right atrial pressure
RA_{Pmin}	mmHg	$[-2,6]^{16}$	5.7	Minimal right atrial pressure
RA_{Pmean}	mmHg	$[-1,8]^{16}$	7.2	Mean right atrial pressure
$RV_{\text{I}-\text{EDV}}$	mL/m^2	$[44,80]^{29}$	65.8	Indexed right ventricular end diastolic volume
$RV_{\text{I}-\text{ESV}}$	mL/m^2	$[19,46]^{29}$	28.1	Indexed right ventricular end systolic volume
RV_{EF}	%	$[44,71]^{29}$	57.3	Right ventricular ejection fraction
RV_{Pmax}	mmHg	$[15,28]^{16}$	27.4	Maximal right ventricular pressure
RV_{Pmin}	mmHg	$[0,8]^{16}$	3.9	Minimal right ventricular pressure
SAP_{max}	mmHg	$[-,140]^{20}$	112.2	Systolic systemic arterial pressure
SAP_{min}	mmHg	$[-,80]^{20}$	62.0	Diastolic systemic arterial pressure
PAP_{max}	mmHg	$[15,28]^{16}$	25.8	Systolic pulmonary arterial pressure
PAP_{min}	mmHg	$[5,16]^{16}$	16.0	Diastolic pulmonary arterial pressure
PAP_{mean}	mmHg	$[10,22]^{16}$	20.6	Mean pulmonary arterial pressure
PWP_{min}	mmHg	$[1,12]^{16}$	11.2	Minimal pulmonary wedge pressure
PWP_{mean}	mmHg	$[6,15]^{16}$	11.8	Mean pulmonary wedge pressure
SVR	mmHg $\cdot$ min/L	$[11.3,17.5]^{16}$	15.7	Systemic vascular resistance
PVR	mmHg $\cdot$ min/L	$[1.9,3.1]^{16}$	1.9	Pulmonary vascular resistance

TABLE 4. List of model outputs, their units of measure, the echocardiographic ranges for a healthy individual and the values returned by the 0D cardiocirculatory model with the reference setting of parameters.

(e.g. the maximum or the mean) of the unknowns $\mathbf{x}$ and the parameters $\check{\mathbf{u}}$:

$$\check{\mathbf{y}} = \mathbf{g}(\{\mathbf{x}(t; \mathbf{x}_0, \check{\mathbf{u}}), t \in [24T_{\text{HB}}, 25T_{\text{HB}}]\}; \check{\mathbf{u}}).$$

The model outputs are defined in Table 4, where a subscript "I-" refers to the indexed volumes, i.e., normalized by the body surface area (BSA). In general, a subset $\mathbf{y}$ of the model outputs is used as observational data in model (1), depending on the specific application.

We determine the reference setting of parameters $\check{\mathbf{u}}^{\text{R}}$ to represent an ideal healthy individual (Table 3). Specifically, we set HR $= 75$ bpm[49] and adjust the values of the other parameters based on modifications of literature values,[1,10,43] ensuring that the model outputs $\check{\mathbf{y}}$ fall within echocardiograpic ranges characteristic of a healthy individual (Table 4).

We perform a global sensitivity analysis (GSA) on the $\check{\mathbf{u}}$ parameters of the 0D cardiocirculatory model indicated in Table 3 to identify those that most influence a set of observational data $\mathbf{y}$. This allows to reduce the dimensionality of the parameter space by focusing on the most significant parameters. We compute total Sobol' indices[34] to evaluate the impact of each model parameter on the variance of each observational data y_{j}, for $j = 1, \ldots, O$, accounting for both first-order and higher-order interactions among parameters:

$$S_{\text{i}}^{\text{j,T}} = 1 - \frac{\text{Var}_{\check{p}(\check{\mathbf{u}}_{\sim\text{i}})}[\mathbb{E}_{\check{p}(\check{u}_{\text{i}})}[y_{\text{j}}|\check{\mathbf{u}}_{\sim\text{i}}]]}{\text{Var}_{\check{p}(\check{\mathbf{u}})}[y_{\text{j}}]},$$

where $\check{p}$ is the sampling pdf of the parameters and $\check{\mathbf{u}}_{\sim\text{i}}$ represents the set of all parameters except for $\check{u}_{\text{i}}$. We estimate the total Sobol' indices using Saltelli's method,[30] allowing the

Parameter	EA_{LV}	$V_{\mathrm{U,LV}}$	R_{AR}^{SYS}	C_{AR}^{SYS}	R_{VEN}^{SYS}	C_{VEN}^{SYS}	R_{AR}^{PUL}	C_{AR}^{PUL}	HR
Modification	$+40\%$	-10%	$+50\%$	-40%	$+5\%$	-5%	$+10\%$	-10%	$+10\%$

TABLE 5. List of further modifications in the parameters ranges for the GSA to model hypertension. The modifications affect the lower bound (-) or the upper bound (+) of the ranges.

parameters $\check{\mathbf{u}}$ to vary in a hypercube. The number of samples for the Saltelli's method increases linearly with the number of parameters N_{u} and a user-defined value $N \in \mathbb{N}$ as $2N(N_{\mathrm{u}} + 1)$. We set $N = 2^{12}$. Since the sensitivity analysis depends on the parameter ranges, we will specify them for the specific application (sections 3.3.1 and 3.3.2). We will solve Bayesian inverse problems for the parameters $\mathbf{u} \in \mathbb{R}^{\mathrm{D}}$ associated with at least one Sobol index greater than 0.1. All the parameters not selected will be kept fixed at their reference value.

Remark. The heart rate (HR) is usually measured in clinical practice. In our framework there are two possibilities to treat HR: ignoring the data and estimating it by means of the UQ-VAE; developing one decoder that can handle different values for HR. Even if the second approach reduces the dimension of the parameter space by one, it would require the training of an encoder for each value of HR because the map $\mathcal{F}$ of the mathematical model (1) would change with HR. Therefore, we restrain ourselves to the first approach.

3.3.1. *Systemic hypertension test case.* We solve the Bayesian inverse problem for the 0D cardiocirculatory model in the context of systemic hypertension.[27] The chosen observational data are:

$$\mathbf{y} = [CI, LV_{\mathrm{I-ESV}}, LV_{\mathrm{EF}}, LV_{\mathrm{Pmax}}, SAP_{\mathrm{max}}, SAP_{\mathrm{min}}].$$

These observational data are directly affected by pressure increases in the systemic circulation.

The hypercube used for the sensitivity analysis is built around the reference parameter setting $\check{\mathbf{u}}^{\mathrm{R}}$, with variations of $\pm 25\%$ relative to $\check{\mathbf{u}}^{\mathrm{R}}$ to account for intervariability between individuals. Additionally, the parameters listed in Table 5 allow for further adjustments to the lower and upper bounds of the ranges to model systemic hypertension. Based on the total Sobol' indices (Figure 10), the selected parameters are:

$$\mathbf{u} = [EA_{\mathrm{LV}}, R_{\mathrm{AR}}^{SYS}, C_{\mathrm{AR}}^{SYS}, R_{\mathrm{VEN}}^{SYS}, HR].$$

Since all entries of the parameter vector $\mathbf{u}$ of the 0D cardiocirculatory model are positive, we work with their logarithms to formulate an unconstrained inverse problem. We assume the parameter vector to be distributed as a lognormal random variable $\log U \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$. We set the prior mean as $\boldsymbol{\mu}_{\mathrm{pr}} = \log \mathbf{u}^{\mathrm{R}}$, where $\mathbf{u}^{\mathrm{R}}$ is the vector of reference values for $\mathbf{u}$ and $\log$ is applied element-wise. We let Γ_{pr} to be a diagonal matrix. During the sensitivity analysis, for $i = 1, \ldots, D$, each entry u_{i} of the parameter vector varies in the interval $[a_{\mathrm{i}} u_{\mathrm{i}}^{\mathrm{R}}, b_{\mathrm{i}} u_{\mathrm{i}}^{\mathrm{R}}]$ with $a_{\mathrm{i}}, b_{\mathrm{i}} \in \mathbb{R}_{+}$, $a_{\mathrm{i}} < b_{\mathrm{i}}$, thus $\log u_{\mathrm{i}} \in [\log a_{\mathrm{i}} + \log u_{\mathrm{i}}^{\mathrm{R}}, \log b_{\mathrm{i}} + \log u_{\mathrm{i}}^{\mathrm{R}}]$. We set $(\Gamma_{\mathrm{pr}})_{\mathrm{i,i}} = (\log b_{\mathrm{i}} - \log a_{\mathrm{i}})^2 / 12$.

We generate a dataset $\{(\mathbf{u}^{(\mathrm{m})}, \mathbf{y}^{(\mathrm{m})})\}_{\mathrm{m}=1}^{\mathrm{M}}$, where $M \in \mathbb{N}$ and $\mathbf{u}^{(\mathrm{m})}$ is sampled from $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$. Let $\tilde{\mathbf{y}}^{(\mathrm{m})}$ be the noiseless observations of the solution of (28) with parameter $\mathbf{u}^{(\mathrm{m})}$. Then, $\mathbf{y}^{(\mathrm{m})} = \tilde{\mathbf{y}}^{(\mathrm{m})} + \boldsymbol{\varepsilon}^{(\mathrm{m})}$, for $m = 1, \cdots, M$, where $\boldsymbol{\varepsilon}^{(\mathrm{m})}$ is sampled from $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{E}}, \Gamma_{\mathrm{E}})$ with $\boldsymbol{\mu}_{\mathrm{E}} = \mathbf{0}$ and Γ_{E} diagonal with $(\Gamma_{\mathrm{E}})_{\mathrm{i,i}} = (\eta \max_{m} |\tilde{y}_{\mathrm{i}}^{(\mathrm{m})}|)^2$, for $i = 1, \cdots, O$ and $\eta \in \mathbb{R}$. Observe that differently from section 3.1, $(\Gamma_{\mathrm{E}})_{\mathrm{i,i}}$ depends on

Sobol indices (S_T)

	CI	$LV_{L\text{-}ESV}$	LV_{EF}	LV_{Pmax}	SAP_{max}	SAP_{min}
EA_{LA}	0.00	0.00	0.00	0.00	0.00	0.00
EB_{LA}	0.00	0.00	0.00	0.00	0.00	0.00
$V0_{LA}$	0.00	0.00	0.00	0.00	0.00	0.00
EA_{LV}	0.01	0.66	0.44	0.01	0.01	0.00
EB_{LV}	0.01	0.00	0.00	0.01	0.01	0.00
$V0_{LV}$	0.00	0.07	0.05	0.00	0.00	0.00
EA_{RA}	0.00	0.00	0.00	0.00	0.00	0.00
EB_{RA}	0.00	0.00	0.00	0.00	0.00	0.00
$V0_{RA}$	0.00	0.00	0.00	0.00	0.00	0.00
EA_{RV}	0.01	0.00	0.00	0.00	0.00	0.00
EB_{RV}	0.03	0.00	0.00	0.01	0.01	0.00
$V0_{RV}$	0.00	0.00	0.00	0.00	0.00	0.00
R_{min}	0.00	0.00	0.00	0.00	0.00	0.00
R_{max}	0.00	0.00	0.00	0.00	0.00	0.00
R_{AR}^{SYS}	0.01	0.20	0.15	0.55	0.56	0.69
C_{AR}^{SYS}	0.00	0.03	0.02	0.20	0.20	0.16
L_{AR}^{SYS}	0.00	0.00	0.00	0.00	0.00	0.00
R_{VEN}^{SYS}	0.79	0.05	0.05	0.21	0.20	0.04
C_{VEN}^{SYS}	0.02	0.00	0.00	0.01	0.01	0.01
L_{VEN}^{SYS}	0.00	0.00	0.00	0.00	0.00	0.00
R_{AR}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00
C_{AR}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00
L_{AR}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00
R_{VEN}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00
C_{VEN}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00
L_{VEN}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00
HR	0.11	0.00	0.29	0.00	0.00	0.12

FIGURE 10. Total Sobol' indices for 0D cardiocirculatory model in the hypertension case.

the specific observational data because they are associated with different quantities. We fix $\eta = 0.05$.

We use a decoder with 5 hidden layers and 250 neurons per layer trained on a dataset of $M = 4096$ samples. The encoder consists of 3 hidden layers and 250 neurons per layer and is trained on a subset of 100 samples of the decoder dataset.

We generate 2^{17} samples from the distribution $\mathcal{N}(\boldsymbol{\mu}_{\text{pr}}, \Gamma_{\text{pr}})$ to estimate (24) and (25).

The UQ-VAE mean $\boldsymbol{\mu}_{\text{post}}^{(m)}$ estimates the posterior mean $\mathbb{E}_{P_{U|Y}(\mathbf{u}|\mathbf{y}^{(m)})}[\mathbf{u}]$ with an average relative error less than 2.0% on 100 test samples (Figure 11). The UQ-VAE covariance $\Gamma_{\text{post}}^{(m)}$ estimates the posterior covariance $\text{Cov}_{P_{U|Y}(\mathbf{u}|\mathbf{y}^{(m)})}(\mathbf{u})$ with an average relative error less than 0.5%. Both the posterior and the UQ-VAE means achieve similar average relative errors on the true parameters $\mathbf{u}^{(m)}$.

The chosen observational data $\mathbf{y}$ are related to the practical case of systemic hypertension. We now test the UQ-VAE using 25 equally spaced observations of all the 12 unknowns of (28) in $[24T_{\text{HB}}, 25T_{\text{HB}}]$ ($O = 300$). We perform a new sensitivity analysis to account for the different observational data and we select the parameters:

$$\mathbf{u} = [EA_{\text{LA}}, EB_{\text{LA}}, EA_{\text{LV}}, EB_{\text{LV}}, EB_{\text{RA}}, EA_{\text{RV}}, EB_{\text{RV}}, V_{\text{U,RV}},$$
$$R_{\text{AR}}^{\text{SYS}}, C_{\text{AR}}^{\text{SYS}}, R_{\text{VEN}}^{\text{SYS}}, C_{\text{VEN}}^{\text{SYS}}, R_{\text{AR}}^{\text{PUL}}, C_{\text{AR}}^{\text{PUL}}, L_{\text{AR}}^{\text{PUL}}, HR].$$

The definition of the prior distribution and of the dataset are analogous to the previous case. The noise distribution has mean $\boldsymbol{\mu}_{\text{E}} = \mathbf{0}$ and diagonal covariance matrix Γ_{E}. Γ_{E} is such that all the observations of the same unknown of (28) are related to the same variance obtained by taking the maximum of the absolute values of all the corresponding samples
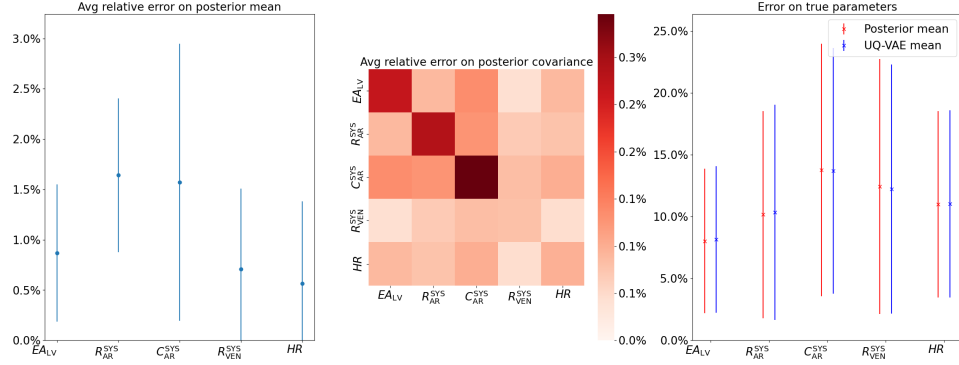
FIGURE 11. From left to right: average and standard deviation of the relative errors of the UQ-VAE mean $\boldsymbol{\mu}_{\text{post}}^{(\text{m})}$ on the posterior one $\mathbb{E}_{p_{U|Y}(\mathbf{u}|\mathbf{y}^{(\text{m})})}[\mathbf{u}]$; average relative error of the UQ-VAE covariance $\Gamma_{\text{post}}^{(\text{m})}$ on the posterior one $\text{Var}_{p_{U|Y}(\mathbf{u}|\mathbf{y}^{(\text{m})})}[\mathbf{u}]$; average and standard deviations of the relative errors of the posterior and UQ-VAE means on the true parameters $\mathbf{u}^{(\text{m})}$. The test are run for 100 samples for the systemic hypertension case and the relative errors and standard deviations are computed element-wise.

and observations:

$$(\Gamma_{\text{E}})_{i,i} = \left(\eta \max_{\substack{m \\ 1+25\text{mod}(i-1,25) \leq k \leq 25(\text{mod}(i-1,25)+1)}} |\tilde{y}_{\text{k}}^{(\text{m})}| \right)^2 .$$

We generate 2^{19} samples from the distribution $\mathcal{N}(\boldsymbol{\mu}_{\text{pr}}, \Gamma_{\text{pr}})$ to estimate (24) and (25).

In this case, the UQ-VAE mean is more accurate than the posterior mean in estimating the true parameters (Figure 12). As in Figure 6, this is because the increase in the number of observational data corresponds to a decrease in the regularity of $p_{U|Y}(\mathbf{u}|\mathbf{y})$ and a poorer estimate of (24) and (25). Consequently, also the estimates of the posterior mean and covariance matrix are worse than Figure 11 with average relative errors less than 15% and 1.6%, respectively. Moreover, the accuracy of the UQ-VAE on the true parameters is unaffected by the higher number of parameters to estimate with respect to the previous case.

3.3.2. *Ventricular septal defect test case.* We solve the Bayesian inverse problem for the 0D cardiocirculatory model in the context of ventricular septal defect.[7,36] This congenital defect is characterized by a hole in the septum (wall) between the two ventricles. To account for the blood flow through the ventricular septal defect, the 0D cardiocirculatory model requires a modification in the differential equations for the volumes of the left and right ventricles:

(29)
$$\begin{cases} \dot{V}_{\text{LV}}(t) = Q_{\text{MV}}(t) - Q_{\text{AV}}(t) - Q_{\text{VSD}}(t), \\ \dot{V}_{\text{RV}}(t) = Q_{\text{TV}}(t) - Q_{\text{PV}}(t) + Q_{\text{VSD}}(t), \end{cases}$$
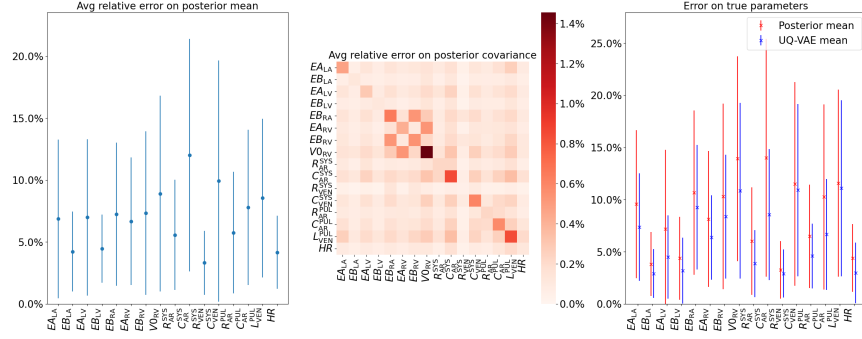
FIGURE 12. From left to right: average and standard deviation of the relative errors of the UQ-VAE mean $\boldsymbol{\mu}_{\text{post}}^{(m)}$ on the posterior one $\mathbb{E}_{p_{U|Y}(\mathbf{u}|\mathbf{y}^{(m)})}[\mathbf{u}]$; average relative error of the UQ-VAE covariance $\Gamma_{\text{post}}^{(m)}$ on the posterior one $\text{Var}_{p_{U|Y}(\mathbf{u}|\mathbf{y}^{(m)})}[\mathbf{u}]$; average and standard deviations of the relative errors of the posterior and UQ-VAE means on the true parameters $\mathbf{u}^{(m)}$. The test are run for 100 samples for the systemic hypertension case and the relative errors and standard deviations are computed element-wise. The data are 25 equally spaced observations of all the 12 unknowns of (28).

where:

$$Q_{\text{VSD}}(t) = \frac{p_{\text{LV}}(t) - p_{\text{RV}}(t)}{R_{\text{VSD}}},$$

with R_{VSD} representing the resistance to the blood flow through the ventricular septal defect. We use as parameter the radius of the ventricular septal defect r_{VSD} with a reference value of $0.6 r_{\text{AV}} = 0.9$ cm, where $r_{\text{AV}} = 1.5$ cm is the aortic annulus radius.[7] We compute R_{VSD} by assuming that the length of the ventricular septal defect and the thickness of the aortic annulus are equal to L and applying two times the Poiseuille law:

$$R_{\text{VSD}} = \frac{8\mu L}{\pi r_{\text{VSD}}^4} = R_{\text{min}} \left(\frac{r_{\text{AV}}}{r_{\text{VSD}}} \right)^4,$$

where μ denotes the blood viscosity.

The observational data are chosen as follows:

$$\begin{aligned} \mathbf{y} = [&LV_{\text{Pmax}}, LV_{\text{Pmin}}, RA_{\text{Pmean}}, RV_{\text{Pmax}}, RV_{\text{Pmin}}, SAP_{\text{max}}, SAP_{\text{min}}, \\ &PAP_{\text{max}}, PAP_{\text{min}}, PAP_{\text{mean}}, PWP_{\text{mean}}, PVR, Q_{\text{P}}, Q_{\text{S}}], \end{aligned}$$

where Q_{P} and Q_{S} are the mean flows through the pulmonary and aortic valves. These two observational data are particularly important in the case of ventricular septal defects because their ratio $Q_{\text{P}}/Q_{\text{S}}$ is typically around 1 in healthy individuals, indicating that the same amount of blood flows through the pulmonary and systemic circulation. In the case of ventricular septal defect, this ratio increases beyond one, meaning that some of the oxygenated blood from the lungs flows to the right ventricle, bypassing the systemic circulation and failing to supply oxygen to the body's cells.

The hypercube used for the sensitivity analysis is built around the reference setting of parameters $\check{\mathbf{u}}^{\text{R}}$, accounting for intervariability between individuals by allowing a variation

Sobol indices (S_T)

	LV_{pmax}	LV_{pmin}	RA_{pmean}	RV_{pmax}	RV_{pmin}	SAP_{max}	SAP_{min}	PAP_{max}	PAP_{min}	PAP_{mean}	PWP_{mean}	PVR	Qp	Qs
EA_{LA}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB_{LA}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
$V0_{LA}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EA_{LV}	0.01	0.03	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB_{LV}	0.07	0.45	0.02	0.00	0.01	0.07	0.02	0.01	0.03	0.01	0.15	0.00	0.02	0.04
$V0_{LV}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EA_{RA}	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB_{RA}	0.00	0.00	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$V0_{RA}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EA_{RV}	0.02	0.00	0.06	0.01	0.16	0.02	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.02
EB_{RV}	0.01	0.00	0.25	0.04	0.27	0.01	0.00	0.04	0.02	0.04	0.02	0.00	0.05	0.02
$V0_{RV}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R_{min}	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
R_{max}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R_{AR}^{SYS}	0.10	0.01	0.00	0.03	0.01	0.10	0.31	0.03	0.03	0.03	0.03	0.00	0.03	0.03
C_{AR}^{SYS}	0.08	0.01	0.00	0.02	0.00	0.08	0.15	0.02	0.01	0.02	0.02	0.00	0.01	0.01
L_{AR}^{SYS}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R_{VEN}^{SYS}	0.15	0.06	0.17	0.11	0.08	0.15	0.01	0.11	0.11	0.13	0.14	0.00	0.09	0.27
C_{VEN}^{SYS}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L_{VEN}^{SYS}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R_{AR}^{PUL}	0.00	0.01	0.00	0.03	0.01	0.00	0.00	0.04	0.20	0.12	0.04	0.99	0.06	0.00
C_{AR}^{PUL}	0.00	0.00	0.01	0.03	0.01	0.00	0.00	0.05	0.10	0.00	0.00	0.00	0.00	0.00
L_{AR}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R_{VEN}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00
C_{VEN}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L_{VEN}^{PUL}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
r_{VSD}	0.57	0.10	0.25	0.71	0.34	0.57	0.22	0.67	0.46	0.64	0.45	0.00	0.66	0.45
HR	0.01	0.31	0.21	0.02	0.11	0.01	0.30	0.03	0.04	0.01	0.14	0.00	0.07	0.16

FIGURE 13. Total Sobol' indices for 0D cardiocirculatory model in the ventricular septal defect case.

of $\pm 25\%$ with respect to $\check{\mathbf{u}}^R$. Based on the sensitivity analysis results (Figure 13), the selected parameters are:

$$\mathbf{u} = [EB_{LV}, EA_{RV}, EB_{RV}, R_{AR}^{SYS}, C_{AR}^{SYS}, R_{VEN}^{SYS}, R_{AR}^{PUL}, r_{VSD}, HR].$$

Since all entries of the parameter vector $\mathbf{u}$ of the 0D cardiocirculatory model are positive, we work with their logarithms to formulate an unconstrained inverse problem. We assume that the parameter vector follows a lognormal distribution $\log U \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$. We set the prior mean as $\boldsymbol{\mu}_{\mathrm{pr}} = \log \mathbf{u}^R$, where $\mathbf{u}^R$ is the vector of reference values for $\mathbf{u}$ and $\log$ is applied element-wise. We let Γ_{pr} to be a diagonal matrix. During the sensitivity analysis, for $i = 1, \ldots, D$, each entry u_i of the parameter vector varies within the range $[0.75 u_i^R, 1.25 u_i^R]$, so that $\log u_i \in [\log 0.75 + \log u_i^R, \log 1.25 + \log u_i^R]$. We set $(\Gamma_{\mathrm{pr}})_{i,i} = (\log 1.25 - \log 0.75)^2/12$.

We generate a dataset $\{(\mathbf{u}^{(m)}, \mathbf{y}^{(m)})\}_{m=1}^M$, where $M \in \mathbb{N}$ and $\mathbf{u}^{(m)}$ is sampled from $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$. Let $\tilde{\mathbf{y}}^{(m)}$ be the noiseless observations of the solution of (28) modified with (29) with parameter $\mathbf{u}^{(m)}$. Then, $\mathbf{y}^{(m)} = \tilde{\mathbf{y}}^{(m)} + \boldsymbol{\varepsilon}^{(m)}$, for $m = 1, \cdots, M$, where $\boldsymbol{\varepsilon}^{(m)}$ is sampled from $\mathcal{N}(\boldsymbol{\mu}_E, \Gamma_E)$ with $\boldsymbol{\mu}_E = \mathbf{0}$ and Γ_E diagonal with $(\Gamma_E)_{i,i} = (\eta \max_m |\tilde{y}_i^{(m)}|)^2$, for $i = 1, \cdots, O$ and $\eta \in \mathbb{R}$. We fix $\eta = 0.05$.

We use a decoder with 5 hidden layers and 250 neurons per layer trained on a dataset of $M = 4096$ samples. The encoder consists of 3 hidden layers and 250 neurons per layer and is trained on a subset of 100 samples of the decoder dataset.

We generate 2^{17} samples from the distribution $\mathcal{N}(\boldsymbol{\mu}_{\mathrm{pr}}, \Gamma_{\mathrm{pr}})$ to estimate (24) and (25).

The UQ-VAE mean $\boldsymbol{\mu}_{\mathrm{post}}^{(m)}$ estimates the posterior mean $\mathbb{E}_{\mathrm{P}_{U|Y}(\mathbf{u}|\mathbf{y}^{(m)})}[\mathbf{u}]$ with an average relative error less than 2.0% on 100 test samples (Figure 14). The UQ-VAE covariance $\Gamma_{\mathrm{post}}^{(m)}$ estimates the posterior covariance $\mathrm{Cov}_{\mathrm{P}_{U|Y}(\mathbf{u}|\mathbf{y}^{(m)})}(\mathbf{u})$ with an average relative error
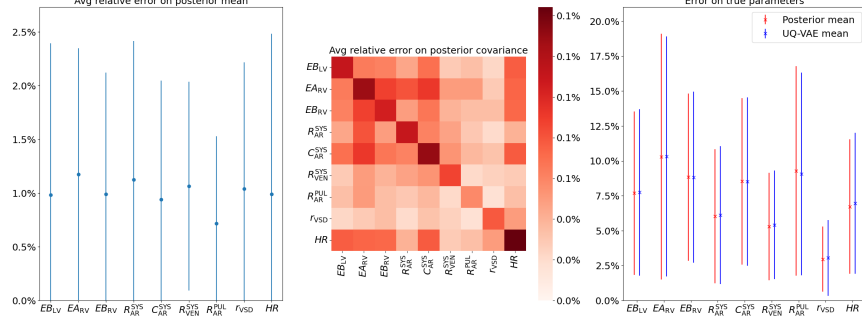
FIGURE 14. From left to right: average and standard deviation of the relative errors of the UQ-VAE mean $\boldsymbol{\mu}_{\text{post}}^{(\text{m})}$ on the posterior one $\mathbb{E}_{\mathrm{p}_{\mathrm{U}|\mathrm{Y}}(\mathbf{u}|\mathbf{y}^{(\text{m})})}[\mathbf{u}]$; average relative error of the UQ-VAE covariance $\Gamma_{\text{post}}^{(\text{m})}$ on the posterior one $\text{Var}_{\mathrm{p}_{\mathrm{U}|\mathrm{Y}}(\mathbf{u}|\mathbf{y}^{(\text{m})})}[\mathbf{u}]$; average and standard deviations of the relative errors of the posterior and UQ-VAE means on the true parameters $\mathbf{u}^{(\text{m})}$. The test are run for 100 samples for the ventricular septal defect case and the relative errors and standard deviations are computed element-wise.

	EB_{LV}	EA_{RV}	EB_{RV}	$R_{\text{AR}}^{\text{SYS}}$	$C_{\text{AR}}^{\text{SYS}}$	$R_{\text{VEN}}^{\text{SYS}}$	$R_{\text{AR}}^{\text{PUL}}$	r_{VSD}	HR
$\mathbf{u}$	0.11	0.70	0.087	0.66	1.1	0.32	0.10	0.79	82
$\boldsymbol{\mu}_{\text{post}}$	0.09	0.64	0.074	0.62	1.0	0.32	0.09	0.82	82
Relative error	0.14	0.08	0.15	0.07	0.06	0.01	0.10	0.04	0.01

TABLE 6. True and estimated parameter, along with their relative error, for a test sample of the 0D cardiocirculatory model with VSD.

less than 0.2%. Both the posterior and the UQ-VAE means achieve similar average relative errors on the true parameters $\mathbf{u}^{(\text{m})}$.

We perform forward uncertainty quantification of a single test sample. We use a Sobol' sequence to generate 2^{16} parameter samples from the approximate posterior distribution $q_\phi(\mathbf{u}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_{\text{post}}, \Gamma_{\text{post}})$. For each parameter sample, we run the 0D cardiocirculatory model with the VSD and compute the mean and standard deviation of the resulting time transients pressures and volumes in the four cardiac chambers. We compare the means and standard deviation of the estimated pressure-volume (PV) loops with the true PV loops (Figure 15). PV loops are indicators of cardiac dysfunction and their accurate estimation provides valuable clinical insights into a patient's condition. The estimated PV loops for the right atrium and both ventricles are highly accurate, with smaller uncertainties observed in the left ventricle. The estimation for the left atrium is less accurate compared to the other chambers. This discrepancy may be attributed to the high estimation error in EB_{LV} (Table 6) which affects both the pressure and volume in the left atrium. Another possible explanation is that the true posterior distribution may not be Gaussian, as we assumed in our approximation, potentially introducing a bias in the estimated distribution.
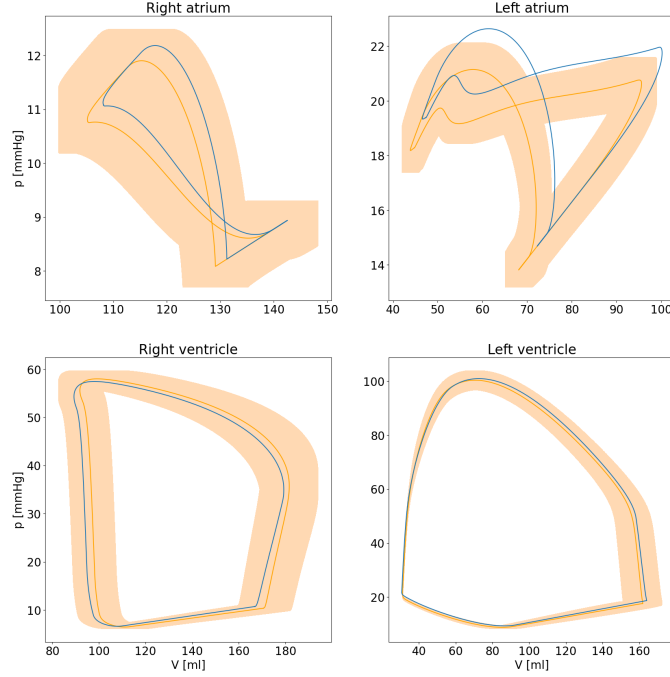
FIGURE 15. Comparison between the true PV loops (blue) and the mean (orange) of those estimated by sampling the parameters from the approximate posterior distribution $q_\phi(\mathbf{u}|\mathbf{y})$. The shaded area represents the standard deviation of the estimated PV loops.

4. CONCLUSIONS

We proposed an improvement (in the form of a novel loss function) of Uncertainty Quantification Variational AutoEncoders (UQ-VAE) for solving Bayesian inverse problems. We introduced a new loss function that significantly reduces the computational cost of a single evaluation, compared to the loss function proposed in Ref. 42, while preserving a relationship between its minimum $(\hat{\boldsymbol{\mu}}, \hat{\Gamma})$ and $(\mathbf{u}_{\mathrm{MAP}}, \Gamma_{\mathrm{Lap}})$ when the forward map $\mathcal{F}$ in (1) is affine (Theorem 1 and Theorem 2) and extending the result to the non affine case.

We showed the capabilities of the UQ-VAEs to approximate both the mean and covariance of the posterior distribution in the cases of the Poisson problem (section 3.1), a non linear problem (section 3.2) and a 0D cardiocirculatory model for systemic hypertension (section 3.3.1) and for ventricular septal defect (section 3.3.2).

With few observational data, the UQ-VAE estimates of the true parameters are as accurate as the posterior mean. With a large number of observational data, the true parameters were estimated better using the UQ-VAE mean rather than the posterior one, due to the latter being poorly approximated by the sample mean in the case of low regularity of the posterior distribution. A similar behavior was observed when the noise level η decreased from high to low. These results show that the UQ-VAE approach is effective in estimating model parameters by solving inverse problems, even for low regular posterior distributions.

The main advantage of a trained UQ-VAEs is in performing inverse uncertainty quantification on the estimated parameters in a short computational time. UQ-VAEs estimated

posterior covariance matrices, with average relative errors less than 1.6% for the 0D cardio-circulatory model and a test set of 100 samples. The average relative error on the posterior variance in the case of the Poisson problem for a test set of 100 samples was less than 11%. For this problem, the UQ-VAE estimated about $42,000$ coefficients $(D+D(D+1)/2$, with $D = 17^2)$ for the mean and the covariance matrix, maintaining a good accuracy.

The proposed novel loss function significantly reduced the training time compared to the previous formulation, while also improving accuracy in estimating $(\mathbf{u}_{\mathrm{MAP}}, \Gamma_{\mathrm{Lap}})$. This reduction in computational cost enables the solution of high dimensional Bayesian inverse problems that were previously computationally infeasible.

A limitation of UQ-VAEs arises when the map $\mathcal{F}$ is approximated by a NN ψ. In this case, the decoder is trained on noiseless observational data and the approximation error is also computed using such data, which is only feasible within an in silico framework. In contrast, the encoder is trained using noisy observational data. When the map $\mathcal{F}$ is known, noiseless observational data are no longer required: an advantage in realistic scenarios where observational data are inherently noisy.

Moreover, the decoder is trained using a supervised approach, with paired input and output data, whereas the encoder is trained in a semi-supervised manner, requiring only observational data and prior knowledge about the distribution of the parameters. In real case scenarios, the observational data of a phenomenon are directly measured, while the prior knowledge is represented by a mathematical model describing the phenomenon. The prior mean can be fixed to a reference setting of the model parameters and the prior covariance can be defined by a variation of the prior mean, as we have done in section 3. Therefore, a semi-supervised approach for training the decoder, which requires only observational data, would broaden the applicability of UQ-VAEs, as it would eliminate the need for in silico generated parameters and observational data.

Finally, we approximate the posterior distribution using a Gaussian distribution, which is not correct for general maps $\mathcal{F}$ from parameters to observational data. This limitation can be addressed by using hierarchical VAEs, which enable sampling from more expressive approximate posteriors.[19,35,44]

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

[1] A. Albanese, L. Cheng, M. Ursino and N. W. Chbat, An integrated mathematical model of the human cardiopulmonary system: model development, *American Journal of Physiology-Heart and Circulatory Physiology* **310** (2016) H899–H921.

[2] T. Bayes, An essay towards solving a problem in the doctrine of chances, *Philosophical transactions* **53** (1763) 370–418.

[3] T. Bui-Thanh, O. Ghattas, J. Martin and G. Stadler, A computational framework for infinite-dimensional bayesian inverse problems part i: The linearized case, with application to global seismic inversion, *SIAM Journal on Scientific Computing* **35** (2013) A2494–A2523.

[4] R. H. Byrd, P. Lu, J. Nocedal and C. Zhu, A limited memory algorithm for bound constrained optimization, *SIAM Journal on scientific computing* **16** (1995) 1190–1208.

[5] F. Caforio, F. Regazzoni, S. Pagani, E. Karabelas, C. Augustin, G. Haase, G. Plank and A. Quarteroni, Physics-informed neural network estimation of material properties in soft tissue nonlinear biomechanical models, *Computational Mechanics* **75** (2024) 487–513.

[6] E. Chong, *An Introduction to Optimization* (Wiley, Hoboken, 2013).

[7] W. Dakkak, M. H. Alahmadi and T. I. Oliver, Ventricular septal defect, StatPearls Publishing, October 2024.

[8] Y. Daon and G. Stadler, Mitigating the influence of the boundary on pde-based covariance operators., *Inverse Problems & Imaging* **12**.

[9] J. Deasy, N. Simidjievski and P. Lió, Constraining variational inference with geometric jensen-shannon divergence, Advances in Neural Information Processing Systems, 2020, `https://proceedings.ne urips.cc/paper_files/paper/2020/file/78719f11fa2df9917de3110133506521-P aper.pdf`.

[10] L. Dedè, F. Regazzoni, C. Vergara, P. Zunino, M. Guglielmo, R. Scrofani, L. Fusini, C. Cogliati, G. Pontone and A. Quarteroni, Modeling the cardiac response to hemodynamic changes associated with COVID-19: a computational study, *Mathematical Biosciences and Engineering* **18** (2021) 3364–3383.

[11] M. Evans and T. Swartz, *Approximating integrals via Monte Carlo and deterministic methods* (OUP Oxford, Oxford, 2000).

[12] A. Formaggio, M. De Luca, S. Borrelli, G. Putame, N. De Vita, F. Minelli, F. Della Corte, R. Vaschetto, A. L. Audenino, C. Olivieri et al., A multidomain 0d model for continuous positive airway pressure ventilation circuit design: Validation and applications, *Biocybernetics and Biomedical Engineering* **45** (2025) 170–180.

[13] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010.

[14] H. Goh, S. Sheriffdeen, J. Wittmer and T. Bui-Thanh, Solving bayesian inverse problems via variational autoencoders, Proceeding of Machine Learning Research, 2nd Annual Conference on Mathematical and Scientific Machine Learning, September 2021.

[15] Y.-C. Hung, A review of monte carlo and quasi-monte carlo sampling techniques, *Wiley Interdisciplinary Reviews: Computational Statistics* **16** (2024) e1637.

[16] J. W. Hurst, R. C. Schlant, C. E. Rackley, E. H. Sonnenblick and N. K. Wenger, *The heart, arteries e veins* (McGraw-Hill Inc., New York, 1990).

[17] K. H. Jin, M. T. McCann, E. Froustey and M. Unser, Deep convolutional neural network for inverse problems in imaging, *IEEE transactions on image processing* **26** (2017) 4509–4522.

[18] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, January 2014.

[19] D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever and M. Welling, Improved variational inference with inverse autoregressive flow, Advances in neural information processing systems, 2016, `ht tps://proceedings.neurips.cc/paper_files/paper/2016/file/ddeebdeefdb7e 7e7a697e1c3e3d8ef54-Paper.pdf`.

[20] R. M. Lang, L. P. Badano, V. Mor-Avi, J. Afilalo, A. Armstrong, L. Ernande, F. A. Flachskampf, E. Foster, S. A. Goldstein, T. Kuznetsova et al., Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the american society of echocardiography and the european association of cardiovascular imaging, *European Heart Journal-Cardiovascular Imaging* **16** (2015) 233–271.

[21] H. Li, J. Schwab, S. Antholzer and M. Haltmeier, Nett: Solving inverse problems with deep neural networks, *Inverse Problems* **36**.

[22] A. Lucas, M. Iliadis, R. Molina and A. K. Katsaggelos, Using deep neural networks for inverse problems in imaging: beyond analytical methods, *IEEE Signal Processing Magazine* **35** (2018) 20–36.

[23] J. A. Nelder and R. Mead, A simplex method for function minimization, *The computer journal* **7** (1965) 308–313.

[24] F. D. M. Neto and A. J. da Silva Neto, *An introduction to inverse problems with applications* (Springer Science & Business Media, London, 2012).

[25] H. V. Nguyen and T. Bui-Thanh, Tnet: A model-constrained tikhonov network approach for inverse problems, *SIAM Journal on Scientific Computing* **46** (2024) C77–C100.

[26] F. Nielsen, A family of statistical symmetric divergences based on jensen's inequality, December 2010.

[27] W. H. Organization, *Global report on hypertension: the race against a silent killer* (World Health Organization, Geneva, 2023).

[28] F. Regazzoni, M. Salvador, P. C. Africa, M. Fedele, L. Dedè and A. Quarteroni, A cardiac electromechanical model coupled with a lumped-parameter model for closed-loop blood circulation, *Journal of Computational Physics* **457** (2022) 111083.

[29] L. G. Rudski, W. W. Lai, J. Afilalo, L. Hua, M. D. Handschumacher, K. Chandrasekaran, S. D. Solomon, E. K. Louie and N. B. Schiller, Guidelines for the echocardiographic assessment of the right heart in adults: a report from the american society of echocardiography: endorsed by the european association of echocardiography, a registered branch of the european society of cardiology, and the canadian society of echocardiography, *Journal of the American society of echocardiography* **23** (2010) 685–713.

[30] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola, Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index, *Computer physics communications* **181** (2010) 259–270.

[31] M. Salvador, F. Kong, M. Peirlinck, D. W. Parker, H. Chubb, A. M. Dubin and A. L. Marsden, Digital twinning of cardiac electrophysiology for congenital heart disease, *Journal of the Royal Society Interface* **21**.

[32] M. Salvador, M. Strocchi, F. Regazzoni, C. M. Augustin, L. Dede', S. A. Niederer and A. Quarteroni, Whole-heart electromechanical simulations using latent neural ordinary differential equations, *NPJ Digital Medicine* **7**.

[33] I. M. Sobol, On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Computational mathematics and mathematical physics* **7** (1967) 86–112.

[34] I. M. Sobol, Sensitivity estimates for nonlinear mathematical models, *Math. Model. Comput. Exp.* **1** (1993) 407–414.

[35] C. K. Sønderby, T. Raiko, L. Maaløe, S. K. Sønderby and O. Winther, Ladder variational autoencoders, Advances in neural information processing systems, 2016, https://proceedings.neurips.cc/paper_files/paper/2016/file/6ae07dcb33ec3b7c814df797cbda0f87-Paper.pdf.

[36] D. E. Spicer, H. H. Hsu, J. Co-Vu, R. H. Anderson and F. J. Fricker, Ventricular septal defect, *Orphanet journal of rare diseases* **9** (2014) 1–16.

[37] A. M. Stuart, Inverse problems: a bayesian perspective, *Acta numerica* **19** (2010) 451–559.

[38] T. Sutter, I. Daunhawer and J. Vogt, Multimodal generative learning utilizing jensen-shannon-divergence, Advances in Neural Information Processing Systems, 2020, https://proceedings.neurips.cc/paper_files/paper/2020/file/43bb733c1b62a5e374c63cb22fa457b4-Paper.pdf.

[39] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (VH Winston and Sons, Washington DC, 1977).

[40] G. G. Tong, C. A. S. Long and D. E. Schiavazzi, Invaert networks: A data-driven framework for model synthesis and identifiability analysis, *Computer Methods in Applied Mechanics and Engineering* **423** (2024) 116846.

[41] G. G. Tong, C. A. Sing-Long and D. E. Schiavazzi, Invaert networks for amortized inference and identifiability analysis of lumped-parameter haemodynamic models, *Philosophical Transactions A* **383** (2025) 20240215.

[42] A. Tonini and L. Dede', Enhanced uncertainty quantification variational autoencoders for the solution of bayesian inverse problems, February 2025.

[43] A. Tonini, F. Regazzoni, M. Salvador, L. Dede', R. Scrofani, L. Fusini, C. Cogliati, G. Pontone, C. Vergara and A. Quarteroni, Two new calibration techniques of lumped-parameter mathematical models for the cardiovascular system, *International Journal for Numerical Methods in Engineering* **126** (2025) e7648.

[44] A. Vahdat and J. Kautz, Nvae: A deep hierarchical variational autoencoder, Advances in neural information processing systems, 2020, https://proceedings.neurips.cc/paper/2020/file/e3b21256183cf7c2c7a66be163579d37-Paper.pdf.

[45] H. Van Nguyen, T. Bui-Thanh and C. Dawson, Taen: a model-constrained tikhonov autoencoder network for forward and inverse problems, *Computer Methods in Applied Mechanics and Engineering* **446** (2025) 118245.

[46] U. Villa, N. Petra and O. Ghattas, hippylib: an extensible software framework for large-scale deterministic and bayesian inverse problems, 2016, version 2.1.0.

[47] U. Villa, N. Petra and O. Ghattas, hippylib: an extensible software framework for large-scale deterministic and bayesian inverse problems, *Journal of Open Source Software* **3** (2018) 940.

[48] U. Villa, N. Petra and O. Ghattas, Hippylib: An extensible software framework for large-scale inverse problems governed by pdes: Part i: Deterministic inversion and linearized bayesian inference, *ACM Trans. Math. Softw.* **47** (2021) 1–34.

[49] G. Q. Zhang and W. Zhang, Heart rate, lifespan, and mortality risk, *Ageing research reviews* **8** (2009) 52–60.

[50] Y. Zhu, J. Zhang, A. Gangrade and C. Scott, Label noise: Ignorance is bliss, Advances in Neural Information Processing Systems, 2024, `https://proceedings.neurips.cc/paper_files/paper/2024/file/d3696c79d572c995a74eac78037551a8-Paper-Conference.pdf`.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

Tonini, A.; Bui-Thanh, T.; Regazzoni, F.; Dede', L; Quarteroni, A.
Improvements on uncertainty quantification with variational autoencoders

Gimenez Zapiola, A.; Boselli, A.; Menafoglio, A.; Vantini, S.

55/2025 Gimenez Zapiola, A.; Boselli, A.; Menafoglio, A.; Vantini, S.
Hyper-spectral Unmixing algorithms for remote compositional surface mapping: a review of the state of the art

54/2025 Tomasetto, M.; Williams, J.P.; Braghin, F.; Manzoni, A.; Kutz, J.N.
Reduced order modeling with shallow recurrent decoder networks

53/2025 Zecchi, A. A.; Sanavio, C.; Perotto, S.; Succi, S.
Telescopic quantum simulation of the advection-diffusion-reaction dynamics

51/2025 Tomasetto, M.; Braghin, F., Manzoni, A.
Latent feedback control of distributed systems in multiple scenarios through deep learning-based reduced order models

50/2025 Bonetti, S.; Botti, M.; Antonietti, P.F.
Conforming and discontinuous discretizations of non-isothermal Darcy–Forchheimer flows

49/2025 Zanin, A.; Pagani, S.; Corti, M.; Crepaldi, V.; Di Fede, G.; Antonietti, P.F.; the ADNI
Predicting Alzheimer's Disease Progression from Sparse Multimodal Data by NeuralODE Models

48/2025 Temellini, E.; Ballarin, F.; Chacon Rebollo, T.; Perotto, S.
On the inf-sup condition for Hierarchical Model reduction of the Stokes problem

47/2025 Gimenez Zapiola, A.; Consolo, A.; Amaldi, E.; Vantini, S.
Penalised Optimal Soft Trees for Functional Data