J. Hauth, X. Huan, B.Y. Zhou, N.R. Gauger, M. Morelli, A. Guardone
Correlation Effects in Bayesian Neural Networks for Computational Aeroacoustics Ice Detection
In: AIAA Scitech 2020 Forum, Orlando, FL, USA, 6-10 Jan. 2020, ISBN: 978-1-62410-595-1, Paper AIAA 2020-1414, doi:10.2514/6.2020-1414

Correlation Effects in Bayesian Neural Networks for Computational Aeroacoustics Ice Detection

Jeremiah M. A. Hauth* and Xun Huan†
University of Michigan, 1231 Beal Ave, Ann Arbor, MI 48109-2133, USA

Beckett Y. Zhou‡ and Nicolas R. Gauger§
TU Kaiserslautern, Bldg 34, Paul-Ehrlich-Strasse, 67663 Kaiserslautern, Germany

Myles Morelli¶ and Alberto Guardone‖
Politecnico di Milano, Building B12 Campus Bovisa, Via La Masa, 34 20156 Milano, Italy

In-flight rotor icing presents a serious problem in the operation of rotorcraft in cold climates, as complex ice shapes can significantly degrade the aerodynamic performance and handling characteristics of rotorcraft. Reliable real-time detection of ice formation is thus a critical enabling technology in improving rotorcraft safety. In this paper, we continue our previous work to explore a novel approach towards developing a real-time in-flight ice detection system using computational aeroacoustics and Bayesian neural networks (BNNs). We focus on our use of BNNs constructed from the simulated aeroacoustics dataset to enable rapid predictions of aerodynamic performance indicators together with quantified uncertainty. Specifically, we investigate the effectiveness and tradeoffs among several approximate Bayesian inference techniques for training BNNs: (Gaussian) mean-field variational inference (MFVI), full-covariance variational inference (FCVI), and Stein variational gradient descent (SVGD). We find the correlations in weight uncertainty to be low for this application and with our current BNN setup, although MFVI (which ignores correlations altogether) noticeably under-predicts the variance. SVGD is both computationally fast and captures non-Gaussian and correlation structures, appearing to be a well-suited method for the situation currently considered.

I. Introduction

In-flight rotor icing presents a serious problem for the operation of rotorcraft in cold climates, where complex ice shapes can significantly degrade the aerodynamic performance and vehicle handling characteristics leading to accidents and even fatalities [1–3]. Reliable real-time detection of ice formation is therefore a critical enabling technology for improving rotorcraft safety. With visual assessment of ice nearly impossible under high rotor speeds, recent studies started exploring the use of acoustic signatures for detecting rotor blade ice [4, 5]. In a recent paper [6], we proposed a novel approach towards developing a real-time in-flight ice detection system using computational aeroacoustics (CAA) and Bayesian neural networks (BNNs). In particular, the use of BNNs allowed machine learning (ML) predictions to also offer quantified uncertainty, reflecting the quality and credibility of the predicted values. However, constructing a BNN involves performing Bayesian inference on a very high-dimensional (thousands to millions or more) problem, which is extremely challenging and where we initially ameliorated by adopting independent (mean-field) and Gaussian approximations to the uncertainty distributions. In this paper, we focus on investigating the errors and effects due to the previously neglected correlation structures, and explore new computationally practical methods that can also retain correlation.

Under this framework, an icing simulation code based on a fully unsteady collection efficiency approach is coupled with the aeroacoustic solver in the open-source software suite SU2 [7], in order to compute far-field noise signatures corresponding to iced rotor blades in various icing conditions. While these numerical simulations offer high-fidelity

*PhD Candidate, Department of Mechanical Engineering, Student Member AIAA, hauthj@umich.edu
†Assistant Professor, Department of Mechanical Engineering, Member AIAA, xhuan@umich.edu
‡Research Scientist, Chair for Scientific Computing, Member AIAA, yuxiang.zhou@scicomp.uni-kl.de
§Professor, Chair for Scientific Computing, Associate Fellow AIAA, nicolas.gauger@scicomp.uni-kl.de
¶PhD Candidate, Department of Aerospace Sciences and Technologies, mylescarlo.morelli@polimi.it
‖Professor, Department of Aerospace Sciences and Technologies, alberto.guardone@polimi.it

flow and acoustic solutions based on physical principles, their utilization for on-board real-time ice detection remains impractical. This is due to the fact that after an acoustic signal is collected, an inverse problem needs to be solved to infer the corresponding ice profile that generated this acoustic signal, followed by an aerodynamic computation from that estimated ice profile to obtain the relevant performance indicators such as the rotor blade lift and moment coefficients. Each numerical solution of the unsteady Reynolds-averaged Navier Stokes (RANS) equations even under a simplified two-dimensional representation of a pitching rotor blade, currently takes minutes to hours to complete despite being parallelized with tens of processors. The computational cost would increase further significantly when realistic three-dimensional rotor blades are simulated with scale-resolving methods such as large eddy simulations. Compounded by the inverse problem that is inherently multi-query and requiring possibly thousands or more forward solutions, the overall computation time would be orders of magnitude higher than real-time requirements. We thus seek accelerations through ML techniques, namely neural networks (NNs).

NNs are powerful data-driven modeling tools with the potential for accurate predictions given particularly complex underlying trends, and their usage is rapidly growing in applications of engineering and computational aerospace sciences, including for rotorcraft aeroacoustics. In this paper, we aim to construct NNs offline from a database of computational fluid dynamics (CFD) simulations, for the purpose of rapid online predictions of flight performance characteristics from rotor acoustic signatures under different degrees of rotor icing. Traditional training of a NN utilizes an optimization framework to fit the training data, which is generally achieved through a variation of the stochastic gradient descent algorithm [8, 9]. Upon completion, best values are obtained for the NN's weight and bias terms (collectively referred to as parameters in this paper), and correspondingly the NN is able to produce a *single-value* output prediction given an input. In a traditional NN, no uncertainty information is available to indicate the quality and credibility of that output number. Indeed, we would expect the prediction uncertainty to vary depending on the architecture of the NN; the quantity, quality, and informativeness of the training data; and the input regime. It is critical that we understand the degree to which we can trust NN predictions and as such we must give appropriate distributions to the predicted quantities.

To model the uncertainty inherent in a NN prediction, we employ a particular class of NNs called Bayesian NNs, where network parameters are treated as random variables with associated uncertainty distributions instead of as single-valued. Under a Bayesian framework, the NN parameters are assigned prior probability density functions (PDFs), and these PDFs are updated accordingly with available training data to form posterior PDFs. A Monte Carlo (MC) sampling from these PDFs can then yield a distribution of predictions reflecting the uncertainty induced by the training data.

In this paper we investigate multiple computational techniques for constructing the posterior PDFs of NN parameters, each with tradeoffs between accuracy and computational efficiency. We start by presenting preliminary results based on uncorrelated parameter PDFs computed using mean-field variational inference with indpendent Gaussian approximations as a baseline. We then investigate variational families that retain correlation, such as full-covariance variational inference and Stein variational gradient descent. Our goal is then to understand the importance of correlation in this ice detection application and its effect on the prediction uncertainty of our BNN output. With this knowledge, we demonstrate efficient constructions of BNN models that predict flight performance metrics given an acoustic signal, along with an accurate portrayal of the prediction uncertainty.

This paper is organized as follows. Section II briefly describes the governing physics as well as numerical solvers to simulate the ice accretion, fluid flow, and aeroacoustics for a rotor blade airfoil. Section III presents detailed formulation and computational techniques for constructing the BNN through solving a Bayesian inference problem on the network weights. Discussions and comparisons of BNN performance with varying degrees of correlation retention are shown in Sec. IV. The paper ends with conclusion remarks in Sec. V.

II. Ice Detection Problem and Simulation Workflows

We briefly describe the ice detection problem, and the solvers and workflows for generating training data; more details can be found in [6]. Our overall goal is to create a real-time acoustics-based ice detection system using a ML model trained offline from a database of high-fidelity physics-based simulations. As illustrated in Fig. 1, after an acoustic signal is collected, an inverse problem needs to be solved to infer the corresponding ice profile that generated this acoustic signal, followed by an aerodynamics computation from that ice profile to obtain the relevant performance metrics such as the rotor blade lift and moment coefficients. The intensive computational demand for each flow solve compounded by the multi-query requirement of an inverse problem makes real-time speeds nearly impossible to achieve. We thus seek accelerations through ML to "skip" the inverse problem, and create a mapping directly from acoustic

signals to the performance indicators.

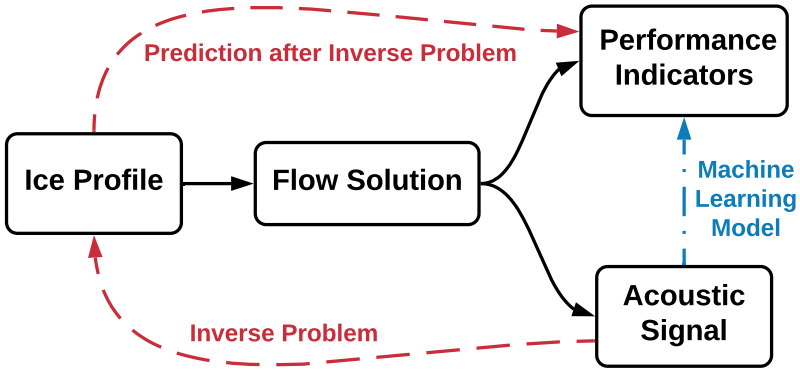


Fig. 1 Workflow routes illustrating the physics causation (black solid), traditional inverse problem approach (dashed red), and proposed machine learning model (dotted-dash blue).

The training dataset thus composes of acoustic signals and their corresponding performance indicators. To generate this database, we perform a series of simulations with SU2 [7], an open-source software in C++ and Python created for multi-physics simulation and design. As a proof-of-concept demonstration in this paper, our flows are two-dimensional and modeled by unsteady RANS accompanied by the Menter shear-stress transport (SST) turbulence model. They are solved via a finite volume method on unstructured meshes using a standard edge-based data structure on a dual grid with control volumes constructed through a median-dual vertex-based scheme, and together with implicit dual-time stepping achieve second-order accuracy in space and time.

The ice profiles are generated via the following procedure: first, super-cooled water droplet impingement locations and collection efficiency are computed using a Lagrangian based particle tracking solver called PoliDrop [10]; then, ice accretion is calculated by an in-house code PoliMIce [11]; finally, ice shapes are computed using an unsteady multi-step approach, whereby non-linear ice accretion is accounted for by iteratively updating the surface solution on which the ice accretes. Once the ice profiles are created, the iced airfoil shapes are then subject to a CAA analysis using the permeable-surface Ffowcs-Williams-Hawkings (FWH) formulation in SU2 [12]. The permeable FWH formulation allows fluid to flow through the fictitious surface Γ_p. Details of the flow field are then extracted and the noise source is then propagated to the far-field. This implementation is shown in the schematic from Fig. 2. The computational domain is thus divided into two separate regions: the near-field CFD region Ω_1, and the far-field CAA region Ω_2. The position of the FWH surface lies $\frac{3}{4}$ chords length from the airfoil trailing edge. The position of the observer locations is chosen based on where noise would likely be perceived on a conventional main rotor/tail rotor helicopter, which is directly below the main rotor.

A database of 101 high-fidelity simulations is created, corresponding to iced airfoils with different degrees of icing severity. These data will be used to construct the ML model.

Inputs: The inputs to the ML model will be the far-field noise signal, which can be obtained from the CAA computations (e.g., Fig. 3). The raw signals will then be pre-processed with a frequency analysis to obtain the power spectral density (PSD) (e.g., Fig. 4). Lastly, the PSD will be discretized into 151 equal-width bins for the 0–140 Hz range. This 151-dimensional vector will be the input into the ML model.

Outputs: The output predictions of the ML model will be aerodynamic performance metrics. From the airfoil geometry and the CFD simulations, we extract two flight performance metrics: the coefficient of lift C_L and coefficient of moment C_M. These quantities are calculated at each time step as the airfoil undergoes a pitching movement ranging from an angle of attack between $-1°$ to $11°$ (e.g., Fig. 5). From the hysteresis loop of each test case, the minimum, maximum, and mean values of C_L and C_M are recorded. This 6-dimensional vector will be the output prediction from the ML model.

III. Bayesian Neural Networks and Variational Inference

A NN maps input x to output $\hat{y}$: we write $\hat{y} = f(x)$ where $\hat{y}$ (with hat) denotes the NN *prediction*. While the interior composition of an NN can be highly complicated (e.g., convolutional NNs), for this paper we focus on densely connected

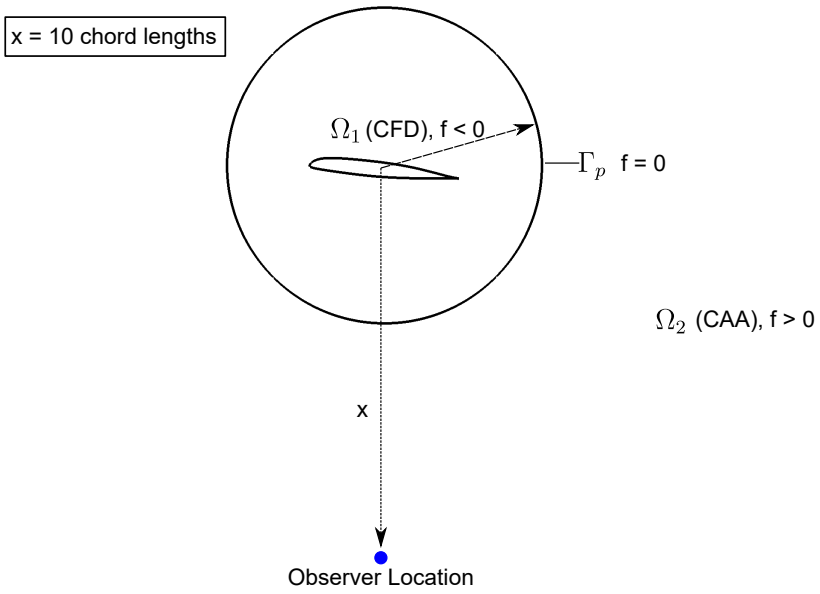


Fig. 2 Schematic of the permeable control surface Γ_p separating the CFD and CAA domains and the relative observer locations.

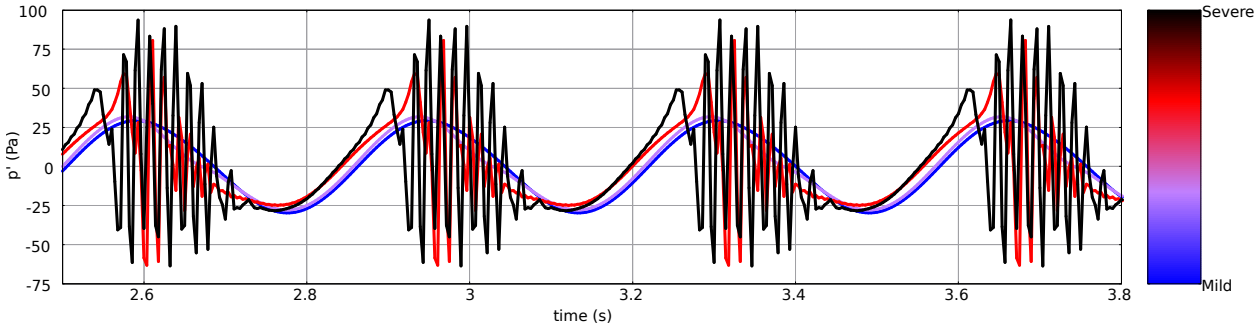


Fig. 3 Far-field noise signal of mildly iced and severely iced airfoils.

feed-forward NNs for simplicity. For example, Fig. 6a shows a graphical representation of such a NN with an input layer, two hidden layers, and an output layer. A good choice on the number of layers, number of nodes in each layer, and activation functions—together referred to as hyperparameters—often requires trial-and-error and experience. In this work, we pre-select reasonable hyperparameter settings and do not conduct a systematic comparison of performance under different choices. Once the network architecture is decided, the collection of all trainable weights and bias terms (we denote them by w) constitute the free parameters to be tuned. Accordingly, we update our notation to $\hat{y} = f(x; w)$.

Given N training data points in the form of input-output pairs $(x_T, y_T) = \{x_n, y_n\}_{n=1}^N$, NN training typically involves choosing w to minimize the least squares loss function:

$$w^* = \arg\min_w \left\{ \frac{1}{N} \sum_{n=1}^{N} [f(x_n; w) - y_n]^2 \right\}. \tag{1}$$

The optimization is often carried out with stochastic gradient descent [8, 9]. However, regardless of how Eq. (1) is solved, its solution returns a single-valued w^*. Consequently, the prediction at a given input x is also single-valued: $\hat{y} = f(x; w^*)$, and it does not carry uncertainty information that would be affected by the quantity and quality of the training data.

A BNN is similar to a NN, but treats w as random variables with associated PDFs that represent the uncertainty on

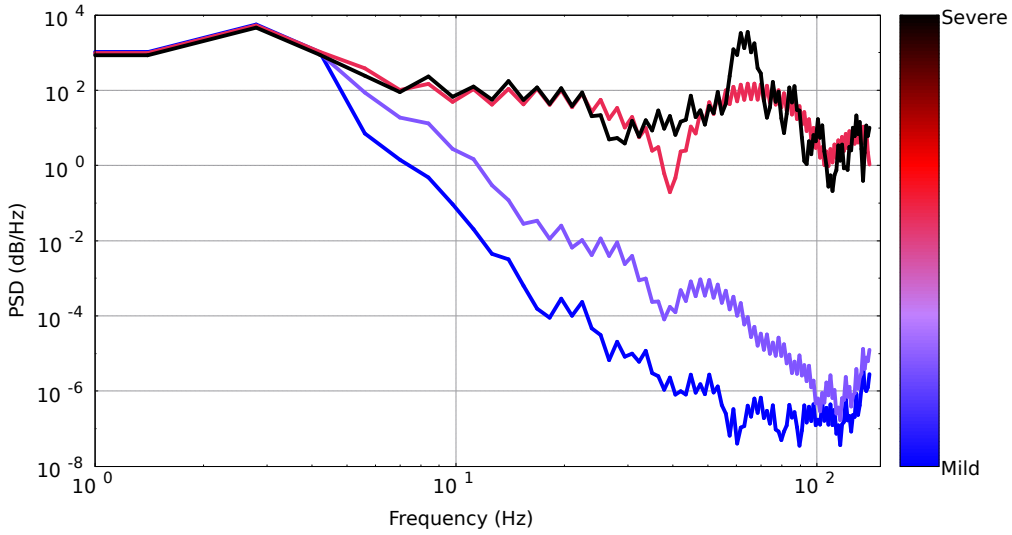
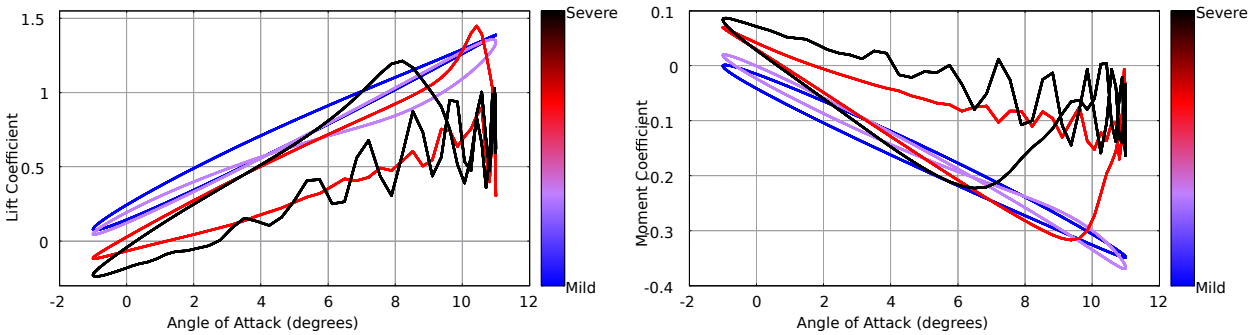


Fig. 4 Far-field noise power spectral density of the mildly iced and severely iced airfoils.



(a) Lift Coefficient Hysteresis Loop

(b) Moment Coefficient Hysteresis Loop

Fig. 5 Performance characteristics of mildly iced and severely iced airfoils.

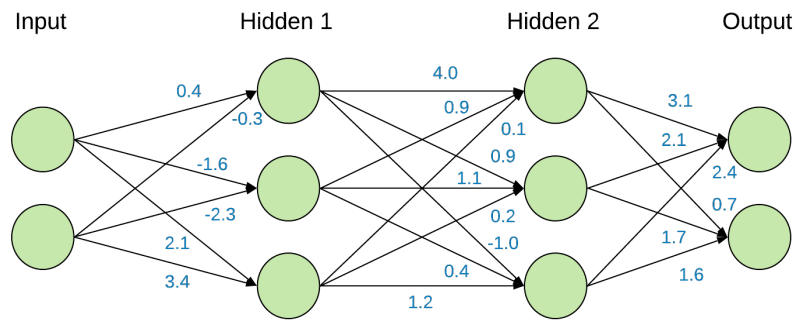
w (Fig. 6b). When training data become available, these PDFs are updated through Bayes' theorem:

$$p(w|x_T, y_T) = \frac{p(y_T|x_T, w)p(w)}{p(y_T|x_T)}, \tag{2}$$

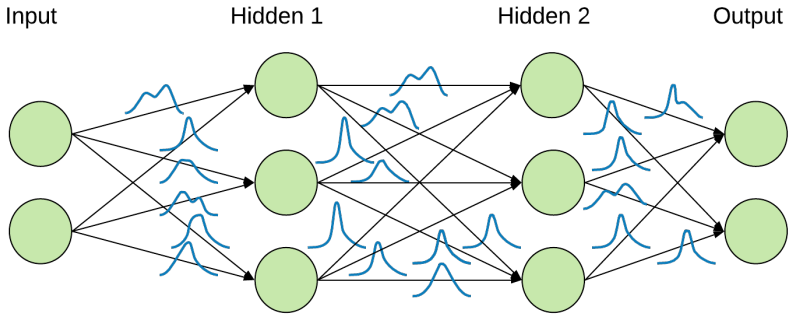
where $p(w)$ is the prior PDF on the weight parameters and we also assumed $p(w|x_T) = p(w)$ (i.e., the prior uncertainty should not change from knowing only the input values of the training data), $p(y_T|x_T, w)$ is the likelihood function, $p(w|x_T, y_T)$ is the posterior PDF, and $p(y_T|x_T)$ is the Bayesian evidence. Constructing a BNN then entails computing the posterior $p(w|x_T, y_T)$—that is, our updated uncertainty on w given the training dataset (x_T, y_T).

Characterizing the full $p(w|x_T, y_T)$ can be extremely challenging since w may easily reach thousands- or even millions-dimensional for deep NNs arising from complex engineering and science applications. While the go-to algorithms for Bayesian inference—Markov chain Monte Carlo methods (e.g. [13])—can generate samples from the true posterior PDF, high dimensionality (beyond around 100) quickly renders them intractable. We thus explore a series of *variational inference (VI)* algorithms in this paper, based on seeking good approximations of the posterior PDF and thereby turning the sampling task into a more tractable optimization problem.

The aim of VI is to approximate the posterior PDF with another PDF from a parametric (and usually standard-form) family that is "close" in some sense. Denoting the approximating family to be $q(w;\theta)$ parameterized by θ, the Kullback-Leibler (KL) divergence is often adopted to quantify the degree of dissimilarity between $q(w;\theta)$ and the true



(a) Neural network



(b) Bayesian neural network

Fig. 6 NN and BNN, each with an input layer, two hidden layers, and an output layer. Weights are shown in blue as deterministic values in NN, and as probability density functions in BNN.

posterior $p(w|x_T, y_T)$:

$$D_{\mathrm{KL}}[q(w;\theta) \parallel p(w|x_T, y_T)] = \int_{-\infty}^{\infty} q(w;\theta) \log\left[\frac{q(w;\theta)}{p(w|x_T, y_T)}\right] dw = \mathbb{E}_{q(w;\theta)}\left[\log q(w;\theta) - \log p(w|x_T, y_T)\right]. \tag{3}$$

Solving the VI problem entails finding θ^* that yields the closest posterior approximation:

$$\theta^* = \arg\min_{\theta} D_{\mathrm{KL}}\left[q(w;\theta) \parallel p(w|x_T, y_T)\right]. \tag{4}$$

To numerically solve this, one can derive an equivalent problem minimizing the well-known *evidence lower bound (ELBO)* that can be approximated using an unbiased Monte Carlo estimator:

$$\theta^* = \arg\min_{\theta} \mathbb{E}_{q(w;\theta)}\left[\log q(w;\theta) - \log p(y_T|x_T, w) - \log p(w)\right] \tag{5}$$

$$\approx \arg\min_{\theta} \left\{\frac{1}{M}\sum_{m=1}^{M}\left[\log q\left(w^{(m)};\theta\right) - \log p\left(y_T|x_T, w^{(m)}\right)\right] - \log p\left(w^{(m)}\right)\right\} \tag{6}$$

Here samples $w^{(m)}$ are drawn from $q(w;\theta)$.

In this study we feature comparisons among three VI methods with different degrees of fidelity and computational cost, described below.

Mean-field variational inference (MFVI): A popular choice for $q(w;\theta)$ is the independent (mean-field) Gaussian family [14]:

$$q(w;\theta) = \prod_{k=1}^{K} q(w_k;\theta_k) \tag{7}$$

where K is the total number of parameters in the NN. The choice of independence allows decomposing the joint PDF into a product of marginal PDFs $q(w_k;\theta_k)$, each now following a univariate Gaussian $\mathcal{N}(\mu_k,\sigma_k^2)$ where the parameter $\theta_k = \{\mu_k,\sigma_k\}$ carries the values of mean and standard deviation. The optimization problem often can leverage gradient-based algorithms, with gradient information extracted through back-propagation [14] or automatic differentiation. The main advantage of MFVI is its speed since it only involves a $2K$-dimensional optimization problem. However, it is disadvantaged by its inability to capture parameter correlations and non-Gaussian structures, and a tendency to under-predict the uncertainty [15].

Full-covariance variational inference (FCVI): Alternatively, one may retain the full-covariance Gaussian structure $\mathcal{N}(\mu,\Sigma)$ without any independence assumptions, where now $\theta = \{\mu,\Sigma\}$. To take advantage of and also ensuring the symmetric and positive definite property of the covariance matrix Σ, the lower triangular matrix L from its Cholesky decomposition $\Sigma = LL^T$ is managed and updated through the optimization process. Similar to MFVI, gradient-based optimization algorithms (e.g., [16]) can be deployed under the support of back-propagation. FCVI is higher fidelity compared to MFVI due to its ability to capture correlation effects. However, it is also more expensive now involving a $\frac{K^2+3K}{2}$-dimensional optimization problem (K from μ, and $\frac{K^2+K}{2}$ from L).

Stein variational gradient descent (SVGD): Lastly, we explore Stein variational gradient descent [17], a non-parametric particle based method for Bayesian inference. This method leverages the relationship between the (functional) gradient of objective in Eq. (4) to the Stein discrepancy, the latter which can be approximated using a set of particles. A gradient-descent procedure can then formed to iteratively minimize this gradient, as summarized by

$$w_i^{\ell+1} \leftarrow w_i^\ell + \epsilon_\ell \hat{\phi}^*(w_i^\ell) \qquad \text{where} \qquad \hat{\phi}^*(w) = \frac{1}{m}\sum_{j=1}^{m}\left[k(w_j^\ell,w)\nabla_{w_j^\ell}\log p(w_j^\ell|x_T,y_T) + \nabla_{w_j^\ell}k(w_j^\ell,w)\right], \tag{8}$$

where $w_i^\ell, i = 1,\ldots,m$ represents the i-th particle at the ℓ-th iteration, ϵ_ℓ is the step size, $k(\cdot,\cdot)$ is a positive definite kernel. Furthermore, the gradient of the true log-posterior can be evaluated using

$$\nabla_w \log p(w|x_T,y_T) = \nabla_w\left[\log p(y_T|x_T,w) + \log p(w) - \log p(y_T|x_T)\right] = \nabla_w \log p(y_T|x_T,w) + \nabla_w \log p(w)$$

requiring only likelihood and prior evaluations, since the gradient with respect to the evidence term is zero. The overall effect is an iterative transport of a set of points to match the target distribution $p(w|x_T,y_T)$. Figure 7 provides a visual illustration of the concept in a correlated bivariate Gaussian case. The advantage of SVGD is its ability to capture correlation as well as non-Gaussian structures. However, it may become computational expensive in high dimensions especially with a high number of particles.

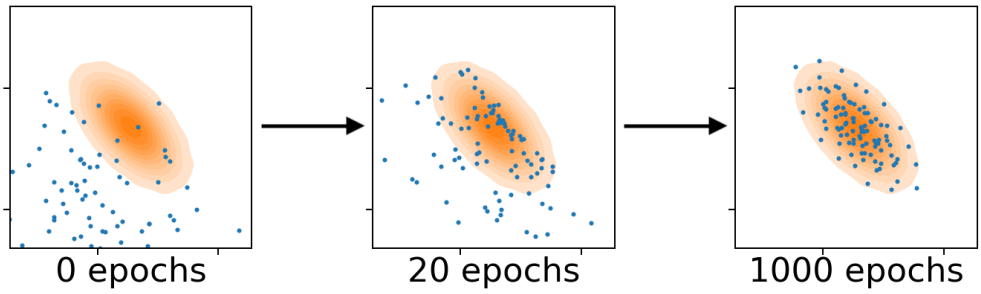


Fig. 7 In this correlated bivariate Gaussian illustration, SVGD particles move to arrive at a better approximation to the posterior distribution.

IV. Results

We present results on two cases. The first will be a test case involving a single-layer linear BNN, resulting in a conjugate Gaussian system and where the exact posterior PDFs can be analytically computed. We use this case as a benchmark to compare the different VI methods against the exact solution, and to motivate the need for capturing correlation effects. The second will be the ice detection application, driven by complex underlying physical processes and where the exact solution is unavailable.

A. Linear-Gaussian test case

Consider a single-layer scalar-output BNN with a linear activation function and no bias term: $\hat{y} = x^T w$ (see Fig. 8). Further assuming an additive Gaussian likelihood model for each of the N training point $y_n = \hat{y}_n + \epsilon_n = x_n^T w + \epsilon_n$, we can assemble the overall N-point system into

$$y = Xw + \epsilon \qquad \text{with} \qquad \epsilon \sim \mathcal{N}(0, \Sigma_\epsilon), \tag{9}$$

where $(x_1, \cdots, x_n)^T = X \in \mathbb{R}^{N\times K}$, $w \in \mathbb{R}^{K\times 1}$, $y, \epsilon \in \mathbb{R}^{N\times 1}$, and $\Sigma_\epsilon \in \mathbb{R}^{N\times N}$. If a Gaussian prior is adopted with $w \sim \mathcal{N}(\mu_0, \Sigma_0)$, then a conjugate system is formed and the posterior PDF can be shown to remain Gaussian: $w|x_T, y_T \sim \mathcal{N}(\mu_\pi, \Sigma_\pi)$ with mean and covariance formulas

$$\mu_\pi = \Sigma_\pi [X^T \Sigma_\epsilon^{-T} y + \Sigma_0^{-1} \mu_0] \tag{10}$$

$$\Sigma_\pi = \left(\Sigma_0^{-1} + X^T \Sigma_\epsilon^{-1} X\right)^{-1}. \tag{11}$$

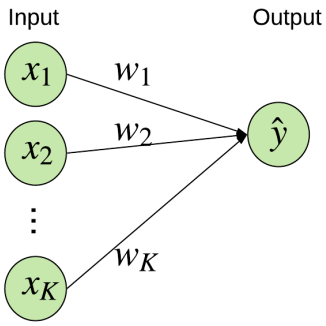


Fig. 8 A single-layer scalar-output NN where the BNN weight posterior PDFs can be computed analytically.

For this test, we generate $N = 100$ synthetic data points of (x, y) pairs under an arbitrarily designated true w^*. Based on these data, we approximate the posterior PDF using the variational inference methods described in Sec. III, and also compare them to the exact posterior dictated by Eqs. (10) and (11). Figure 9 presents the posterior PDF results for a two-dimensional w ($K = 2$) case. The true posterior (blue shade) is positively correlated between the two components of w. MFVI is unable to capture the correlation and only provides an independent Gaussian approximation, and also appears to underpredict the overall uncertainty especially in the elongated axis. FCVI constructs a full covariance multivariate Gaussian and is therefore able to provide an excellent depiction of the correlated posterior. However, we would expect additional approximation errors for FCVI when the true posterior is no longer Gaussian. SVGD is also able to accurately capture the correlation structure of the posterior through its particles, and we expect it to be able to adapt to non-Gaussian shapes as well. Higher dimensional cases have also been tested with similar conclusions, and omitted in this paper for brevity.

B. Ice Detection Application

We now deploy our BNN framework to the ice detection problem as described in Sec. II. The BNN architecture adopted is summarized in Table 1, where we intentionally start with a small network structure in order to keep the number of parameters low. With reference to Eq. (2), the prior is chosen to be independent Gaussian $\mathcal{N}(0, 1^2)$ for each weight, and the likelihood corresponds to an additive Gaussian noise in the form of $y = f(x; w) + \epsilon$ with $\epsilon \sim \mathcal{N}(0, 0.1^2)$. The optimization for Eq. (6) utilizes the Adam algorithm [18] with a learning rate of 0.002 and trained to 1000 epochs. In total there are 996 trainable parameters whose distributions are being inferred.

The BNNs are constructed with 80 training data points, and the remaining 20 points are held out for subsequent testing. Once constructed, 10,000 samples are drawn from the weight posterior PDF, and the output is evaluated for each sample via the NN forming an empirical distribution for the posterior predictive QoIs. Figure 10 summarizes these BNN predictions (figure y-axis) and compares them to the true label values (figure x-axis), with the vertical uncertainty intervals representing ± 1.96 standard deviations of the marginal PDFs. All three methods produced similar trends in terms of predictive uncertainty magnitudes, with the blue $C_{L,max}$ and red $C_{L,min}$ carrying higher uncertainty

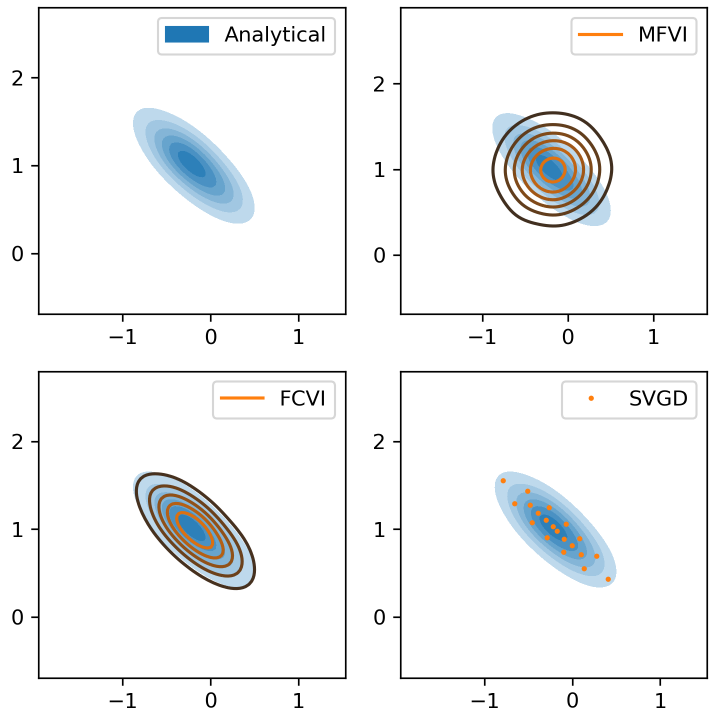


Fig. 9 Comparison of numerical methods for inferring a correlated bivariate Gaussian posterior PDF.

Table 1 BNN architecture

Layer	Name	No. of Neurons	Activation Fcn.
1	input	151	-
2	hidden 1	6	tanh
3	hidden 2	6	tanh
4	output	6	affine

than the other QoIs. Furthermore, MFVI appears to have overall smallest uncertainty levels, followed by FCVI, and with SVGD having the highest. This observation is consistent with MFVI's tendency to under-predict variance due to its inability to portray correlation. Furthermore, it is likely that SVGD's capturing of non-Gaussian forms further induces a more accurate uncertainty representation, albeit it has a separate source of approximation error with its finite number of particles. Figure 11 further presents the joint distributions between every pair of QoIs for a specific test case. The size of these two-dimension PDFs are consistent with the trends in Fig. 10, and furthermore correlations and non-Gaussian forms are also present. We note that MFVI (and FCVI) can produce non-Gaussian and correlated distributions in this figure since these are the posterior predictive PDFs for $f(x;w)$ (while MFVI restricts w to be independent Gaussians, $f(x;w)$ can be highly non-Gaussian due to the nonlinearity of f). Figure 12 illustrates sample pairwise joint posterior PDFs among select NN weight parameters. Here we see that while MFVI and FCVI exhibit some differences in uncertainty sizes, both sets of results appear to be nearly uncorrelated from visual inspection; the non-Gaussian form in the SVGD is evident.

Overall, the correlations in weight uncertainty appear to be low for this application and with our current BNN setup, although a noticeable under-prediction in variance takes place when MFVI (which ignores correlation altogether) is employed. With our implementations, the training times for MFVI, FCVI, and SVGD are approximately 8 min, 80 min, and 15 min, respectively. (We note that these are times for BNN training which can be completed offline on the ground; online in-flight evaluation times of BNNs are extremely fast regardless which training method is used, taking around few milliseconds on a laptop.) Taking into consideration of the computational cost, SVGD seems to be the well-suited

method for the situation currently considered. However, the strength of correlations as well as computational costs may change as different BNN architectures are used and with higher dimensional spaces; these areas will be explored as future work.

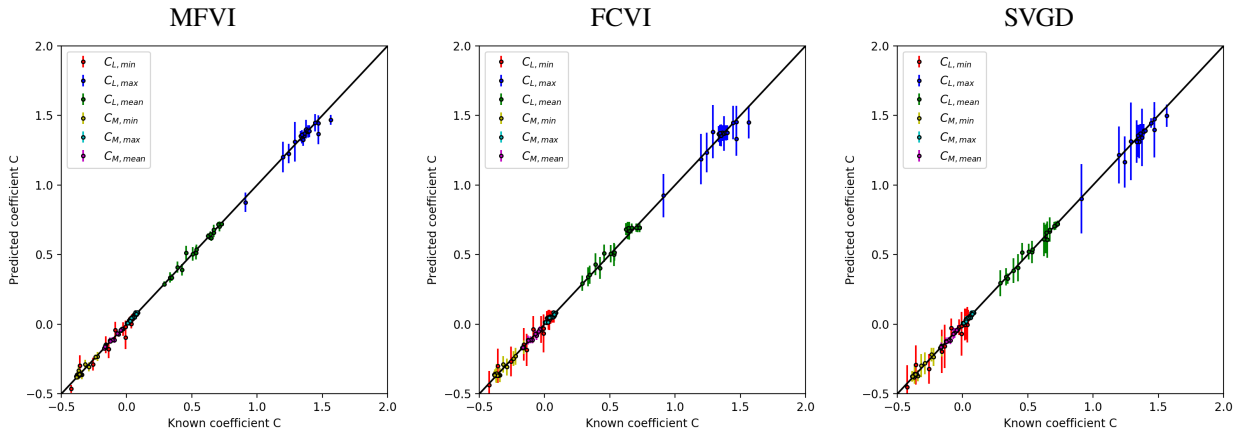


Fig. 10 QoI prediction results for 20 test data (simulations not used for BNN training) based on training with 80 data points. Center dot represents the mean from 10,000 predictions and error bars represent ± 1.96 sample standard deviations.

V. Conclusions

In this paper, we present a proof-of-concept illustration towards developing a real-time in-flight ice detection system based on computational aeroacoustics and Bayesian neural networks (BNNs). BNNs are particularly attractive since they can produce accurate predictions rapidly and thus suitable for real-time usage, and can provide uncertainty information surrounding the predicted values to indicate the quality and credibility of these machine learning computations. This uncertainty information is critical for application of safety related tools, such as icing detection systems. We investigate the effectiveness and tradeoffs among several approximate Bayesian inference techniques for training BNNs: (Gaussian) mean-field variational inference (MFVI), full-covariance variational inference (FCVI), and Stein variational gradient descent (SVGD). We find the correlations in weight uncertainty to be low for this application and with our current BNN setup, although MFVI (which ignores correlations altogether) noticeably under-predicts the variance. SVGD is both computationally fast and captures non-Gaussian and correlation structures, appearing to be a well-suited method for the situation currently considered.

Acknowledgments

The computational resources provided by the RHRK high performance computing center via the 'Elwetritsch' high performance cluster at the TU Kaiserslautern is gratefully acknowledged. The work from Politecnico di Milano has received funding from the European Union's H2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 721920. Further information can be found at the Network for Innovative Training on ROtorcraft Safety (NITROS) project website.

References

[1] NTSB, "National Transportation Safety Board Aviation Accident Final Report," NTSB No. WPR17FA047, December 2017.

[2] NTSB, "National Transportation Safety Board Aviation Accident Final Report," NTSB No. WPR16LA104, June 2018.

[3] NTSB, "National Transportation Safety Board Aviation Accident Final Report," NTSB No. ANC16FA023, January 2018.

[4] Cheng, B., Han, Y., Brentner, K. S., Palacios, J. L., and Morris, P. J., "Quantification of rotor surface roughness due to ice accretion via broadband noise measurement," *Annual Forum Proceedings - AHS International*, Vol. 4, 2014, pp. 2680–2692.

[5] Chen, X., Zhao, Q., Barakos, G. N., and Kusyumov, A., "Numerical analysis of rotor aero-acoustic characteristics for ice detection," *International Journal of Aeroacoustics*, Vol. 18, No. 6-7, 2019, pp. 596–620. doi:10.1177/1475472X19871531.

[6] Zhou, B. Y., Gauger, N. R., Morelli, M., Guardone, A., Hauth, J., and Huan, X., "Towards a Real-Time In-Flight Ice Detection System via Computational Aeroacoustics and Bayesian Neural Networks," *AIAA/ISSMO Multidisciplinary Analysis and Optimization at the AVIATION Forum*, AIAA Paper 2019–3103, Dallas, TX, 2019. doi:10.2514/6.2019-3103.

[7] Economon, T. D., Palacios, F., Copeland, S. R., Lukaczyk, T. W., and Alonso, J. J., "SU2: An Open-Source Suite for Multiphysics Simulation and Design," *AIAA Journal*, Vol. 54, No. 3, 2016, pp. 828–846. doi:10.2514/1.J053813.

[8] LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R., "Efficient BackProp," *Neural Networks: Tricks of the Trade*, edited by G. Montavon, G. B. Orr, and K.-R. Müller, Springer-Verlag Berlin Heidelberg, 2012, pp. 9–48. doi:10.1007/978-3-642-35289-8_3.

[9] Robbins, H., and Monro, S., "A Stochastic Approximation Method," *The Annals of Mathematical Statistics*, Vol. 22, No. 3, 1951, pp. 400–407. doi:10.1214/aoms/1177729586.

[10] Morelli, M., Zhou, B. Y., and Guardone, A., "Simulation and Analysis of Oscillating Airfoil Ice Shapes via a Fully Unsteady Collection Efficiency Approach," *Proceedings of the 75th Annual Forum*, VFS International, Philadelphia, 2019.

[11] Gori, G., Zocca, M., Garabelli, M., Guardone, A., and Quaranta, G., "PoliMIce: A simulation framework for three-dimensional ice accretion," *Applied Mathematics and Computation*, Vol. 267, 2015, pp. 96–107.

[12] Zhou, B., Albring, T. A., Gauger, N. R., Ilario, C., Economon, T. D., and Alonso, J. J., "Reduction of Airframe Noise Components Using a Discrete Adjoint Approach," *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2017–3658, Reston, Virginia, 2017. doi:10.2514/6.2017-3658.

[13] Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (eds.), *Handbook of Markov Chain Monte Carlo*, Chapman and Hall, 2011.

[14] Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D., "Weight Uncertainty in Neural Networks," *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37, 2015, pp. 1613–1622.

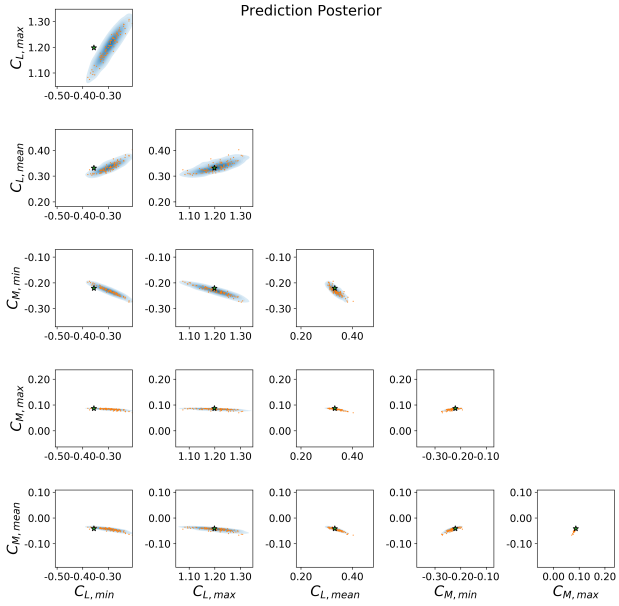
[15] Blei, D. M., Kucukelbir, A., and McAuliffe, J. D., "Variational Inference: A Review for Statisticians," *Journal of the American Statistical Association*, Vol. 112, No. 518, 2017, pp. 859–877. doi:10.1080/01621459.2017.1285773.

[16] Titsias, M. K., and Lázaro-Gredilla, M., "Doubly Stochastic Variational Bayes for non-Conjugate Inference," *Proceedings of the 31st International Conference on Machine Learning*, Vol. 32, Beijing, China, 2014, pp. 1971–1979.

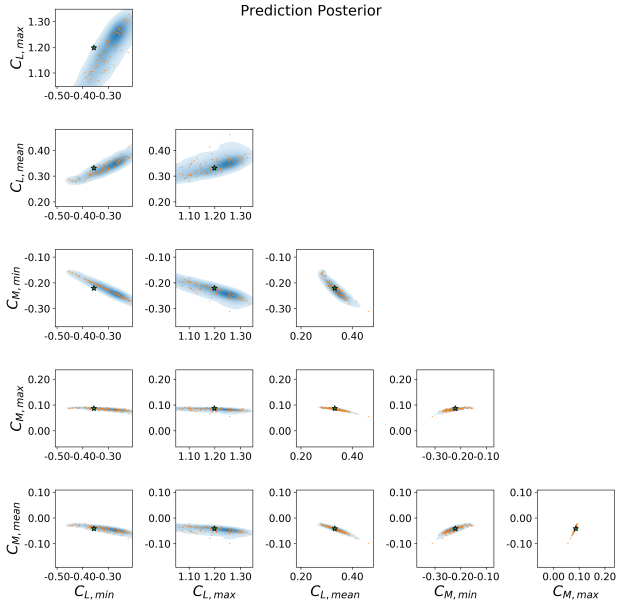
[17] Liu, Q., and Wang, D., "Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm," *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, Barcelona, Spain, 2016, pp. 2378–2386.

[18] Kingma, D. P., and Ba, J., "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, 2015.

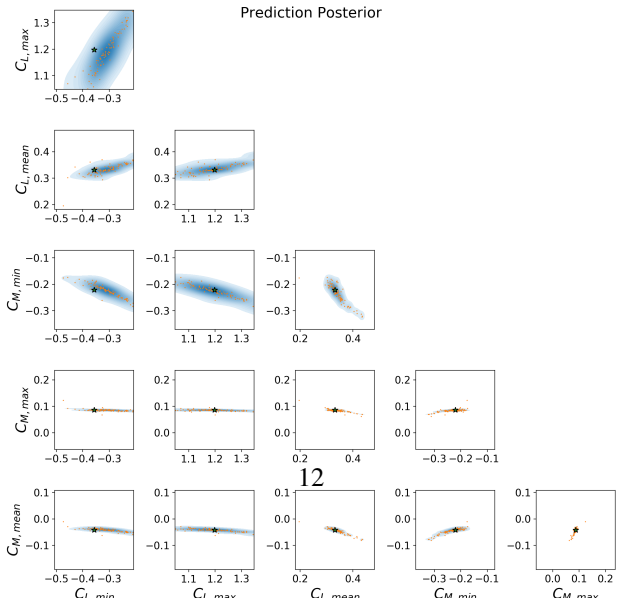
MFVI



FCVI



SVGD



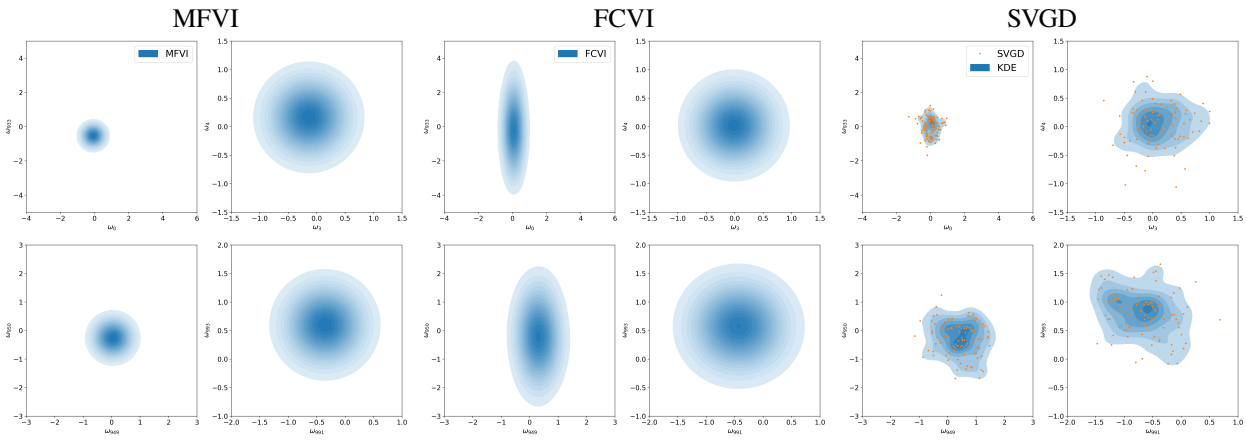


Fig. 12 Pairwise posterior distributions for select weight components from the NN calculated using MFVI, FCVI, and SVGD. These posteriors are constrained to Gaussian distributions in the case of MFVI and FCVI.