A prediction interval method for uncertainty quantification of regression models

Pei Zhang[1], Siyan Liu[1], Dan Lu[1], Guannan Zhang[2], Ramanan Sankaran[1]

[1]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830

[2]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830

Deep Learning for Simulation (simDL), ICLR 2021 Workshop

Summary

This work proposes a novel prediction interval (PI) method for uncertainty quantification (UQ) of regression neural networks (NNs). The method requires no distributional assumption, does not introduce extra hyper-parameters, and can effectively identify out-of-distribution (OOD) samples and quantify their uncertainty.

We demonstrate the advantages of our method in a toy regression task with non-Gaussian noise and two real-world scientific applications, in comparison with two state-of-the-art UQ methods—a quality-driven (QD) approach (Pearce et al., 2018) and a deep ensemble (DE) method (Lakshminarayanan et al., 2017).

Our Method

For a standard regression task, $y = f_w(\boldsymbol{x})$: $\mathbb{R}^d \to \mathbb{R}$, with a given dataset $\mathcal{D}_{train} = \{\boldsymbol{x}_i, y_i\}_{i=1}^N$

Goal: Learn the function $f_w(\boldsymbol{x})$ and the prediction intervals (PIs) to quantify the uncertainty of the prediction

Key idea: Learn $f_w(\boldsymbol{x})$, the upper and lower bounds of the PI separately using three independent NNs.

Step 1: Train $f_w(\boldsymbol{x})$ NN with dataset $\mathcal{D}_{train} = \{\boldsymbol{x}_i, y_i\}_{i=1}^N$ using mean squared error (MSE) loss

Step 2: Obtain two new datasets from $f_w(\boldsymbol{x})$ NN

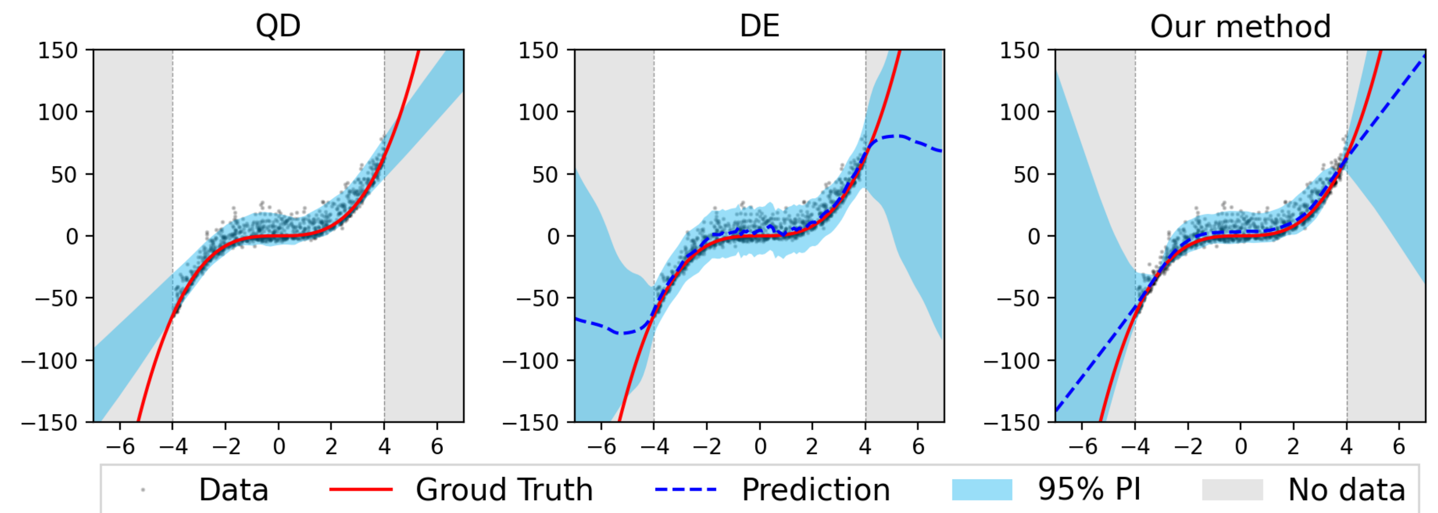
$$\mathcal{D}_{upper} = \{(\boldsymbol{x}_i, y_i - f_w(\boldsymbol{x}_i)) | y_i \geq f_w(\boldsymbol{x}_i), i = 1, \dots, N\}$$
$$\mathcal{D}_{lower} = \{(\boldsymbol{x}_i, f_w(\boldsymbol{x}_i) - y_i) | y_i < f_w(\boldsymbol{x}_i), i = 1, \dots, N\}$$

Step 3: Train two new NNs—$u_{\boldsymbol{\theta}}(\boldsymbol{x})$ and $v_{\boldsymbol{\xi}}(\boldsymbol{x})$—to represent the upper and lower uncertainty profiles with the two datasets $\mathcal{D}_{upper}$ and $\mathcal{D}_{lower}$, respectively, using MSE loss

Step 4: Find two coefficients α and β such that a target percentage γ of training samples are covered by the PI, $[f_w - \beta v_{\xi}, f_w + \alpha u_{\theta}]$

Step 5: Get the final PI as $[f_w - \beta v_{\xi}, f_w + \alpha u_{\theta}]$

A Toy Regression With Non-Gaussian Noise



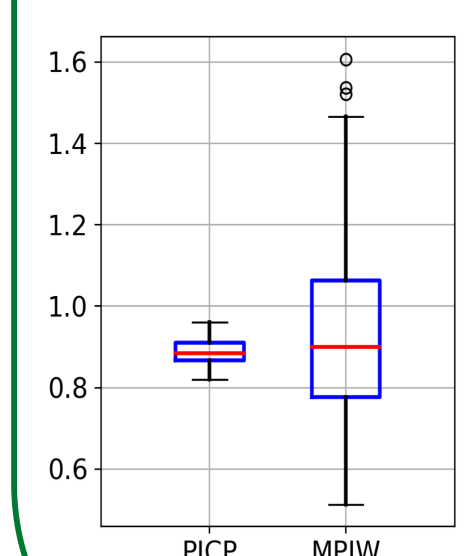
Our method and QD outperform DE by producing tighter bounds on in-distribution (ID) data, as both methods do not impose assumptions on the noise distribution. Our method and DE outperform QD in OOD region by providing more reasonable (wider) PIs.

An Earth System Land Model

	Output 1		Output 2		Output 3		Output 4		Output 5	
	PICP	MPIW	PICP	MPIW	PICP	MPIW	PICP	MPIW	PICP	MPIW
QD	94.6%	2.09	99.2%	2.24	100%	3.09	99.5%	1.76	99.6%	2.33
Our method	90.8%	1.96	90.0%	0.67	91.4%	0.57	91.6%	0.72	90.0%	0.62

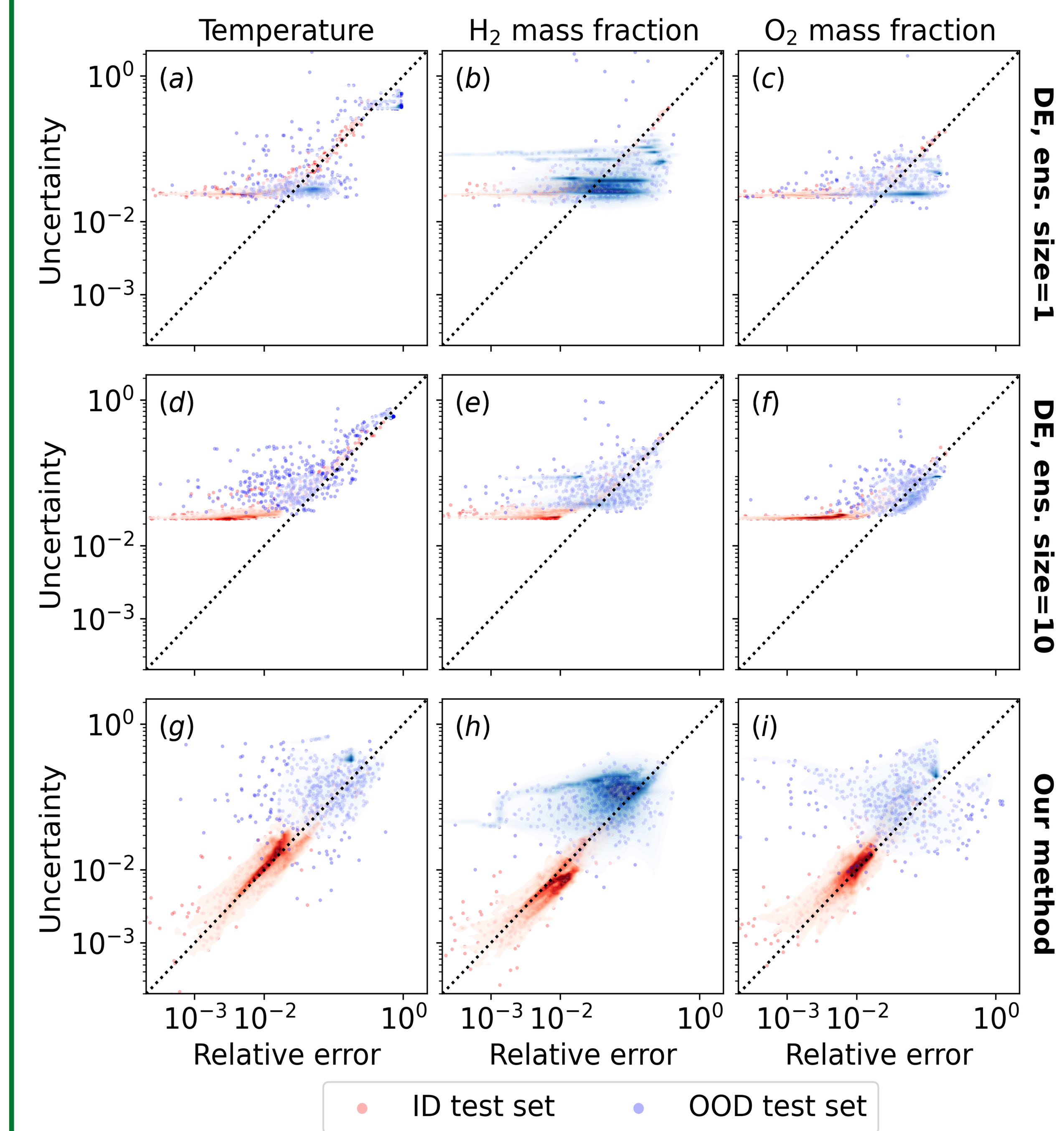
	Output 6		Output 7		Output 8		Output 9		Output 10	
	PICP	MPIW	PICP	MPIW	PICP	MPIW	PICP	MPIW	PICP	MPIW
QD	98.8%	2.48	99.6%	2.15	98.6%	1.94	99.6%	2.12	99.6%	1.75
Our method	90.2%	0.59	90.0%	1.41	90.4%	0.81	90.2%	0.72	90.0%	0.67

PIs are calculated for surrogate models of ten variables from the Earth System Land Model (ELM) simulations. Two metrics—prediction interval coverage probability (PICP) and mean prediction interval width (MPIW)—are used for comparison. A sound PI method should have a PICP close to the targe value with a small MPIW. Our method outperforms QD by providing PICP closer to the 90% target and MPIW on average 60% narrow.



PICP and MPIW provided by QD are sensitive to the hyper-parameters. Our method does not need hyper-parameter fine tuning but QD does.

An OOD-aware Autoencoder-based Combustion Model



PIs are calculated for an autoencoder NN for datasets of syngas CO/H_2 combustion with 12 thermo-chemical state variables. ID test set is from a 0-D reactor and OOD test set is from a 3-D direct numerical simulation (DNS) of turbulent flames.

DE with a single run in the first row fails to capture the difference in uncertainties for ID and OOD samples. DE with 10 runs in the second row shows improved but still limited separation of OOD samples from ID and it fails to produce the uncertainty-error correlation. Our method in the third row shows a strong correlation between the uncertainty and the error, and clearly demonstrates that OOD and ID have different uncertainty magnitudes.

Acknowledgement

This material is based upon work supported in part by the U.S. Department of Energy, Office of Science, Offices of Advanced Scientific Computing Research, under the contract ERKJ352, and by the AI Initiative at the Oak Ridge National Laboratory (ORNL). ORNL is operated by UT-Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725.

References

Tim Pearce, Alexandra Brintrup, Mohamed Zaki, and Andy Neely. High-quality prediction intervals for deep learning: A distribution-free, ensembled approach. In Jennifer Dy and Andreas Krause (eds.), Proceedings of the 35th International Conference on Machine Learning, volume 80 of Proceedings of Machine Learning Research, pp. 4075–4084, Stockholmsmssan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, pp. 64056416, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.