CSVE: Enhancing Uncertainty Quantification in Industrial KPI Prediction

Yiran Ma,a Zhichao Chen,a Xinmin Zhang,a,\* Zhihuan Song,a Manabu Kano,b

aThe State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China

bDepartment of Systems Science, Kyoto University, Kyoto 606-8501, Japan

xinminzhang@zju.edu.cn

Abstract

Data-driven soft sensing technologies have been widely accepted to predict key performance indicators (KPI) automatically and accurately in industrial processes. Despite their utility, quantifying uncertainty—particularly epistemic uncertainty—is frequently neglected. This neglect is noteworthy, as uncertainty offers insights into prediction reliability, which is a crucial concern for industrial operators and engineers. Indeed, uncertainty quality serves as a critical criterion for determining whether it can be safely applied to real-world industrial processes. To address this gap, this work introduces calibrated Stein variational ensemble (CSVE), a method that integrates a formal Bayesian framework with the concept of ensemble. The proposed CSVE method yields calibrated probability density functions as predictions, offering both point predictions and natural uncertainty quantification. Furthermore, for the first time in the context of soft sensors, well-designed metrics are employed to assess uncertainty quality. The effectiveness of the proposed approach is validated through its application to a real-world industrial KPI prediction task. The results reveal that common methods such as bootstrap aggregation and parameter posterior inference methods can display miscalibrated uncertainty, posing reliability risks in industrial applications. In contrast, the proposed approach demonstrably mitigates these issues and enhances the quantification of uncertainty.

**Keywords**: Data-driven modeling, Industrial KPI prediction, Soft sensing, Uncertainty Calibration, Stein Variational Ensemble

* 1. Introduction

Predicting industrial key performance indicators (KPIs) accurately is important for contemporary industrial systems, especially in the realms that require relevant information for downstream tasks. Data-driven soft sensors, represented by discriminative models, have emerged as an innovative solution for predicting KPIs that are traditionally difficult to measure (Kano and Fujiwara, 2013; Zhang et al., 2023). Even though various discriminative models have achieved impressive accuracy in this field, they often neglect quantifying the uncertainty of predictions. The accurate quantification of uncertainty, especially epistemic uncertainty, is crucial in KPI prediction because uncertainty provides insights into the reliability of predictions, enabling more informed decision-making. Moreover, the discriminative models’ limited robustness poses risks to safety-critical applications. This emphasizes the necessity of designing an excellent uncertainty quantification (UQ) method that can integrate with the numerous discriminative models developed for soft sensing.

Considering the lack of stochasticity in discriminative model parameters, bootstrap aggregation (Fortuna et al., 2009) and Monte Carlo (MC) dropout (Cao et al., 2023; Yang et al., 2022) were used in existing studies as straightforward methods to improve robustness or obtain predictive uncertainty. Bayesian Neural Networks (BNNs) based on variational inference (VI) can also be seen in some studies (Lee et al., 2021). However, the existing methods often have inadequate regularization or overly strong assumptions, leading to inadequate exploration in the parameter space. Relative research (Ashukha et al., 2021) has shown that the exploration of different modes in the posterior is crucial for good performance. Therefore, these traditional methods can display miscalibrated uncertainty, posing severe risks to safety-critical applications.

To address these problems, this study introduces the calibrated Stein variational ensemble (CSVE) as a new UQ method for soft sensors. The proposed method integrates a formal Bayesian framework with the concept of ensemble. Meanwhile, the importance of implementing uncertainty calibration for all UQ methods is emphasized. As a result, the proposed methodology yields calibrated probability density functions as predictions, offering better point predictions and better uncertainty quantification. The proposed approach is validated through comparative evaluations on a real-world KPI prediction task in an industrial process, demonstrating its effectiveness in improving epistemic uncertainty quantification.

* 1. Problem Formulation

Data-driven soft sensors are typically constructed and tested on the available historical datasets. However, since the real world is dynamic and variable, their performance may deteriorate gradually without being noticed. This means that relying solely on predictions from these soft sensor models may introduce risks into production processes. Therefore, generating additional information about the reliability of the predictions should be considered. To this end, quantifying uncertainty is a straightforward solution. Consequently, precise uncertainty quantification is as important as predictive accuracy for data-driven models.

**Evaluating Uncertainty Quality.** In this work, besides the negative log-likelihood (NLL) commonly used in Bayesian inference, other metrics that evaluate the uncertainty quality in the regression tasks are employed to assess the performance of the methods. According to (Naeini et al., 2015; Kuleshov et al., 2018; Levi et al., 2022), Expected Calibration Error (ECE) is used to evaluate the quality of credible intervals. Moreover, Expected Normalized Calibration Error (ENCE) (Levi et al., 2022), which is based on the idea that the predicted standard deviation should match the absolute error, is also used as an indicator of uncertainty quality.

* 1. Methodology

As mentioned before, the poor performance of existing UQ methods is caused by inadequate exploration of the parameter space. Stein Variational Gradient Descent (SVGD) (Liu and Wang, 2016), a representative of particle-based variational inference methods (ParVIs), was proposed to address this problem. Mathematically, SVGD simulates the gradient flow that minimizes the KL divergence in the Wasserstein space (Chen et al., 2018). Owing to the kernel used in SVGD, all the particles can affect each other. Therefore, the particles, following the dynamical system ruled by the gradient flow, can make a more comprehensive exploration of the parameter space. Considering this theoretical progress, this study proposes a new UQ method that integrates SVGD with uncertainty calibration to enhance the accuracy and uncertainty quality of industrial KPI predictions. The method proposed in this work is termed as calibrated Stein variational ensemble (CSVE), which comprises two parts: Stein variational ensemble (SVE) and interval-based uncertainty calibration. The details are as follows.

* + 1. Stein Variational Ensemble
			1. Base Model

The ensemble model is composed of multiple base models of the same structure. The principle of constructing the base models is the decomposition of aleatoric (data) uncertainty and epistemic (model) uncertainty. As the posterior approximation performed by SVGD is an estimate of model uncertainty, the data uncertainty should be captured by each base model separately. To guarantee the compatibility of the proposed method, an additional learnable parameter is added to each base model to capture noise inherent in data, and the discriminative model adopted can remain unmodified. Therefore, each base model provides a probability density function (PDF) as its prediction. By aggregating the PDFs provided by the base models, a more comprehensive prediction of the KPI can be obtained.

* + - 1. Update Rule

SVE is composed of base models. The set of parameter vectors of base models is regarded as the particles in the SVGD framework. In each iteration, the gradient of each base model is first computed after a parallel forward pass. Subsequently, the parameter vectors can be updated simultaneously according to the SVGD update rule:

where The asymptotic convergence characteristic of SVGD guarantees that, over time, a set of particles will be obtained that effectively serve as representative samples of the parameter posterior.

* + 1. Intervals-based Uncertainty Calibration

Trained Bayesian models require additional uncertainty calibration to fully leverage their capabilities in uncertainty quantification. After a Bayesian model has been trained using the training data, calibration should be carried out in accordance with its performance on an independent calibration set. Formally, a Bayesian regression model is well-calibrated (Kuleshov et al., 2018) if

where denotes the inverse function of the cumulated density function (CDF), and denotes the indicator function.

Given a calibration set independent of the training set, let denote the predictive CDFs produced by an uncalibrated model. Then, pairs of PCDs are obtained via



Figure 1 The schematic diagram of the blast furnace (Geerdes et al., 2020)

Subsequently, the calibration model is fitted on the dataset such that is calibrated. Finally, the calibrated CDF is used to substitute the uncalibrated CDF for making uncertainty-aware predictions.

* 1. Case Study: Results and Discussion

This section uses the Blast Furnace (BF) ironmaking process as a case study to evaluate the efficacy of the proposed methodology. The BF operates as a sophisticated moving-bed reactor, hosting a range of concurrent activities including complex chemical reactions, mass and heat transfer, alongside the flow of multiphase fluids. This dynamic environment is further detailed in Figure 1, which shows the internal structure of the BF divided into five distinct zones: Throat, Stack (or Shaft), Belly, Bosh, and Hearth. In this setting, raw materials gradually descend towards the lower sections of the furnace, counterbalanced by the upward movement of hot air. The end products of the blast furnace ironmaking process are hot metal and liquid slag, which mainly accumulate in the BF hearth. The extreme conditions prevalent within the BF preclude direct measurements of molten iron quality (Luo et al., 2023). Instead, the silicon content in the hot metal is employed as an indirect indicator of the molten iron quality. Under stable operational conditions, a high silicon content is indicative of excessive energy consumption due to increased temperatures, whereas a low silicon content suggests a disruption in standard reaction processes. Therefore, the precise and real-time estimation of the silicon content is crucial for maintaining the quality of the hot metal.

To construct the predictive model, time series data of 10000 samples were collected from the blast furnace ironmaking plant of Baosteel Group in China. The collected data was divided into a training set (64% of the entire dataset), a calibration set (16% of the entire dataset), and a test set (the remaining 20% of the entire dataset). Considering the time-series characteristic of the process data, it is not appropriate to shuffle the dataset. Therefore, the calibration set is selected as 3 small segments interspersed with the training

Figure 2 Calibration curves

set. To predict the silicon content, 105 process variables, such as blast humidity, blast pressure, and gas flow rate, were selected as the secondary variables.

To validate the proposed CSVE, an MLP with the architecture of 105-128-128-128-1 is taken as an example. Bootstrap aggregation, MC dropout, and BNN-SGVI are adopted as baselines, with the same number of parameter samples .

* + 1. Role of Calibration

The intervals-based uncertainty calibration is performed on all the baselines and CSVE. According to the calibration curves (Kuleshov et al., 2018) in Figure 2, it can be concluded that the uncertainty calibration procedure is effective for generating better credible intervals. As interval-based calibration is a simple post-processing method, it is believed that it should become a standard procedure for various UQ methods.

* + 1. Metrics Evaluation

In this part, each method is applied to the uncertainty calibration procedure for fairness. The metrics evaluation results are reported in Table 1. CSVE is not only significantly superior in accuracy metrics (MSE, MAE, R2) compared to other methods but also exceeds in approximation quality (NLL), displaying strong point prediction and uncertainty quantification capabilities. As for calibration performance (ECE, ENCE), the proposed method gives competitive results. It is noteworthy that although MC dropout seems to be proficient in generating precise credible intervals of predictions, it still should not be the first choice for UQ considering its worst accuracy caused by dropout implementation and awful assumption of the posterior form.

Table 1 Metrics evaluation. \* or \*\* marks the methods that our method significantly outperforms at a p-value < 0.05 or p-value < 0.01 via a paired samples t-test, respectively. Bolded results indicate the best in each metric. Underlined results indicate the second-best in each metric. The *standard deviation* is also reported, denoted by the figures to the right of ±.

|  |  |  |
| --- | --- | --- |
| Methods | Accuracy Metrics | Uncertainty Quality Metrics |
| MSE | MAE | R2 | NLL | ECE | ENCE |
| BNN-SGVI (Cali.) | 0.0710±0.013\* | 0.2074±0.023\* | 0.8971±0.019\* | 0.0492±0.063\* | 0.0430±0.023\* | 0.2070±0.114\* |
| MC Dropout (Cali.) | 0.0788±0.003\*\* | 0.2139±0.005\*\* | 0.8858±0.005\*\* | 0.1736±0.032\*\* | **0.0107±0.004** | 0.1038±0.020 |
| Bootstrap (Cali.) | 0.0695±0.005\*\* | 0.2036±0.008\*\* | 0.8994±0.007\*\* | 0.0347±0.030\*\* | 0.0383±0.012\*\* | 0.2429±0.048\*\* |
| CSVE (ours) | **0.0587±0.009** | **0.1852±0.016** | **0.9150±0.013** | **-0.0140±0.023** | 0.0243±0.009 | **0.0962±0.030** |

* 1. Conclusions

In this work, the necessity of improving the uncertainty quantification of soft sensing models in industrial KPI prediction is first emphasized. Then, a new uncertainty quantification method called CSVE is proposed. CSVE is designed based on SVGD and intervals-based uncertainty calibration. Like most deep ensemble methods, the proposed modeling framework can enhance existing differentiable soft sensor models without any modifications to the main model structure. In the case study on the blast furnace ironmaking process, it can be observed that both the uncertainty quality and accuracy metrics of the proposed method are improved due to the precise posterior approximation. Additionally, the experimental results also support the proposal of applying uncertainty calibration as a standard procedure for the UQ methods. Future work will validate the proposed methodology on different models, such as recurrent neural networks and Transformers.

References

Ashukha, A., Lyzhov, A., Molchanov, D., Vetrov, D., 2021, Pitfalls of In-Domain Uncertainty Estimation and Ensembling in Deep Learning.

Cao, L., Zhang, H., Meng, Z., Wang, X., 2023, A parallel GRU with dual-stage attention mechanism model integrating uncertainty quantification for probabilistic RUL prediction of wind turbine bearings. Reliability Engineering & System Safety, 235, 109197.

Chen, C., Zhang, R., Wang, W., Li, B., Chen, L., 2018, A Unified Particle-Optimization Framework for Scalable Bayesian Sampling.

Fortuna, L., Graziani, S., Xibilia, M.G., 2009, Comparison of Soft-Sensor Design Methods for Industrial Plants Using Small Data Sets. IEEE Transactions on Instrumentation and Measurement, 58, 2444–2451.

Geerdes, M., Chaigneau, R., Lingiardi, O., 2020, Modern Blast Furnace Ironmaking: An Introduction (Fourth Edition, 2020). IOS Press.

Kano, M., Fujiwara, K., 2013, Virtual Sensing Technology in Process Industries: Trends and Challenges Revealed by Recent Industrial Applications. Journal of Chemical Engineering of Japan, 46, 1–17.

Kuleshov, V., Fenner, N., Ermon, S., 2018, Accurate Uncertainties for Deep Learning Using Calibrated Regression, in: Proceedings of the 35th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 2796–2804.

Lee, M., Bae, J., Kim, S.B., 2021, Uncertainty-aware soft sensor using Bayesian recurrent neural networks. Advanced Engineering Informatics, 50, 101434.

Levi, D., Gispan, L., Giladi, N., Fetaya, E., 2022, Evaluating and Calibrating Uncertainty Prediction in Regression Tasks. Sensors 22, 5540.

Liu, Q., Wang, D., 2016. Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Luo, Y., Zhang, X., Kano, M., Deng, L., Yang, C., Song, Z., 2023, Data-driven soft sensors in blast furnace ironmaking: a survey. Frontiers of Information Technology & Electronic Engineering, 24, 327–354.

Naeini, M.P., Cooper, G., Hauskrecht, M., 2015, Obtaining Well Calibrated Probabilities Using Bayesian Binning. Proceedings of the AAAI Conference on Artificial Intelligence 29.

Yang, J., Peng, Y., Xie, J., Wang, P., 2022, Remaining Useful Life Prediction Method for Bearings Based on LSTM with Uncertainty Quantification. Sensors 22, 4549.

Zhang, X., He, B., Zhu, H., Song, Z., 2023, Information Complementary Fusion Stacked Autoencoders for Soft Sensor Applications in Multimode Industrial Processes. IEEE Transactions on Industrial Informatics, 1–11.