POLITECNICO DI TORINO
Repository ISTITUZIONALE

Robust Ensemble of Computational Techniques for the Detection of Outliers in the Inverse Uncertainty Quantification with Limited Data

Original
Robust Ensemble of Computational Techniques for the Detection of Outliers in the Inverse Uncertainty Quantification with Limited Data / Pedroni, N.. - ELETTRONICO. - (2023), pp. 19647-19647. (UNCECOMP 2023 5 th ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering Athens, Greece 12-14 June 2023).

Availability:
This version is available at: 11583/2986313 since: 2024-02-24T18:54:57Z

Publisher:
National Technical University of Athens

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

15 November 2026

Robust Ensemble of Computational Techniques for the Detection of Outliers in the Inverse Uncertainty Quantification with Limited Data

Nicola Pedroni*

Politecnico di Torino
Dipartimento di Energia
Corso Duca degli Abruzzi 24, 10129 Torino (Italy)
nicola.pedroni@polito.it

ABSTRACT

In the analysis of safety-critical systems, the outputs of the simulation models can diverge substantially from the actual system dynamics. This discrepancy is due to the aleatory (randomness) and epistemic (lack of knowledge) uncertainties affecting the system behavior and modelling, respectively. In this view, Model Calibration (MC) or Inverse Uncertainty Quantification (IUQ) is of paramount importance.

However, in real engineering applications the data collection can be prohibitively expensive. Thus, the obtained data can be insufficient to accurately represent the true system response. In such situations, IUQ methods may lead to biased results. In addition, due to malfunctioning of sensors and measurement technologies, human errors, external contamination or unexpected deviations in the system behavior, outliers can be found in the collected samples. In these cases, the outcomes of MC or IUQ may be misleading or even erroneous, especially in the presence of limited data [1].

To address this problem, the paper proposes a framework for the robust detection of outliers in the IUQ of the dynamic models of safety-critical systems, in the presence of scarce experimental data. The approach is based on an ensemble of three statistical and artificial intelligence techniques, relying on diverse principles: (i) density-based Isolation Forest (IF) [2], which detects outliers by isolating anomalies from normal points through an ensemble of trees; (ii) Finite Mixture Models (FMMs) [3], that provide an unsupervised probabilistic clustering of the data and identify the outliers as those observations characterized by the smallest likelihood; and (iii) Sparse Stacked AutoEncoders (SSAE) [4], which are trained to reconstruct multivariate data and spot out abnormal samples as those characterized by the largest reconstruction error. Each algorithm is originally tailored to compute anomaly scores for all the available observations, in order to quantify their “degree of outlyingness”. These three (possibly different) rankings are finally combined using the Borda Count to get a unique, robust and reliable ranking.

The proposed ensemble-based approach is embedded within an inverse quantification of mixed probabilistic (aleatory) and set-based (epistemic) uncertainties, in the presence of scarce functional (time series) data. It is tested on two case studies: (i) a toy problem, where the true outliers and uncertainty models are known; (ii) the NASA Langley Uncertainty Quantification Challenge on Optimization Under Uncertainty [5]. The method is applied to data sets with different degrees of outlier contamination (i.e., 1%, 5% and 10%) producing calibration results that are strongly robust to outliers.

References

[1] L.G. Crespo, B. K. Colbert, S. P. Kenny, D. P. Giesy. On the quantification of aleatory and epistemic uncertainty using Sliced-Normal distributions, Systems & Control Letters 134 (2019) 104560.
[2] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation-based anomaly detection, ACM Transactions on Knowledge Discovery from Data 6 (2012) 3.
[3] M.A.T. Figueiredo, A.K. Jain, Unsupervised learning of finite mixture models. In IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 381–396.
[4] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring. Mech Syst Signal Process 115 (2019) 213–37.
[5] L.G. Crespo, S.P. Kenny, The NASA Langley challenge on optimization under uncertainty. Mechanical Systems and Signal Processing 152 (2021) 107405.