# A Probabilistic Machine Learning Pipeline Using Topological Descriptors for Real-Time State Estimation of High-Rate Dynamic Systems

Yang Kang Chua[1], Daniel Coble[2], Arman Razmarashooli[3], Steve Paul[1], Daniel A. Salazar Martinez[3], Chao Hu[1], Austin R.J. Downey[2,4], and Simon Laflamme[3,5]

[1] Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269, USA
[2] Department of Mechanical Engineering, University of South Carolina, Columbia, USA
[3] Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA, 50010, USA
[4] Department of Civil and Environmental Engineering, University of South Carolina, Columbia, USA
[5] Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, 50010, USA

**Abstract.** High-rate systems refer to structures that undergo rapid changes, exhibiting dynamics that undergo changes in short durations, often less than 100 milliseconds. Examples include hypersonic vehicles, active blast mitigation, and ballistic packages. Developing feedback control systems requires state estimations that can be updated on timescales of less than one millisecond. However, due to the nonlinear and non-stationary dynamics of high-rate systems, they entail high uncertainties, posing challenges for predictive modeling. In this study, we propose a probabilistic machine-learning pipeline for estimating the state of a high-rate system. This approach involves applying probabilistic models and topological data analysis techniques to extract features from the datasets obtained from the Dynamic Reproduction of Projectiles in Ballistic Environments for Advanced Research (DROPBEAR) testbed. We examine the design of probabilistic models for structure state estimation, emphasizing the importance of prediction intervals. We evaluate the best model through several performance metrics, such as mean absolute error, Signal to Noise Ratio, and Time Response Assurance Criterion while assessing the quality of predictive uncertainty by creating uncertainty calibration curves and calculating the Expected Confidence Error (ECE). The incorporation of probabilistic machine learning enables decision-makers to make informed decisions under uncertainty, enhancing the practical utility of the pipeline. The pipeline's robustness to signal noise and its ability to handle spurious data are presented and discussed.

**Keywords:** Structural health monitoring · high-rate systems · nonlinear time series, topological data analysis · probabilistic machine learning · uncertainty quantification.

# 1 Introduction

High-rate systems undergo rapid and extreme changes within short timeframes, common in aerospace, automotive safety, and structural engineering. These systems experience dynamic events with magnitudes surpassing 100 $g_n$ and durations under 100 milliseconds [4]. Examples include hypersonic vehicles, active blast mitigation, and ballistic packages [11, 20, 23]. Real-time feedback on structural integrity can enhance survivability. However, this research is challenging due to large uncertainties in external loads, high non-stationarities, heavy disturbances, and unmodeled dynamics from system configuration changes [12].

Various methods, including physics-based and data-based techniques, address these challenges. Physics-based modeling, such as real-time model updating for high-rate dynamics [5] and a model reference adaptive system achieving sub-millisecond computational speeds [24], has proven effective. Data-based techniques involve machine learning models like Long Short-Term Memory (LSTM) for high-rate state estimation [2, 3]. These approaches tackle issues like data scarcity and the complexity of modeling high-rate dynamics [10]. However, data-driven methods often lack uncertainty quantification (UQ), which is crucial for making informed decisions based on prediction confidence levels [17]. Incorporating UQ enhances the reliability and effectiveness of decision-making.

This paper develops a High-Rate Structural State Estimation Pipeline (HR-SSEP), a probabilistic machine learning pipeline for real-time structural state estimation and uncertainty quantification in high-rate dynamic systems. We explore integrating Topological Data Analysis (TDA) for feature extraction. The pipeline's performance is evaluated using laboratory datasets from the Dynamic Reproduction of Projectiles in Ballistic Environments for Advanced Research (DROPBEAR) testbed [13].

The rest of the paper is organized as follows: Section 2 covers the DROPBEAR testbed and Monte Carlo (MC) Dropout. Section 3 details the methodology for developing the machine learning pipeline. Section 4 presents and discusses the results. Section 5 concludes with recommendations for future work.

# 2 Background

This section presents background on the experimental testbench and Monte Carlo Dropout.

## 2.1 DROPBEAR experimental testbed

The Dynamic Reproduction of Projectiles in Ballistic Environments for Advanced Research (DROPBEAR) experimental testbed, shown in Fig. 1, was developed to study high-rate dynamic systems [13]. It comprises a 51 x 6 x 350 mm beam equipped with a single accelerometer (model 393B04 by PCB Piezotronics) mounted at the beam's free edge. The testbed incorporates a movable roller support system, allowing controlled variation in boundary conditions during experiments. This roller follows a predefined profile ranging from 48 mm to 175 mm, initiating vibrations in the beam without requiring extraneous inputs. Experimental tests involved various input profiles, including six square wave inputs, six sinusoidal inputs, and six impulse inputs, each designed to elicit

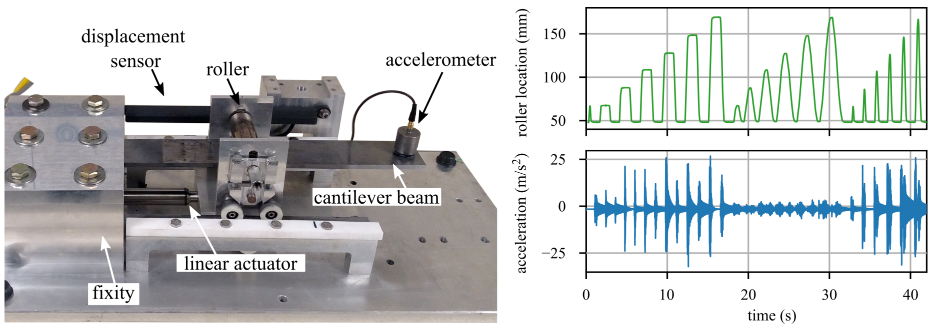



Fig. 1: DROPBEAR experimental setup along with the displacement and acceleration signals [22].

specific structural responses. Data acquisition during experiments was conducted using a 14-bit ADC for the linear transducer (SPS-L225-HALS by Honeywell) and a 24-bit IEPE ADC for acceleration data (NI-9234). These measurements provide insights into the dynamic behavior of structures under ballistic environments, facilitating the development and validation of advanced state estimation techniques. The dataset used in this work is made available through a public repository [22].

### 2.2 MC Dropout

MC dropout, initially introduced as a regularization technique to mitigate overfitting in deep neural networks (DNNs) [19], has emerged as a promising method for approximating posterior predictive distributions in Bayesian neural networks (BNNs)[8].

In MC dropout, dropout layers are added after each fully connected layer of the DNN. During training, these dropout layers introduce randomness by stochastically dropping connections between neurons, creating a randomized sparse network. At test time, multiple forward passes through the model are conducted with different dropout patterns, resulting in an ensemble of predictions. This ensemble can then be leveraged to estimate prediction uncertainty.

One of the key advantages of MC dropout is its simplicity of implementation, requiring minimal modifications to existing DNN architectures. It exhibits low computational cost and scalability, making it applicable to various types of neural networks, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [9, 7].

## 3 Methodology

The methodology for this study involves investigating probabilistic machine learning pipelines for high-rate state estimation using recent research on feature extraction through TDA [18]. An overview of the HR-SSEP is shown in figure 2.

HR-SSEP begins with a low-pass filter to the raw data to enhance quality and reduce noise. After preprocessing, TDA feature extraction is applied to the

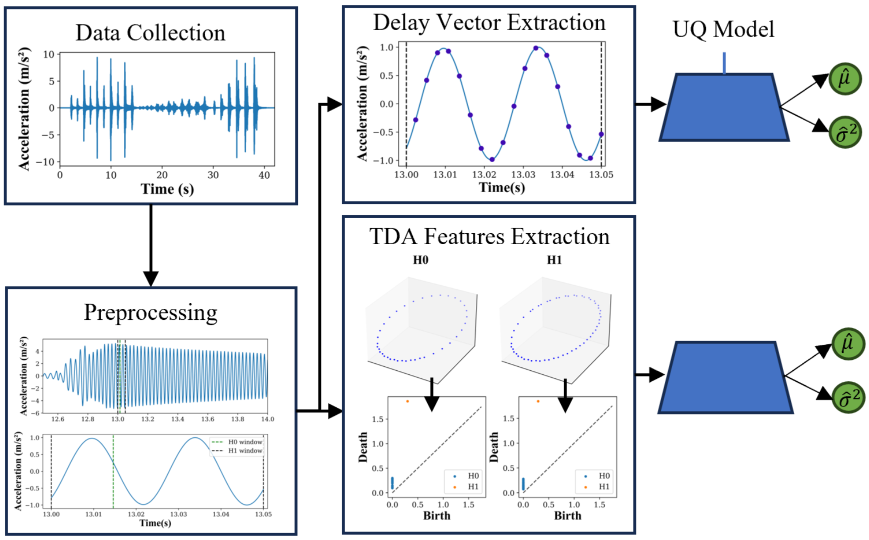



Fig. 2: Overview of the probabilistic machine learning pipeline using TDA features extraction and delay vector extraction

filtered data. Initially, the collected signal data are one-dimensional. To use TDA techniques, the signals must be transformed into point clouds using Takens' embedding [21]. Time series data $x(t)$ is converted into delay vectors $\chi(t) = [x(t), x(t-\tau), x(t-2\tau), \ldots, x(t-(d-1)\tau)]$, where $\tau$ is the time delay and $d$ is the embedding dimension. According to Takens' theorem, proper selection of $d$ and $\tau$ preserves the topological characteristics of the original system. Typically, $d$ is determined using the false nearest neighbor test, and $\tau$ is found using mutual information [15, 14].

To address nonstationarity from the moving boundary condition, two sliding windows are used to extract local TDA features, assuming stationarity within these windows. Given the sampling frequency, maximum frequency ($f_{\max}$), and minimum frequency ($f_{\min}$) of the data, two sliding windows (H0 and H1) are constructed with calculated window sizes. The time delay ($\tau$) is calculated using the equation: $\tau = \frac{0.25}{f_{\max}}$.

This ensures that the embedded signal forms a unit circle at $f_{\max}$, distinguishing shapes at lower frequencies, which resemble ellipses. For H0 feature extraction, the window size is $\frac{1}{f_{\max}} + 2\tau$, and for H1 feature extraction, it is $\frac{1}{f_{\min}} + 2\tau$. These windows are min-max scaled to standardize them, focusing on temporal aspects rather than amplitude. The scaled windows are converted into point clouds using calculated time delays ($\tau$). With an embedding dimension $d$ set to 3, TDA techniques, specifically persistent homology [6], are applied to extract 12 features (6 from H0 and 6 from H1) that numerically represent the point cloud's shape. Feature selection is then performed using correlation heatmap analysis, resulting in 11 selected features. To compare, we use a baseline method of delayed vector extraction, employing the H1 window size and setting the embedding dimension to 11 for a fair comparison with TDA features. Both methods' input data are normalized to ensure consistent scaling, promoting stable training and improved convergence.

The augmented dataset then serves as the training input to the MC Dropout algorithm. For MC Dropout, the number of layers, hidden units, and ensemble size are set to 3 layers with 64 hidden units each and 15 forward passes for making predictions. The model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. The training process was conducted for 1000 epochs, with early stopping based on validation loss to prevent overfitting. The algorithm was trained on standard (STD) movement profiles. Subsequently, the model's generalization was tested using data from stepwise movement (STM) profiles with 10, 30, and 60 steps, and random movement (RND) profiles.

## 4 Results and Discussion

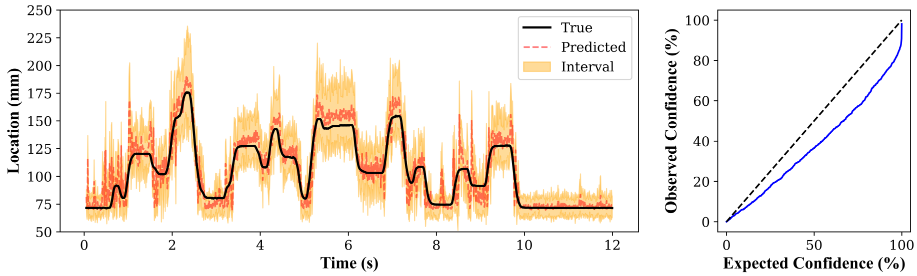
The prediction results for the random movement profile using MC Dropout are shown in Figure 3. The performance of MC Dropout under two different feature inputs is summarized in Table 1. The metrics used for evaluation include mean absolute error (MAE), signal-to-noise ratio measured in decibels ($\mathrm{SNR_{dB}}$), and time response assurance criterion (TRAC). TRAC, which ranges from 0 to 1, measures the correlation between two-time series signals, where a higher value indicates a stronger correlation. The equation for TRAC is given in equation 2 for the reference time series vector y and prediction vector $\hat{y}$ [1]. Additionally, Expected Confidence Error (ECE) is reported, representing the weighted average discrepancy between predicted and actual uncertainties, providing insight into the model's confidence calibration [16]. Furthermore, the average prediction time for each input sample was 278 $\mu$s, meeting the requirement for predictions to be made in under 100 ms.

$$\mathrm{TRAC} = \frac{(y^T \hat{y})^2}{(y^T y)(\hat{y}^T \hat{y})}. \tag{1}$$

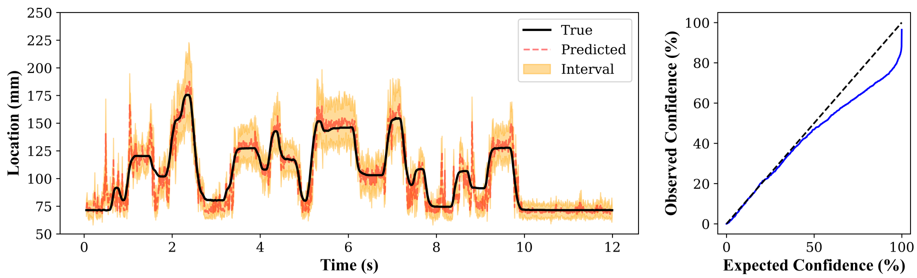
Table 1: Results for MC Dropout

| | Delay vector input | | | | TDA features input | | | |
|---|---|---|---|---|---|---|---|---|
| Data types | MAE (mm) | TRAC | $\mathrm{SNR_{dB}}$ | ECE (%) | MAE (mm) | TRAC | $\mathrm{SNR_{dB}}$ | ECE (%) |
| STD | 8.92 | 0.9875 | 18.15 | 22.03 | 7.74 | 0.9870 | 18.73 | 8.75 |
| STM - 10 | 8.24 | 0.9940 | 21.05 | 6.35 | 6.86 | 0.9933 | 21.55 | 5.52 |
| STM - 30 | 8.81 | 0.9938 | 20.75 | 10.80 | 7.41 | 0.9921 | 20.80 | 2.96 |
| STM - 60 | 9.53 | 0.9931 | 20.26 | 12.98 | 8.29 | 0.9909 | 20.11 | 3.43 |
| RND | 7.95 | 0.9905 | 19.65 | 11.76 | 6.99 | 0.9906 | 20.29 | 6.16 |

The improved performance of the TDA features is evident from the lower MAE, higher TRAC, and higher $\mathrm{SNR_{dB}}$ values, as well as lower ECE (%) values compared to the delay vector input. The lower ECE (%) values associated




(a) Result using Delay vector as input




(b) Result using TDA features as input.

Fig. 3: Prediction results using MC Dropout algorithm and calibration curve (right).

with TDA features indicate better calibration than the delay vector input. This suggests that utilizing topological features extracted through TDA enhances the accuracy and robustness of the MC Dropout model in estimating high-rate states. Overall, these results underscore the potential of TDA-based feature extraction in improving the performance of probabilistic machine learning models for high-rate state estimation tasks.

## 5 Conclusion

In this work, a probabilistic machine-learning pipeline for high-rate state estimation was developed with the combination of TDA feature extraction and traditional feature engineering techniques. TDA feature extraction exhibited superior performance over delay vector extraction by comparing the results with a baseline feature extraction method and a non-probabilistic model. Notably, the Expected Confidence Error (ECE) metric for TDA features was lower, indicating improved calibration compared to the baseline. Despite this, the TDA-based features demonstrated lower MAE, higher TRAC, and higher $\mathrm{SNR}_{\mathrm{dB}}$, underscoring their effectiveness in enhancing model accuracy and robustness.

The findings of this study highlight the promise of TDA-based feature extraction in improving the performance of high-rate state estimation models. Future research may include stabilizing the metric for the probabilistic model by addressing run-to-run variation and exploring additional ways to optimize and generalize the proposed probabilistic machine-learning pipeline. This extension could involve incorporating a forecasting ability into the probabilistic model.

**Acknowledgments.** The authors would like to acknowledge the financial support from the Defense Established Program to Stimulate Competitive Research (DEPSCoR) award number FA9550-22-1-0303, the Air Force Office of Scientific Research (AFOSR) award number FA9550-23-1-0033, and the National Science Foundation awards numbers CCF-1937460 and CCF-2234919. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

## References

1. Avitabile, P., Pingle, P.: Prediction of full field dynamic strain from limited sets of measured data. Shock and Vibration **19**(5), 765–785 (2012). https://doi.org/10.1155/2012/408919, http://dx.doi.org/10.1155/2012/408919
2. Barzegar, V., Laflamme, S., Hu, C., Dodson, J.: Ensemble of recurrent neural networks with long short-term memory cells for high-rate structural health monitoring. Mechanical Systems and Signal Processing **164**, 108201 (Feb 2022). https://doi.org/10.1016/j.ymssp.2021.108201, http://dx.doi.org/10.1016/j.ymssp.2021.108201
3. Coble, D., Satme, J., Kabir, E., Downey, A., Bakos, J., Andrews, D., Huang, M., Moura, A., Dodson, J.: Towards online structural state-estimation with submillisecond latency (03 2023)
4. Dodson, J., Downey, A., Laflamme, S., Todd, M.D., Moura, A.G., Wang, Y., Mao, Z., Avitabile, P., Blasch, E.: High-rate structural health monitoring and prognostics: An overview. In: Madarshahian, R., Hemez, F. (eds.) Data Science in Engineering, Volume 9. pp. 213–217. Springer International Publishing, Cham (2022)
5. Downey, A., Hong, J., Dodson, J., Carroll, M., Scheppegrell, J.: Millisecond model updating for structures experiencing unmodeled high-rate dynamic events. Mechanical Systems and Signal Processing **138**, 106551 (Apr 2020). https://doi.org/10.1016/j.ymssp.2019.106551, http://dx.doi.org/10.1016/j.ymssp.2019.106551
6. Edelsbrunner, H., Harer, J., et al.: Persistent homology-a survey. Contemporary mathematics **453**(26), 257–282 (2008)
7. Gal, Y., Ghahramani, Z.: Bayesian convolutional neural networks with bernoulli approximate variational inference (2015). https://doi.org/10.48550/ARXIV.1506.02158, https://arxiv.org/abs/1506.02158
8. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning (2015). https://doi.org/10.48550/ARXIV.1506.02142, https://arxiv.org/abs/1506.02142
9. Gal, Y., Ghahramani, Z.: A theoretically grounded application of dropout in recurrent neural networks (2015). https://doi.org/10.48550/ARXIV.1512.05287, https://arxiv.org/abs/1512.05287
10. Guo, L., Guo, Y., Billings, S.A., Coca, D.: Approximate observability of infinite dimensional bilinear systems using a volterra series expansion. Systems & Control Letters **75**, 20–26 (2015)
11. Hallion, R.P., Bedke, C.M., Schanz, M.V.: Hypersonic weapons and us national security: A 21st century breakthrough (2016), http://docs.wixstatic.com/ugd/a2dd91b7016a5428ff42c8a21898ab9d0ec349.pdf

12. Hong, J., Laflamme, S., Dodson, J., Joyce, B.: Introduction to state estimation of high-rate system dynamics. Sensors **18**(2), 217 (Jan 2018). https://doi.org/10.3390/s18010217, http://dx.doi.org/10.3390/s18010217
13. Joyce, B., Dodson, J., Laflamme, S., Hong, J.: An experimental test bed for developing high-rate structural health monitoring methods. Shock and Vibration **2018**, 1–10 (Jun 2018). https://doi.org/10.1155/2018/3827463, http://dx.doi.org/10.1155/2018/3827463
14. Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical Review A **45**(6), 3403–3411 (Mar 1992). https://doi.org/10.1103/physreva.45.3403, http://dx.doi.org/10.1103/PhysRevA.45.3403
15. Kreer, J.: A question of terminology. IRE Transactions on Information Theory **3**(3), 208–208 (Sep 1957). https://doi.org/10.1109/tit.1957.1057418, http://dx.doi.org/10.1109/TIT.1957.1057418
16. Kuleshov, V., Fenner, N., Ermon, S.: Accurate uncertainties for deep learning using calibrated regression (2018). https://doi.org/10.48550/ARXIV.1807.00263, https://arxiv.org/abs/1807.00263
17. Nemani, V., Biggio, L., Huan, X., Hu, Z., Fink, O., Tran, A., Wang, Y., Zhang, X., Hu, C.: Uncertainty quantification in machine learning for engineering design and health prognostics: A tutorial (2023). https://doi.org/10.48550/ARXIV.2305.04933, https://arxiv.org/abs/2305.04933
18. Razmarashooli, A., Chua, Y.K., Barzegar, V., LIU, H., Laflamme, S., Hu, C., Downey, A.R., Dodson, J.: Topological data analysis for real-time extraction of time series features. STRUCTURAL HEALTH MONITORING 2023 (2023)
19. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research **15**(56), 1929–1958 (2014), http://jmlr.org/papers/v15/srivastava14a.html
20. Stein, C., Roybal, R., Tlomak, P., Wilson, W.: A review of hypervelocity debris testing at the air force research laboratory. Space Debris **2**(4), 331–356 (2000). https://doi.org/10.1023/b:sdeb.0000030024.23336.f5, http://dx.doi.org/10.1023/b:sdeb.0000030024.23336.f5
21. Takens, F.: Detecting strange attractors in turbulence, p. 366–381. Springer Berlin Heidelberg (1981). https://doi.org/10.1007/bfb0091924, http://dx.doi.org/10.1007/BFb0091924
22. Vereen, A., Downey, A., Dodson, J., Moura, A.G.: Dataset-8-dropbear-acceleration-vs-roller-displacement (Aug 2023), https://github.com/High-Rate-SHM-Working-Group/Dataset-8-DROPBEAR-Acceleration-vs-Roller-Displacement
23. Wadley, H., Dharmasena, K., He, M., McMeeking, R., Evans, A., Bui-Thanh, T., Radovitzky, R.: An active concept for limiting injuries caused by air blasts. International Journal of Impact Engineering **37**(3), 317–323 (Mar 2010). https://doi.org/10.1016/j.ijimpeng.2009.06.006, http://dx.doi.org/10.1016/j.ijimpeng.2009.06.006
24. Yan, J., Laflamme, S., Hong, J., Dodson, J.: Online parameter estimation under non-persistent excitations for high-rate dynamic systems. Mechanical Systems and Signal Processing **161**, 107960 (2021)