2026 NDIA MICHIGAN CHAPTER
GROUND VEHICLE SYSTEMS ENGINEERING
AND TECHNOLOGY SYMPOSIUM
APPLIED ARTIFICIAL INTELLIGENCE (AAI) TECHNICAL SESSION
AUG. 11-13, 2026 - NOVI, MICHIGAN

TERRAIN-AWARE PROBABILISTIC PREDICTION OF SUSPENSION COMPONENT FORCES IN UNMANNED GROUND VEHICLES

Yang Kang Chua[1], Mohammad Mundiwala[1], Xudong Wang[2], Matthew Castanier, PhD[3], Zhen Hu, PhD[2], and Chao Hu, PhD[1]

[1]School of Mechanical, Aerospace, and Manufacturing Engineering, University of Connecticut, Storrs, CT 06269, USA
[2]Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, MI 48128, USA
[3]U.S. Army Combat Capabilities Development Command (DEVCOM) Ground Vehicle Systems Center, United States , Warren, MI 48397, USA

ABSTRACT

Unmanned ground vehicles (UGVs) operating in unstructured environments must account not only for terrain traversability but also for terrain-induced loads that affect component durability. Existing path planning approaches primarily consider obstacle avoidance and mobility, while neglecting cumulative structural degradation due to repeated loading. High-fidelity physics-based simulations can capture these effects but are computationally prohibitive for real-time applications. This study investigates machine learning-based surrogate models for predicting vehicle component reaction forces from terrain height sequences generated using a controlled parametric terrain formulation. Both feed-forward and recurrent neural network architectures are evaluated, and ensemble-based probabilistic techniques are incorporated to quantify predictive uncertainty. Results show that the ensemble long short-term memory (Ens-LSTM) model achieves the lowest prediction error (mean absolute error of 0.621 kN) while maintaining narrow 95% prediction intervals (3.22–4.28 kN). A simpler ensemble feed-forward network (Ens-NN) achieves comparable accuracy (0.626 kN) with reduced model complexity. These results demonstrate that data-driven surrogate models can provide accurate and uncertainty-aware force predictions, enabling the integration of structural reliability considerations into fatigue-aware path planning for UGVs.

Citation: Yang Kang Chua, Mohammad Mundiwala, Xudong Wang, Matthew Castanier, Zhen Hu, and Chao Hu, "Terrain-Aware Probabilistic Models for Vehicle Component Force Prediction," In *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium* (GVSETS), NDIA, Novi, MI, Aug. 11-13, 2026.

1. INTRODUCTION

Unmanned ground vehicles (UGVs) have gained increasing attention for their ability to operate in harsh or repetitive environments where human intervention is difficult or unsafe [1, 2]. These systems are widely deployed in military, industrial, and agricultural applications, including reconnaissance missions, off-road navigation, and field operations where vehicles frequently traverse highly veraible terrain [3-5]. To address these challenges, extensive research focuses on terrain-aware path planning through mobility analysis. Traditional path planning approaches rely on graph-based or sampling-based algorithms to generate feasible paths while minimizing distance or travel cost [6, 7]. More recent studies incorporate terrain traversability and environmental uncertainty into planning frameworks to ensure that UGVs can successfully complete the identified paths with certain pre-specified likelihoods [8, 9]. While these approaches improve navigation reliability, they primarily evaluate terrain from a mobility perspective and overlook potential mechanical components degradation caused by dynamic loads experienced by vehicle components during traversal. A critical step toward incorporating mechanical degradation into path planning is predicting the resulting reaction forces induced by a given path.

In this work, we develop a probabilistic machine learning (ML)-based framework to predict vehicle component reaction forces using terrain height history. Vehicle-terrain interaction data are generated using high-fidelity vehicle dynamics simulations, and multiple learning architectures are evaluated to capture the relationship between terrain geometry and structural response. To improve prediction reliability, probabilistic modeling techniques are incorporated to quantify uncertainty in the predicted forces. The resulting framework provides fast, uncertainty-aware predictions of terrain-induced loads, enabling efficient surrogates for fatigue-aware path planning. An overview of the proposed framework is shown in Fig. 1.

The main contributions of this work are summarized as follows:

- Generation of vehicle-terrain interaction data and component reaction forces from a vehicle-terrain simulation for data-driven modeling.
- A comparative study of multiple ML architectures to evaluate their ability to predict terrain-induced component forces from terrain height history.
- Incorporation of probabilistic ML techniques to enable uncertainty-aware force predictions for improved reliability in path planning.

Overall, this work establishes an efficient surrogate framework for terrain-induced load prediction and provides a foundation for fatigue-aware path planning that accounts for vehicle structural reliability, with the potential to be integrated with terrain traversability considerations.

2. BACKGROUND

This section provides an overview of the simulation environment and probabilistic modeling framework used in this study. First, we describe the high-fidelity vehicle-terrain simulation environment implemented using PyChrono. Next, we introduce the neural network ensemble approach used for uncertainty-aware force prediction.

2.1. Vehicle Mobility Simulation

Vehicle dynamics and terrain interactions are simulated using `PyChrono`, the Python interface to the Chrono multi-physics engine [10, 11]. High-fidelity multibody simulation tools such as Chrono have been

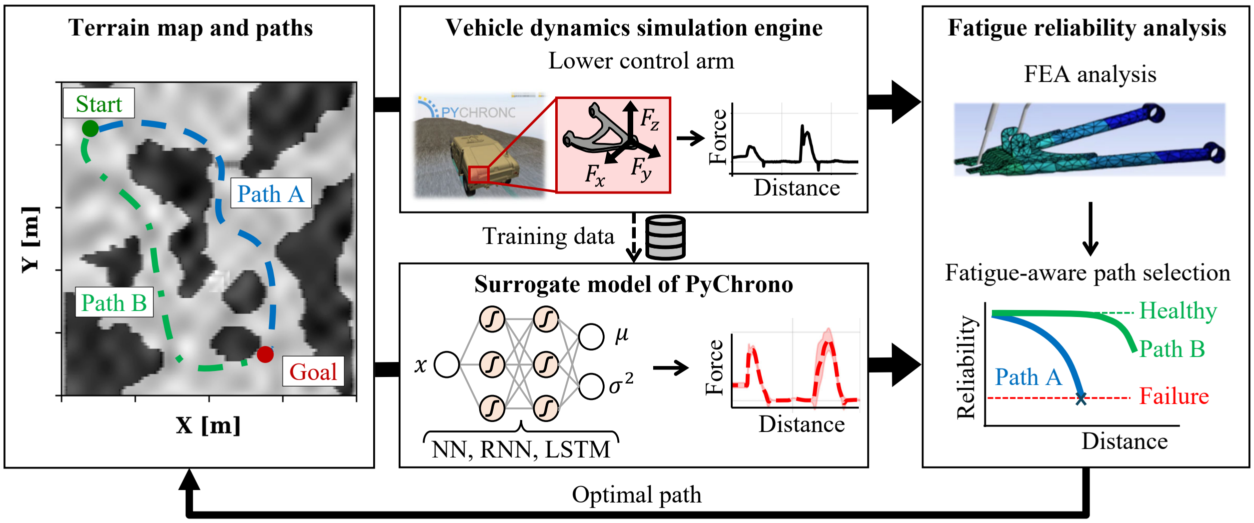

Figure 1: Overview of the proposed fatigue-aware path planning framework. Terrain profiles extracted from candidate paths are passed through a surrogate model to predict terrain-induced component forces, which are then used for fatigue reliability evaluation to select an optimal path.

widely used in terramechanics and off-road mobility studies for modeling vehicle–terrain interactions and generating simulation data for data-driven modeling and control development [12, 13, 14].

In this work, PyChrono is used to simulate vehicle traversal over parameterized terrain height profiles, capturing the resulting lower control arm reaction forces. These simulated force histories serve as ground truth data for training and evaluating the machine learning surrogate models.

2.2. Neural Network Ensemble

A neural network ensemble (NNE) is a collection of independently trained neural networks whose predictions are combined to improve overall model performance. Ensembles are commonly used to reduce overfitting, improve generalization, and provide more robust predictions compared to a single network [17].

For an ensemble containing E networks, a network e predicts both a mean $\mu_i(\mathbf{x})$ and a variance $\sigma_i^2(\mathbf{x})$ for a response of interest y at input $\mathbf{x}$ [18]. The overall predictive mean and variance of the ensemble are computed as:

$$\mu(\mathbf{x}) = \frac{1}{E}\sum_{i=1}^{E} \mu_i(\mathbf{x}) \tag{1}$$

and

$$\begin{aligned}\sigma^2(\mathbf{x}) = \underbrace{\frac{1}{E}\sum_{e=1}^{E} \sigma_i^2(\mathbf{x})}_{\text{Aleatoric uncertainty}} \\ + \underbrace{\frac{1}{E}\sum_{e=1}^{E} \mu_i^2(\mathbf{x}) - \mu^2(\mathbf{x}).}_{\text{Epistemic uncertainty}}\end{aligned} \tag{2}$$

The predictive mean in Equation (1) is the average of the individual network means. The predictive variance in Equation (2) is decomposed into two components: the aleatoric uncertainty, representing the average predictive variance from the individual networks, and the epistemic uncertainty, which captures the variance of the predicted means across the ensemble members. Each network is trained using the negative log-likelihood loss [19], which encourages both accurate predictions and calibrated uncertainty estimates:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}\log\sigma_i^2 + \frac{(y_i-\mu_i)^2}{2\sigma_i^2}\right) \quad (3)$$

where N is the number of training samples and y_i is the corresponding ground-truth value for a a given $\mathbf{x}$.

3. METHODOLOGY

This section presents the methodology used to study surrogate modeling for predicting vehicle component reaction forces from terrain height history. We begin by outlining the terrain generation approach used to construct a diverse dataset. We then introduce the ML models evaluated in this study. Finally, we detail the metrics used for performance evaluation and uncertainty quantification (UQ).

3.1. Terrain Generation

To systematically evaluate vehicle responses under controlled terrain variations, 1D terrain profiles were generated using a parametric hump model adapted from [15]. Minor modifications were introduced to better control hump geometry and ensure compatibility with the simulation framework. The resulting formulation for terrain elevation $z(x)$ [m] at position x [m] is given by

$$\begin{aligned} z(x) = \max(&\sqrt{R^2-(x-a)^2} \\ &+ H - R,\, 0) \end{aligned} \quad (4)$$

where

$$R = \frac{L^2}{8H} + \frac{H}{2}. \quad (5)$$

Here, x denotes the longitudinal coordinate along the terrain profile, the parameter a denotes the center location of the hump, L [m] is the hump width, and H [m] is the hump height.

To generate datasets for model development and evaluation, 50 training paths of equal total length were created, each containing 8 evenly spaced humps along the path. Hump parameters were sampled from predefined uniform ranges, with $H \sim \mathcal{U}(0.1,\, 0.4)$ [m] and $L \sim \mathcal{U}(3,\, 8)$ [m]. Only combinations satisfying $L/H > 16$ were retained to avoid unrealistic curvature and excessive vertical accelerations during vehicle traversal. Sampling was repeated as necessary to ensure that each path contained exactly 8 valid humps. In addition to the training set, 20 validation paths were generated for hyperparameter tuning. These validation terrains follow the same overall spacing pattern as the training terrains but include small positional perturbations in hump locations, introducing minor variability.

Two independent datasets were constructed for testing. The first (Test 1) consists of 50 paths with the same evenly spaced configuration as the training dataset, enabling evaluation under structured terrain conditions. The second dataset (Test 2) contains 50 paths with 8–11 randomly spaced humps per path. Example terrain height profiles and corresponding vertical reaction forces (F_z) from the training and test sets are shown in Figure 2. Distinctive force spikes occur each time the vehicle traverses a hump, followed by a damped wave pattern as the vehicle descends, illustrating the vehicle's dynamic response to terrain variations.

The vehicle-terrain interaction was simulated using the Soil Contact Model (SCM) deformable terrain model implemented in `PyChrono`. The soil parameters were selected to represent moderately stiff off-road terrain conditions and are summarized in Table 1. Vehicle motion along the generated paths was controlled using a path-following driver with PID-based steering and speed regulation.

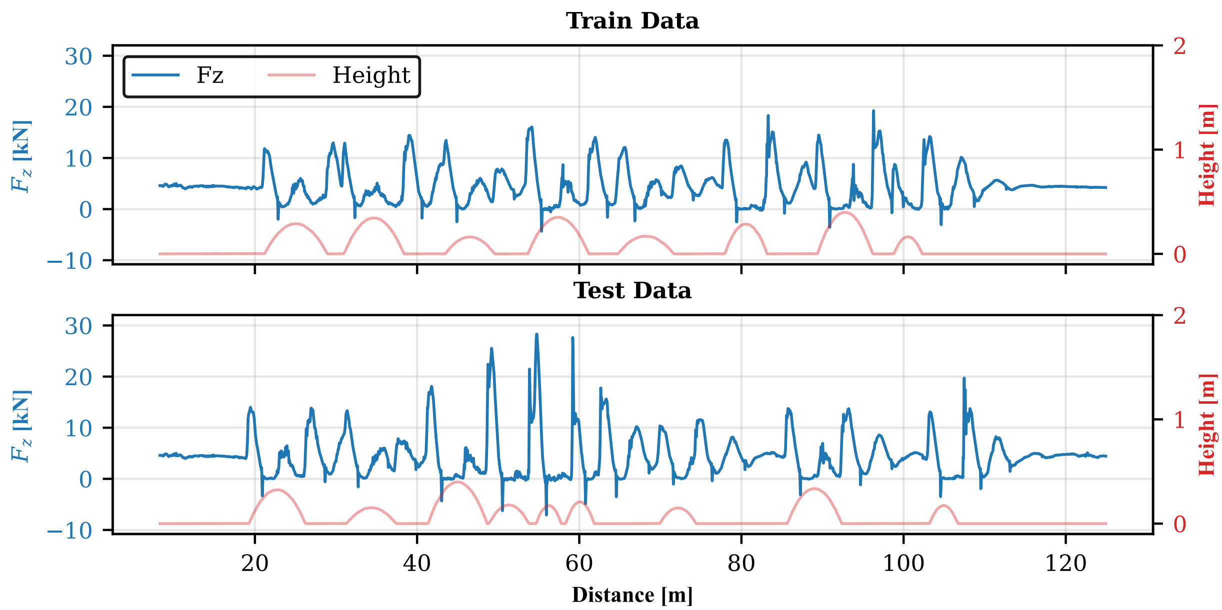

Figure 2: Vertical reaction force of the lower control arm (F_z) and terrain height profile along the traversed terrain distance for training (top) and testing (bottom) datasets.

The steering controller uses a look-ahead distance of 1.0 m with gains $K_p = 0.5$, $K_i = 0.05$, and $K_d = 0.05$. The target vehicle speed was set to 10 m/s, and speed tracking was achieved using PID gains $K_p = 3.2$, $K_i = 0.02$, and $K_d = 0.05$.

3.2. Machine Learning Models

Recurrent neural networks (RNNs) process sequential data by maintaining a hidden state that encodes information from previous time steps. Long short-term memory (LSTM) networks extend this architecture by introducing internal state vectors, $\mathbf{h}_t$ and $\mathbf{c}_t$, which improve retention of temporal information and mitigate vanishing gradient issues [20]. A single-step forward pass of an LSTM layer is given by:

$$f_t = \sigma(W_f x_t + U_f \mathbf{h}_{t-1} + \mathbf{b}_f) \quad (6)$$

$$i_t = \sigma(W_i x_t + U_i \mathbf{h}_{t-1} + \mathbf{b}_i) \quad (7)$$

$$o_t = \sigma(W_o x_t + U_o \mathbf{h}_{t-1} + \mathbf{b}_o) \quad (8)$$

$$\tilde{\mathbf{c}}_t = \tanh(W_c x_t + U_c \mathbf{h}_{t-1} + \mathbf{b}_c) \quad (9)$$

$$\mathbf{c}_t = f_t \odot \mathbf{c}_{t-1} + i_t \odot \tilde{\mathbf{c}}_t \quad (10)$$

$$\mathbf{h}_t = o_t \odot \tanh(\mathbf{c}_t) \quad (11)$$

where σ and $\tanh$ are activation functions applied elementwise, and $\odot$ denotes the Hadamard product. The hidden state $\mathbf{h}_t \in \mathbb{R}^H$ represents the internal encoding of the terrain sequence, where H is the number of hidden units in the LSTM layer. The input at each time step, $x_t \in \mathbb{R}$, is the terrain height at the current location, which is a scalar. To capture the recent terrain profile, the model processes a sequence of these scalars, $\mathbf{x}_t = [x_{t-L+1}, x_{t-L+2}, \ldots, x_t]$, where L is the length of the input history.

The final hidden state is passed through a fully connected layer to produce the predicted force vector:

$$\hat{\mathbf{y}}_t = \mathbf{V}\mathbf{h}_t + \mathbf{c}, \quad (12)$$

where $\hat{\mathbf{y}}_t \in \mathbb{R}^3$ represents the predicted vehicle component forces (F_x, F_y, F_z), $\mathbf{V} \in \mathbb{R}^{3\times H}$ is the output weight matrix, and $\mathbf{c} \in \mathbb{R}^3$ is the bias vector. The model maps the internal hidden state of the LSTM directly to the predicted forces, without using previous force values as inputs.

In this study, LSTM models serve as the primary architecture, while feed-forward

Table 1: SCM terrain model parameters used in the vehicle-terrain simulation. [16]

Parameter	Symbol	Value
Bekker pressure-sinkage (frictional)	K_ϕ	1×10^9
Bekker pressure-sinkage (cohesive)	K_c	5×10^8
Bekker exponent	n	2.8
Soil cohesion	c	950 Pa
Friction angle	ϕ	37.5°
Janosi shear coefficient	k	0.048 m
Elastic stiffness	K_e	1×10^{10} Pa/m
Terrain damping coefficient	C_d	3×10^4 Pa·s/m

neural networks are used as baseline models for comparison. All models are trained using the terrain sequences described in Section 3.1, and their hyperparameters are optimized using Optuna to ensure systematic evaluation of predictive performance (see Appendix A.1 for final values). After hyperparameter optimization, all models are converted into NNEs with an ensemble size of $E = 5$, as discussed in Section 2.2. The aggregated ensemble predictions provide the mean and uncertainty estimates of the vehicle component forces.

To formalize this task, the prediction problem is defined as learning a mapping from terrain height sequences to the resulting vehicle structural response. Each model takes as input a sequence of terrain height values representing the ground profile along the traversal path and predicts the corresponding reaction force at the vehicle lower control arm, enabling rapid estimation of component forces from terrain data.

3.3. Evaluation Metrics

We evaluate predictive performance and uncertainty quality using mean absolute error (MAE) and expected calibration error (ECE). These metrics assess deterministic accuracy and probabilistic calibration, respectively.

Mean Absolute Error MAE measures the average absolute deviation between predicted and true force vectors across all test samples. It is defined as:

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_1 \qquad (13)$$

where $\mathbf{y}_i \in \mathbb{R}^3$ and $\hat{\mathbf{y}}_i \in \mathbb{R}^3$ denote the true and predicted force vectors for sample i, respectively, and $\|\cdot\|_1$ represents the ℓ_1 norm, corresponding to the sum of absolute errors across the three force components (F_x, F_y, F_z). Here, N denotes the total number of samples in the dataset. Each sample corresponds to a terrain sequence window used as input to the model and its associated force prediction. MAE therefore provides a direct and interpretable measure of the average prediction error magnitude across all force directions.

Expected Calibration Error ECE quantifies the discrepancy between predicted confidence levels and empirical coverage, providing a measure of probabilistic calibration [21]. Calibration is evaluated by comparing predicted confidence intervals with the observed frequency of ground-truth force vectors falling within those intervals. ECE is defined as:

$$\mathrm{ECE} = \sum_{j=1}^{K} w_j \left| \hat{c}_j - c_j \right|, \qquad (14)$$

where K is the number of confidence bins partitioning the interval $[0, 1]$, w_j is the weight associated with bin j (typically proportional to the number of samples in the bin), $\hat{c}_j$ denotes the empirical coverage in bin j, and c_j is the corresponding nominal confidence level. Lower ECE values indicate better agreement between predicted uncertainty and observed coverage.

To further assess calibration quality, we plot a calibration curve by graphing the empirical coverage $\hat{c}_j$ against the nominal confidence level c_j for all bins. A perfectly calibrated model aligns with the diagonal line, where predicted confidence matches observed coverage.

4. RESULTS

This section provides a detailed analysis of the four models evaluated in this work. All models use a neural network ensemble framework to predict the mean and variance of the vehicle's lower control arm force.

- Ens-NN-NoHist: An ensemble of feedforward neural networks that uses terrain height at time t as input, serving as a baseline without temporal context.
- Ens-RNN: An ensemble of RNNs that uses 200 steps of terrain height history as input.
- Ens-NN: An ensemble of feedforward neural networks that uses 200 steps of terrain height history as input.
- Ens-LSTM: An ensemble of LSTM networks that uses 200 steps of terrain height history as input.

These models are compared in their ability to predict vehicle lower control arm forces from terrain height history. Before training, both the input terrain heights and the target force outputs are normalized to improve training stability.

Two types of test terrains are considered, as described in Section 3.1. Test 1 corresponds to structured terrain with evenly spaced humps, while Test 2 features randomized terrain with irregular spacing and clustered hump formations. The results are presented in two parts. First, we analyze prediction errors across all models using MAE, highlighting the influence of terrain history and model architecture on performance. Second, we evaluate the quality of uncertainty estimation, examining how well each model quantifies confidence in its force predictions. Each model was trained and evaluated 10 times to account for run-to-run variations due to random initialization and stochastic optimization.

Before analyzing aggregate errors and uncertainty metrics, we first illustrate the predictive performance of the Ens-LSTM model on a representative average test case (Test 038). Figure 3 shows the force components F_x, F_y, and F_z predicted by the Ens-LSTM, overlaid with the corresponding ground-truth forces and 95% prediction intervals. The model captures the main force spikes as the vehicle traverses each hump, while also reflecting the residual damping effects as the vehicle descends the terrain features.

4.1. Error Analysis

Figure 4 reports the MAE values for all models across Test 1 and Test 2 terrains. Ens-NN-NoHist only uses the current height at time t, serving as a baseline to evaluate the impact of history information.

The results show that Test 1 generally exhibits lower errors than Test 2, as expected, since the models were trained on terrain profiles similar to Test 1. Models that incorporate terrain history (Ens-NN, Ens-RNN, and Ens-LSTM) significantly outperform Ens-NN-NoHist, confirming that

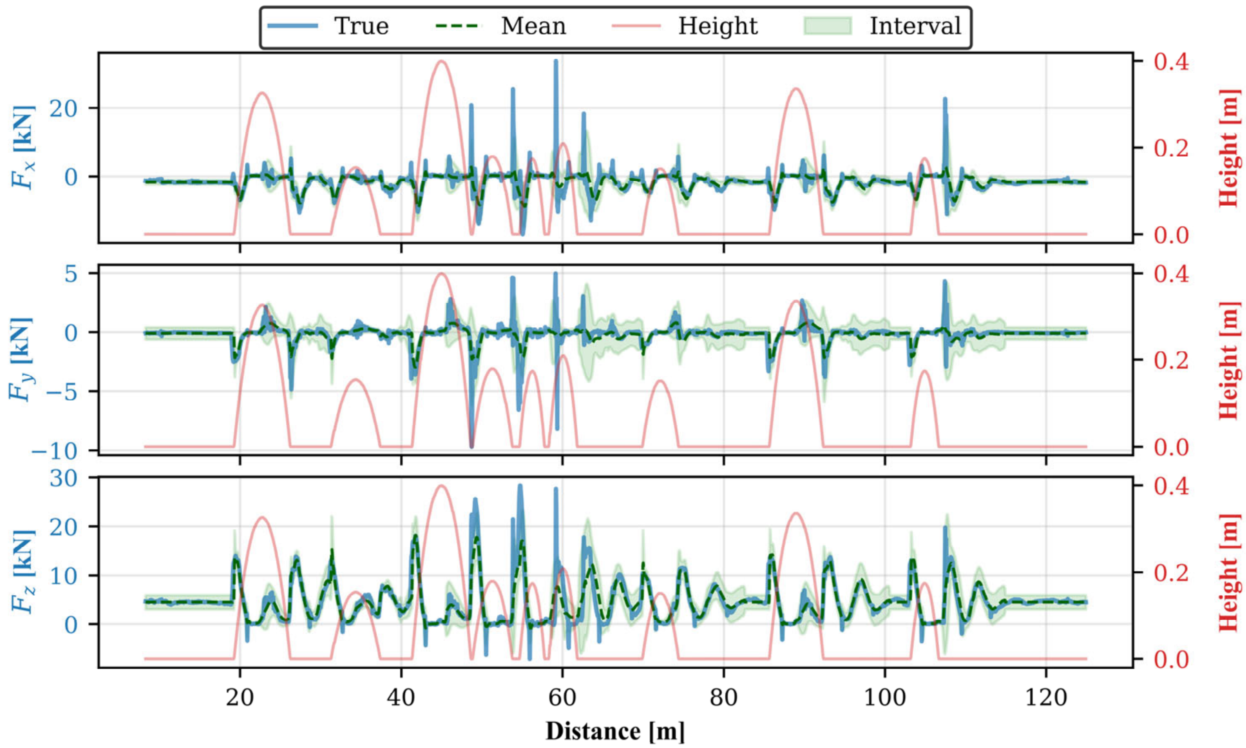

Figure 3: Example prediction of F_x, F_y, F_z for Ens-LSTM on the average test case (Test 038). Blue lines represent ground truth measurements, while green dashed lines show the model's mean predictions. The shaded green region indicates the 95% prediction interval, and the red line shows the terrain height profile.

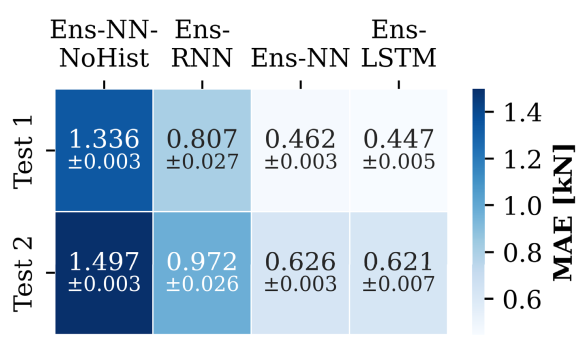

Figure 4: MAE comparison across models on Test 1 and Test 2 terrains. Ens-NN achieves near-optimal performance with approximately four times fewer parameters than Ens-LSTM.

prior terrain information is critical for accurate force prediction.

Among the models, Ens-LSTM achieves the lowest MAE (0.447 kN for Test 1 and 0.621 kN for Test 2), with Ens-NN closely matching its performance (0.462 kN and 0.626 kN, respectively). Despite having substantially more parameters than the feed-forward Ens-NN (117,382 vs. 25,734), the Ens-LSTM provides only marginal improvements in accuracy. In contrast, Ens-RNN exhibits higher errors (0.807 kN and 0.972 kN). These results indicate that incorporating terrain history significantly improves prediction accuracy compared with the Ens-NN-NoHist baseline, while more complex recurrent architectures provide limited additional benefit for this task.

Overall, the critical factor for accurate force prediction is access to a sufficiently long terrain history. While LSTM can theoretically leverage temporal dependencies, the marginal gain over Ens-NN suggests that explicit sequential modeling provides limited additional benefit. Consequently, Ens-NN offers the best trade-off between accuracy and efficiency, demonstrating that a properly parameterized feed-forward network can effectively map terrain history to vehicle component forces. An additional example corresponding to a more challenging terrain configuration is provided in Appendix A.2, illustrating model predictions for a test path with relatively larger prediction errors.

4.2. Uncertainty Quantification and Interval Analysis

Accurate force predictions alone are insufficient for safety-critical vehicle applications, as over- or underestimation of component loads can lead to unsafe designs or overly conservative operation. Uncertainty quantification provides an estimate of confidence in model predictions, enabling engineers to assess risk, make informed design decisions, and integrate predictions into fatigue-aware path planning. In this work, we evaluate UQ through calibration metrics, coverage analysis, and interval width, focusing on the more challenging Test 2 terrain with randomized humps. Results for Test 1 are provided in Appendix A.3.

To evaluate the quality of uncertainty predictions, we compute the ECE for each model and force component. Figure 5 presents the ECE values for Test 2. Ens-LSTM achieves the lowest ECE for F_x and F_y (5.306% and 16.082%), indicating the best calibration for these components, while Ens-NN-NoHist achieves the lowest ECE for F_z (5.079%). This highlights that Ens-LSTM provides the most reliable uncertainty estimates for lateral and longitudinal forces, whereas Ens-NN-NoHist performs best for vertical reaction force predictions.

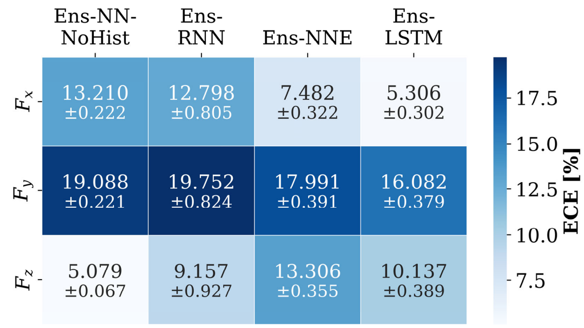

Figure 5: ECE heatmap for uncertainty calibration of all models on Test 2 terrain. Each cell reports the ECE value for the predicted lower control arm force components (F_x, F_y, and F_z).

Calibration curves provide further insight into the nature of the uncertainty estimates. Figure 6 shows the predicted confidence levels versus empirical coverage for Test 2. Ideally, model predictions should lie along the diagonal, indicating perfect calibration between predicted confidence and observed coverage. In practice, most models lie slightly above the diagonal across several confidence levels, indicating under-confident predictions where uncertainty intervals are somewhat wider than required. While perfect calibration is preferred, mild underconfidence is generally safer than overconfidence in engineering applications, as it reduces the risk of underestimating critical loads during vehicle component design and fatigue analysis.

For the F_x component, Ens-NN-NoHist lies closer to the diagonal than the more complex models, despite exhibiting higher MAE. This indicates that its predicted confidence levels align more closely with empirical coverage. In contrast, the more accurate models tend to produce slightly conservative uncertainty estimates, resulting in empirical coverage that exceeds the predicted confidence levels and therefore appears more under-confident.

To complement calibration analysis, we evaluate predicted uncertainty intervals in terms of empirical coverage and interval

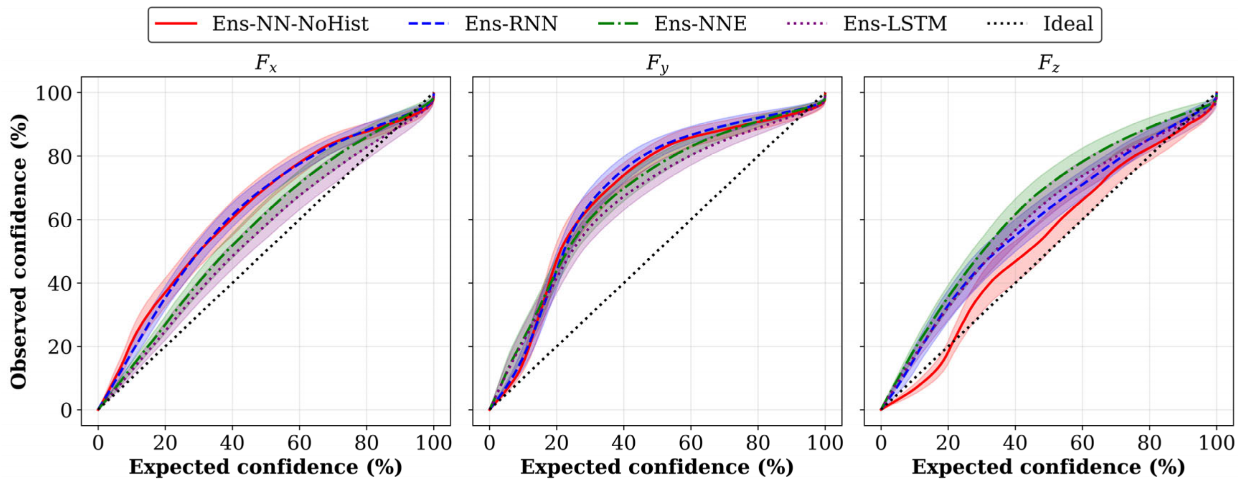

Figure 6: Calibration curves for Test 2 terrain. The left, middle, and right panels correspond to F_x, F_y, and F_z, respectively. The diagonal line represents perfect calibration. Curves above the diagonal indicate under-confident predictions, while curves below indicate over-confident predictions. Shaded regions represent ±1 standard deviation across 10 runs to illustrate run-to-run variation.

width. Each model outputs a mean force and an associated prediction interval, with a nominal target coverage of 95%. Empirical coverage is calculated as the fraction of ground-truth values falling inside the predicted intervals, and interval width is the average distance between the upper and lower bounds. Figure 7 reports mean coverage (%) and mean interval width for each force component and model.

All models achieve coverage values near the target 95%, but the average width of the prediction intervals varies. Ens-LSTM provides the narrowest intervals for all force components (F_x: 3.22 kN, F_y: 1.53 kN, F_z: 4.28 kN), making it the most informative choice despite slightly lower coverage for some components (F_x: 92.7%, F_z: 92.8%). Ens-NN-NoHist achieves comparable coverage for F_x and F_y (93.5% and 94.6%) but with substantially wider intervals (8.70 kN and 2.85 kN), which are less desirable in practice. For F_z, Ens-NN achieves slightly above-target coverage (95.0%) with a relatively narrow interval (5.05 kN). Ens-RNN exhibits intermediate performance across all components, producing moderate interval widths and coverage values (e.g., F_x: 94.2% coverage, 8.792 kN interval), reflecting a compromise between informativeness and reliability.

These results illustrate that ECE alone does not fully characterize uncertainty quality, as it depends on the chosen confidence threshold. In practice, narrower intervals with coverage near or above the target are preferred, since they provide both confidence and precision. In this setting, Ens-LSTM achieves this balance across all components, delivering efficient and reliable uncertainty estimates suitable for deployment.

5. DISCUSSION

The results highlight several insights regarding the relationship between terrain history, model architecture, and predictive uncertainty. In terms of prediction accuracy, the feed-forward Ens-NN achieves performance comparable to the LSTM-based model despite its simpler structure. This suggests that the terrain height history primarily encodes spatial patterns such as bump spacing, clustering, and cumulative elevation changes. These geometric features can be extracted directly from

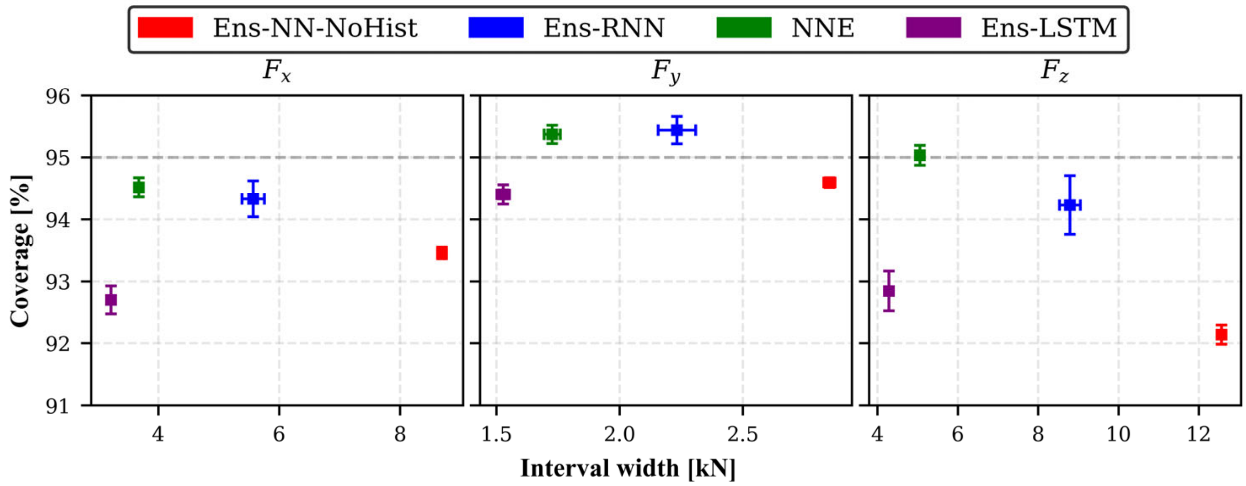

Figure 7: Uncertainty quantification trade-off analysis for Test 2 terrain across all three force components (F_x, F_y, F_z). Each subplot shows the relationship between prediction interval width (x-axis) and empirical coverage (y-axis) for the four models, with error bars indicating standard deviation. The horizontal dashed line at 95% marks the nominal coverage target.

the terrain profile without requiring explicit temporal sequence modeling. Consequently, sequential architectures such as LSTM provide only marginal improvements for this problem, indicating that terrain-to-force prediction behaves more like a spatial pattern recognition task than a temporal dynamics problem.

While Ens-LSTM produces the narrowest prediction intervals, slight over-confidence is observed for certain components when evaluated using empirical coverage. In contrast, the feed-forward Ens-NN achieves empirical coverage close to the nominal 95% level for the reaction force component while maintaining relatively narrow intervals, indicating well-calibrated uncertainty estimates. Importantly, elevated prediction uncertainty often occurs when the terrain history differs from patterns seen during training. In practical applications, such uncertainty signals are valuable for path planning, where regions with high predictive uncertainty may indicate unfamiliar terrain conditions and trigger more conservative vehicle behavior.

6. CONCLUSIONS

This work investigated machine learning models for predicting lower control arm forces from terrain height profiles with quantified predictive uncertainty. Four architectures were evaluated: Ens-NN-NoHist, Ens-RNN, Ens-NN, and Ens-LSTM. The results demonstrate that incorporating terrain history improves prediction accuracy, while Ens-LSTM delivers the most efficient uncertainty intervals. In contrast, the feed-forward Ens-NN provides well-calibrated uncertainty estimates, particularly for vertical reaction forces.

Overall, terrain history primarily encodes spatial patterns critical for force prediction, and the associated uncertainty estimates indicate regions where predictions may be less reliable. These insights can inform terrain-aware vehicle control and path planning. Future work will explore integrating the predictive model into real-time vehicle decision-making and extending it to more complex terrains and additional vehicle responses.

7. REFERENCES

[1] S. Ge et al., ”Retracted: The Use of Intelligent Vehicles and Artificial Intelligence in Mining Operations: Ethics, Responsibility, and Sustainability,” in IEEE Transactions on Intelligent Vehicles, vol. 8, no. 2, pp. 1021-1024, Feb. 2023, doi: 10.1109/TIV.2023.3246118

[2] S. Ge et al., ”Making Standards for Smart Mining Operations: Intelligent Vehicles for Autonomous Mining Transportation,” in IEEE Transactions on Intelligent Vehicles, vol. 7, no. 3, pp. 413-416, Sept. 2022, doi: 10.1109/TIV.2022.3197820

[3] J. E. Naranjo, M. Clavijo, F. Jiménez, O. Gómez, J. L. Rivera and M. Anguita, ”Autonomous vehicle for surveillance missions in off-road environment,” 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden, 2016, pp. 98-103, doi: 10.1109/IVS.2016.7535371

[4] Q. Liu, L. Zhao, Z. Tan, and W. Chen, “Global path planning for autonomous vehicles in off-road environment via an A-star algorithm,” International Journal of Vehicle Autonomous Systems, vol. 13, no. 4, p. 330, 2017, doi: https://doi.org/10.1504/ijvas.2017.087148

[5] T. Oksanen and A. Visala, “Coverage path planning algorithms for agricultural field machines,” J. Field Robot., vol. 26, no. 8, pp. 651–668, 2009. https://doi.org/10.1002/rob.20300

[6] J. Bai, W. Zhu, S. Liu, L. Xu, and X. Wang, “Path Planning Method for Unmanned Vehicles in Complex Off-Road Environments Based on an Improved A* Algorithm,” Sustainability, vol. 17, no. 11, p. 4805, May 2025, doi: https://doi.org/10.3390/su17114805

[7] T. Xu, “Recent advances in Rapidly-exploring random tree: A review,” Heliyon, vol. 10, no. 11, pp. e32451–e32451, Jun. 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e32451

[8] P. Borges et al., “A Survey on Terrain Traversability Analysis for Autonomous Ground Vehicles: Methods, Sensors, and Challenges,” Field Robotics, vol. 2, no. 1, pp. 1567–1627, Mar. 2022, doi: https://doi.org/10.55417/fr.2022049

[9] J. Yin, Z. Hu, Z. P. Mourelatos, D. Gorsich, A. Singh, and S. Tau, “Efficient Reliability-Based Path Planning of Off-Road Autonomous Ground Vehicles Through the Coupling of Surrogate Modeling and RRT*,” IEEE transactions on intelligent transportation systems, vol. 24, no. 12, pp. 15035–15050, Dec. 2023, doi: https://doi.org/10.1109/tits.2023.3296651

[10] R. Serban, M. Taylor, D. Negrut, and A. Tasora, “Chrono::Vehicle: template-based ground vehicle modelling and simulation,” International Journal of Vehicle Performance, vol. 5, no. 1, p. 18, 2019, doi: https://doi.org/10.1504/ijvp.2019.097096

[11] Alessandro Tasora et al., “Chrono: An Open Source Multi-physics Dynamics Engine,” Springer eBooks, pp. 19–49, May 2015, doi: https://doi.org/10.1007/978-3-319-40361-8_2

[12] S. D. Jadhav et al., “Containerization Approach for High-Fidelity

Terramechanics Simulations," SAE Technical Paper Series, vol. 1, Apr. 2023, doi: https://doi.org/10.4271/2023-01-0105

[13] S. Benatti, A. Tasora, D. Fusai, and D. Mangoni, "A Modular Simulation Platform for Training Robots via Deep Reinforcement Learning and Multibody Dynamics," Proceedings of the 2019 3rd International Conference on Automation, Control and Robots, pp. 7–11, Oct. 2019, doi: https://doi.org/10.1145/3365265.3365274

[14] A. Young et al., "Enhancing Autonomous Navigation by Imaging Hidden Objects Using Single-Photon LiDAR," 2025 IEEE International Conference on Robotics and Automation (ICRA), pp. 4907–4914, May 2025, doi: https://doi.org/10.1109/icra55743.2025.11128292

[15] F. Shi, "Formulation and Simulation for Speed Control Humps," American Journal of Mechanical Engineering, vol. 12, no. 3, pp. 26–34, Sep. 2024, doi: https://doi.org/10.12691/ajme-12-3-1

[16] A. Gallina, R. Krenn, M. Scharringhausen, T. Uhl, and B. Schäfer, "Parameter Identification of a Planetary Rover Wheel-Soil Contact Model via a Bayesian Approach," Journal of Field Robotics, vol. 31, no. 1, pp. 161–175, Sep. 2013, doi: https://doi.org/10.1002/rob.21480.

[17] D. Opitz and R. Maclin, "Popular Ensemble Methods: An Empirical Study," Journal of Artificial Intelligence Research, vol. 11, pp. 169–198, Aug. 1999, doi: https://doi.org/10.1613/jair.614

[18] Balaji Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles," arXiv (Cornell University), Jan. 2016, doi: https://doi.org/10.48550/arxiv.1612.01474

[19] S. Kristoffersson Lind, Z. Xiong, P.-E. Forssén, and V. Krüger, "Uncertainty quantification metrics for deep regression," Pattern Recognition Letters, vol. 186, pp. 91–97, Sep. 2024, doi: https://doi.org/10.1016/j.patrec.2024.09.011.

[20] A. Graves, "Long Short-Term Memory," *Studies in Computational Intelligence*, vol. 385, pp. 37–45, 2012, doi: https://doi.org/10.1007/978-3-642-24797-2_4.

[21] V. Kuleshov, N. Fenner, and S. Ermon, "Accurate Uncertainties for Deep Learning Using Calibrated Regression," arXiv.org, Jun. 30, 2018. https://arxiv.org/abs/1807.00263

8. CONTACT INFORMATION

Yang Kang Chua
Ph.D. Student
School of Mechanical, Aerospace, and Manufacturing Engineering
University of Connecticut
Storrs, CT 06269, USA
Email: yangkang.chua@uconn.edu

9. ACKNOWLEDGMENT

This work was partially supported by the Digital Design Research, Analysis, and Manufacturing Center (D2REAM) at the University of Connecticut under Cooperative Agreement W56HZV-21-2-0001 with the US Army DEVCOM Ground Vehicle

Systems Center (GVSC), through the Virtual Prototyping of Autonomy Enabled Ground Systems (VIPR-GS) program. Any opinions, findings, or conclusions in this paper are those of the authors and do not necessarily reflect the sponsor's views.

10. ACRONYMS

AGV	Autonomous Ground Vehicle
ECE	Expected Calibration Error
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
NNE	Neural Network Ensemble
RNNs	Recurrent Neural Networks
SCM	Soil Contact Model
UQ	Uncertainty Quantification

APPENDIX

A.1. Hyperparameter Details

Hyperparameters for each model were optimized using Optuna. The final selected configurations, along with fixed parameters used across all models, are summarized in Table A.1.

Table A.1: Final hyperparameters for each model after Optuna tuning. Batch size (128) and number of layers (4) are fixed for all models.

Model	Hidden Size	Dropout	Learning Rate
Ens-NN-NoHist	512	0.031	2.42e-4
Ens-RNN	128	–	1.78e-4
Ens-NN	64	0.032	1.68e-4
Ens-LSTM	64	–	6.64e-4

A.2. Additional Prediction Example for Challenging Terrain

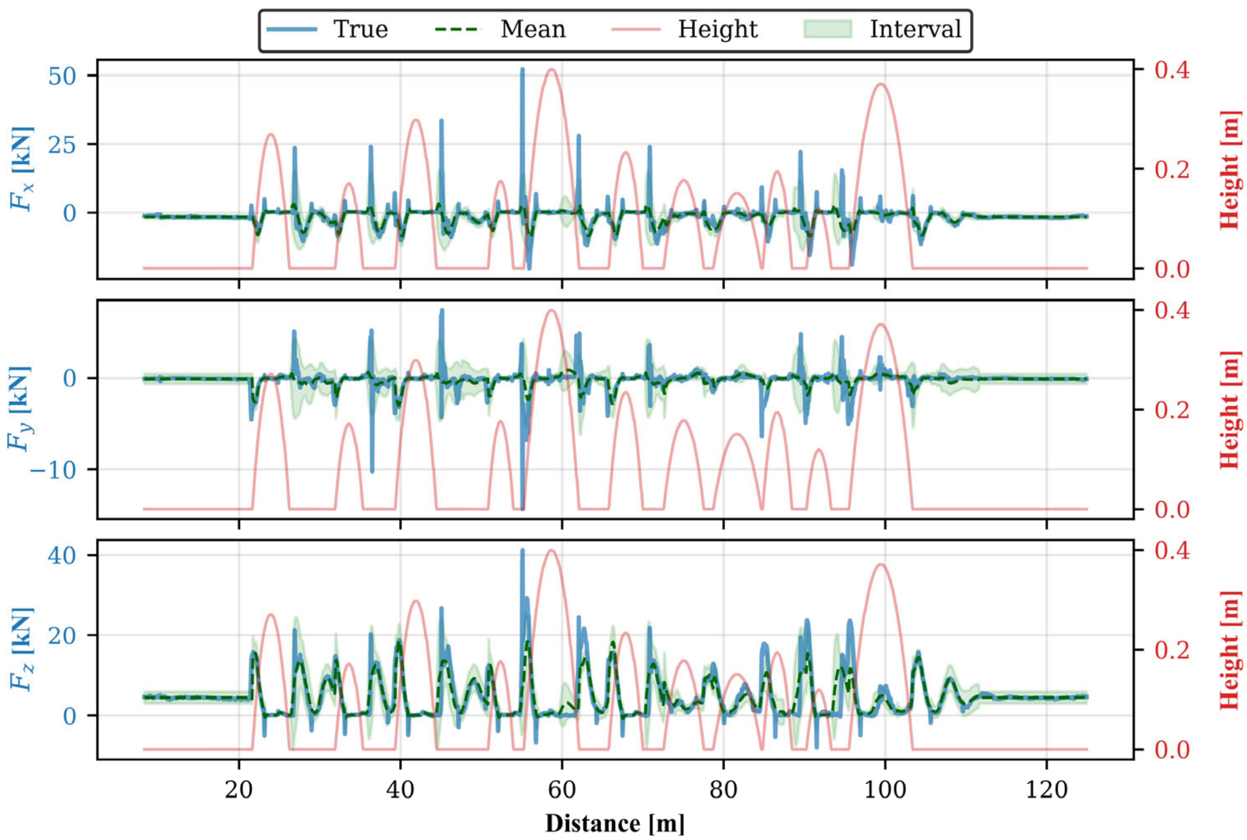

Figure A.1: Ens-LSTM predictions of lower control arm forces (F_x, F_y, F_z) on Test 045. Blue lines show ground truth, green dashed lines show model mean, and the shaded area indicates the 95% prediction interval. The red line shows terrain height.

This appendix section presents an additional prediction example for a more challenging terrain configuration. Figure A.1 shows a test path with relatively larger prediction errors, with an overall MAE of 0.863 kN. In this case, the terrain contains a larger number of humps within a limited distance (11 humps), resulting in shorter spacing between consecutive humps. This configuration produces closely spaced force spikes and stronger dynamic interactions, making the prediction task more challenging for the model.

A.3. Uncertainty Quantification Performance on Test 1

This appendix reports the uncertainty quantification performance for all models on Test 1 terrain. Results include expected calibration error (ECE), calibration curves, and empirical coverage/interval widths.

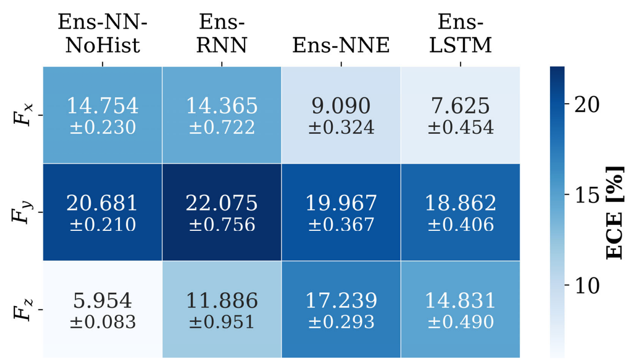

Figure A.2: ECE heatmap for uncertainty calibration of all models on Test 1 terrain. Each cell reports the ECE value for the predicted lower control arm force components (F_x, F_y, and F_z).

Figure A.2 presents the ECE heatmap for Test 1 terrain. Each cell reports the ECE value for the predicted lower control arm force components (F_x, F_y, and F_z) across all models. Ens-NN-NoHist achieves the lowest ECE for F_z (5.954%), while Ens-LSTM provides the best calibration for F_x (7.625%) and F_y (18.862%). This visualization highlights differences in uncertainty calibration across models and force components for structured terrain.

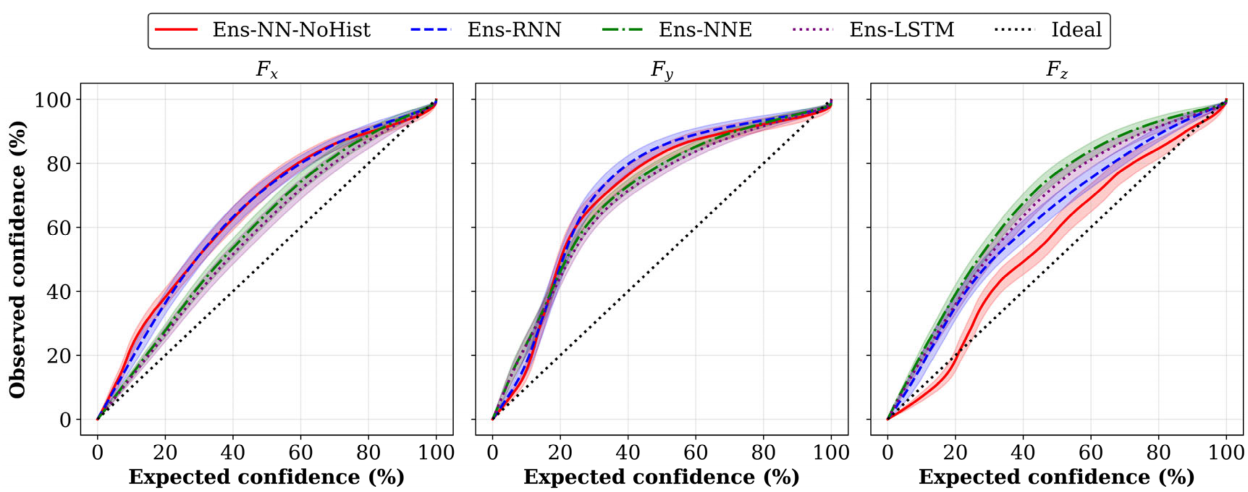

Figure A.3: Calibration curves for all models on Test 1 terrain. The left, middle, and right panels correspond to F_x, F_y, and F_z, respectively. The diagonal line represents perfect calibration, while curves above the diagonal indicate under-confident predictions. Shaded regions show ±1 standard deviation across 10 runs.

Figure A.3 presents the calibration curves for all models on Test 1 terrain. The left, middle, and right panels correspond to F_x, F_y, and F_z, respectively. Curves above the diagonal indicate under-confident predictions, while curves below indicate over-confident predictions. Shaded regions show ±1 standard deviation across 10 runs, illustrating run-to-run variability. Overall, all models tend to be slightly under-confident, with Ens-NN-NoHist for F_x lying closest to the diagonal.

Finally, we conclude the analysis by comparing empirical coverage and prediction interval width for a nominal target coverage of 95% on Test 1 terrain. Figure A.4 reports the actual coverage (%) achieved by each model as well as the corresponding mean interval widths for all force components.

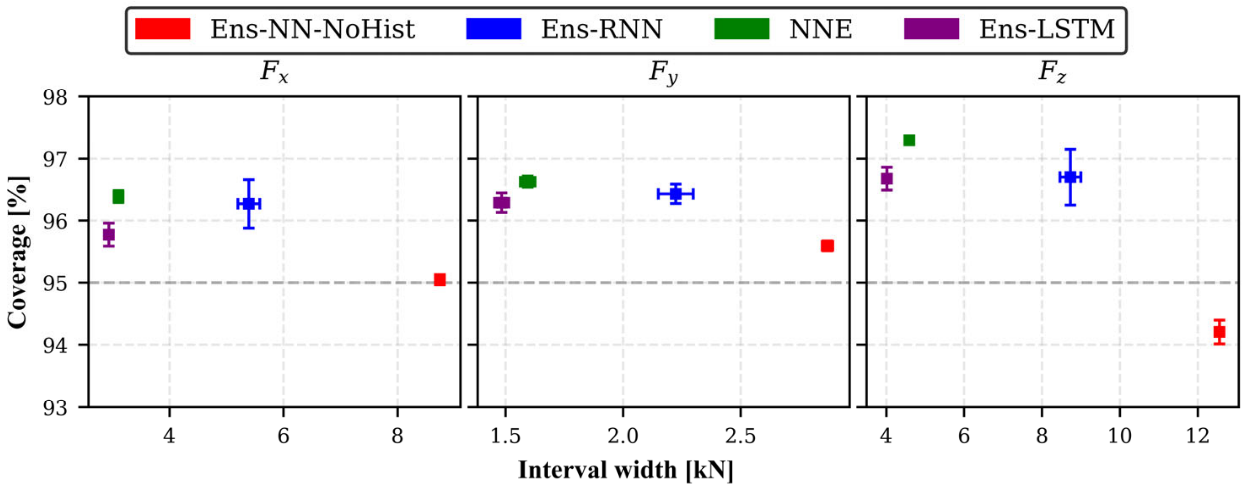

Figure A.4: Uncertainty quantification trade-off analysis for Test 1 terrain across all three force components (F_x, F_y, F_z). Each subplot shows the relationship between prediction interval width (x-axis) and empirical coverage (y-axis) for the four models, with error bars indicating standard deviation. The horizontal dashed line at 95% marks the nominal coverage target.

Since Test 1 terrain closely resembles the training data, all models achieve coverage values near the target. Ens-LSTM provides the narrowest intervals across all components (F_x: 2.94 kN, F_y: 1.48 kN, F_z: 4.02 kN), making it the most informative choice, while Ens-NN-NoHist achieves comparable coverage but with substantially wider intervals, especially for F_x and F_z. Ens-RNN and NNE show intermediate performance, with coverage slightly above the target and moderate interval widths. These results highlight that, even when models are well-calibrated in terms of coverage, interval width remains a critical factor for practical interpretability and decision-making.