Proceedings of the 4th International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2026.37507

Structural Reliability Assessment Using Probabilistic Models and Engineering Data

Mingdong Ma

Emma Willard School, Troy, USA
mingdongma7@gmail.com

Abstract. Structural reliability assesses the probability that a structure will satisfy specified safety and serviceability requirements. This assessment is challenging because loads, material properties, deterioration, and measurements are subject to uncertainty. This review explains how probability models and engineering data can address that problem. The review follows a decision-first framework. It starts with the limit state and the time period. It then links each source of uncertainty to a suitable method. By synthesizing recent engineering studies within this framework, the review also identifies recurring gaps in failure data, model-error treatment, time-varying uncertainty, surrogate-model uncertainty, and reporting consistency. The main methods include the first-order reliability method (FORM), the second-order reliability method (SORM), Monte Carlo simulation, Bayesian updating, surrogate models, and active learning. Recent studies show that field data can improve bridge and building models and that surrogate and active-learning methods can reduce the number of costly high-fidelity simulations. Each method also has limitations that depend on the structure of the reliability problem. A local method can miss another failure region. A data update can fit noise. A surrogate can be accurate on average and still be wrong near failure. Therefore, engineers need checks at every step. They should state the decision, test the input model, compare methods, and report model error. No one method fits every case. A useful result must be accurate enough, clear enough, and tied to a real engineering choice.

Keywords: structural reliability, probabilistic models, Bayesian updating, surrogate models, engineering data

1. Introduction

A structure does not face one fixed load or one fixed strength. A bridge carries different traffic each day. A building faces changing wind, snow, and use. Materials also vary from one part to another. Age adds more change. Corrosion can remove steel. Cracks can grow. Repairs can change how forces move through the structure. Engineers need a way to include these changes in a safety check. Structural reliability gives them that way. It links uncertain inputs to a clear failure condition. It then quantifies the probability of reaching that failure condition. Design codes already manage uncertainty. They use load factors, strength factors, and standard load cases. These rules are useful because engineers can apply them in the same way. However, prescribed safety factors do not explicitly quantify the contribution of each source of uncertainty. They also cannot show how new

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.37507

data alter a reliability assessment. An old bridge may have inspection records and sensor data. A new bridge may have test records and better material data. A reliability study can use this evidence. The study can also show when the evidence is weak. Recent bridge reviews show that this issue matters for repair and inspection choices [1].

This paper uses a decision-first view. The review does not begin with a list of methods. It begins with the choice that an engineer must make. The choice may concern repair, testing, or continued use. The engineer then defines failure and a time period. Next, the engineer lists the uncertain inputs.

The engineer can choose a calculation method after those steps. This order keeps the model tied to the real problem. It also makes weak assumptions easier to see.

Across the methods and engineering studies reviewed here, a recurring gap is that method development, data updating, validation, and the final engineering decision are often treated as separate parts of the reliability process. This review therefore contributes a decision-first synthesis that connects these elements in one work path. The synthesis also highlights five recurring research needs: limited true failure data, explicit treatment of model error, time-varying uncertainty, uncertainty in surrogate predictions near failure, and more consistent reporting and benchmarking. Framing these issues together helps clarify not only how reliability can be calculated, but also when a result is sufficiently supported to guide an engineering action. Figure 1 summarizes the framework used throughout the review.

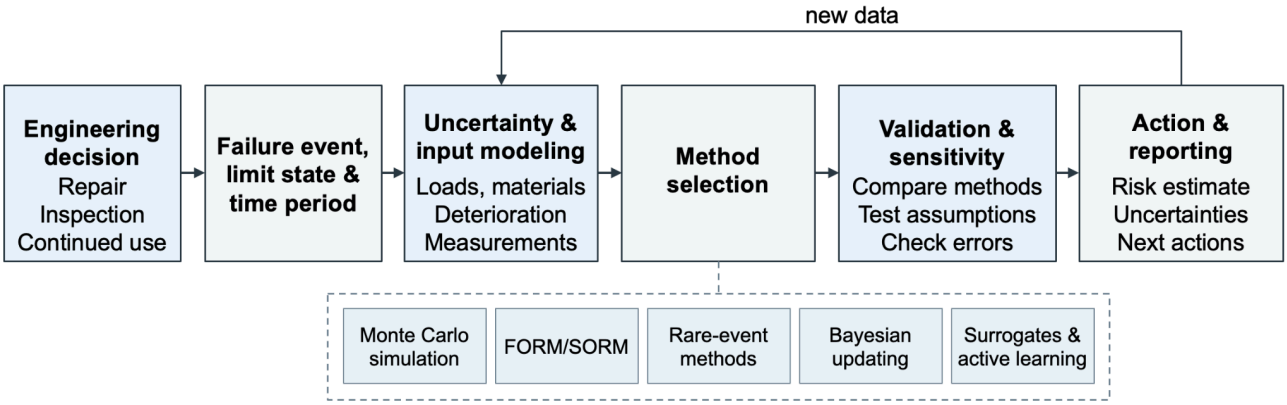

Figure 1. Decision-first framework for structural reliability assessment (picture credit: original)

The review follows this order. Section 2 builds the reliability problem from the decision, the limit state, and the input data. Section 3 explains how the main methods answer the problem. The section starts with direct simulation. It then moves to faster local methods, rare-event methods, data updates, and learning models. Section 4 groups recent studies by the lesson they give. This order differs from a simple list of structure types. Section 5 gives a practical work path and sets out the main research needs. Section 6 gives the main conclusions.

2. Building the reliability problem

2.1. Decision context and limit-state definition

A reliability number has value only when it supports a choice. An engineer may need to decide if a bridge can stay open. The engineer may need to choose an inspection date. The engineer may also need to compare two repair plans. Each choice needs a clear event of concern. Collapse is one event.

DOI: 10.54254/2753-8818/2026.37507

Too much movement is another event. Fatigue cracking can be a third event. The engineer should define this event before choosing a probability method.

A limit-state function provides a mathematical representation of the failure event. The function usually compares resistance with demand. Under the sign convention used here, a positive value of the limit-state function indicates the safe domain. A zero value marks the boundary. A negative value means that failure has taken place. The failure probability is the chance of a negative value. The reliability index provides a standardized measure related monotonically to failure probability, with larger values generally corresponding to lower failure probability. Hasofer and Lind gave this index a clear geometric form [2]. Their index measures the shortest distance to the failure boundary in a standard probability space.

The time period also needs a clear statement. A one-year bridge risk is not the same as a fifty-year bridge risk. A member risk is also not the same as a system risk. One beam may fail without causing the full structure to fail. So the report must name the event, the part of the structure, and the time period. These details may look basic. However, they control the meaning of the final number.

2.2. Uncertainty and input modeling

The next step is to list the uncertain inputs. Resistance inputs may include steel strength, concrete strength, dimensions, stiffness, and connection behavior. An older structure may add corrosion, crack size, or repair quality. Demand inputs may include traffic, wind, snow, heat, earthquakes, and impact. The model should include inputs that can change the decision. The model does not need every possible detail. Too many weak inputs can make the result harder to check.

Engineers often split uncertainty into two types. Aleatory uncertainty means natural variation. Traffic will vary even after many years of counting. Epistemic uncertainty means missing knowledge. Limited test data are a common source of epistemic uncertainty. A rough structural model can cause it too. More data can cut epistemic uncertainty. More data cannot remove all natural variation. This split helps the engineer choose the next action. Natural variation often stays inside the probability model. Missing knowledge may call for a test, an inspection, or a wider range of results.

The shape of each probability distribution matters. Two models can have the same mean strength. However, the models can give very different chances of low strength. The link between variables matters too. Wind loads from one storm may rise together. Material samples from one batch may share the same bias. An independence rule can miss a bad input mix. Conversely, imposing a strong dependence without empirical or physical justification can distort the estimated failure probability. So engineers should use data or a physical reason for each link. They should test more than one reasonable case when the link is not clear.

Measurements need their own error model. A sensor does not show the exact bridge state. Noise can change the reading. Heat can affect several sensors. Drift can look like damage.

If measurement error is omitted from the observation model, Bayesian updating may misattribute sensor noise or drift to changes in structural parameters. The updated model may then look too certain. So the observation model should include known errors. The report should also say what the sensor cannot measure.

2.3. Sensitivity analysis of influential inputs

A sensitivity check can make the model smaller and clearer. The check asks how each input changes the result. The first-order reliability method (FORM) gives a local view near one main failure point.

DOI: 10.54254/2753-8818/2026.37507

Global sensitivity analysis evaluates input influence over the full specified uncertainty range rather than only near one design point. Radoń and Zabojszcza used both types in an engineering example [3]. Their study shows a practical use. Engineers can prioritize more accurate measurement of influential input variables and spend less effort on inputs that have little effect.

Sensitivity results still need care. An input may look unimportant near one failure point. The same input may matter in another part of the space. A result can also change when inputs depend on each other. So engineers should not use one ranking as a final truth. They should use it as a guide for more checks. A simple ranking is most useful when the model, the data range, and the decision are stated beside it.

3. Choosing a way to estimate and update risk

3.1. Direct simulation as a clear baseline

Monte Carlo simulation uses a direct set of steps. Each simulation draws one realization of the uncertain input variables from their prescribed probability distributions. The structural model gives one response. The computer repeats this process many times. The failure probability is estimated by the proportion of simulated samples that fall within the failure domain. This method can use a nonlinear model. It can also use a model that has no simple equation. The method is easy to explain. So it often gives a good check for a more complex method.

The main problem is rare failure. A sample may contain no failures when the true risk is very small. Estimating very small failure probabilities with crude Monte Carlo simulation may require an impractically large number of samples. However, a large sample may cost too much when each bridge model takes hours to run. The analyst should also report the sampling error. One failure probability does not show how much the answer changes from one run to another. A confidence range can make this numerical error clear.

3.2. Fast local approximation for simple failure surfaces

FORM transforms the inputs to a standard probability space and then finds the most likely point on the failure boundary. It locally approximates the limit-state surface by a tangent hyperplane at the design point, reducing the number of required model runs. FORM also indicates which inputs matter locally near that point. The method works well when one smooth region controls failure. However, FORM can miss another failure region and can also have difficulty with sharp or non-smooth boundaries.

The second-order reliability method (SORM) improves the FORM approximation by incorporating the local curvature of the limit-state surface near the design point. This change can improve accuracy for nonlinear limit-state surfaces. However, SORM remains a local approximation and may still miss a distant or disconnected failure region. FORM and SORM are therefore most useful as fast estimates when the relevant failure region is sufficiently well characterized. A simulation-based check can help determine whether the local approximation is too simple.

3.3. Rare-event sampling near failure

Rare-event methods move model runs toward the area that matters. Importance sampling draws more points near failure. Likelihood-ratio weights are then used to correct for sampling from the alternative importance distribution. Subset simulation uses another idea. It splits one rare event into

DOI: 10.54254/2753-8818/2026.37507

several events that are less rare [4]. Each event moves the sample closer to failure. This method can work well in a large input space.

These methods save work, but they need settings. Importance sampling needs a useful sampling region. Subset simulation requires intermediate conditional-probability thresholds and a transition-sampling rule. A poor setting can give an unstable answer. So the analyst should repeat the run. The analyst should also change key settings. A second method can help when the failure shape is not known. The final report should show these checks and not only the best number.

3.4. Bayesian model updating with engineering data

Bayesian updating starts with what the engineer knows before a new test. This knowledge forms the prior distribution. Design files, material records, old tests, and judgment can shape the prior. A likelihood then links possible model values to the new data. The posterior distribution combines prior information with the likelihood of the observed data through Bayes' theorem. The posterior keeps the uncertainty that remains. This feature is useful because one best-fit value can hide many other possible values.

Engineering data come from many places. Material tests give strength and stiffness data. Load tests give strain and movement under a known load. Vibration tests give frequencies and mode shapes. Sensors give long records under traffic and weather. Inspections give crack, corrosion, and repair data. Each source shows only part of the structure. So the update should say which model values the data can support. The update should also keep a range for values that the data cannot separate.

A full data update may need many costly model runs. Yang and coauthors reduced this cost with a bridge substructure and new sample points [5]. Zeng, Kim, and Qin used sensitivity checks and a Kriging model in a dynamic update [6]. Both studies used a smaller or faster model for repeated work. However, the faster model must keep the response that the data measure. A local bridge model can still depend on the full bridge. A fast model can also smooth out an important response.

A good update must predict data that were not used to fit the model. The test can use another load case. It can also use another sensor or time period. A close fit to the training data is not enough. Model values can absorb an error in supports, materials, or loads. The posterior may then look narrow for the wrong reason. A separate prediction test can find this problem.

3.5. Surrogate learning with explicit validation

A surrogate model is a computationally inexpensive approximation of the input-output relationship of a high-fidelity structural model. Common choices include response surfaces, Kriging, support vector methods, and neural networks. The surrogate learns from a set of full model runs. It can then give many low-cost predictions. Reliability work gives the surrogate a hard task. The model must be right near the failure boundary. A small error there can change many cases from safe to failed. A low average error may not show this problem.

Active learning adds selected high-fidelity model runs as the surrogate is refined. One established Kriging-based approach uses the predicted response and local uncertainty to select new samples near the estimated failure boundary [7]. The high-fidelity model then provides the reference response at the selected point, and the surrogate is updated. This cycle concentrates expensive evaluations in regions that are informative for the failure probability rather than in areas that are clearly safe or clearly failed.

DOI: 10.54254/2753-8818/2026.37507

Newer studies use other learning models. Bai and Song used physics to guide neural-network samples near failure [8]. Nguyen and coauthors used a flexible model and a range for prediction error [9]. Xue, Su, and Li used support vector regression with a new sample rule [10]. Their tests included more than one failure point and separate failure regions. These studies show that sample choice can matter as much as model choice.

The analyst should keep surrogate error separate from structural uncertainty. A large Monte Carlo sample does not correct a poor surrogate. The surrogate should be tested at new high-fidelity points near the final failure boundary, and the estimated failure probability should be tracked as the training set grows. Stability of the estimate is useful evidence, but convergence of the estimated failure probability does not prove that every disconnected failure region has been identified. Exploratory high-fidelity evaluations away from the identified boundary can help detect missed failure regions.

4. Lessons from recent engineering studies

4.1. Local updating and whole-structure validation

Bridge models show why the scale of an update matters. Yang and coauthors focused their update on a smaller bridge region [5]. This choice saved full-bridge runs. It also matched the place where the data gave the most information. However, the local region still received forces from the rest of the bridge. A wrong boundary condition could move the error into the updated values. So the final values need a whole-bridge check.

Chen and coauthors studied a steel-concrete girder with a large load test [11]. Their detailed model gave a closer link between model values and material behavior. This added detail can make the updated parameters more closely connected to physically meaningful material behavior. However, more detail also adds more unknown values. One test may not separate all of them. Several value sets may match the same measured movement. So model detail should grow only when the data can support it.

These studies give one shared lesson. Engineers should choose the smallest model that still keeps the behavior of interest. They should then test that model at a larger scale. This two-level check can save time and still protect the system meaning of the result.

4.2. Decision relevance of data updating

Qu and Sun used long-term structural health monitoring data from a cable-stayed bridge. Their model dynamically predicted extreme main-girder stress, updated a warning threshold as new data arrived, and estimated time-varying reliability indices [12]. This result shows a useful move from a fixed rule to a data-based rule. However, changes in sensor measurements may arise from temperature, traffic, noise, or drift rather than structural damage. The warning model must distinguish these effects as far as the available data allow. A model that adapts too quickly can raise false alarms, while one that adapts too slowly can miss rapid structural changes.

Wang and coauthors studied an existing masonry structure. They updated resistance and future load probability models and then combined component-level results into an overall safety grade using member weights [13]. This step connected reliability calculations to a practical assessment outcome. However, the final grade also depended on weighting and grading rules that are modeling choices rather than quantities determined directly by the data. These choices should therefore be reported together with the resulting safety probability or grade.

DOI: 10.54254/2753-8818/2026.37507

Alzbutas and Dundulis linked a finite-element model with probability software [14]. Their work showed how a detailed safety check can carry uncertain inputs through the full model. It also showed a long chain of possible errors. The physical model can be wrong. The probability settings can be weak. Data-transfer and interface errors can also occur when information is passed between software tools. So engineers must check the full work chain before they use the final risk in a decision.

The main point is simple. A data update should lead to a clear action or a clearer choice. The update may change an inspection date, a load limit, a repair plan, or a warning level. If the update only changes a model value, the engineer still needs to explain why that change matters.

4.3. Physical fidelity in fast reliability methods

Fatigue provides a demanding test for fast reliability methods because damage accumulates over many load cycles. The result depends on the load history, material properties, crack-growth or damage model, and associated uncertainties. Qian, Wei, and Huang used active-learning Kriging for fatigue reliability [15]. Their method placed new high-fidelity evaluations near the uncertain failure boundary and reduced the cost of the probability calculation. However, computational acceleration cannot compensate for a poorly specified load, crack-growth, or damage model.

Methodological studies often evaluate learning-based reliability methods on analytical benchmark functions and simplified mechanical examples. These tests help because a known answer can show if a method finds more than one failure region. However, a field structure is less clean. The field data may be missing. The support conditions may change. The sensor may drift. So a method should pass a clean test first. It should then pass a test with real data problems.

A fair speed claim needs more than a count of full model runs. One study may spend a large amount of time on training or on a search set. Another study may hide that cost. So a comparison should report full model calls, total time, and error in the final probability. The study should also repeat the run. A method has little value if it is fast but gives a different answer each time.

Taken together, these studies show that the main challenge is not simply choosing the most advanced reliability method. Reliable decisions require three elements to be considered together: physical fidelity at the relevant structural scale, validation that exposes model and data error, and a clear link from updated reliability estimates to an engineering action. Faster models, richer data, or more detailed simulations can each improve one part of the analysis, but none alone guarantees a better decision. These common lessons motivate the research gaps discussed in Section 5.

5. A practical path and future needs

5.1. A clear work path for engineers

A practical study can follow five steps. First, the engineer states the decision, the failure event, and the time period. Next, the engineer builds a short list of important uncertain inputs. The engineer records the data source and error for each input. Then the engineer chooses a method that fits the model cost and the failure shape. After that, the engineer checks the result with another method or new data. Finally, the engineer reports both the risk and the main reasons for uncertainty.

The method should match the problem. FORM can give a fast first answer for one smooth failure area. Direct Monte Carlo can give a clear answer when the model is cheap. Rare-event simulation can help when failure is very unlikely. Bayesian updating can use tests and sensor data. A surrogate can help when one full model run costs too much. Method selection should be based on accuracy,

DOI: 10.54254/2753-8818/2026.37507

computational cost, interpretability, and problem requirements rather than novelty alone. A simple method may be better when it gives enough accuracy and a clearer reason for the result.

The final check should focus on the decision. The engineer should vary plausible input distributions, dependence assumptions, and model settings in sensitivity analyses. The engineer should also test another load case when data are available. If the decision changes under a small and reasonable model change, the result is not strong. The engineer may need more data or a safer choice. If the decision stays the same across the checks, the result has more support.

5.2. Outstanding research challenges

The synthesis above points to five recurring research gaps. The first is the lack of true failure data. Safe structures spend most of their lives far from failure. Sensors mainly record normal use, weather, and noise. A model can fit this common data and still be wrong in the far tail. Simulated failures can help. However, they carry the rules of the simulation model. Future studies should mark which data are measured and which data are simulated. They should also state where the model moves beyond observed cases.

The second problem is model error. Every structural model leaves something out. The supports may be too simple. The material rule may be wrong. The damage shape may be unknown. A Bayesian update can hide these errors inside new model values. Future work should give model error its own term. It should also use data from a different test for the final check.

The third problem is change over time. Traffic, climate, use, and repairs may not follow the past. A smooth trend may miss a sudden change. Future models should update as new data arrive. They should also use more than one future case when the path is not clear. A bridge model may need a normal growth case, a high-growth case, and a repair case. These cases show the effect of future choices.

The fourth problem is poorly quantified surrogate-model error. A neural network or support vector regression model can appear accurate on average while still having large or poorly calibrated errors near failure. Physics-based constraints can guide sample selection and limit implausible outputs. Bai and Song showed one way to use physics near the failure surface [8]. Future models should also evaluate uncertainty estimates near the failure boundary. Reported predictive intervals should be calibrated so that empirical coverage is consistent with the stated confidence level.

The fifth problem is weak reporting. Papers often use different tests for speed and accuracy, which makes methods difficult to compare. Future studies should report input definitions, data sources, distribution choices, dependence assumptions, stopping rules, computational cost, and validation results. Shared benchmarks should include missing data and model error rather than relying only on idealized analytical functions.

Future systems should also connect new information to action. A system could compare the value of an inspection with its cost and could support decisions about sensor configuration, load limits, or repair timing. However, the system should preserve the engineer's role in interpreting evidence and consequences. It should make clear what the data support, what remains uncertain, and which assumptions drive the recommendation. This separation can make the resulting action easier to audit and defend.

6. Conclusion

Structural reliability assessment connects uncertain loads, resistance, deterioration, model error, and observations to a clearly defined limit state. Its practical value depends on stating the decision

DOI: 10.54254/2753-8818/2026.37507

context, failure event, reference period, and system boundary before calculating a probability. These definitions determine how the estimated failure probability and reliability index should be interpreted. They also provide a consistent basis for comparing alternative models and updating decisions when new engineering data become available. This decision-centered formulation makes assumptions and uncertainty sources easier to audit. The work should start with a decision. The engineer should then define the event and the time period. This order gives meaning to the failure probability and the reliability index. It also helps ensure that the reliability analysis remains directly connected to an engineering decision.

Each method answers a different need. FORM and SORM give fast local answers. Monte Carlo gives a direct answer but may need many runs. Rare-event methods spend more runs near failure. Bayesian updating uses tests and field data. Surrogates and active learning cut the cost of slow models. However, every method needs checks. The engineer must test the input model, numerical settings, data errors, and model errors.

Recent studies show that data can improve bridge, building, and fatigue decisions. They also show that a fast or detailed model can still give false confidence. No one method is best for every structure. The best method fits the failure shape, data quality, model cost, and effect of a wrong choice. The final result should be simple enough to trace. It should also be strong enough to guide a real engineering action.

References

[1] Primo, G. S. P., Silva, R., Evangelista, F., Jr., & Oliveira, M. H. (2025). Statistical reliability analysis for assessing bridge structural integrity: A review paper. *Infrastructures, 10*(7), Article 156. https://doi.org/10.3390/infrastructures10070156

[2] Hasofer, A. M., & Lind, N. C. (1974). Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division, 100*(1), 111–121. https://doi.org/10.1061/JMCEA3.0001848

[3] Radoń, U., & Zabojszcza, P. (2025). The application of structural reliability and sensitivity analysis in engineering practice. *Applied Sciences, 15*(1), Article 342. https://doi.org/10.3390/app15010342

[4] Au, S.-K., & Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics, 16*(4), 263–277. https://doi.org/10.1016/S0266-8920(01)00019-4

[5] Yang, S.-H., Yi, T.-H., Qu, C.-X., Zhang, S.-H., & Li, C. (2023). Adaptive sampling-based Bayesian model updating for bridges considering substructure approach. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 9*(3), Article 04023024. https://doi.org/10.1061/AJRUA6.RUENG-1077

[6] Zeng, J., Kim, Y. H., & Qin, S. (2023). Bayesian model updating for structural dynamic applications combing differential evolution adaptive metropolis and kriging model. *Journal of Structural Engineering, 149*(6), Article 04023070. https://doi.org/10.1061/JSENDH.STENG-10837

[7] Echard, B., Gayton, N., & Lemaire, M. (2011). AK-MCS: An active learning reliability method combining kriging and Monte Carlo simulation. *Structural Safety, 33*(2), 145–154. https://doi.org/10.1016/j.strusafe.2011.01.002

[8] Bai, Z., & Song, S. (2023). Structural reliability analysis based on neural networks with physics-informed training samples. *Engineering Applications of Artificial Intelligence, 126*, Article 107157. https://doi.org/10.1016/j.engappai.2023.107157

[9] Nguyen, T.-T., Dang, V.-H., Hoang, D.-M., Pham, X.-D., Nguyen, T.-H., & Dinh, V.-T. (2024). Robust active learning framework for structural reliability analysis using uncertainty quantification and flexible meta-model. *Structures, 63*, Article 106465. https://doi.org/10.1016/j.istruc.2024.106465

[10] Xue, G., Su, M., & Li, J. (2025). A novel active learning method combining adaptive support vector regression and Monte Carlo simulation for structural reliability assessment. *Buildings, 15*(22), Article 4183. https://doi.org/10.3390/buildings15224183

[11] Chen, D., Kim, C.-W., Zhou, X., & Strauss, A. (2025). Bayesian updating of nonlinear constitutive models for steel–concrete composite girder bridges using large-scale load test data. *Journal of Bridge Engineering, 30*(1), Article 04024100. https://doi.org/10.1061/JBENF2.BEENG-6848

DOI: 10.54254/2753-8818/2026.37507

[12] Qu, G., & Sun, L. (2024). Performance prediction for steel bridges using SHM data and Bayesian dynamic regression linear model: A novel approach. *Journal of Bridge Engineering, 29*(7), Article 04024044. https://doi.org/10.1061/JBENF2.BEENG-6435

[13] Wang, F., Zhang, J., Huang, S., Yin, B., & Wu, L. (2025). Probabilistic model optimization and safety assessment methods for existing masonry structures. *Buildings, 15*(20), Article 3716. https://doi.org/10.3390/buildings15203716

[14] Alzbutas, R., & Dundulis, G. (2024). Probabilistic assessment of structural integrity. *Axioms, 13*(3), Article 154. https://doi.org/10.3390/axioms13030154

[15] Qian, H.-M., Wei, J., & Huang, H.-Z. (2023). Structural fatigue reliability analysis based on active learning Kriging model. *International Journal of Fatigue, 172*, Article 107639. https://doi.org/10.1016/j.ijfatigue.2023.107639