Practice Points NUMBER 31

# Seeing the Big Picture: Machine Learning Tools in Building Preservation Decisions

REBECCA NAPOLITANO
JOE KALLAS

Imagine managing a portfolio of 200 historic buildings in a dense urban district. You have climate data from 15 weather stations spanning two decades, structural surveys documenting crack patterns and material decay, soil moisture readings, data on temperature fluctuations, and documentation of every intervention. The result is a rich but complex dataset: thousands of observations across dozens of variables. The challenge is not a lack of data but determining where to focus attention and how to extract meaningful insights from it.

This scenario is increasingly common in preservation practice. Digital documentation tools, environmental sensors, and GIS systems have become standard. Variation in building types, sizes, and construction periods further compounds this complexity. Dimensionality reduction techniques help distill complex datasets to what matters. Instead of examining each variable alone, they reveal underlying relationships and quantify which factors most influence outcomes. This data-driven approach helps mitigate confirmation bias by testing assumptions against the entire dataset, revealing patterns that intuition alone might miss.

Data-driven analysis is not new to preservation; we have done quantitative work for decades.[1] What has changed is scale and accessibility. Modern machine learning algorithms process datasets that were impossible to handle twenty years ago; cloud-based platforms like Google Colab eliminate the need for specialized software or expensive infrastructure; and most importantly, we can now make these tools accessible to practitioners who have never written code.[2]

This *Practice Point* introduces two complementary techniques: factor analysis, which identifies underlying patterns among correlated variables, and feature importance analysis, which ranks variables according to their ability to predict specific outcomes. Factor analysis helps you understand the structure of your data, discovering, for example, that humidity, rainfall, and groundwater levels form a coherent "moisture-related" pattern. Feature importance analysis addresses which variables best predict a specific outcome (e.g., condition rating). At the component level, these methods identify which environmental conditions most affect specific materials or structural elements; at the building scale, they prioritize interventions based on factors threatening overall integrity; at the district or city scale, they allocate resources across multiple sites by identifying shared vulnerabilities and site-specific risks. These tools are not a replacement for professional judgment, but a means of enhancing it, helping practitioners extract more value from the data they already collect and make more confident, evidence-driven decisions.

We begin with project planning to help you assess whether these methods are appropriate for your data and objectives. Then we explain the mechanics of factor analysis and feature importance, followed by a discussion of visualization methods and interpretation. Each step of the process is supported by interactive computational notebooks using a synthetic dataset representing 200 heritage buildings and a range of variables that guide implementation. We close with a discussion of challenges and best practices when using these methods.

## INTERACTIVE NOTEBOOKS

The authors have developed five Google Colab notebooks to provide hands-on training for practitioners. The synthetic dataset includes 200 heritage buildings with 17 variables, including construction year, district, material type, foundation type, environmental conditions (temperature, rainfall, humidity, freeze–thaw cycles, and soil moisture), deterioration indicators (crack width and salt deposition), spatial attributes (latitude and longitude), and outcome measures (condition rating and intervention urgency).

Start with Data Preparation and proceed sequentially.

**Data Preparation (00_data_preparation.ipynb)**
Handles initial data cleaning and preprocessing. Includes file upload, data dictionary reference, handling missing values, one-hot encoding, and standardization. Also provides an optional multicollinearity check using Variance Inflation Factors (VIF).
**Output:** processed_data.csv.

**Simple Statistical Methods (01_simple_statistics.ipynb)**
Covers foundational statistical techniques including correlation analysis, simple and multiple regression, group comparisons, and VIF-based multicollinearity diagnostics with suggested fixes. Best for small datasets (fewer than 50 observations), limited variables (two to five), or when transparent, easily interpretable methods are needed for stakeholder communication.

**Factor Analysis (02_factor_analysis.ipynb)**
Implements exploratory factor analysis workflows including Bartlett's test, KMO test, scree plot evaluation, and factor extraction with Varimax rotation. Also discusses methodological choices (Factor Analysis vs. PCA; Varimax vs. Promax).
**Output: factor_scores.csv and factor_loadings.csv.**

**Feature Importance (03_feature_importance.ipynb)**
Builds and evaluates machine learning models for feature ranking, including Random Forest and Gradient Boosting. Incorporates train-test splitting (with attention to temporal and spatial structure), and model evaluation and comparison. Also introduces alternative models (Ridge, Lasso, Elastic Net, SVM, Neural Networks, KNN) with guidance on when to use each.
**Output: feature_importance.csv and model_comparison.csv.**

**Visualization (04_visualization.ipynb)**
Provides exploratory and publication-ready visualizations such as correlation heatmaps, distribution plots, boxplots, pairplots, summary statistics, and spatial risk mapping.
**Output: High-resolution (300 DPI) PNG figures suitable for reports and publications.**

All notebooks include beginner-friendly explanations, error handling, troubleshooting guidance, and suggested next steps.

Scan this QR code or visit github.com/rkn2/practicePointFeatureImp to access all interactive computational notebooks and datasets.

## Project Planning: Scoping

Before you open a computational notebook or run algorithms, plan. Just as you would not start a structural stabilization project without assessing the building's condition and understanding failure mechanisms, do not jump into data-driven analysis without defining your objectives and evaluating your data. Prerequisites for data inputs and practitioner skills are summarized in Table 1.

**Table 1. Prerequisites for Effective Application of Machine Learning in Heritage Projects**

| Category | Requirement | Description |
|---|---|---|
| **Data Inputs** | Structured Dataset | Data organized in rows (buildings or components) and columns (variables). |
| | Sample Size | Recommended 30+ observations for reliable trends; 100+ for robust feature importance. |
| | Consistency | Variables measured using consistent units and protocols across the dataset (e.g., all cracks measured in mm). |
| **Practitioner Skills** | Domain Expertise | Ability to identify relevant variables (e.g., distinguishing rising damp from condensation) and interpret the physical meaning of statistical results. |
| | Data Literacy | Familiarity with basic spreadsheet operations, including CSV formatting and data cleaning. |
| | Introductory Technical Knowledge | Participants are guided through pre-written Python scripts. Prior coding experience is helpful for deeper customization but not strictly required to complete the baseline analysis. |

Dimensionality reduction methods are best suited to interdisciplinary teams. Preservation planners can use feature importance to prioritize maintenance resources across large municipal inventories, while structural engineers and architects gain quantitative tools to identify dominant drivers of deterioration. Conservators can further refine these analyses by correlating environmental exposure data with specific decay mechanisms. Key interdisciplinary skills required include data curation (collecting consistent field data) and collaborative interpretation (combining statistical outputs with physical evidence).

The workflow follows three stages (Fig. 1). First, data preparation organizes your dataset, handles missing values, and standardizes variables for fair comparison. Second, analysis applies algorithms to identify patterns (factor analysis) or rank importance (feature importance). Third, interpretation translates numerical outputs into preservation insights.

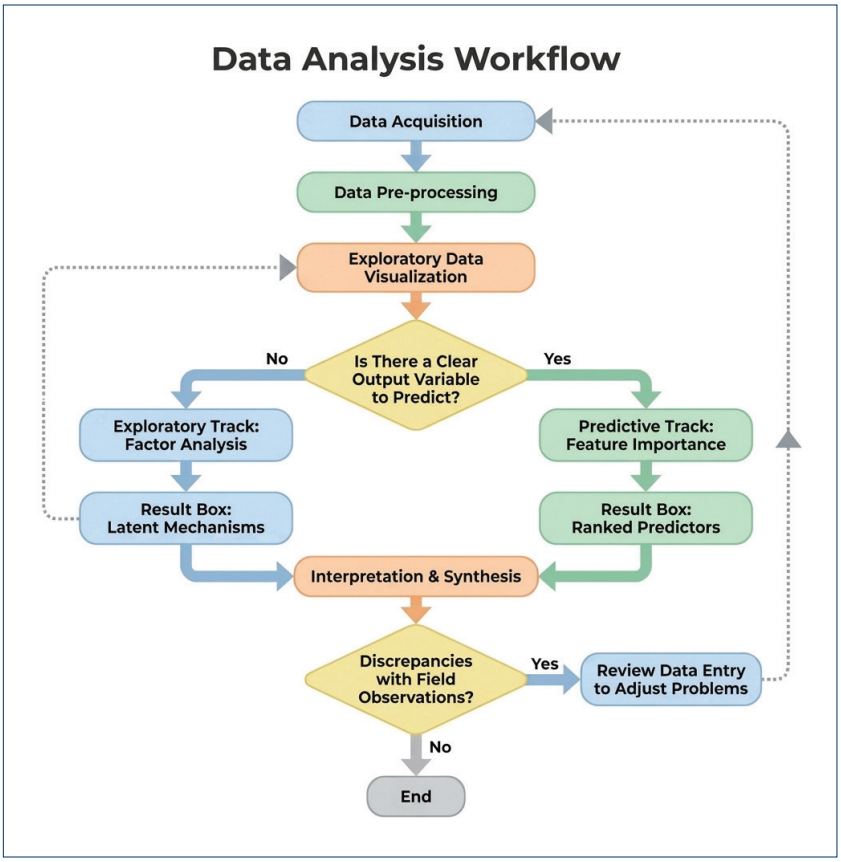


**Fig. 1.** *Proposed framework for data-driven prioritization in heritage preservation. All graphics courtesy of the authors.*

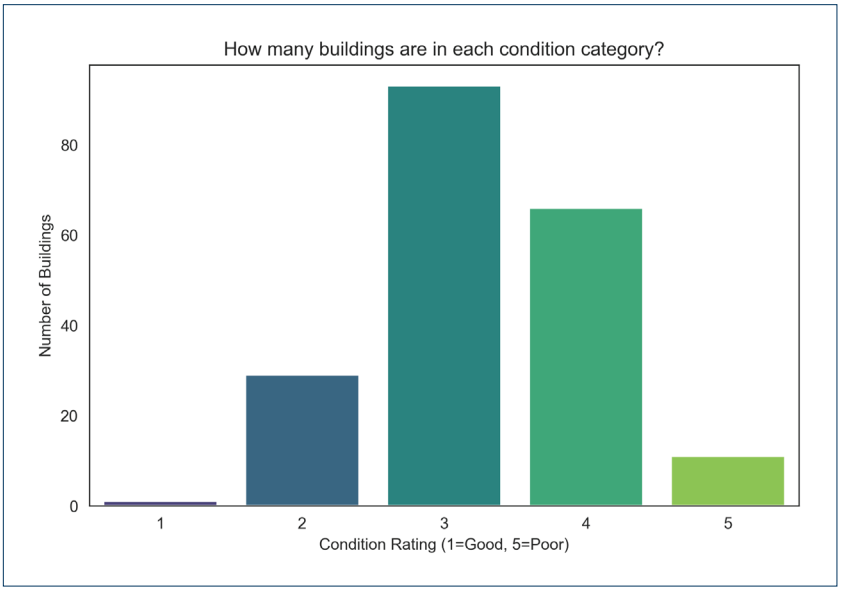


**Fig. 2.** *Distribution of building condition ratings in the sample dataset. Understanding the balance of your data is essential before starting analysis.*

## Data Preparation and Simple Statistical Methods

Dimensionality reduction techniques are powerful but not appropriate for every challenge. They excel in complex situations with many variables, unclear relationships between factors, and uncertainty about which parameters matter most.[3] Practical experience and statistical literature suggest minimum requirements. For factor analysis, aim for at least five to ten observations per variable to achieve a stable factor structure.[4] For Random Forest models, 30 observations is the absolute minimum, but 100 or more observations yield more reliable importance rankings.[5] With smaller samples, results become unstable.[6]

Beyond quantity, data quality matters enormously. Variables should be measured using consistent protocols, instruments, and units; inconsistent measurement methods introduce noise that obscures real patterns. Heritage data present specific challenges, including temporal autocorrelation (measurements over time are not independent), spatial dependencies (buildings in the same district share exposures), and survivorship bias (severely deteriorated buildings may have been demolished, removing extreme cases from your dataset). These issues can lead to spurious correlations. Where possible, account for temporal structure by including time lags or using time-series methods, and for spatial structure by including location variables or using spatial statistics. Watch for outliers—extreme or erroneous data points that distort results. Examine them carefully before removing them; in heritage data, extremes may represent significant events, such as storms or structural failures, that are worth understanding rather than discarding. The distribution of condition ratings should be assessed for imbalance, and condition categories should be clearly defined (Figs. 2 and 3).

Before applying dimensionality reduction techniques, examine your data's structure through three complementary statistical visualizations to identify patterns and potential issues. Start by examining correlation patterns to identify which variables move together (Fig. 4). For a more detailed view of relationships, pairwise scatter plots reveal non-linear patterns and outliers that simple correlation coefficients miss (Fig. 5). Before proceeding to analysis, check for multicollinearity using Variance Inflation Factors (VIF), which identify redundant variables that should be removed or combined. These steps are demonstrated in the companion notebooks.

After these data preparation steps are complete, the workflow moves to the parallel tracks of exploratory and predictive analysis, which will ultimately converge on interpretation.

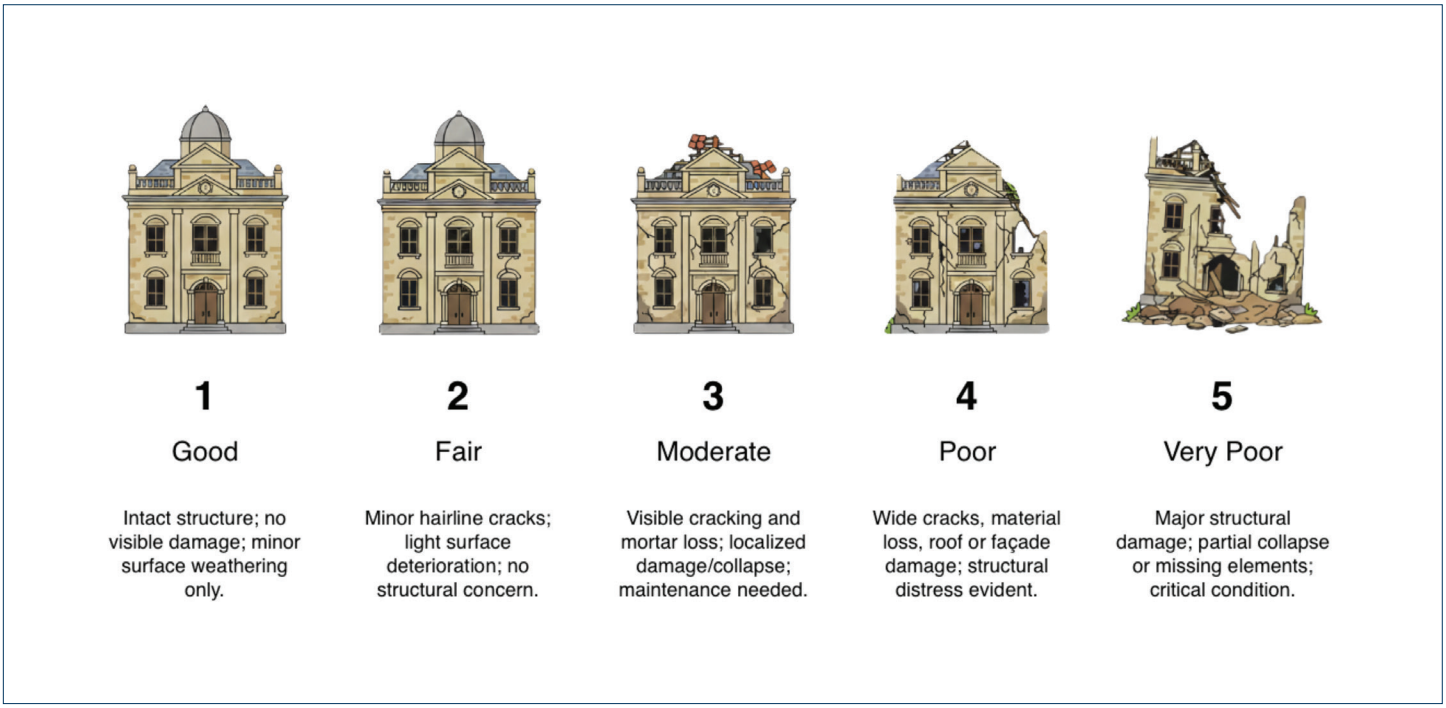


**Fig. 3.** *Visual representation of the five-level building condition rating scale used in this study, ranging from intact structures to severe structural damage and partial collapse.*

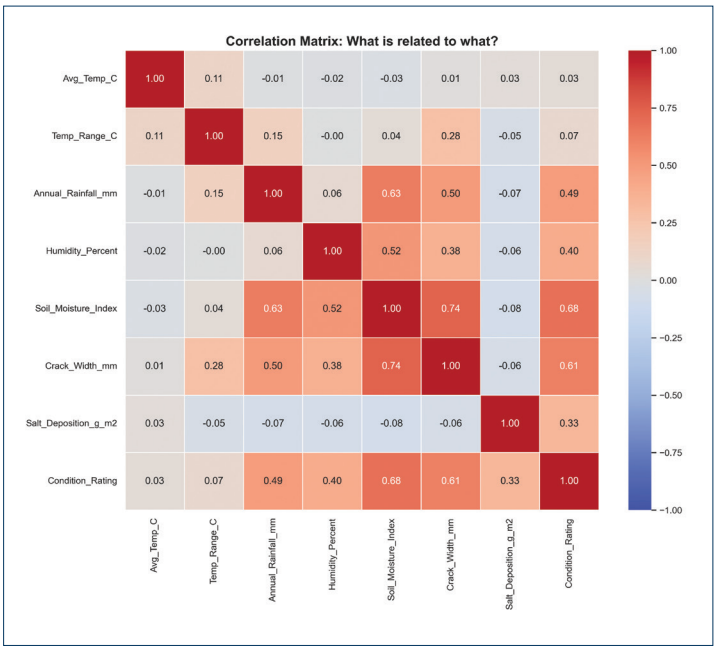


**Fig. 4.** *Correlation matrix heatmap showing relationships between variables. Darker red indicates strong positive correlation, while darker blue indicates strong negative correlation.*

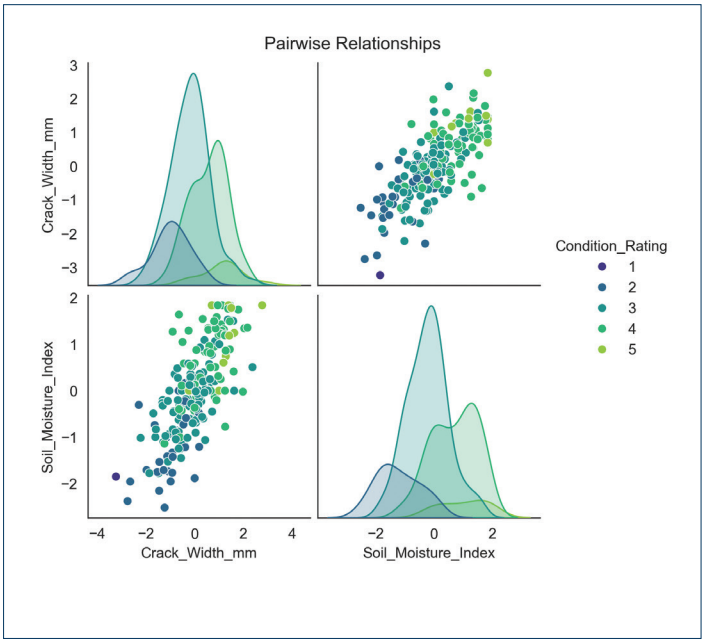


**Fig. 5.** *Pairwise scatter plots showing relationships among key continuous variables. The diagonal shows distributions of individual variables, while off-diagonal plots reveal bivariate relationships. Note that all variables have been standardized for comparability, which is why some variables show negative values.*

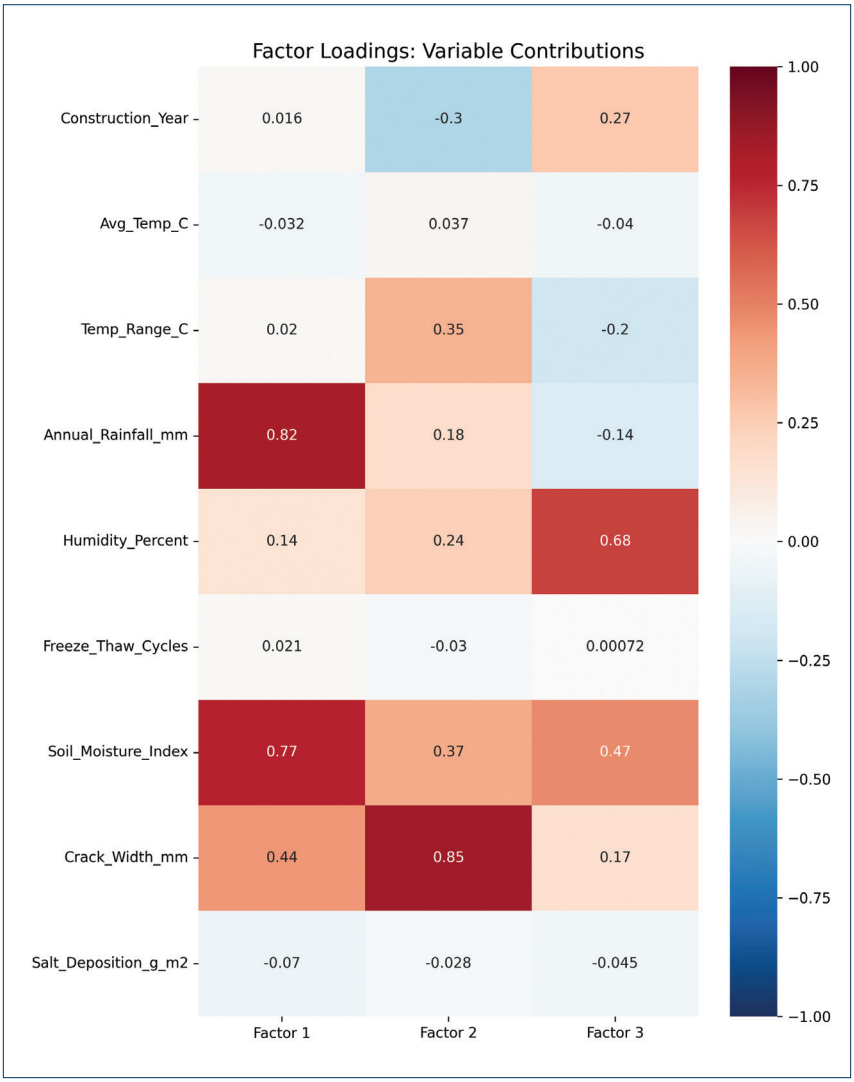


**Fig. 6.** *Factor loadings heatmap showing three extracted factors. Factor 1 captures the groundwater environment in terms of annual rainfall and soil moisture, Factor 2 isolates structural damage represented by crack width, and Factor 3 represents humidity. This separation allows us to test if the environment actually predicts damage.*

## Exploratory Track: Factor Analysis

**What it does.** Factor analysis explores patterns and relationships among parameters that affect structures, but these reflect correlational, not causal, mechanisms.[7] When you measure many variables, such as temperature, humidity, wind exposure, material properties, and structural dimensions, these variables are rarely independent. Temperature and humidity vary together with seasonal patterns; wind exposure and salt deposition co-occur at coastal sites; and soil moisture and groundwater levels fluctuate in coordination.

Factor analysis identifies clusters of variables that vary together, revealing hidden structures.[8] It also identifies and extracts factors associated with structural behavior. These factors represent overarching patterns that account for observed correlations among parameters. Understanding why these correlations exist requires domain expertise; the statistical method only reveals their presence.

Think of factors as categories or groups of conditions that collectively correlate with structural behavior. These conditions are not considered alone but as part of broader, interrelated themes. Instead of separately tracking multiple moisture-related variables (perhaps rainfall, humidity, condensation, groundwater level, and soil saturation), factor analysis might reveal that these cluster into two meaningful factors: “atmospheric moisture” and “ground moisture.” This consolidation simplifies subsequent analysis while preserving essential information. Each factor corresponds to a specific set of related conditions; parameters within the same factor share common characteristics in how they co-vary with structural impacts. The technique assigns a loading to each variable for each factor—a weight indicating how strongly that variable contributes to that factor. High loadings (typically above 0.6–0.7) indicate strong relationships; low loadings suggest the variable is not part of that factor’s pattern.

Factor analysis reduces dimensionality by condensing numerous parameters into fewer meaningful factors. If you begin with 30 variables, factor analysis might identify five to eight underlying factors that capture 70–80 percent of the variation. This simplification focuses attention on critical conditions associated with structural changes. Rather than monitoring and responding to 30 different variables, you focus conservation efforts on the handful of factors that show the strongest associations with deterioration patterns. The results are interpretable factors that represent underlying structural behavior drivers. These factors are derived mathematically; the preservation professional gives them meaning based on domain expertise. In other words, the algorithm identifies which variables cluster together; the professional names the factors. A factor with high loadings on temperature range, freeze–thaw cycles, and thermal expansion might be named “thermal stress.” A factor loading heavily on wind speed, rainfall intensity, and exposed surface area might be termed “weather exposure.”

**When to use it.** Factor analysis is particularly valuable in several preservation scenarios. First, it is useful when you have collected many variables but are uncertain which ones are most relevant or how they relate to each other. Second, it is useful when ongoing monitoring collects numerous parameters, but you need a smaller, more communicable set of metrics for practical decision-making. Third, it is useful when identifying deterioration mechanisms, especially when multiple environmental or structural stresses may be acting on a building and you want to understand which combinations of stresses are most significant. Lastly, it is useful when comparing multiple buildings or sites to determine whether they share similar patterns of stress.

Factor analysis is less appropriate when you have fewer than 10 variables, when your variables are largely independent, or when you already have strong theoretical knowledge of which variables matter most. In such cases, feature importance analysis may be more appropriate.

**How to use it.** Factor analysis involves several steps.[9] First, variables are standardized so that differences in units do not affect their relative influence. Correlations among variables are then calculated, after which factors are extracted to account for the observed relationships. These factors are ordered by the amount of variance they explain, with the first accounting for the greatest share, followed by subsequent factors.

Diagnostic tools then determine how many factors to retain. A scree plot shows where additional factors stop adding meaningful information. You interpret and name factors based on variables with strong loadings (typically 0.6 or higher). If soil moisture, rainfall, and crack width load on one factor, it may be interpreted as "groundwater-driven deterioration" (Fig. 6). Each observation receives a factor score showing how strongly it exhibits each factor.

The companion notebook guides you through the complete workflow: data upload, statistical suitability testing, scree plot visualization, automatic factor number suggestion, Varimax rotation, and interpretation of factor loadings.

## Predictive Track: Feature Importance

**What it does.** While factor analysis identifies groups of related variables, feature importance analysis ranks individual variables by their predictive power for specific outcomes (e.g., deterioration severity, maintenance costs, and intervention timing).[10] The technique trains machine learning models—such as Random Forest or Gradient Boosting—to learn relationships between input variables and target outcomes. These ensemble methods construct multiple decision trees and aggregate results to determine which parameters matter most.[11] Unlike simpler statistical approaches, these models automatically capture variable interactions.[12] High humidity alone may not predict deterioration, but combined with temperature fluctuations, it may strongly predict decay.

After training, the models estimate the predictive power lost if each variable were removed. Variables whose removal causes large performance drops are deemed more important. Importance scores range from 0 to 1, with higher values indicating greater predictive power.

The ranking generated by these models guides the prioritization of conditions and interventions. Parameters identified as highly significant should be prioritized in preservation strategies.[13] If the analysis reveals that groundwater level is the strongest predictor of foundation deterioration, monitoring and managing groundwater can become the top priority.

Feature importance analysis is an iterative process in which models are trained and evaluated using historical data and domain expertise. The process uses part of your data to build the model (the training set) and reserves part to test how well predictions work on new data (the test set). This ensures that the rankings are reliable for decision-making rather than simply describing patterns specific to a single dataset.

Feature importance contributes to dimensionality reduction by filtering out less influential variables rather than grouping them (as in factor analysis). This simplifies both interpretation and data collection by focusing future monitoring and modeling on the variables that drive outcomes.

**When to use it.** Feature importance analysis is particularly valuable in several scenarios. First, use it when you have a clear outcome to predict. The technique requires that you have both input variables (e.g., environmental conditions, material properties, and structural characteristics) and a target (outcome) variable that you want to predict or explain (e.g., deterioration rating, intervention urgency, or structural performance). If you have only input variables without an outcome variable, factor analysis may be more appropriate.

Second, this technique shines when prioritization is the goal. When you need to rank variables to guide resource allocation, monitoring priorities, or intervention strategies, feature importance can help answer questions like "Which three factors should we monitor most closely?" and "Where should we spend our limited maintenance budget?"

Third, use this technique when complex relationships are suspected. If multiple variables interact nonlinearly, traditional regression assumes linear relationships and requires you to specify interactions explicitly. In contrast, Random Forests and Gradient Boosting automatically detect nonlinear patterns and interactions.

Lastly, this method matters when predictive accuracy counts. It is ideal for situations in which you want not just to understand relationships but also to predict future outcomes, such as "Given current conditions, which buildings will require intervention in the next five years?"

Feature importance is less appropriate in several cases: when you lack outcome data, when you have fewer than 30 observations, or when relationships are simple enough that traditional statistical methods suffice.[14] It is also less effective when variables are highly collinear—when several variables measure nearly the same thing—because the model may distribute their importance unevenly. When variables are highly correlated—that is, multicollinear—feature importance rankings become unstable. You can check for multicollinearity by calculating VIF for each variable. Values above 10 indicate severe collinearity.[15] In such cases, consider removing redundant variables (for example, retain annual rainfall but remove monthly rainfall averages), using factor scores from factor analysis

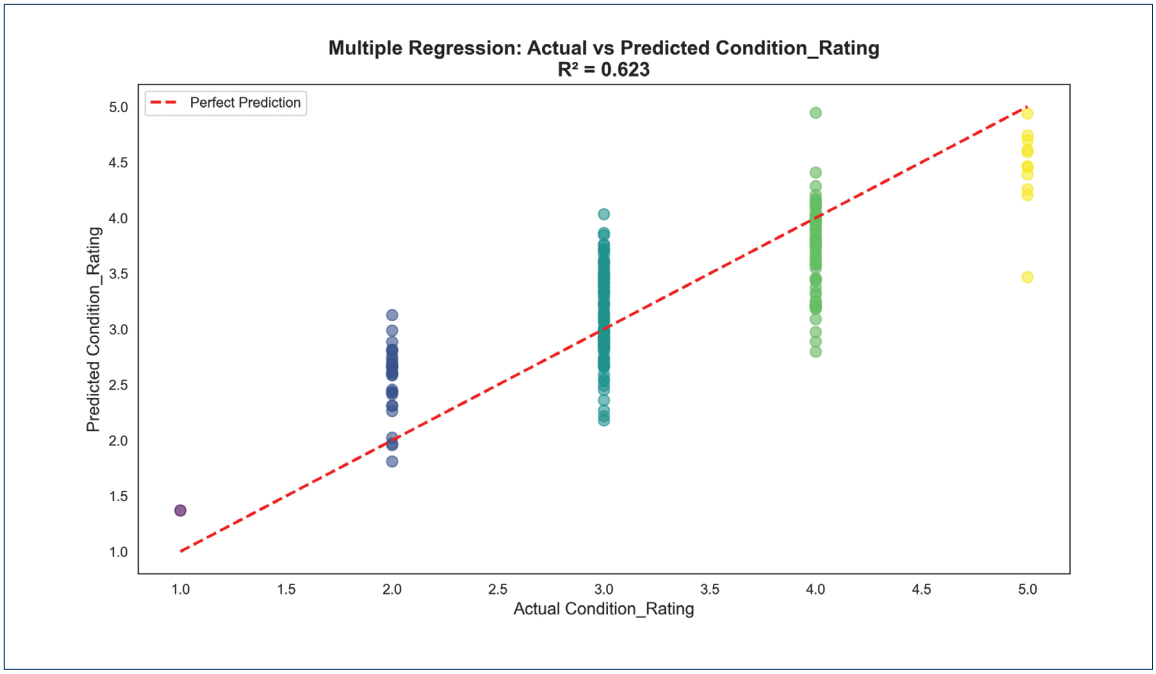


**Fig. 7.** *Actual and predicted values from a regression model. Points clustering closely around the red dashed line indicate accurate predictions, while those far from the line represent buildings whose values were predicted inaccurately.*

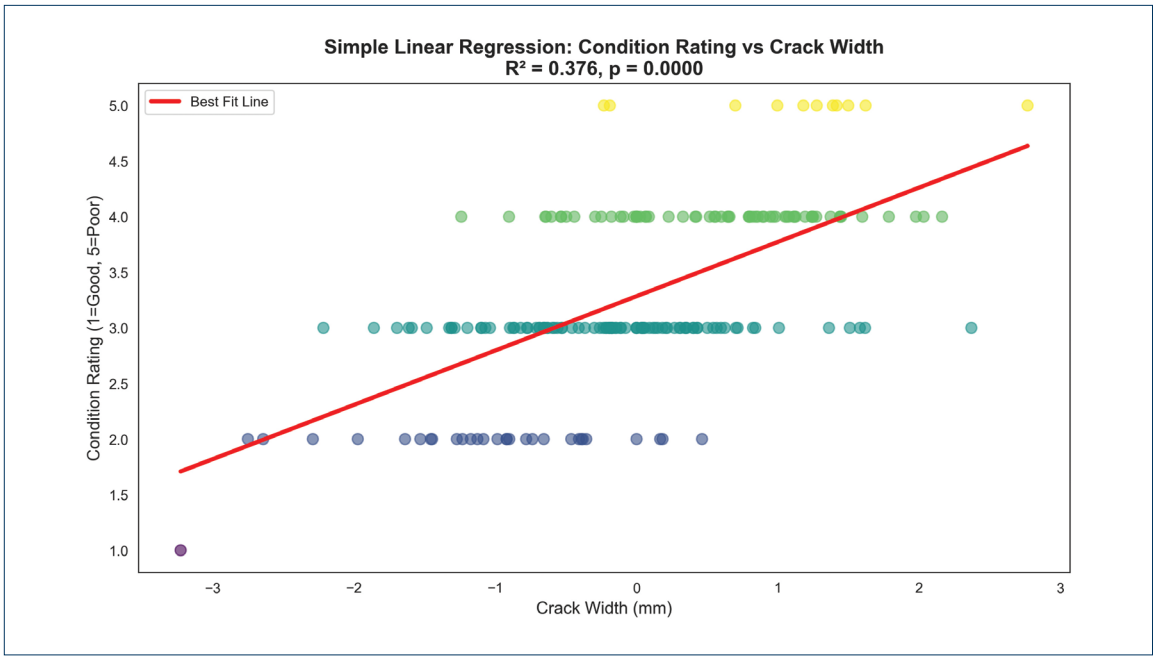


**Fig. 8.** *Simple linear regression of condition rating and crack width, with scattered points demonstrating the insufficient explanatory power of single variables and demonstrating why multivariate approaches are necessary.*

as inputs, or accepting that importance rankings may vary while focusing only on the top three to five features that consistently rank high across different model runs.

The interactive notebooks implement both Random Forest and Gradient Boosting models, which work similarly but have different strengths. We recommend running both models and comparing their feature importance rankings. If both models identify the same top variables as most important, you can be confident in those findings. If the models disagree substantially, that suggests either data quality issues or complex relationships that merit closer investigation.

**How to use it.** While the tutorial notebook largely automates the mechanics of feature importance analysis, interpretation of the results requires a clear understanding of the underlying logic. The process begins with defining a target variable, which must be cast as a numerical or ordinal

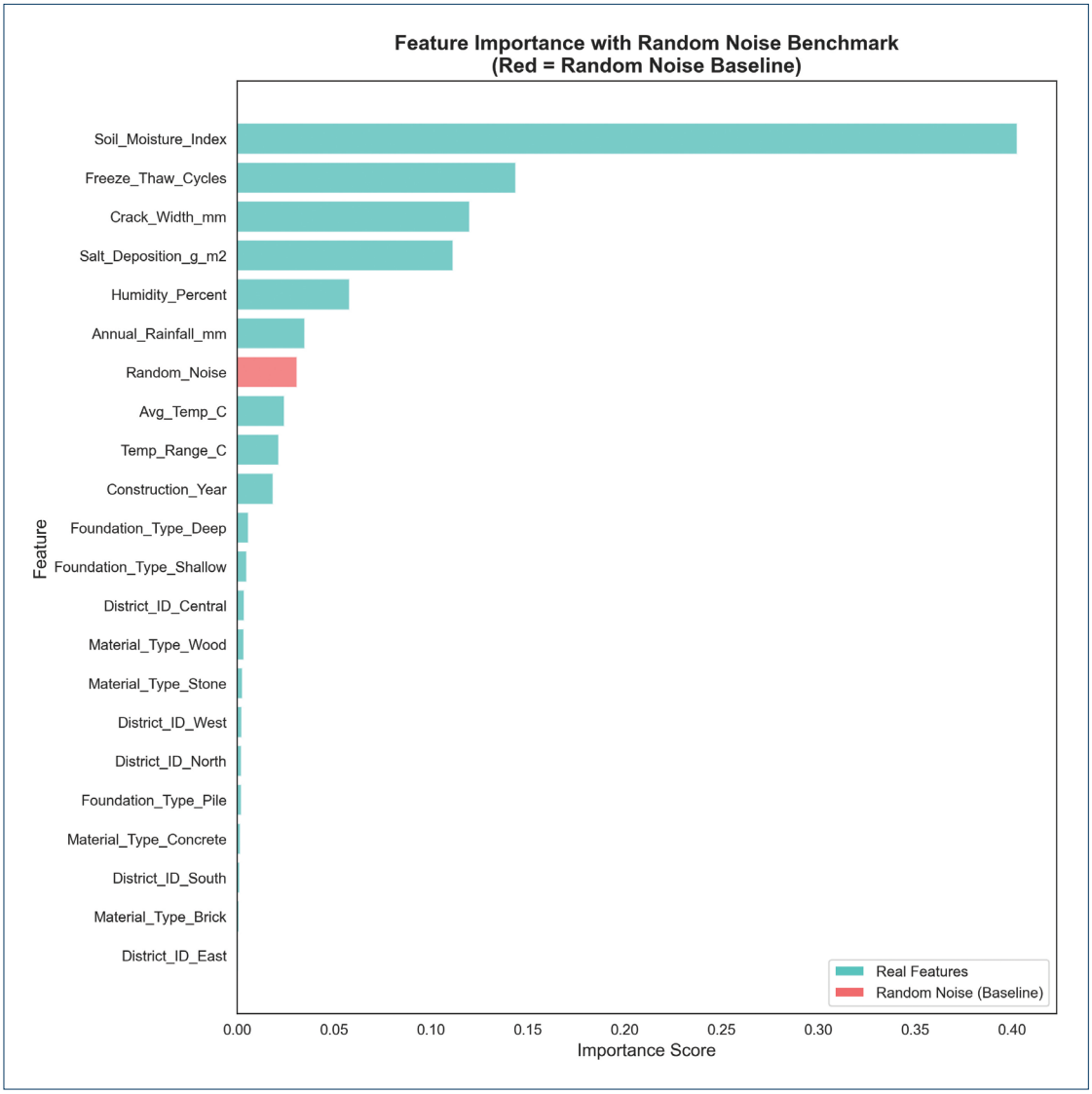


**Fig. 9.** *Feature importance ranking with random noise benchmark (shown in red). Real features scoring above the noise baseline demonstrate genuine predictive power.*

variable; this can be a measurable outcome such as damage severity, deterioration rate, or intervention cost. Simultaneously, feature variables representing environmental factors, material characteristics, or geometric properties are selected and standardized to ensure consistent numeric formatting.

A distinction in heritage data analysis is the strategy for splitting data into training and testing sets. Unlike standard random splits, which can introduce data leakage, heritage datasets often necessitate temporal or spatial partitioning. This may involve training on historical periods (e.g., 2010–2017) and testing on subsequent years (2018–2020), or training on specific geographic districts to test generalization to other districts.

Once partitioned, ensemble algorithms such as Random Forests and Gradient Boosting are employed. These methods construct multiple decision trees to capture the complex, nonlinear interactions inherent in the built environment. Validation is then performed to assess prediction accuracy against domain-specific benchmarks. In the context of heritage datasets, an $R^2$ value between 0.4 and 0.6 typically indicates useful relationships, whereas a value above 0.7 is excellent but uncommon (Fig. 7). Single-variable approaches, by contrast, demonstrate limited explanatory power (Fig. 8).

Similarly, a Mean Absolute Error (MAE) of 0.5 rating points on a standard 1–5 scale is considered useful for prioritization, even if it lacks the precision required for individual assessments. To ensure the stability of these rankings, five-fold or 10-fold cross-validation is recommended.[16] Finally, feature importance is computed by measuring the performance loss that occurs when a specific variable is removed (permutation importance).[17] These results, ideally visualized as bar charts, should be interpreted in conjunction with preservation expertise.

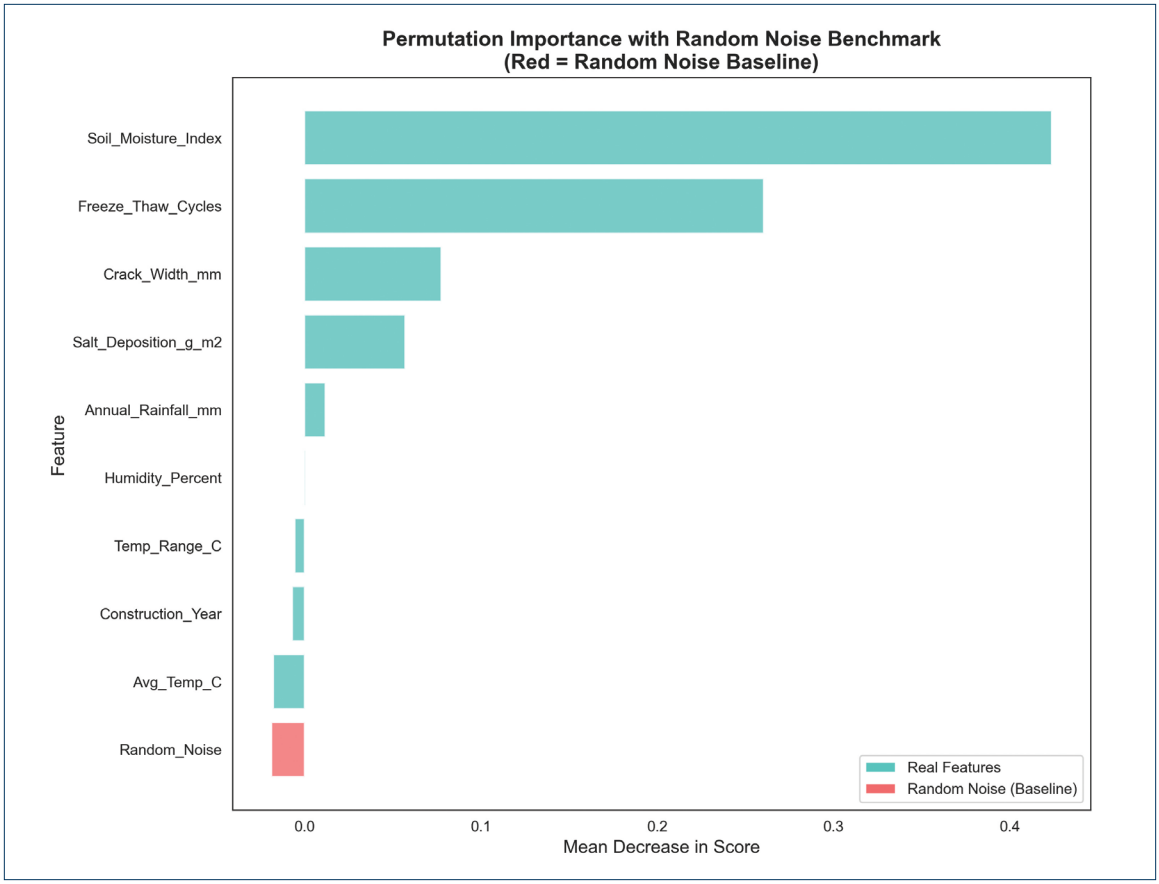


**Fig. 10.** *Permutation importance scores with random noise benchmark (shown in red). This validation technique confirms that the ranked variables truly contribute to predictions rather than being artifacts of the model's internal structure. Features scoring above the noise baseline demonstrate reliable predictive power.*

To ensure that identified features represent genuine predictive relationships rather than spurious correlations, we include a random noise variable as a baseline benchmark. A random variable drawn from a standard normal distribution is added to the feature set before model training. Any real feature scoring below this noise variable indicates weak or unreliable predictive power and should be excluded from interpretation. This approach provides an objective threshold for distinguishing meaningful predictors from statistical artifacts (Fig. 9). To validate that importance rankings reflect true predictive power rather than model artifacts, permutation importance measures performance loss when each variable is shuffled (Fig. 10).

Finally, comparing multiple model types justifies algorithm selection and reveals whether simpler approaches might suffice. Furthermore, using the uncertainty estimates provided by algorithms such as Random Forests allows for reporting predictions with appropriate confidence intervals (e.g., "2.5 ± 0.4"), resulting in more transparent analytical output.

Never treat computational outputs as infallible; always validate results through site inspections and material testing. If model outputs conflict with field observations, reexamine your data. Feature importance provides rankings of influence, not proof of causation. High importance indicates correlation, not necessarily direct causation. Combine computational results with field expertise and view rankings as guides for investigation, not definitive conclusions. High-importance variables often cluster together, suggesting broader mechanisms such as moisture-related stress. The value lies in prioritization: deciding where to focus monitoring or intervention.

The companion notebook walks through model training, validation, and interpretation, including train-test splitting with temporal and spatial considerations, explanations of hyperparameters, dual model training using Random Forests and Gradient Boosting, performance evaluation, and feature importance visualization. The notebook includes model comparison tables as well as troubleshooting guidance for common issues and saves all results as CSV files.

## Combining the Methods

Factor analysis and feature importance are complementary. Factor analysis reveals hidden structures like groups of related parameters representing broader processes like "moisture exposure" or "thermal stress." Feature importance quantifies how strongly each variable predicts specific outcomes. Used together, they can transform complex datasets into actionable insights.

In terms of protocol, you can begin with factor analysis to identify underlying factors and inform hypotheses about relevant physical processes. Then you can apply feature importance to rank specific variables driving observed outcomes. When variables grouped by factor analysis also rank highly in feature importance, confidence in interpretation can increase, whereas discrepancies can prompt further investigation.

Although the workflow in this article is presented generically, closely related approaches combining dimensionality reduction with feature importance analysis have already been successfully applied in several real post-disaster heritage contexts. In earlier work following the 2020 Beirut port explosion, dimensionality reduction strategies were used to manage large sets of correlated architectural, structural, and exposure-related variables, enabling the identification of dominant patterns in damage behavior before applying supervised learning models.[18] While that work relied on feature clustering and selection rather than classical exploratory factor analysis, the objective was the same: to reduce complex, interrelated datasets into a smaller number of interpretable drivers that could meaningfully be linked to physical damage mechanisms. Feature importance analysis then quantified which attributes within those reduced sets most strongly influenced observed damage levels, revealing that social occupancy, exposure conditions, and geometric characteristics often outweighed traditionally emphasized descriptors.

A related methodology was later applied in southern Ukraine following the 2023 Kakhovka Dam breach, which caused flooding that affected historic masonry buildings across the Kherson region.[19] In that study, dimensionality reduction was implemented through hierarchical clustering of correlated features to address multicollinearity and stabilize feature-importance rankings, rather than through formal factor extraction, though the interpretation of correlated variables was guided by factor-based reasoning. The resulting reduced feature set enabled clearer interpretation of dominant damage drivers—such as flood impact duration, proximity to the breach, foundation type, maintenance condition, social occupancy, and building orientation relative to flow direction—while reducing the risk that correlated inputs would distort feature-importance rankings. Across these applications, dimensionality reduction and feature importance served complementary roles: the former structured complex datasets into interpretable groupings, while the latter translated those groupings into ranked, decision-relevant priorities that could inform rapid post-disaster assessment and intervention planning.

## Visualization and Interpretation

Running algorithms represents only the initial phase; the more significant challenge involves translating quantitative outputs into decisive action. In practice, the results of these analyses are intended to inform collaborative decision-making among preservation architects, structural engineers, mechanical, electrical, and plumbing specialists, and building envelope experts, rather than to prescribe solutions from a single disciplinary perspective. While computational models excel at highlighting patterns, their findings require validation through professional judgment and comparison with field observations.

When translating model outputs into intervention priorities, visualizations help identify patterns that inform decision-making (Fig. 11). To operationalize these results, practitioners can rank sites by risk using factor scores or predicted outcomes, establishing specific thresholds that trigger intervention. However, these data-driven priorities must be confirmed via site visits, as models may overlook recent repairs or specific local conditions. Furthermore, final decisions must weigh statistical risk against practical constraints like cost, technical feasibility, and cultural significance. Computational analysis should be viewed as a tool that supports, rather than supplants, core preservation principles. For illustrative scenarios demonstrating convergent evidence, model failure, and high uncertainty, refer to the “Interpretation Guide” in the feature importance notebook.[20]

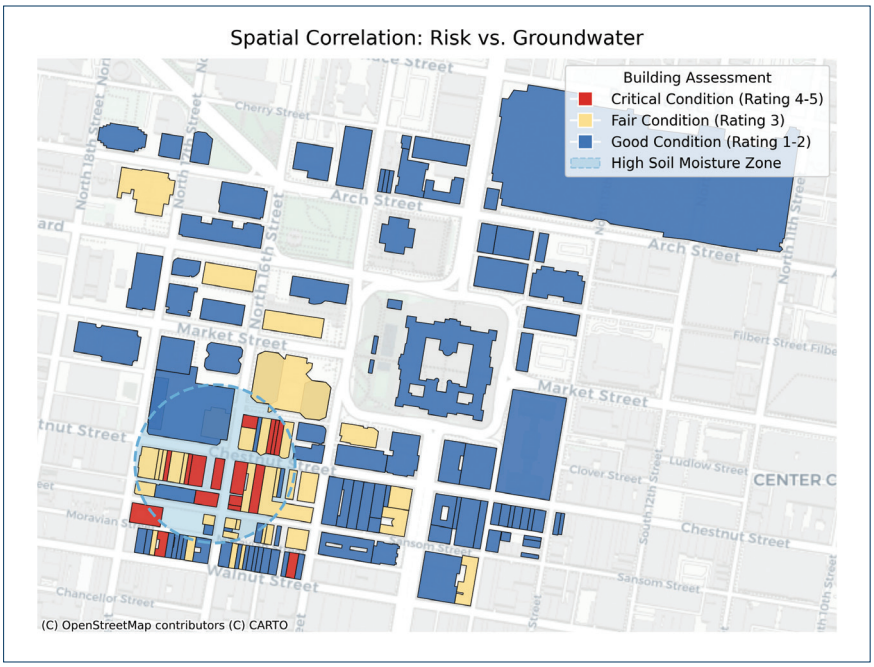


**Fig. 11.** *Synthetic spatial correlation between building risk and environmental factors. The clustering of high-risk buildings (red) overlaps with the high groundwater zone (blue shading), indicating a shared environmental driver. This visualization supports district-level interventions (e.g., drainage improvements) rather than isolated building repairs.*

Discrepancies between model outputs and field observations are not uncommon in heritage practice, particularly when data are heterogeneous or temporally misaligned. Rather than indicating model failure, these conflicts often reveal data gaps or recent site changes.

When model outputs conflict with field observations or domain knowledge, follow these steps. First, quantify the disagreement. How large is the discrepancy? If a model ranks a building as low-risk but you observe active deterioration, document specifically what the model missed (e.g., recent water infiltration or localized spalling). Second, check data quality. Were the variables measured correctly for that building? Common issues include sensors placed in unrepresentative locations, measurements taken before recent damage, data entry errors, or missing variables that could explain the discrepancy. Third, consider temporal mismatches. Models trained on historical data may not capture recent changes. A building may have been in good condition when measured but deteriorated since. Conversely, recent repairs may have improved the condition beyond what historical patterns predict. Fourth, evaluate model limitations. Is the disagreement within the model's expected error range? Fifth, weigh evidence by strength. As a rough guide, assign approximate weights as follows:

- Direct field observation by a qualified professional: 40 percent
- Model prediction with $R^2 > 0.7$ and a large sample: 30 percent
- Historical deterioration records: 20 percent
- Model prediction with $R^2 < 0.5$ or small sample: 10 percent

These are guidelines, not rules; adjust based on data quality and expertise available. Finally, make a decision and document your reasoning. If field evidence outweighs model output, trust the field evidence and investigate why the model failed, as this reveals data gaps or model limitations. If model output is strong and field evidence is ambiguous, use the model to prioritize further investigation. Always document your reasoning for future reference and accountability.

## Challenges and Tips

Applying dimensionality reduction and feature importance analyses in real preservation projects requires care and judgment. Heritage data are rarely ideal since they are often sparse, heterogeneous, and collected under varying conditions over long time spans. The goal is not to produce perfect models, but to extract useful patterns that support informed decision-making. Recognizing the limitations of both data and algorithms helps ensure that computational analysis enhances, rather than distorts, professional expertise.

Heritage datasets are often incomplete, with missing values scattered throughout. Imputation methods can help, but you should interpret results cautiously when more than 10–15 percent of your data requires imputation. Complex models may fit noise rather than signal, a problem known as overfitting. The best defense is validation: always test your model on data it has never seen. If performance drops sharply on test data compared to training data, your model has likely memorized idiosyncrasies rather than learned generalizable patterns.

A highly accurate model has little value if you cannot explain its reasoning to practitioners or stakeholders. This matters especially in preservation, where decisions must often be justified to review boards, funding agencies, or the public. Favor transparent algorithms over black-box models when possible. Computational results should complement (not replace) condition surveys and material analyses. The most robust decisions combine algorithmic insights with field observations and expert judgment.

Document your data sources, methods, and parameter settings so others can reproduce and verify your results. Transparency builds trust in data-driven preservation decisions and protects you if results are later questioned. What matters at the component level (e.g., moisture content in a specific wall) may differ from what matters at the district scale (e.g., groundwater rise affecting multiple buildings). Always interpret findings within the appropriate spatial and temporal context.

Models require periodic updating as new data accumulate. Retraining is needed when new data exceed 20–30 percent of your original training set or annually if you collect continuous monitoring data. Prediction errors should be monitored over time. If MAE increases by 50 percent or more, your model may have become outdated due to changing conditions, a phenomenon called concept drift. For datasets spanning decades, recent 10–15-year rolling windows often perform better than cumulative data because they prioritize current conditions over historical patterns that may no longer apply.

As the field continues to confront increasing data availability and complexity, integrating analytical methods with traditional expertise offers a path toward more transparent, consistent, and evidence-informed decision-making. Future work may expand these models across larger and more diverse datasets, refine interpretability, and further align computational outputs with the practical realities of conservation practice.

**Rebecca Napolitano** is an assistant professor of architectural engineering at Pennsylvania State University. She leverages machine learning to create more efficient and accurate methods for documenting and diagnosing heritage buildings. By automating the analysis of images and 3D data, her research helps better monitor, understand, and safeguard irreplaceable heritage sites worldwide. Rebecca can be reached at nap@psu.edu.

**Joe Kallas** is an engineer, architect, and cultural heritage expert contributing to UNESCO-led recovery efforts in crisis-affected regions worldwide, including

Beirut, Ukraine, and Syria. His work integrates AI, 3D assessment, and digital tools to support local communities in documenting damage, planning emergency responses, and protecting threatened historic urban environments. Joe can be reached at jkallas@sgh.com.

## Notes


1. Stephen D. Mikesell, "Historic Preservation That Counts: Quantitative Methods for Evaluating Historic Resources," *The Public Historian* 8, no. 4 (1986): 61–74, https://doi.org/10.2307/3377500; Richard Veillon, *State of Conservation of World Heritage Properties—A Statistical Analysis (1979–2013)* (UNESCO World Heritage Centre, 2014), whc.unesco.org/en/documents/134872.

2. Veronika Samborska, "Scaling Up: How Increasing Inputs Has Made Artificial Intelligence More Capable," *Our World in Data* (2025), ourworldindata.org/scaling-up-ai; Google, *Google Colaboratory* (2025), colab.research.google.com; Trinh Nguyen, "Top 10 No Code Machine Learning Platforms to Use in 2024" (2024), neurond.com/blog/no-code-machine-learning-platforms.

3. Aasim Ayaz Wani, "Comprehensive Review of Dimensionality Reduction Algorithms: Challenges, Limitations, and Innovative Solutions," *PeerJ Computer Science* 11 (2025), doi:10.7717/peerj-cs.3025; Imola K. Fodor, *A Survey of Dimension Reduction Techniques* (Lawrence Livermore National Laboratory, 2002), computing.llnl.gov/sites/default/files/148494.pdf.

4. Karen Grace-Martin, "How Big of a Sample Size Do You Need for Factor Analysis?" (2021), theanalysisfactor.com/sample-size-needed-for-factor-analysis.

5. Huazhen Wang et al., "An Experimental Study of the Intrinsic Stability of Random Forest Variable Importance Measures," *BMC Bioinformatics* 17 (2016): 60, doi:10.1186/s12859-016-0900-5.

6. M. Luz Calle and Víctor Urrea, "Letter to the Editor: Stability of Random Forest Importance Measures," *Briefings in Bioinformatics* 12, no. 1 (2011): 86–89, doi:10.1093/bib/bbq011.

7. Iván Palomares Carrascosa, "Understanding Correlation and Causation" (2024), statology.org/understanding-correlation-and-causation; Jim Frost, "Correlation vs. Causation: Understanding the Differences" (2023), statisticsbyjim.com/basics/correlation-vs-causation.

8. David Goretzko et al., "Exploratory Factor Analysis: Current Use, Methodological Developments and Recommendations for Good Practice," *Current Psychology* 40, no. 6 (2021): 3510–3521, doi:10.1007/s12144-019-00300-2.

9. Brett Williams et al., "Exploratory Factor Analysis: A Five-Step Guide for Novices," *Journal of Emergency Primary Health Care* 8, no. 3 (2010), journals.sagepub.com/doi/pdf/10.33151/ajp.8.3.93.

10. Mirka Saarela and Susanne Jauhiainen, "Comparison of Feature Importance Measures as Explanations for Classification Models," *SN Applied Sciences* 3, no. 272 (2021), doi:10.1007/s42452-021-04148-9.

11. Leo Breiman, "Random Forests," *Machine Learning* 45, no. 1 (2001): 5–32, doi:10.1023/A:1010933404324.

12. Jerome H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics* 29, no. 5 (2001): 1189–1232, doi:10.1214/aos/1013203451; Zhengze Zhou and Giles Hooker, "Unbiased Measurement of Feature Importance in Tree-Based Methods," *ACM Transactions on Knowledge Discovery from Data* 15, no. 2 (2021): 1–21, doi:10.1145/3429445; Sumanta Basu et al., "Iterative Random Forests to Discover Predictive and Stable High-Order Interactions," *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 8 (2018): 1943–48, doi:10.1073/pnas.1711236115.

13. Joe Kallas and Rebecca Napolitano, "Understanding Critical Masonry Building Attributes Shaping Vulnerability to Blast Loads: Data-Driven Insights from the 2020 Beirut Explosion," *International Journal of Disaster Risk Reduction* 110 (2024): 104640, doi:10.1016/j.ijdrr.2024.104640.

14. Joe Kallas and Rebecca Napolitano, "Understanding Key Building Attributes Influencing Flood Vulnerability in Masonry Structures After the 2023 Kakhovka Dam Breach in Ukraine," *ACM Journal on Computing and Cultural Heritage* 18, no. 3 (2025), doi:10.1145/3744743; Joe Kallas and Rebecca Napolitano, "Beyond Field Surveys: Understanding the Role of 3D Spatial Attributes for Data-Driven Blast Vulnerability Assessment of Masonry Buildings," *International Journal of Disaster Risk Reduction* 128 (2025): 105672, doi:10.1016/j.ijdrr.2025.105672; Jayita Gulati, "Logistic vs. SVM vs. Random Forest: Which One Wins for Small Datasets?" (2025), machinelearningmastery.com/logistic-vs-svm-vs-random-forest-which-one-wins-for-small-datasets/; Fiona K. Ewald et al., *A Guide to Feature Importance Methods for Scientific Inference*, arXiv preprint arXiv:2404.12862 (2024), arxiv.org/abs/2404.12862.

15. Pennsylvania State University STAT 462, "Detecting Multicollinearity Using Variance Inflation Factors" (2025), online.stat.psu.edu/stat462/node/180.

16. Meenu Bhagat and Brijesh Bakariya, "A Comprehensive Review of Cross-Validation Techniques in Machine Learning," *International Journal of Smart and Archived Technology* 16 (2025): 2229–7677, doi:10.1234/ijsat.2025.01.1305.

17. Saarela and Jauhiainen, "Comparison of Feature Importance Measures as Explanations for Classification Models."

18. Kallas and Napolitano, "Understanding Critical Masonry Building Attributes Shaping Vulnerability to Blast Loads"; Kallas and Napolitano, "Beyond Field Surveys."

19. Kallas and Napolitano, "Understanding Key Building Attributes Influencing Flood Vulnerability in Masonry Structures After the 2023 Kakhovka Dam Breach in Ukraine."

20. Rebecca Napolitano and Joe Kallas, "Dimensionality Reduction for Heritage Preservation: Interactive Companion Notebooks," GitHub repository, github.com/rkn2/practicePoint (2025).; GitHub, 2025.


**The Association for Preservation Technology International**
1930 18th St. NW
Suite B2, #1153
Washington, DC 20009