Verification and Validation under Uncertainty Applied to Finite Element Models of Historic Masonry Monuments

Sezer Atamturktur
Architectural Engineering Department
The Pennsylvania State University, University Park, PA 16802

ABSTRACT: The goal of this paper is to take a step towards illustrating the application of tools collectively referred to as Verification and Validation (V&V) to improve the predictive accuracy of large-scale Finite Element (FE) simulations of the National Cathedral, D.C. For this purpose, we first complete a series of code and solution verification studies. “Code verification” quantifies errors due to a potentially deficient implementation of mathematical models in a computer code. “Solution verification” quantifies the numerical error introduced by discretizing mathematical models on a computational mesh. Next, we integrate large amounts of experimental and computational information collected from the choir vaults of the Cathedral. Measurement uncertainty is assessed from the replicated experiments. In parallel, a design of computer experiments is analyzed by perturbing model parameters, which enables us to explore variability of the model parameter domain. Both from the physical measurements and computer experiments, comparative features are extracted probabilistically as mean and variance statistics. Using a Phenomenon Identification and Ranking Table (PIRT), the uncertain parameters that are candidates for calibration are ranked based on the sensitivity of test-analysis comparative features. Parameters whose influence on the output of interest is not significant are disqualified from being calibrated. Once the comparative features and calibration parameters are defined, Bayesian inference is used to compound our prior knowledge about the calibration parameters together with experimental observations collected from vibration testing. Prior probability distribution incorporates expert judgment while the variance of measured features account for the experimental uncertainty. Bayesian inference, then, results in updated knowledge of the calibration parameters in the form of a posterior probability distribution. The point of this exercise is not to obtain a calibrated FE model that identically matches the experimental data. It is, rather, to better understand where modeling uncertainty originates from and to obtain model predictions that are statistically consistent with the measurements and their uncertainty. The complete V&V “story” is told of how to establish confidence in the model of an existing structure.

1. INTRODUCTION

The requirement for periodic structural assessment of historic monuments is primarily driven by the need to rectify the weakening effects of aging. Among all historical buildings, the monuments composed of structurally engaged arches, ribs, webs and buttresses are generally the most commonly encountered and, yet, the most demanding structures to maintain. The implementation of the Finite Element (FE) method for structural assessment of these complex historic buildings emerges as a convenient and fast tool. However, one must bear in mind that the predictive ability of a FE model is tied to the accuracy of numerical analysis, correctness of the input definition and appropriateness of the assumptions established about the physical

1 Ph.D. candidate, The Pennsylvania State University. Mailing: Architectural Engineering Department, The Pennsylvania State University, 104 Engineering Unit A, University Park, PA 16802. Phone: 814-863-9093. E-mail: hsa109@psu.edu.
phenomena of interest. A model with low predictive capability yields erroneous results, which may in turn be detrimental to the maintenance and management of the structure. Historic masonry monuments that challenge the intuition of even the best of engineers are especially prone to this problem.

Engineers are often forced to simplify structures when defining the inter-element connectivity of the structural components by making educated assumptions. For instance, the connectivity of transverse and longitudinal walls in a historic structure depends on many factors, such as contact pressure, surface friction, existing cracks, as well as the unique elastic behavior of each stone unit. However, the in-built boundary connectivity options in general-purpose FE packages include translational and rotational restraints without giving options for the implementation of the underlying physics. An attempt to model joint friction complicates the model due to the unknown friction coefficients (Hills et al, 2004). Thus, although approximate, implementing the boundary conditions built in FE packages remains the only plausible option to the practicing engineer. As seen, the boundary conditions of historic masonry monuments are difficult to model. Nor can we measure these phenomena at the required levels of fidelity and accuracy. There are many similar instances routinely experienced during the development of a FE model that eventually restrict its predictive capability and force arbitrary premises to enter the analysis.

In this work, we attempt the complicated undertaking of defining the poorly known model parameters through inferences made when predictions of a FE model are compared to experimental measurements. The difference, however, between such stochastic inference and conventional calibration, or FE model updating, must be made clear.

**Deterministic Model Calibration:** The conventional approach to achieving predictive accuracy is to maximize the correlation of model predictions and test measurements. This technology is usually deterministic in nature and the predictive capability of a FE model is improved through comparison to a single experimental dataset. Models developed in this way can be considered to provide snapshot representations of reality (Dascotte, 2004). In this discussion, we refer to this approach as the “**deterministic model calibration**.” Deterministic methods, by definition, overlook the uncertainty inherent in any aspect of testing, model development and parameter calibration. However, imprecisely known structural properties unavoidably introduce uncertainty. For instance, determining precise constitutive properties of historic masonry units and mortar mix is hard to achieve. Analysis of material samples or coupons, if allowed, is an available resource. These tests, however, only represent a fraction of the total volume of materials used in the construction. Likewise, models of material behavior, regardless of how elaborate they are, only represent an approximation of reality. Aside from the unknown statistics of unit-to-unit variability of stone and mortar, there is also inherent uncertainty in the information regarding material properties. Managing these sources of uncertainty must indeed be an inherent component of calibration and validation. An overview of methods that deal with the propagation of input uncertainty to output predictions is given by a group of scientists in structural dynamics (Mace, Worden and Mason, 2005).

Aside from the uncertainty of structural properties, calculations and measurements themselves also exhibit uncertainty as illustrated by two benchmark studies. The first study has shown that when multiple analysts develop numerical models of the very same structure, but use different software, element types and mesh sizes, they obtain varying results (Imregun and Ewins 1988). The second study has shown that, similar to computer models, physical measurements are also prone to uncertainty. Coleman and Steele (1999) presented the readings of 24 students from a thermometer to be distributed around a mean value of 97.2° while the true temperature was 96°.

**Stochastic Model Calibration:** Stochastic approaches evolved from the growing awareness of inherent uncertainty in physical measurements and FE modeling. While determinist model
updating aims at reaching a direct match between the analytically and experimentally derived comparative features, stochastic model updating defines the parameters probabilistically and aims to reach a statically accurate model, not necessarily one that best-fits the experimental datasets. Unlike the deterministic approach that “looks” at a single snapshot point in a cloud of realizations, stochastic methods have a strong statistical basis. They can be considered as seeking the correlation of two clouds of points, where the experimental scatter comes from repeated experiments and the analytical scatter comes from repeated computer runs. Stochastic methods need to pay attention to the distance between the centers of gravity of these two clouds of points and to account for their scatters. Throughout the discussion, we refer to this approach as the “stochastic model calibration.”

We begin with Table 1 as the starting point to write the list of potential sources of modeling uncertainty in a FE simulation. It also starts to define possible remedies to mitigate the effect that this uncertainty may have on model predictions. We use the theory of probability to model and propagate uncertainty because it is a mature mathematical theory applicable to the variability in material properties, boundary conditions, and experimental testing that we need to deal with. Advancing model calibration and validation beyond their determinist nature is a goal of this study.

<table>
<thead>
<tr>
<th>Source of Uncertainty</th>
<th>Remedy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input parameter uncertainty</td>
<td>Probabilistic calibration of parameters against experiments</td>
</tr>
<tr>
<td>Inadequacy of the code</td>
<td>Code verification activities</td>
</tr>
<tr>
<td>Experimental error</td>
<td>Probabilistic treatment of repeated experiments</td>
</tr>
<tr>
<td>Numerical uncertainty</td>
<td>Solution verification activities</td>
</tr>
<tr>
<td>Variability of model parameters</td>
<td>Probabilistic Treatment of repeated computer runs</td>
</tr>
</tbody>
</table>

This study is applied to the National Cathedral in Washington, D.C. By considering all knowable sources of uncertainty and error, we intend to obtain an FE model of the vaults of the National Cathedral that can be used not only to make predictions, but also to gain a better understanding of predictive accuracy and uncertainty. To do so, repeated experiments and multiple computer runs need to be integrated within the framework of statistical decision-making. It is imperative to note that we cannot possibly verify and validate every aspect of the model or quantify every single source of uncertainty. V&V activities, by their very nature, cannot prove the “correctness” of a mathematical model; instead V&V should be considered as a way to illustrate the absence of proof that the model is incorrect (Hemez 2007). We have decided, for this publication, to start by going after the most significant modeling assumptions that are discussed in the next section.

These tasks are neither inexpensive nor rapid enough to be immediately implemented in the civil engineering community. However, disseminating the findings of the present study, and many others to be completed in the future, will help to demystify the structural analysis of historic monuments.

2. THE WASHINGTON NATIONAL CATHEDRAL

The National Cathedral was designed as authentic Gothic revival architecture in the early 20th century. Similar to medieval schedules, the construction of the cathedral was interrupted several times over a century and it was eventually completed in year 1990. The construction technique of the cathedral closely followed the medieval techniques of using quarried stone without any reinforcement (Washington National Cathedral, 2004). The quality of construction, however, was better controlled compared to its medieval counterparts. We believe this resulted
in better-quality experimental measurements, as compared to similar vibration tests conducted on original medieval structures such as Beverly Minster, U.K., or Santa Maria Novella Basilica, Italy, and as discussed in Section 6.

Figure 1. The main nave of the National Cathedral, Washington, D.C. (Left: Cross-section through the main nave; Right: Interior view of the nave.)

The choir of the cathedral is vaulted with even-level fan vaults elevated above the stone piers. The vaults are composed of stone ribs and webbing supported by the concrete surcharge, stone piers, walls and buttresses, as illustrated on the left of Figure 1. (Picture courtesy of the National Cathedral.) The present study focuses on vaults covering the choir of the cathedral.

3. BACKGROUND REVIEW

The application of computer predictive simulations to masonry structures immediately raises questions about the validity of their predictions. Mark (1982) initiated one of the earliest efforts to confirm the applicability of the FE method to the analysis of masonry structures by comparing photo-elastic test measurements collected on small-scale, plastic samples of Gothic vaults to FE model predictions. A quarter century later, the applicability of the FE method to various structures has long been accepted in the field of engineering. One difficulty, however, remains the definition of physically sound and appropriate structural property values. When dealing with unreinforced masonry, there are numerous opportunities to misinterpret the actual structure, build an unsuitable model and obtain completely irrelevant or incorrect solutions.

Early calibration activities applied to masonry monuments date back to the late 1980’s. These efforts were based on a visual approach and constituted a mere comparison of crack locations to analytical estimates of tensile zones. See, for example, Mark and Hutchinson, 1986. Visual comparison, however, concentrates on a few locations in a large-scale, masonry building and is of limited effectiveness. It is also susceptible to significant error, especially when support settlements are present in the historic masonry structure. In some masonry structures, destructive and non-destructive tests that focus on stress, strain or deflection under artificial loading, have remedied the problems related to the visual methods. See, for instance, Boothby et al., 1998. Methods based on in situ strain or deflection measurements, while successful when applied to masonry bridges, are impractical for larger masonry structures, such as cathedrals, due to the difficulty in sufficiently loading the structure to achieve a detectable
response spectrum. Because of this practical infeasibility, several scholars have attempted to examine the behavior of scaled laboratory test specimens. One example is discussed by Creazza et al. 2002. The concept of testing scaled specimens has the potential to illustrate the structural behavior for regimes that cannot be tested in real life, for instance, collapse mechanisms. The scaled test specimens, however, represent only a portion of a computational model and thus, this procedure has the drawback of overlooking the actual elastic restraints exerted by adjacent elements. Accordingly, the alternative load distribution paths within the structure are also absent from the analysis.

In the last three decades, the dynamic identification of structural response and model calibration have extended to wide-ranging fields with different practical implementations. Although limited in number, there are studies on masonry structures that are found useful to the present study. For instance, Sortis et al. (2005) calibrated the FE model of a two-story, load-bearing stone masonry building using experimentally obtained modal parameters. Gentile and Saisi (2007) investigated structural damage occurring in a historic masonry tower. Findings published in these studies support the assertion that vibration-based model updating methods offer a potential to deliver useful information about the state of a masonry structure.

The studies reviewed convey one common message: until the model is correlated with real-life data, its predictions must be treated with due caution. As such, verifying and validating numerical models based on rigorous scientific tools (instead of mere engineering judgment) is a field of prime importance. These studies also illustrate that, when properly integrated with the eventual application, experimentation plays an indispensable role in model validation.

4. FINITE ELEMENT MODEL DEVELOPMENT

To develop the FE model, all structural members that significantly contribute to the behavior of interest must be included. Figure 2 illustrates a model, where the behaviors targeted are the static, quasi-static and dynamic responses of the vaulted upper portion of the cathedral.

![Figure 2. Illustration of the full bay model that indicates element and material types.](image-url)
Because this investigation focuses on the response of the vaulted upper portion, it is deemed suitable to model components above the springing level. Components immediately adjacent to the vaults, walls, surcharge and piers also characterize the structural behavior of vaults. At the same time, however, it is of prime importance to simplify the structure to reduce the CPU time it takes to calculate the FE solution. This can be achieved, for example, by replacing the effects of buttresses with translational restraints, as done in this study. The model illustrated in Figure 2 offers a good compromise between fidelity to the actual load paths and fast, turn-around time.

The National Cathedral is a recent, medieval-like construction, which offers the advantage that a full set of drawings are available to the authors. Thus, the geometry of the vaults is derived from these documents. Moldings and decorations typically added onto structural components of Cathedrals require simplification to create the geometry. It is necessary to avoid poor aspect ratios when meshing the geometry, especially at the intersections of multiple ribs. The physical geometry of the piers and ribs of vaults are simplified to rectangles that preserve the area and moments of inertia of the original cross section. Figure 3 depicts the original and simplified cross sections of ribs.

![3-a) Highly ornamented rib. 3-b) Simplified rib geometry.](image)

*Figure 3. Property-preserving simplification of the rib geometry.*

Although, the error introduced by this simplification is minimal and can easily be justified by the improved aspect ratios of the mesh, it cannot be remedied via calibration. Thus, it is important to quantify the effect of geometry simplification on the responses of interest and consider the trade-off between geometric fidelity and ease to mesh the computational domain. Our guiding principle is to simplify the geometry unless the geometrical details impact the dynamic response that the model must capture. We investigated the effect of geometry simplification on the rib maximum displacement. With a simplified geometry, the model predicts a maximum deformation that is within 2% of predictions made with the original, high-fidelity geometry.

In most FE packages, modeling capabilities are available to represent a structural dynamics response using a variety of finite element types. Among all element types available in ANSYS, the SHELL93 elements are deemed to be the most suitable to model the stone webbing. The quadratic shape functions and mid-nodes of SHELL93 make it appropriate to represent curved surfaces without losing accuracy. To model the ribs, piers, surcharge and vault, the 10-node SOLID95 element is used. Because they possess mid-nodes, these finite elements are again the most suitable three-dimensional structural elements to represent curved surfaces.

The stress-strain law of masonry is largely non-linear. Masonry units exhibit inelastic behavior in both tension due to cracking and compression due to irreversible softening effects. In common applications, masonry assemblies are anisotropic and inhomogeneous due to the mortar joints. When analyzing the global response of the structure, modeling the real behavior of a masonry assembly is nearly impossible, and thus approximations and assumptions are made to describe the materials. The present study is initiated with *a priori* assumption of linearly elastic, isotropic
constitutive behavior for the masonry assembly. Table 3 lists the values of modulus of elasticity and density for various structural components. One objective of validation study is to assess the extent to which this hypothesized material model is acceptable.

### Table 2. Our prior knowledge of material properties of limestone and concrete.

<table>
<thead>
<tr>
<th>Structural Component</th>
<th>Material Type</th>
<th>Modulus of Elasticity (E)</th>
<th>Density ((\rho))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walls and columns</td>
<td>Indiana limestone</td>
<td>8-to-21 GPa</td>
<td>2,100 kg/m³</td>
</tr>
<tr>
<td>Vault ribs and webbing</td>
<td>Indiana limestone</td>
<td>8-to-21 GPa</td>
<td>2,100 kg/m³</td>
</tr>
<tr>
<td>Surcharge</td>
<td>Concrete</td>
<td>24 GPa</td>
<td>2,100 kg/m³</td>
</tr>
<tr>
<td>Walls</td>
<td>Brick</td>
<td>8 GPa</td>
<td>1,800 kg/m³</td>
</tr>
</tbody>
</table>

The modeled portions of the structure are supported elastically by the portions excluded from the model. Hence, another assumption is needed while quantifying the physical properties of these supports and representing them as idealized boundary conditions. Because many physical responses, such as deflection, are highly sensitive to the boundary conditions, their precise definitions are crucial to prediction accuracy. When a two-dimensional arch, for example, is modeled with a fixed boundary instead of hinged conditions, the changes brought to the maximum displacement and support reactions are similar to those that would result from varying the modulus of elasticity (or density) of the homogenized stone and mortar assembly by a factor of 100%. A simple error in boundary condition can therefore have a large effect on the accuracy of predictions.

Once the boundary conditions are assigned, the initial model is ready for analysis. The loading condition for which the response is desired is defined based on what is known about the structure and what is expected to be learned from the model. Loading conditions take, for example, the form of abutment movement, gravity loading, wind loading or earthquake loading.

### 5. VERIFICATION ACTIVITIES

Verification invariably deals with numerical uncertainty caused by programming mistakes, deficient implementations of algorithms and models and the spatial and temporal discretization of continuous equations. Code verification activities quantify errors due to a potentially deficient implementation of algorithms and mathematical models in a computer code. Mesh refinement and solution verification quantify the numerical error introduced by discretizing the mathematical models on a computational mesh. Both code and solution verification must be applied to the aspects of the model that have significant influence on the accuracy of solutions of a particular simulation. What this implies is that verification is application specific. We further state that a validation exercise, regardless of the goodness of its results, cannot be considered meaningful until a satisfactory level of verification has been achieved.

#### 5.1 Code Verification

Code verification establishes that the code is, for the most part, bug-free and numerical algorithms are implemented correctly. Code verification involves the comparison of approximate calculations with the known solutions of partial differential equations that the code is solving. The difficulty is that exact, closed-form solutions are only available for a select few cases when the physics, geometry, boundary and initial conditions and forcing functions are simple.

Since the validation assessment proposed relies on vibration measurements, it is particularly important to confirm the ability of ANSYS to accurately predict the vibration characteristics of a thin shell (SHELL93) and solid (SOLID95) deformable body. The case study is selected to be an arch structure as the curvature of an arch introduces coupling between extensional, torsional
and flexural modes. The coupling is very relevant to the present study and cannot be observed with simpler structures, such as straight beams.

In thin shell theory, the fundamental concept is the substitution of stress resultants and couples in lieu of integration through the thickness. This assumption holds accurate when the thickness is less than one tenth of the radius of curvature of the reference surface. When deflections are small and a geometric non-linearity is not introduced, the theory of thin shells has proven to yield accurate solutions. Complete explanations of the theory are available from a large body of literature among which we cite Kraus 1967 and Soedel 2004. The stone webbing of vaults of the cathedral is thin (approximately 0.1 m) and approximately 1/100th the radius of curvature of its vaults. Thus, thin shell elements provide an appropriate representation of the vault webbing.

To carry out the verification of SHELL93 finite elements implemented in ANSYS, a reference arch, that has a 6-meter radius, 1-meter depth and 0.1-meter thickness, is selected. Approximate, but still highly accurate, solutions for the natural frequencies of flexural modes of circular arches are obtained from Blevins, 1993. These closed-form solutions are compared to FE predictions to assess the error introduced by SHELL93 elements that approximate the governing differential equations.

![Figure 4. Correlation between closed-form and FE-predicted natural frequencies.](image)

Figure 4 compares closed-form and FE solutions of natural frequency for the first four resonant modes of the arch. The figure illustrates that the agreement is excellent because the pairs of closed-form and FE solutions are located almost perfectly on the 45-degree line. Figure 5 gives a visual comparison of closed-form and FE-predicted mode shapes for these four modes. The analysis indicates that the SHELL93 element performs accurately when compared to a closed-form solution. It increases our confidence that the code can be used without the risk of providing erroneous solutions for a mechanical behavior that would be key to the validation study.

### 5.2 Solution Verification

As we discussed in the previous section, numerical methods always provide approximate results. Mathematical idealization, that is, the choice of element types in finite element analysis, is the first level of approximation. The second level is introduced by the discretization in time and space of differential equations. The theory of FE modeling states that, when the mesh size Δx is reduced, approximate solutions should converge to the exact solution. This is known as
asymptotic convergence. However, the sufficiency of mesh size to achieve a desired level of numerical error is always restricted by the available computational resource. Solution verification is necessary to quantify the difference between discrete solutions obtained when the problem is run at mesh resolution \( \Delta x \) and asymptotically converged solutions (Hemez, 2007).

(a) Mode-1, closed-form solution.  
(b) Mode-1, FE model solution.  

(a) Mode-2, closed-form solution.  
(b) Mode-2, FE model solution.  

(a) Mode-3, closed-form solution.  
(b) Mode-3, FE model solution.  

(a) Mode-4, closed-form solution.  
(b) Mode-4, FE model solution.  

Figure 5. Comparison of closed-form and FE mode shape deflections for modes 1-to-4.

Similar to code verification, solution verification is performed by comparing model solutions to a reference solution. The reference solution can be obtained by various methods, one of which is to perform multiple runs of the same problem with successively refined meshes, then, using an extrapolation technique to estimate the solution that would be obtained if the calculation could...
be carried out with “infinite” resolution, that is, $\Delta x \to 0$. The main two techniques that support solution verification activities are the development of solution error Ansatz models and the Grid Convergence Index (GCI). Solution error Ansatz models describe the properties of asymptotic convergence of discrete solutions, while the GCI can be used to estimate bounds of numerical uncertainty (Roache, 1994). Strictly speaking, these techniques apply only to runs performed within the regime of asymptotic convergence.

Table 3. Results of the mesh refinement of an arch discretized with SHELL93 elements.

<table>
<thead>
<tr>
<th></th>
<th>Coarse Mesh (16 Elements)</th>
<th>Medium Mesh (64 Elements)</th>
<th>Fine Mesh (256 Elements)</th>
<th>Extra-fine Mesh (1,024 Elements)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode-1 Frequency</td>
<td>0.09328 Hz</td>
<td>0.09317 Hz</td>
<td>0.09317 Hz</td>
<td>0.09317 Hz</td>
</tr>
<tr>
<td>Mode-2 Frequency</td>
<td>0.28596 Hz</td>
<td>0.28458 Hz</td>
<td>0.28449 Hz</td>
<td>0.28448 Hz</td>
</tr>
<tr>
<td>Mode-3 Frequency</td>
<td>0.58312 Hz</td>
<td>0.57486 Hz</td>
<td>0.57429 Hz</td>
<td>0.57426 Hz</td>
</tr>
</tbody>
</table>

The solution error Ansatz model is an equation that describes the rate at which the discrete solution $y(\Delta x)$, obtained by running the calculation at mesh size $\Delta x$, converges to the reference solution $y^{\text{Reference}}$. The equation takes a functional form such as $|y^{\text{Reference}} - y(\Delta x)| = \beta \cdot (\Delta x^p)$ where the symbol $\beta$ is a pre-factor coefficient and “p” denotes the rate-of-convergence. The triplet $(y^{\text{Reference}}, \beta; p)$ represents the unknowns of the Ansatz equation and a mesh refinement study with a minimum of three runs provides enough equations to estimate these unknowns.

It can be verified that the rate-of-convergence of the numerical method can be estimated as:

$$p = \log \left( \frac{y(\Delta x_M) - y(\Delta x_C)}{y(\Delta x_F) - y(\Delta x_M)} \right) / \log(R),$$

where $\Delta x_C$, $\Delta x_M$ and $\Delta x_F$ refer to the coarse-mesh, medium-mesh and fine-mesh element sizes, respectively. Symbol $R$ denotes a refinement ratio that is assumed, without loss of generality, to be constant, that is, $R = \Delta x_C / \Delta x_M = \Delta x_M / \Delta x_F$. When the mesh size is not uniform due to the irregularity of a mesh, it can be approximated as $\Delta x = (V/N_E)^{1/3}$ where “V” denotes the area or
volume of the computational domain, “Ne” is the total number of finite elements and “D” denotes the dimensionality of the problem (D = 2 for a 2D geometry and D = 3 for a 3D geometry).

A mesh refinement study is carried out for the arch geometry of section 5.1. Finite elements are halved in each dimension of the mesh, which means that the refinement ratio is R = 2 and the number of elements between any two successive runs is multiplied by a factor of four. Table 3 lists the resonant frequencies of modes 1-to-3 for the four meshes analyzed. When the rate-of-convergence is estimated using either the first group of three (coarse, medium, fine) runs or the second group of three (medium, fine, extra-fine) runs, the value of p = 2 is consistently obtained. It means that convergence is 2nd-order as \( \Delta x \to 0 \), which matches the formal order of accuracy of SHELL93 elements. (These elements define quadratic shape functions, hence, \( p_{\text{Theory}} = 2 \).)

The reference that these discrete solutions converge to, as \( \Delta x \to 0 \), can be estimated from:

\[
y_{\text{Reference}} = \frac{R^p y(\Delta x_F) - y(\Delta x_M)}{R^p - 1},
\]

as 0.09317 Hz for the first resonant frequency. It means that, should the coarsest mesh size be used to run the calculation, one expects a 0.12% over-estimation of the first natural frequency. Likewise, running the calculation at the coarsest mesh size over-estimates the second natural frequency by 0.5% only. These levels of numerical error (due to mesh size) are small compared to the variability of measurements that results from vibration tests. We conclude that, as long as a mesh size finer than \( \Delta x \leq 30 \text{ cm} \) is used to discretize the FE model of the cathedral, numerical error will not pose problem.

The analysis of solution convergence yields two important conclusions. The first one is that the finite element (SHELL93) that is so critical to our numerical simulation of the vibration response of the National Cathedral performs according to expectation. It increases our confidence that the code works “correctly.” The second conclusion is that the element size (\( \Delta x \approx 30 \text{ cm} \)) selected to perform the FE simulations is adequate in terms of providing an overall level of solution error that remains much lower than experimental error and uncertainty.

### 6. DYNAMIC EXPERIMENTS

The experimental data used for model calibration are obtained in the form of acceleration responses of the vaults due to impact hammer excitation. Model 393A03, uni-axial seismic accelerometers manufactured by PCB Piezotronics, Inc., with a frequency range of 0.5-to-2,000 Hz and a sensitivity of 1 Volt/g are used. The impact hammer technique is selected to provide the necessary force required to excite the vaults sufficiently above the ambient vibration level and obtain a high signal-to-noise ratio. Figure 6 illustrates the delivery of the excitation for vibration testing. A 5.5 kg PCB model 086D20 instrumented sledge-hammer capable of applying a peak force of 22 kN is found to be suitable for this testing. During the vibration tests, typical excitation amplitudes applied to the National Cathedral choir vaults are about 2.5 kN.

The accelerometer layout on the choir vault is shown in Figure 7. Points 1 (crown), point 3 (longitudinal rib), point 8 (diagonal rib) and point 12 (transverse rib) are selected to be excitation locations. Typical time-domain measurements of impulse excitation and vibration response are illustrated in Figure 8.

Signals are processed and recorded with a Dactron data acquisition system manufactured by LDS Test and Measurement, Ltd. The record length and sampling frequency are adjusted to 1,024 points and 187.5 Hz so that the response of the vaults attenuates within the time frame of 5.4 seconds. The use of a windowing function is, thus, avoided. Signals are averaged a total of five times to estimate the coherence functions.
Figure 6. The hammer operator in action.

Figure 7. The measurement grid.

Figure 8. Typical time history measurements collected during vibration testing.

8-a) Hammer impact, input signal.
8-b) Acceleration vibration response.

Figure 9. Reciprocity check between excitation/measurement points 1 and 12.
The general assumptions of modal analysis are that the system is linear, stationary, and time-invariant. The structure of interest can easily be considered to assume a stationary response. The time varying effects are primarily caused by environmental variations, such as temperature and moisture. These effects are minimal for the duration of vibration testing. The assumption of linearity, however, remains to be verified. The most commonly encountered technique to do so is through a reciprocity check. Figure 9 compares the response measured at point 12 due to an excitation at point 1 to the response measured at point 1 due to an excitation at point 12. It is concluded that the assumption of linearity is justified up to 40 Hz, as long as the testing involves force levels between 500-700 lb and response levels are less than 0.1 g.

The main issue arising from a hammer excitation is the inability to maintain a constant excitation during averages. Averaging requires exciting at constant force and constant angle with respect to the vibration surface. Transient methods, such as impact excitation, also have the problem of exciting the structure at varying response levels in a very short duration and inducing non-linear behavior. It adds uncertainty to the measurements, which needs to be identified and eliminated.

As stated earlier, the excitation levels are varied between 500-700 lb during this series of tests. Differences in Frequency Response Functions (FRFs) and coherence functions caused by this uncontrolled variation of excitation levels are examined. Figure 10 shows the frequency-domain responses and corresponding coherence functions as the excitation level is varied from 500 lb to 700 lb. Because it can be observed that little difference occurs between these two tests, it is concluded that the structure maintains linearity for the range of excitation forces considered.

7. MODEL CALIBRATION AND VALIDATION

One could argue that a model is nothing but a series of equations which map some input parameters to some output predictions. Certainly one must be highly confident in the model and confidence often comes from comparing simulation predictions to experimental measurements. In the context of model validation, measurements serve two distinct purposes. First, they give measured responses used to revise and improve our imprecise knowledge about some of the modeling assumptions or input parameters. The second purpose is to verify, through the fidelity between predictions and measurements, that the model captures the phenomena of interest. The former is widely known as “calibration,” while the latter is known as “validation.”

Even though it is clear that calibration and validation are inherently connected, we take the view that validation is a comprehensive task which encompasses calibration activities. This section summarizes the calibration and validation activities applied to the National Cathedral.
7.1 Defining the Validation Domain

The intended use of the model directs the entire validation exercise and, thus, must be defined carefully in the early stage of the study. The main purpose of FE models of masonry churches is to calculate the displacement and stress response due to static, quasi-static and dynamic loading. If possible, this should be achieved while taking the development of cracks and formation of hinges into consideration.

For this reason, the validation domain defined in study is limited to the amplitude of excitation. This is because, as discussed in Section 6, the excitation level is likely to vary significantly. For the numerical simulations to be useful, the FE model needs to be appropriate over the range of loading amplitudes that the structure responds to. Performing experiments with excitation levels of 500 lb and 700 lb provides a minimum of two settings that cover the validation domain.

7.2 Selection of Comparative Indicators

The comparative features are the condensed forms of the oversized raw data reduced down into manageable pieces of information. The selection of comparative features depends on the computer model, its intended use and available experimental data. It is what makes feature selection a foundational issue that greatly influences the rest of the model validation process.

As long as the structure maintains a linear behavior, modal parameters offer a convenient and physically meaningful way of comparing predictions to measurements. In this study, the linearity assumption proved to hold through reciprocity and linearity checks and, thus, modal parameters are selected as comparative features. This preference is also motivated, to a large extent, by the ease with which they can be extracted from measurements and computed from a linear, FE model. Nonetheless, nothing prevents calibration to be applied with other types of features.

The first nine resonant modes are extracted from signals collected on the choir vaults of the National Cathedral. Modal parameter estimates are obtained using the Eigen-value Realization Algorithm (ERA) from repeated tests for four different excitation points. ERA is a time-domain system identification method that can be automated to a great extent, after an initial phase of searching for the most appropriate settings using stabilization diagrams. Automated processing of time-domain signals is an advantage when it is desired, like here, to process a large number of responses in a consistent manner.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Identified, Resonant Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>σ (Hz)</td>
<td>0.296</td>
</tr>
<tr>
<td>σ/μ (%)</td>
<td>3.15%</td>
</tr>
</tbody>
</table>

(Legend: Identification performed with ERA and an excitation at the crown and response at Point 4.)

Statistics of natural frequencies obtained by applying the excitation at the crown and measuring the response at Point 4 are presented in Table 4. It is observed that the structural response features well-separated modes, relatively low levels of damping (not shown) and a variability of modal frequencies that ranges from 1-to-4%, approximately.

Going back to Section 5, this range of experimental variability of natural frequencies legitimates the choice made for discretization of the FE model. Remember that running the calculation with element sizes of $Δx = 30$ cm provides a level of solution error that ranges from 0.1-to-$\frac{1}{2}$%. This error remains much less than the combination of environmental and identification variability, as shown in Table 4, therefore, providing a defendable justification of our modeling choices.
7.3 Selection of Updating Parameters

Success of model updating depends not only in the appropriate selection of comparative features, in our case modal parameters, but also in the ability of adjust the “correct” parameters. The selection of an updating parameter should be based on a criterion that combines, on one hand, the uncertainty with which this parameter is known and, on the other hand, the sensitivity of the desired predictions to the parameter. Attempting to update a parameter that is well-known (low uncertainty) is a waste of resource. Attempting, on the other hand, to update a parameter that exhibit little influence on the response (low sensitivity) leads to numerical ill-conditioning.

It is common practice to tackle this ranking and selection of parameters using a combination of engineering judgment and input-output correlation, known as effect screening. Effect screening identifies input parameters that control the prediction variability. It is used to rank parameters based on their importance, or “statistical significance.” Several tools have been developed to assist in this selection and an extensive discussion can be found in Blackwell et al. (1999).

The Parameter Identification and Ranking Table (PIRT) is one such tool that has proven to be a useful, systematic approach when applied to the licensing of nuclear power plants. It is used, for example, to rank modeling assumptions and input parameters based on their effect on safety criteria. Inspired from the PIRT, a list of parameters, also known as “factors,” is developed that are believed to influence the prediction of modal parameters. The list is primarily composed of: (a) material properties of stone, brick, and concrete; (b) inter-element connectivity between the components; and (c) geometric properties of the vault. Table 5 lists the parameters investigated.

In the remainder, we summarize how the assessment of uncertainty and sensitivity (columns 3 and 4) are populated and how we arrive at a final decision (column 5) regarding steps taken to improve our knowledge of these parameters.

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Symbol</th>
<th>Uncertainty</th>
<th>Sensitivity</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young's modulus of limestone</td>
<td>E1</td>
<td>High</td>
<td>High</td>
<td>Calibrate</td>
</tr>
<tr>
<td>Young's modulus of brick</td>
<td>E2</td>
<td>High</td>
<td>Low</td>
<td>Eliminate</td>
</tr>
<tr>
<td>Young's modulus of concrete</td>
<td>E3</td>
<td>Medium</td>
<td>Low</td>
<td>Eliminate</td>
</tr>
<tr>
<td>Density of limestone</td>
<td>d1</td>
<td>High</td>
<td>High</td>
<td>Calibrate</td>
</tr>
<tr>
<td>Density of brick</td>
<td>d2</td>
<td>High</td>
<td>Low</td>
<td>Eliminate</td>
</tr>
<tr>
<td>Density of concrete</td>
<td>d3</td>
<td>Medium</td>
<td>Low</td>
<td>Eliminate</td>
</tr>
<tr>
<td>Thickness of webbing</td>
<td>t1</td>
<td>Low</td>
<td>Low</td>
<td>Eliminate</td>
</tr>
<tr>
<td>Thickness of ribs</td>
<td>t2</td>
<td>Low</td>
<td>Low</td>
<td>Eliminate</td>
</tr>
<tr>
<td>Stiffness constant of spring</td>
<td>K1</td>
<td>High</td>
<td>High</td>
<td>Calibrate</td>
</tr>
<tr>
<td>Stiffness constant of spring</td>
<td>K2</td>
<td>High</td>
<td>High</td>
<td>Calibrate</td>
</tr>
<tr>
<td>Stiffness constant of spring</td>
<td>K3</td>
<td>High</td>
<td>Low</td>
<td>Eliminate</td>
</tr>
<tr>
<td>Stiffness constant of spring</td>
<td>K4</td>
<td>High</td>
<td>Low</td>
<td>Eliminate</td>
</tr>
<tr>
<td>Stiffness constant of spring</td>
<td>K5</td>
<td>High</td>
<td>Low</td>
<td>Eliminate</td>
</tr>
</tbody>
</table>

The list of 13 parameters (E1, E2, E3, d1, d2, d3, t1, t2, K1, K2, K3, K4, K5) shown in Table 5 is the starting point of our investigation. The thickness parameters (t1, t2) are well characterized experimentally, which justifies removing them from the list. Triplets (E1, E2, E3) and (d1, d2, d3) denote homogenized Young's modulus (E) and density (d) parameters of the masonry assembly and support conditions at the peripheries of the model. The model also includes springs used to simplify the boundary conditions. Because these springs are “fictitious,” their stiffness constants (K1, K2, K3, K4, K5) are highly uncertain and need to be included in the analysis.
After screening out the two thickness coefficients, the list reduces to a total of 11 parameters. A two-level, full factorial design defined for effect screening, where each parameter is set to either its minimum or maximum bound, requires \(2^{11} = 2,048\) computer runs. Because the full-factorial design would require too much computing time to analyze, we assume that material properties and stiffness constants are uncoupled and investigate their effects separately. While very crude, this assumption reduces the number of FE model runs to \(2^6 + 2^5 = 96\) runs only. To defend this approach, it must be noted that this is a preliminary step in terms of identifying the parameter sensitivity. Once the parameters are reduced to the final list for uncertainty quantification and model calibration, we advocate analyzing a higher-strength design to investigate the potential coupling between these parameters.

Figure 11-a) Used for boundary conditions. 11-b) Used for material properties. Figure 11. Simplified, quarter-bay models used for statistical effect screening.

Figure 11-a illustrates the simplified, quarter-bay model developed to investigate the effects that stiffness coefficients (K1, K2, K3, K4, K5) have on the prediction of modal parameters. Likewise, the simplified model portrayed in Figure 11-b is developed to investigate the effects that material properties (E1, d1, E2, d2, E3, d3) exercise on modal parameters. It suffices, for the purpose of effect screening, to define a quarter-bay model because symmetry dictates that the influence of parameters on the modal response is the same as the one of a full-size model. Running quarter-bay models provides the additional advantage of a further reduction in computing time.

12-a) Screening of (K1, K2, K3, K4, K5). 12-b) Screening of (E1, d1, E2, d2, E3, d3). Figure 12. Results of main effect for natural frequencies and effective masses.
A two-level, full-factorial design is used to investigate the influence that parameters exercise on the first five modes. Response features analyzed for these modes are the resonant frequencies and effective masses in translational \((x, y, z)\) and rotational \((xy, xz, yz)\) directions. It gives a total of seven response features per mode. Analyzing the effective masses is a way to understand the influence of model parameters on mode shapes that, since they are vectors and not scalars, are too high-dimensional to analyze conveniently.

Results are presented in Figure 12-a for spring constants \((K_1, K_2, K_3, K_4, K_5)\) and Figure 12-b for material properties \((E_1, d_1, E_2, d_2, E_3, d_3)\). The figures show the values of \(R^2\) statistics for each one of the seven features (frequencies, effective masses) analyzed. A large \(R^2\), relative to the others, indicates that the corresponding parameter is significant in terms of controlling the variability of response features. The analysis reveals that frequencies and effective masses of the quarter-bay models are not significantly affected by spring stiffness constants \((K_3, K_4, K_5)\). Parameters \(K_1\) and \(K_2\) are much more influential, as can be seen in Figure 12-a. It is also found that the material properties of brick \((E_2, d_2)\) and concrete \((E_3, d_3)\) are not as influential as those of limestone \((E_1, d_1)\), see Figure 12-b.

These sensitivities are reported in column 4 of Table 5. The last step of a PIRT-like analysis is to combine the uncertainty and sensitivity columns to decide what to do about each parameter. For example, a parameter that is known well (little uncertainty) can be eliminated. Likewise, one that exhibits a low-to-medium level of uncertainty but little sensitivity can be eliminated because, although it is uncertain, its variability will not significantly influence model predictions. Table 5 shows that 9 of the initial 13 parameters are eliminated from analysis. It means that they can be kept constant and equal to their nominal values when running the FE model. The parameters that remain in Table 5 are \((E_1, d_1, K_1, K_2)\) because they combine high levels of uncertainty and sensitivity. To refine our knowledge of these four parameters, it is decided to feed them to the uncertainty quantification and calibration procedure of Section 7.5.

### 7.4 Test Analysis Correlation

Test Analysis Correlation (TAC), by definition, involves the comparison of experimental and numerical responses. TAC metrics are, to a large extent, dictated by the features selected and the purpose of the comparison. A TAC metric can be as simple as taking the difference between two scalar-valued features; it can also be as complex as statistical correlation (Hemez, 2007). Because our definition of a validated model is one that reproduces the variability of responses that would be measured experimentally, we are drawn towards defining a statistical metric that is capable of accounting for both experimental and simulation uncertainty. (See Section 7.5.)

![Figure 13. Test analysis correlation for the first experimental mode shape.](image)
With the inception of TAC, one can start to make inferences about model imprecision, where it comes from and how to revise the model to correct it accordingly. Qualitative comparison of the first experimental and predicted mode shapes (from the quarter-bay FE model of Figure 11-a) reveals that the choir walls provide some degree of vertical restraint at the point arch windows. Figure 13 illustrates that this restraint is overlooked in the initial model, since the displacements predicted by the model are about twice the value of those measured, hence, indicating that the FE model is locally too flexible.

It is decided to revise the model by adding linear springs along the connection of the choir walls and pointed arches. Figure 14 depicts the revised FE model used for uncertainty quantification and calibration. Because new boundary springs are introduced, their stiffness constants need to be added to the four parameters selected previously. The list of parameters used for calibration becomes \((E_1, d_1, K_1, K_2, K_6)\) where \(K_6\) is the stiffness constant of the new boundary springs.

**Figure 14. The revised FE model used for uncertainty quantification and calibration.**

Modal parameters define convenient and physically meaningful features for TAC and calibration provided that modes are, first, paired in the correct sequence. Modal pairing can be difficult due to the incompleteness of measurements that introduces spatial aliasing of the vectors. Spatial aliasing refers to the fact that higher-order mode shapes may appear as, and correlate well with, lower-order mode shapes due to spatial incompleteness (lack of instrumentation). This is visible in Figure 15 where the first four identified mode shapes are matched to FE-predicted vectors. In Figure 15-c), for example, Modal Assurance Criterion (MAC) values close to 50% are obtained for the correlation between the third identified vector and five different FE mode shapes.

The pairing of measured and FE-predicted mode shapes is handled by implementing a two-step procedure. Because the excitation and response of the structure are occurring in the vertical direction, the first step is to restrict the selection of FE-predicted mode shapes to those that are primarily composed of vertical motion. This is achieved by focusing on a percentage contribution of vertical movement to the entire modal deflection. Vectors for which the vertical movement constitutes 15% or less of this ratio are eliminated from further consideration. The second step is to assess modal pairing with the MAC, as is usually done in structural dynamics.
illustrates the correlation of identified and FE-predicted mode shapes reached with this two-step procedure. Top figures depict the spatial correlation between mode shape vectors using red, star symbols to denote the identified vectors and solid, black lines to represent the FE vectors. Bottom figures show MAC values between the first four experimental mode shapes and first 25 FE mode shapes. The final results of modal pairing are reported in Table 6.

15-a) MAC values of identified mode 1. 15-b) MAC values of identified mode 2. 15-c) MAC values of identified mode 3. 15-d) MAC values of identified mode 4.

Figure 15. Comparison of measured and predicted mode shapes and MAC values.

It can be observed in Table 6 that the first FE mode is not matched to any experimental mode. This is due to the combination of, first, not being able to excite and measure all structural modes during vibration testing and, second, restricting the FE vectors to those that exhibit high levels of vertical motion. In the uncertainty quantification and calibration procedure of Section 7.5, correlation between measured and FE-predicted modes, which defines the goodness-of-fit function of the model for Bayesian inference, is restricted to the modal pairing of Table 6. This, however, raises another issue of practical importance.

<table>
<thead>
<tr>
<th>Predicted Modes</th>
<th>Identified Modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Mode 1</td>
<td>Test Mode 2</td>
</tr>
<tr>
<td>FE Mode 1</td>
<td></td>
</tr>
<tr>
<td>FE Mode 2</td>
<td>X</td>
</tr>
<tr>
<td>FE Mode 3</td>
<td></td>
</tr>
<tr>
<td>FE Mode 4</td>
<td>X</td>
</tr>
<tr>
<td>FE Mode 5</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Pairing between the identified and FE-predicted modes.
When parameters of the model are perturbed during calibration, the order in which FE modes are predicted may swap. Mode shapes resulting from a parameter perturbation may also become linear combinations of those listed in Table 6 for the nominal model. Clearly one cannot expect modal pairing to remain unchanged as the optimization algorithm explores the design space of parameter values. The danger is that it could yield wrong goodness-of-fit functions and, in turn, incorrect uncertainty quantification and calibration results. To remedy this challenge, the FE mode shapes are transformed onto their proper orthogonal subspace with a Singular Value Decomposition (SVD). Doing so bypasses the issue altogether because, even if mode shapes change order or become linear combinations, the subspace that they define remains unchanged.

7.5 Characterization of Modeling Parameters

There are two types of calibration strategies that differ in the formulation through which they improve the models to make them match measurements better. A first type is the parameter calibration approach that captures the inaccuracy of model parameters. Finite element updating techniques, for the most part, are parameter calibration approaches that optimize model parameters to minimize a TAC cost function. The second type is the bias correction approach that attempts to represent the inadequacy of physics models. These methods compensate for the systematic bias between measurements and model predictions, due to the fact that the model is missing a significant component in terms of representing the physics involved. These two fundamental concepts are combined in the context of Bayesian inference in the landmark study of Kennedy and O’Hagan (2000, 2001). Their approach can simultaneously calibrate model parameters and correct bias. The method provides a statistically meaningful comparison of measurements and predictions; it also incorporates the uncertainty associated with each of them. We adopt a later implementation of the method by Higdon et al. (2008), which is deeply rooted in the following relation:

\[
y_{\text{obs}}(x) = y_{\text{sim}}(x, \theta) + \delta(x) + \varepsilon(x). \tag{3}
\]

In the above relation, \(y_{\text{obs}}(x)\) and \(y_{\text{sim}}(x, \theta)\) are the measurements and predictions, \(\delta(x)\) corresponds to the discrepancy term that represents the systematic bias and \(\varepsilon(x)\) represents the random experimental error. In Kennedy and O’Hagan’s formulation the experimental error is defined as a zero mean, Gaussian random variable. In equation (3), \(x\) represents the controlled variables, while \(\theta\) denotes the best-possible (although unknown) solution for the calibration parameters.

Equation (3) is the basis to search for the values of calibration parameters \(\theta\) such that the predictions match measurements as best as possible. This search rapidly becomes computationally expensive, especially if the FE calculation takes more than a few minutes to run. This difficulty is addressed by substituting fast-running emulators to the FE model \(y_{\text{sim}}(x, \theta)\) and discrepancy term \(\delta(x)\). Emulators, also known as surrogate models, represent the input-output relationship in a purely mathematical manner. Using them means that all information about the underlying physical behavior, material properties, load connectivity, etc., is lost to solely focus on mapping the inputs \((x, \theta)\) to outputs \(y_{\text{sim}}(x, \theta)\). The main advantage is that emulators can be analyzed very rapidly compared to a run the FE model. Gaussian Process Model (GPM) emulators are used in this work. GPM are popular in statistical sciences because they require few assumptions and are capable of approximating many types of non-linear functions. The GPM/SA software (Gaussian Process Modeling and Sensitivity Analysis) is used to calculate sensitivities, propagate uncertainty and calibrate parameters \(\theta\). GPM/SA is described, and applied to other studies, in Atamturktur et al. (2008) and Hemez et al. (2008).

To characterize the prediction uncertainty, GPM/SA explores the joint probability distribution of calibration variables \((K_1, K_2, K_6, E_1, d_1)\) with a Markov-chain random walk that is based on a simple but effective principle: model predictions that better match the measurements originate from combinations of calibration variables that tend to be visited more frequently by the random walk. After performing a sufficient number of Markov Chain Monte Carlo (MCMC) iterations, selected to 50,000 here, the statistics of calibration variables visited are computed to estimate the (unknown) joint probability distribution. This distribution quantifies the uncertainty of parameters \((K_1, K_2, K_6, E_1, d_1)\); it is used to estimate the discrepancy term \(\delta(x)\) and quantify the uncertainty with which the FE model predicts the modal response.
Table 7 defines the prior knowledge of calibration parameters. Uniform probability distributions are assigned within these expected ranges. Ranges for the elasticity modulus and density of stone units are estimated based on studies previously published in the literature. Ranges for the linear spring constants are estimated considering the thrust the springs are expected to carry under the self weight of the Cathedral and number of springs per unit area. After defining the priors, a Latin Hypercube maxi-min design of computer experiments is analyzed to provide to GPM/SA the datasets it needs. Given the time and computing resource available, it is decided to perform 50 runs of the model. For each run, the FE model is evaluated using a combination of parameters (K1, K2, K6, E1, d1) defined by the design and FE-predicted modes are paired to the experimental data. This collection of 50 predictions, together with the four identified modes and statistical model of measurement error, define the datasets fed to the GPM/SA software.

Table 7. Lower and upper bounds of calibration parameters of the FE model.

<table>
<thead>
<tr>
<th>FE Model Parameter</th>
<th>Minimum Bound</th>
<th>Maximum Bound</th>
<th>Type of Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stiffness constants of spring (K1)</td>
<td>1 MPa</td>
<td>2 MPa</td>
<td>Uniform</td>
</tr>
<tr>
<td>Stiffness constants of spring (K2)</td>
<td>0.1 MPa</td>
<td>0.2 MPa</td>
<td>Uniform</td>
</tr>
<tr>
<td>Stiffness constants of spring (K6)</td>
<td>0.2 MPa</td>
<td>0.4 MPa</td>
<td>Uniform</td>
</tr>
<tr>
<td>Young’s modulus of stone (E1)</td>
<td>8 GPa</td>
<td>16 GPa</td>
<td>Uniform</td>
</tr>
<tr>
<td>Density of stone (d1)</td>
<td>1,500 kg/m³</td>
<td>2,500 kg/m³</td>
<td>Uniform</td>
</tr>
</tbody>
</table>

Figure 16. Sensitivity of the four resonant frequencies to parameters (E1, d1, K1, K2, K6).
(x = dummy; θ₁: spring K1; θ₂ = spring K2; θ₃ = spring K3; θ₄ = modulus E1; θ₅ = density d1)

The estimation of posterior distribution parameters (E1, d1, K1, K2, K6) is preceded by a “burn-in” phase. A total of 100 iterations of the Markov-chain random walk are performed, with seven different step sizes, such that the final statistics of posterior probability do not depend on the initial condition defined in Table 7. Burn-in samples are disregarded and not used to estimate the final, posterior distribution. Burn-in also serves the purpose of estimating the optimum step size used for the “real” exploration. This optimum size relates to the
correlation length needed to best-fit the covariance matrix of the GPM emulator. Performing more burn-in trials helps to better optimize the step size but it is costly. Based on past experience and the fact that the FE simulation that we seek to approximate is “well-behaved,” seven trials are deemed sufficient. The procedure is implemented in MATLAB™.

Sensitivities of the first four frequencies to each one of the five calibration parameters (E1, d1, K1, K2, K6) are illustrated in Figure 16. This result is analogous to an analysis-of-variance, where a lower value of the correlation coefficient \( \rho \) means that the corresponding parameter is significant in terms of explaining how the frequency varies. (In other words, sensitivity to the parameter is high.) The symbol “x” used in Figure 16 denotes a dummy parameter that it not used in the analysis but needed to operate the software. The first two resonant frequencies are highly sensitive to all five calibration parameters. Only the Young’s modulus of stone (E1) significantly influences the third frequency and, likewise, only the spring constant K1 influences the forth frequency. These findings can be used to learn more precisely which parameter influences which resonant mode. They also prove useful to interpret the estimation of posterior distribution discussed next.

Figure 17 illustrates the posterior distribution function that results from 50,000 MCMC iterations. The entire function is estimated by the random walk. However, representing graphically a five-dimensional object is not possible. Figure 17 shows, instead, marginal probabilities on the main diagonal and bi-variate distributions in the off-diagonal boxes \((\theta_1; \theta_2), \ldots, (\theta_5; \theta_5)\). The first observation is that the initial marginal distributions of Table 7 remain, for the most part, unchanged. One exception is the spring constant K3 for which the final marginal is clearly non-uniform. It points to preferred values for parameter K3, in terms of better matching the test data. Knowing this information helps us reduce the overall uncertainty with which the resonant modes of the structure are predicted by the FE model.

The second observation from Figure 17 is the correlation between some of the parameters. Parameters E1 and d1 for the material properties of stone are strongly correlated, as indicated by the diagonally-dominant probability contour for pairs \((\theta_4; \theta_5)\) or \((\theta_5; \theta_4)\). This is expected because it is well-known that a modulus of elasticity (E1) can be traded-off against a density (d1) to influence the value of a resonant frequency. Also
discovered is a trade-off between spring constant $K_2$ and Young’s modulus $E_1$. (And, of course, the inverse trade-off between $K_2$ and $d_1$ can be observed too.) It means that the resonant modes can be changed by either modifying the value of spring constant $K_2$ or adjusting the modulus of elasticity of stone. Discovering these trade-offs is important, again, to reduce prediction uncertainty and also learn the mechanism(s) that can be activated to revise the FE model for improved fidelity-to-data.

Table 8. Statistics of calibration parameters after performing Bayesian inference.

<table>
<thead>
<tr>
<th>FE Model Parameter</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stiffness constants of spring $(K_1)$</td>
<td>1.48 MPa</td>
<td>0.29 MPa</td>
</tr>
<tr>
<td>Stiffness constants of spring $(K_2)$</td>
<td>0.151 MPa</td>
<td>0.029 MPa</td>
</tr>
<tr>
<td>Stiffness constants of spring $(K_6)$</td>
<td>0.305 MPa</td>
<td>0.064 MPa</td>
</tr>
<tr>
<td>Young’s modulus of stone $(E_1)$</td>
<td>12.24 GPa</td>
<td>2.34 GPa</td>
</tr>
<tr>
<td>Density of stone $(d_1)$</td>
<td>2,034 kg/m$^3$</td>
<td>284 kg/m$^3$</td>
</tr>
</tbody>
</table>

Table 8 lists the mean and standard deviation statistics inferred from the posterior distributions of Figure 17. These values can be used to further refine the parameter calibration, especially if additional experimental information becomes available. It can be observed by comparing the prior ranges of Table 7 to the +/- one-sigma posteriors of Table 8 that the overall uncertainty is reduced significantly. For example, the prior range of 8 GPa for the Young’s modulus of stone $(E_1)$ is reduced to almost half since +/- $1-\sigma \approx 4.68$ GPa. Because parameter $E_1$ is amongst the most influential ones (see Table 5 and Figure 12-b), “tightening” these input uncertainty bounds significantly reduces the prediction uncertainty of modal response. It can also be seen that the first four natural frequencies are not sufficient to constraint the probability distribution of parameters such as the spring constant $K_1$. Including higher order modes, or other response features, may enable a better definition of these posterior probability distributions.

8. CONCLUSION

Testing an existing building, per se, yields useful information about its characteristics. However, only by integrating these measurements with computerized models is it possible to gain a thorough understanding of structural behavior. This work develops and applies a procedure that integrates information coming from, on one hand, finite element analysis and, on the other hand, experimental modal analysis. Emphasis is not placed on model calibration, which the optimization of model parameters such that predictions better match the measurements. Rather, it is rather on quantifying prediction uncertainty and attempting to genuinely improve the predictive capability. These verification and validation activities are applied to the development of finite element models to predict the vibration response of the Washington National Cathedral, D.C.

The many sources of uncertainty in physical experimentation and computer modeling limit the applicability of deterministic methods, leaving statistical inference as the only plausible option for model validation. We apply a Bayesian inference framework that accounts for uncertainty in testing and modeling to the problem of correlating the measured vibration response of the National Cathedral to finite element predictions. The results obtained illustrate how the sources of modeling uncertainty can be better characterized.

Although the benefits of uncertainty quantification and calibration may not be as obvious in civil engineering as they are in disciplines where prototyping and mass production are common, better techniques to assess modeling strategies can ultimately serve the civil engineering community with an improved understanding of computer modeling. It is especially true when applied to the monitoring of historic monuments that, because of their very nature, cannot be subjected to full-scale inspection or experimental testing. Successful repair and retrofit schemes ultimately depend on the development of a verified and validated simulation capability to better understand the behavior of historic monuments.
ACKNOWLEDGMENTS

The author wishes to express her gratitude to Dave Finley, undergraduate student at Pennsylvania State University, for his keen assistance during the tests. Also recognized for their valuable contributions are Onur Soysal, graduate student at University of Buffalo, for his technical support in the development of scripts, and Linda Hanagan (Penn State University) for lending the equipment for in situ modal testing. The author is also grateful to Cetin Unal, Modeling and Simulation project leader of the Global Nuclear Energy Partnership program at Los Alamos National Laboratory (LANL), for his support. The kind assistance of Brian Williams, LANL, and access he provided to the GPM/SA software are acknowledged and greatly appreciated.

BIBLIOGRAPHICAL REFERENCES


