

Improving Knowledge of Structural System Behavior through Multiple Models

Ian F.C. Smith¹ and Sandro Saitta²

Abstract

A system identification and model updating methodology that accounts for factors influencing the reliability of identification is proposed. An important aspect of this methodology is the generation of a population of candidate models. This paper presents an analysis of error sources that are used to define model populations. A case study illustrates the need for such an approach even when a single conservative model has been appropriate for design. Data mining techniques such as principal component analysis and k-means clustering combined to interpret model predictions. These methods are useful for estimating the dependability of system identification.

Introduction

¹Prof. of Civil Engineering, F. ASCE, IMAC, Struct. Eng. Inst., Station 18, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. E-Mail: ian.smith@epfl.ch.

²Grad. Res. Assist. in Computer Science, IMAC, Struct. Eng. Inst., Station 18, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. E-Mail: sandro.saitta@epfl.ch.

A systematic approach to interpretation of measurement data employs methodologies developed in the field of system identification (Ljung 1999). System identification involves determining the state of a system and values of system parameters through comparisons of predicted and observed responses. When the forms of relationships between observable quantities and system parameters are known, regression techniques are useful for identifying system parameters. However, these techniques are rarely applicable to structural engineering because closed form relationships between system parameters and responses are often not available. Most structures are analyzed using numerical methods. Strategies that compute the values of finite element model parameters through matching predicted responses with measured values are called finite element model updating or model calibration methods. Fully instantiated models that have values for all parameters are obtained through these procedures.

A survey of model updating procedures is given in (Robert-Nicoud et al. 2005c). This paper summarizes the findings of this survey. Many previous studies propose methods for modifying stiffness coefficients that predict dynamic properties of structures, for example (Friswell and Mottershead 1995). Such work is typically aimed at supporting evaluation of earthquake damaged

structures (Chaudhary et al. 2000). Proposals for interpretation of static measurements are few and they have involved minimizing the difference between measured and analytical quantities from a given finite element model (Liu and Chian 1997; Reich and Park 2001). The number of unknown variables is fixed. Models that have varying numbers of degrees of freedom and consequently, different sets of variables are not accommodated in such approaches.

Dependability of system identification has rarely been studied in previous research. A large number of models might predict responses that reasonably match observations. Therefore, procedures that match measured and predicted responses might lead to identification of the wrong model. This is compounded by inevitable errors in modeling assumptions and inaccuracies during measurements. A system identification methodology that takes these factors into account has been developed (Robert-Nicoud 2003; Robert-Nicoud et al. 2005c). A key aspect of this methodology is the generation of a population of candidate solutions in the feasible domain whose objective function values lie below a threshold. An indication of the dependability of identification is obtained through an examination of the characteristics of the population.

Since engineers may be overwhelmed with managing results from multiple models for the

same structure, data mining techniques (Hand et al. 2001; Tan et al. 2006) are examined in this paper. Data mining is an active field of research that is situated at the junction between statistics and computer science. Several recent textbooks (Hand et al. 2001; Webb 2002; Tan et al. 2006) cover this research area. Data mining is usually applied to tasks such as recognition of images (Wilking and Roefer 2004), characters (Sempere and Lopez 2003) and speech (Zhou et al. 2005). In recent years, data mining has been used in engineering as well (Washio 2007; Alonso et al. 2004; Grossman et al. 2001). Examples of applications include oil production prediction (Nguyen and Chan 1999), traffic pattern recognition (Yan et al. 2005), composite joint behavior (Shirazi Kia et al. 2005) and joint damage assessment (Yun et al. 2001).

While data mining is mainly used as a predictive tool, in this paper, the focus is interpretation of spaces of possible models. In (Saitta et al. 2005), data mining techniques have been applied to models to extract knowledge. Although the information obtained has been useful, it is limited to confirming existing relationships between model parameters. No information related to the space of possible models is provided. Furthermore, no guidelines are provided to use the results and the limitations of such a methodology have not been clearly evaluated. For example, it may be useful

for the engineer to know how many clusters there are in the space of possible models.

For these assessments, clustering techniques such as k-means (Webb 2002) are useful. Even though clustering is often proposed for various applications by the data mining community, it is not straightforward; there are many open research issues. K-means clustering has been successfully applied in domains such as speech recognition (Picone 1990), decision support (Packhama et al. 2005) and gene expression data (Chan et al. 2006). While there are well accepted methods for evaluating predictive models such as cross-validation (Webb 2002), clustering of possible models has not been investigated and quantitative methods are not available for evaluating this task. The criterion for assessing the capability of algorithms is subjective and dependent on the final goal of the knowledge discovery task.

Principal component analysis (PCA) (Jolliffe 2002) generate a set of principal components so that a small number of principal components (linear combination of variables) explain most of the variability of the data. It has been successfully applied in domains such as dimension reduction in financial time series (Lendasse et al. 2001) and micro-arrays data (Lexin and Hongzhe 2004). Hybrid data mining methods are proposed in the literature, for example, (Pan et al. 2005; Xu

et al. 2005). Most work combines data mining methods for better predictions. For example, (Ding and He 2004) propose a combination of PCA and k-means to improve the prediction accuracy. However, the visualization improvement is not taken into account.

This paper presents concepts of multi-model system identification and examines the dependability of identification. Ideas behind the use of multiple models are then presented. A case study is used to illustrate the relevance of the approach. Finally, data mining and its usefulness for system identification is introduced. Engineers are thus better able to improve decision making for system identification and downstream tasks such as further measurement, preventative maintenance and structural replacement.

System Identification

This paper is founded on results from eight years of research in model-based system identification (Robert-Nicoud et al. 2005c; Robert-Nicoud et al. 2005a; Smith 2005; Robert-Nicoud et al. 2000; Raphael and Smith 1998). For the purposes of this paper, identification dependability is the likelihood that one of the candidate models, which are obtained through system identification, corresponds to reality. Dependability is poor when many models produce the same response at

measured locations. This observation motivated a proposal for measurement system configuration using entropy in order to maximize identification dependability (Robert-Nicoud et al. 2005b). This paper discusses parameters that influence dependability once the measurement system has been configured.

Models are identified through matching measurement data with model predictions. Model calibration involves minimization of the difference between predictions and measurement data through identification of good values of model parameters. This strategy is based on the assumption that the best match with observations is obtained only when correct values of model parameters are identified. This assumption is flawed; there are several factors that could influence the results. These are discussed below.

Errors influence the dependability of system identification. Various types of errors may compensate for each other such that bad model predictions match measured values. The following definitions are used in this description: measurement error (e_{meas}) is the difference between real and measured quantities in a single measurement. Modeling error (e_{mod}) is the difference between the prediction of a given model and that of the model that accurately represents the real behavior.

Modeling errors have three principal sources e_1 , e_2 and e_3 (Raphael and Smith 2003b). Source e_1 is the error due to the discrepancy between the behavior of the mathematical model and that of the real structure. Source e_2 is introduced during the numerical computation of the solution of the partial differential equations representing the mathematical model. Source e_3 is the error due to the assumptions that are made during the simulation of the numerical model. Typical assumptions are related to the choice of boundary conditions and model parameters such as material properties, for example E and I .

Source e_3 might further be separated into two parts, e_{3a} due to assumptions related to the structure of the model (typically assumptions related to boundary conditions), and e_{3b} due to wrong values of model parameters. Although, in reality it might be impossible to separate these different sources, it is important to distinguish between these errors since usually, the only error source that is recognized by traditional model calibration techniques is e_{3b} . Furthermore, this error classification helps understand the concept of error compensation that is described later in this paper.

The value of e_1 is low when mathematical models are accurate. High model complexity usually involves parameters that are difficult to quantify. For the purposes of this paper, the value of e_1 is

taken to be zero. The value e_2 has a finite non-zero value (for example, they are typically less than 4% for prismatic beams when they are modeled using linear elastic beam theory). The 4% value is an estimation found from an analysis of finite element simulations (Robert-Nicoud et al. 2005c).

The value of e_2 may be reduced through strategies such as modeling with a finer discretization of the structure at “hot spots”, using elements that are more appropriate for modeling the phenomenon and employing more sophisticated interpolation methods. Such strategies have been the subject of countless studies in computational mechanics. Also, this error may be reduced using advanced methods, such as boundary methods and mesh-less models.

Model calibration techniques aim to reduce the value of e_{3b} through estimating the values of model parameters and recent proposals include e_{3a} through inclusion of boundary conditions (Robert-Nicoud et al. 2004). These techniques are only reliable when the values of e_{meas} , e_1 and e_2 are small. While such conditions may exist in laboratories, they are rarely present in practical situations.

If x_a is the real value of a behavior quantity such as deflection, x_{meas} is the measured value and x_c is the value computed using a model, the following relationships are obtained:

$$x_a = x_{meas} + e_{meas} \quad (1)$$

$$x_a = x_c + e_1 + e_2 + e_3 \quad (2)$$

Model calibration procedures minimize the absolute value of the difference between x_{meas} and x_c . The difference between x_{meas} and x_c is known as the residue r . Rearranging the terms in 1 and 2, r is written as

$$r = |x_{meas} - x_c| = |e_1 + e_2 + e_3 - e_{meas}| \quad (3)$$

Thus, model calibration techniques minimize the quantity $(|e_1 + e_2 + e_3 - e_{meas}|)$ instead of minimizing e_3 , which is the error due to incorrect values of model parameters and boundary conditions.

It is possible to choose wrong values of model parameters such that the errors compensate each other to provide a small value of the residue in Equation 3. Furthermore, since e_3 is the sum of

e_{3a} and e_{3b} , a low value of e_3 can be obtained through compensating effects of wrong boundary conditions and model parameters even when e_1 , e_2 and e_{meas} have low values.

When a single measurement is used, evaluation of the degree of match between the measured and the predicted value is performed by computing the residue. When several measurements are used, an appropriate function is needed for combining the residues at all measurement points into a single number. The term “distance metric” refers to a measure of distance between two n-dimensional vectors. The root mean square error (RMSE) is one such metric. RMSE is calculated using the following equation:

$$RMSE = \sqrt{\frac{\sum_i (x_{i,c} - x_{i,m})^2}{n}} \quad (4)$$

where $x_{i,m}$ is the value measured at the i-th measurement point and $x_{i,c}$ is the corresponding value computed using the model. Other examples of distance metrics are given below:

$$Total\ relative\ deviation = \sqrt{\sum_i \left(\frac{x_{i,c} - x_{i,m}}{x_{i,m}}\right)^2} \quad (5)$$

$$\text{Gaussian distance metric} = \sum_i (1 - e^{-\left(\frac{x_{i,c} - x_{i,m}}{x_{i,m}}\right)^2}) \quad (6)$$

Models are evaluated by computing the deviation between their predicted responses and measurements using a distance metric such as the RMSE. The distance metric is used as the objective function (or cost) to be minimized during model calibration. A model is identified as a candidate if the value of the metric is less than a threshold. The threshold is fixed through a consideration of the accuracy of the measurement system, e_{meas} , as well as through an estimate of modeling errors (e_1 , e_2 and e_3). The model calibration algorithm minimizes the distance metric and identifies candidate models that have values less than a specified threshold.

If the model response matches measurements at all points, the distance computed using any metric is zero. When there is no exact match, the value of the distance computed using each metric is likely to be non zero. Consequently, the global minimum of the objective function might also be non zero. This is because each distance metric assigns a different level of importance to individual measurement points. Equation 4 involves equal importance to all measurement points. Therefore, the minimum corresponds to values of parameters that result in the same level of error at all points.

Equation 5 assigns more importance to small measurement values than Equation 4. For example, if the value of a measurement is close to zero, even a small deviation in the value predicted by a model will cause the objective function to assume a high value due to the division by a small number. Such models may be rejected by the model calibration routine even when Equation 6 is chosen as the objective function. However, they are still likely to be accepted if RMSE is used as the objective function, since the value of RMSE is not affected significantly by small values.

The distance metric defined by Equation 6 demonstrates a behavior that is midway between Equation 4 and Equation 5. Since relative deviations from measured values are used in Equation 6, more importance is attributed to small values compared with Equation 4. However, even large deviations do not result in a distance value greater than 1.0.

Since the global minimum depends on the choice of the distance metric, this metric determines the contents of the set of candidate models. The dependence of the set of candidate models on the distance metric affects the dependability of identification when the modeling errors e_1 , e_2 and e_{3a} are non-zero.

The measurement system often has the most influence on the dependability of system identi-

fication. In general, system identification using more measurements is more reliable than identification with a few measurements. Models containing wrong assumptions might even predict responses that match measurements exactly when only a few measurements are available.

When more measurements are added, it becomes increasingly difficult to obtain a low value of the cost function since model predictions may not match all measurement points. In the absence of modeling errors the location of the global minimum does not change when more measurements are added (Figure 1). In the presence of modeling errors, the locations of global minima shift as more measurements are added. This is because wrong values of model parameters might compensate for the effects of wrong assumptions and other errors.

Multiple Models

Compositional modeling is a framework for constructing adequate device models by composing model fragments selected from a model fragment library (Falkenhainer and Forbus 1991). Model fragments partially describe components and physical phenomena. A complete model is created by combining a set of fragments that are compatible. For modeling the behavior of structures, fragments represent support conditions, material properties, geometric properties, nodes, elements and

loading. Assumptions are explicitly represented in model fragments so that the model composition module generates only valid models that are compatible with the assumptions chosen by users.

Model composition makes it possible to search for models containing varying numbers of degrees of freedom. There is no need to formulate an optimization problem in which the number of variables is fixed a-priori. Models are automatically generated by combining model fragments and are analyzed by the finite element method in order to compare their predictions with measurements.

This research is part of the development of a more general system identification methodology that accounts for factors that influence the dependability of identification discussed in the preceding section. An important aspect of the methodology is the use of a stochastic global search and optimization algorithm for the selection of a population of candidate models whose predictions match measurements as closely as possible. Mathematical optimization techniques that make use of derivatives and sensitivity equations are not used because search is performed among sets of model classes that contain varying numbers of parameters and multiple local minima have been observed in the search space.

An indication of the dependability of system identification is obtained through an examination

of the characteristics of the population. If model characteristics show wide variation, it means either that many models match measurements and parameter values might have compensated for the effects of incorrect assumptions modeling or that the measurement system is inadequate. On the other hand, if solutions are located in narrow, well-defined regions of the search space, parameter estimates are likely to be accurate.

A schematic of the process is shown in Figure 2. This paper discusses the steps marked in bold. Users input measurement data and specify a set of modeling assumptions. The model selection process identifies a set of candidate models whose predictions are below a threshold that is determined through an evaluation of errors. A feature extraction module extracts characteristics of these models. Data mining techniques are used to cluster models. These techniques are described next.

A stochastic global search algorithm called PGSL (Raphael and Smith 2003a) is used to minimize the cost function that evaluates the difference between measurements and model predictions. PGSL performs global search through sampling the solution space using a probability density function (PDF). At the beginning of search, a uniform PDF is assumed for the entire search space

so that solutions are generated purely randomly. When good solutions are found, probabilities in those regions are increased so that more intense sampling is carried out in the neighborhood of good solutions. The central assumption of PGSL is that better sets of points are found in the neighborhood of good sets of points.

The search space is gradually narrowed down so that convergence is achieved. PGSL has been successfully applied to optimization problems involving highly non-linear objective functions containing a large number of local minima. It has been empirically observed that the number of evaluations of the objective function required by PGSL for finding the global minimum does not increase as rapidly as other search techniques such as genetic algorithms (Raphael and Smith 2003a).

In the context of system identification of structures, search variables are assumptions that are made to create complete models. These assumptions are related to the condition of the structure such as the presence of cracks and support settlements, as well as values of parameters such as Young's modulus and rigidity of joints. PGSL proposes values of variables according to the current PDF.

The model composition module creates complete models using values of search variables that represent assumptions. These models are analyzed by the finite element method in order to compare the predicted responses with measured values. The cost function evaluates the degree of match between the predicted and measured responses. The value of the cost function is used by PGSL to update the PDF. At the end of each search, a model that predicts a good match with measurements is obtained.

It was shown earlier that the location of the global minimum shifts in the presence of modeling and measurement errors. Therefore, an accurate computation of the global minimum is not necessary and in fact, could result in solutions that are far from reality. Hence the search task is reformulated such that any solution whose objective function value lies below a threshold is considered to be acceptable. A population of solutions is obtained by repeating search several times.

The threshold is computed using an approximate estimate of modeling and measurement errors.

This is done as follows: from Equation 3,

$$r \leq |x_c| + |x_{meas}| = |e_1 + e_2 + e_3| + |e_{meas}| \quad (7)$$

since errors could be positive or negative. Therefore,

$$r_{best} \leq e_{model}^{est} + e_4^{est} \quad (8)$$

where, r_{best} is the residue of a model containing the best values of parameters, e_{model}^{est} is an estimate of modeling errors and e_4^{est} is an estimate of measurement errors. Equation 7 is rewritten in terms of the cost function as

$$cost_{acceptable} \leq threshold_{model} + threshold_{measurement} \quad (9)$$

where $threshold_{model}$ is the cost obtained by substituting e_{model}^{est} for the residue in the cost function, and $threshold_{measurement}$ is the cost obtained by substituting e_4^{est} for the residue in the expression for the cost function. e_4^{est} is related to the accuracy of the sensor and is obtained from in situ measurement under known conditions. In general, e_{model}^{est} is not easy to determine precisely and

a typical value of 4% might be assumed for linear finite element methods. More accurate estimates are obtained through calibrating models using measurements taken under controlled conditions. More details can be found in (Robert-Nicoud 2003).

Using Equation 9, the threshold of the cost function for terminating the search is calculated as

$$threshold = threshold_{model} + threshold_{measurement} \quad (10)$$

A model involving the right set of assumptions and correct values of parameters will have a cost function value equal to this threshold when errors due to mathematical modeling and measurements are equal to estimated values and have the same sign. Therefore, all models whose cost function values are below this threshold are equally eligible to be candidates since the exact values of errors are unknown.

Case Study

In order to demonstrate the need for the approach described in this paper, a bridge that was designed by Maillart in 1933, see Figure 3, is taken for a case study. Still standing today, the Schwandbach Bridge is an early example of a deck stiffened open-spandrel arch and has been named by Billing-

ton (Billington 1979) to be "One of the two or three most beautiful concrete bridges ever built".

The elliptic horizontal ground-plan curve that is supported by a vertical curved thin-walled arch is also an example of daring structural engineering that has inspired engineers for over seventy years.

Considering the aesthetic and technical value of this bridge, it is understandable that preservation is a priority. As a result, the structure is inspected periodically and has been the subject of many verifications as codes have improved, for example, (Salvo 2006). Deflection measurements have not been carried out since the 1930s and while the bridge shows no visible evidence of deterioration, the question of taking measurements arises periodically. Traditionally in Switzerland, bridges are measured for changes in deflection at mid-span during load tests. A single model (usually the design model) is used with the deflection measurement and the loading to determine values for parameters that have some uncertainty, such as the elastic modulus multiplied by the moment of inertia, EI . However, this bridge is too complex for such rudimentary model-calibration strategies.

Figure 4 shows the boundary conditions that were used in the analysis at the design stage. While such assumptions are acceptable at the design stage for achieving safety and serviceability, they are not appropriate for interpreting measurements. Structures do not behave this way in reality.

For example, there is no physical hinge at the extremities of the vertical spandrel elements. These joints cannot be assumed to be fixed either since cracking may reduce joint stiffness. Furthermore, not all joints are expected to have the same stiffness due to factors such as relative slenderness and varying locations on the structure. Maillart would have argued correctly that the structural design would not change significantly if these elements had been analyzed with hinges or not.

However, interpretations of measurements on existing structures have completely different objectives than designs and as a result, conservative design models are usually not appropriate for structural identification. Table 1 shows the vertical midspan deflection of the bridge in cm (positive upwards) as calculated by a three dimensional finite element model of the complete bridge. Three cases are studied: the design load used in the original calculations, a positive temperature increase of $20C^{\circ}$ (self-weight also included) and an asymmetric load of $25.5kN/m^2$ between point 1 and 2 (see Figure 4). The design load has three parts $L1$, $L2$ and $L3$. $L1$ is the self-weight. $L2$ models traffic loading (uniformly distributed load of $10.2kN/m^2$) and the bitumen (uniformly distributed load of $2.7kN/m^2$). $L3$ is the console load ($4.2kN/m$ on one side of the bridge). Finally, the design load is $(1.35 \times L1) + L2 + L3$.

Each case was analyzed using two assumptions: hinged verticals and fixed verticals. Results of the six calculations are given in Table 1 and the percentage change is calculated in the last column. These numbers are indeed insignificant from a design point of view.

Table 2 shows results for the same cases when the models have fixed joints and when the assumptions are either roller or pinned supports and the ends of the beams. The difference in vertical midspan assumptions are more significant with the highest difference arising from a temperature change. Most importantly, these changes can be additive. For example comparing these two assumptions, the percentage difference in midspan deflection due to a negative temperature change (producing negative deflections) and an asymmetric load is 26%. Many other combinations of many more assumptions are possible and this could result in equally large differences in values for midspan deflection. Most importantly, many combinations of wrong modeling assumptions would be able to correspond very closely to measurements of midspan deflection.

While some assumptions can be verified on-site, many are difficult to evaluate. For example in addition to joint stiffnesses and friction coefficients of old bearings, different concrete properties due to differences in mixes, varying evolution of elastic modulae in different locations and differ-

ential settlement of supports could influence the real behavior of an existing structure. In the case of the Schwandbach Bridge, the number of permutations and combinations of modeling assumptions results in several tens of thousands of possible models. Although this case has important technical and historical attributes, these conclusions are equally valid for most ordinary structures of moderate complexity. Rather than "stab" at one model and hope for the best, this paper proposes explicit treatment of multiple models using the methodology described earlier.

A test example focuses on 20 connections in the Schwandbach bridge as shown at the ends of the verticals in Figure 4. A finite element model of the bridge is created. Connections are modeled as rotational springs. Sensor measurements are simulated using specific stiffness values for all connections. This set of stiffness values is the correct model for the structure that needs to be found using system identification.

Data Mining

Since predictions of many models are likely to fall below the threshold values defined earlier, engineers risk becoming overwhelmed by the data. When data mining techniques (Hand et al. 2001; Webb 2002; Tan et al. 2006) are applied to model parameters, engineers may obtain use-

ful knowledge for system identification tasks. In general, the objective is to inform engineers of the accuracy of their diagnoses given the information that is available. For example, if it is not possible to obtain a unique identification, clusters of possible solutions are proposed. The number of clusters of models is useful information for engineers performing system identification. When the methodology defined above outputs M possible models, it does not mean that there are M completely different models of the structure. These M models might only differ slightly in a few values of parameters while representing the same model group. In other situations, models might have important differences representing distinct classes which are referred to as clusters.

To find these clusters, this paper proposes a combination of two data mining techniques - namely PCA and k-means. The new algorithm is described in the next subsection.

Combining PCA and K-means

PCA is a linear method for dimensionality reduction (Jolliffe 2002). The goal of PCA is to generate a new set of variables - called principal components (PC) - that are linear combinations of original variables. Ultimately, PCA finds a set of principal components that define a feature space. PCs are sorted so that the first components explain most of the variability of the data. A property

of principal components is that each component is mutually uncorrelated. Details of the PCA procedure can be found in (Jolliffe 2002).

K-means is a widely applied clustering algorithm. Although it is simple to understand and implement, it is effective only if applied and interpreted correctly. The k-means algorithm divides the data into K clusters according to a given distance measure. Although the Euclidean distance is usually chosen, other metrics may be more appropriate. The algorithm works as follows. First, K starting points (named centroids) are chosen randomly. The number of clusters, K , is chosen a-priori by the user. Then each point is assigned to its closest centroid. Collections of points form the initial clusters. The centroid of each cluster is then updated using the positions of the points assigned to it. This process is repeated until there is no change of centroid or until point assignments remain the same. A complete description of k-means can be found in (Tan et al. 2006).

In the present work, a point in the data mining algorithm represents a model from the system identification perspective. The methodology - combining PCA and k-means - is described next. Six hundred models are generated for the Schwandbach bridge. Input data for the PCA part of the

methodology are the stiffness values of 5 sets of connections. Sets of connections were obtained from grouping the total set of twenty connections using symmetry and the assumption of equivalent stiffness at both ends of the vertical sprandel elements. For example, stiffness values of connections 1, 10, 11 and 20, Figure 4, are represented by parameter p_1 .

The starting point for PCA is a matrix where each row is a different model and each column contains values of a parameter. First, the PCA procedure outlined above is applied to the model parameters. Using the principal components the complete set of model parameters is transformed. Then, the k-means algorithm is applied to the data in the feature space. The final objective is to see if it is possible to separate models into clusters. An example of outcome of the clustering process is given in Figure 5 where every point is obtained from a model. In the top figure these values are presented using two model parameters. In the bottom figure, the two first principal components are used.

Three clusters of models are now visible. Within each class, the structure is taken to have unique state. The centroid of each cluster defines a possible state of the structure. In the present case, three possible states are given. Values of the five parameters are given in Table 3. State 1 has

a low value for p_2 and p_4 representing a low stiffness value at these positions. Low stiffness values typically indicate cracking at these locations. In state 2, one connection (p_1) has a low stiffness value. Finally in state 3, a low stiffness value is identified for p_3 .

This knowledge is useful for decision making related to subsequent new measurements and further structural identification studies. For example, further stiffness measurements are justified at connections represented by parameters p_1 , p_2 and p_4 since they have different values for each state. Conversely, further stiffness measurements at connections represented by parameters p_3 and p_5 are not justified because this would not lead to effective model discrimination.

Overcoming K-means Limitations

K-means has limitations that are mentioned in (Tan et al. 2006). Two are considered here. Firstly, the number of clusters has to be specified by the user a-priori. Strategies for estimating the number of clusters have been proposed and are summarized in (Webb 2002). One of these methods is chosen here and adapted to the system identification context. Secondly, as stated above, the K initial centroids are chosen randomly. Therefore, running P times can result in P different clustering of the same data. A common technique for avoiding such a problem is described in this Section.

In order to evaluate results obtained with the methodology described above, a score function (SF) is defined. The score function combines two aspects: the compactness of clusters and the distance between clusters. The first notion is referred to as within class distance (WCD) whereas the second is the between class distance (BCD). Since objectives are to minimize the first aspect and to maximize the second aspect, combining the two is possible through maximizing $SF = BCD/WCD$. This idea is related to the Fisher criterion (Hand et al. 2001).

It is important that an engineering meaning in terms of model-based diagnosis can be given to these two distances. They are both directly related to the space of models for the task of system identification using multiple models. The WCD represents the spread of models within one cluster. Since it gives information on the size of the cluster, a high WCD means that models inside the class are widely spread and that the cluster may not reflect physical similarity. The BCD is an estimate of the mean distance between the centers of all clusters and therefore, it provides information related to the spread of clusters. For example, a high BCD value means that classes are far from each other and that the system identification is not currently reliable.

From a system identification point of view, BCD values indicate how different the K situations

are. Values of WCD give overviews of sizes of groups of models. As explained above, the number of clusters - in our application, this is the number of classes of models - for a data set is unknown.

The strategy to determine the most reliable number of clusters is to run the procedure for P different predefined number of clusters. The criterion used to see if the number of cluster is appropriate is the same as the score function described above. The higher the value of the criterion, the more suitable the number of clusters.

The second weakness of the procedure is the random choice of the K first centroids. A solution is to run the algorithm N times and to average the value of the score function. Therefore, randomness is controlled by N . The score function defined above thus serves two purposes. First, it gives an idea of the performance of the clustering procedure. Second, it allows choice of a realistic value for the number of clusters. Although values could have physical significance, this number must be interpreted with care. Reducing the random effect of the procedure is done through several runs of the algorithm to compute the score function value. Further work into new formulations of the score function is underway.

Conclusions

A system identification methodology that accounts for factors influencing the dependability of identification has been proposed. The importance and justification of a multi-model approach have been explained and then illustrated with a real case study. A study of the applicability of the methodology to a real structure has resulted in the following conclusions:

- Interpreting measurements is an inverse engineering task that may have many possible solutions.
- The presence of errors in modeling and in measurements further increases the number of possible solutions. Single value optimization methods, traditionally used in model updating strategies, are not generally useful.
- Generation of model instances that provide predictions below a composite error threshold is proposed to account for modeling and measurement errors in a systematic fashion
- Data mining techniques may be useful for grouping models into clusters thus providing information to the engineer about possible model classes. Clusters may indicate different possible types of behavior of a structure, thus guiding subsequent decision making related to new measurements and further system identification studies.

Future work involves the use of more advanced clustering methods while finding a more suitable distance measure. Furthermore, other data mining methods will be studied to discover hidden relationships among model parameters. Additional challenges include, automatic feature selection and accommodating data containing varying sets of parameters. Finally, a range of data mining methods will be combined into a framework for supporting system identification.

Acknowledgments

This work is part of the current results of several years of funding by the Swiss National Science Foundation and the Commission for Technology and Innovation. The authors would like to thank B. Raphael for his support, and E. Bruehwiler, P. Kripakaran, S. Ravindran and A. Salvo for their assistance with the Schwandbach Bridge case study.

References

- Alonso, C., Rodriguez, J., and Pulido, B. (2004). Enhancing consistency based diagnosis with machine learning techniques. *Lecture Notes in Computer Science*, 3040, 312–321.
- Billington, D. (1979). *Robert Maillart's Bridges*. Princeton University Press.

- Chan, Z., Collins, L., and Kasabov, N. (2006). An efficient greedy k-means algorithm for global gene trajectory clustering. *Expert Systems with Applications*, 30(1), 137–141.
- Chaudhary, M., Abe, M., Fujino, Y., and Yoshida, J. (2000). Performance evaluation of two base-isolated bridges using seismic data. *Journal of Structural Engineering*, 116(10), 1187–1195.
- Ding, C. and He, X. (2004). K-means clustering via principal component analysis. In *Proceedings of the 21st International Conference on Machine Learning*, ACM Press. ACM International Conference Proceeding Series, 29.
- Falkenhainer, B. and Forbus, K. (1991). Compositional modeling: Finding the right model for the job. *Artificial Intelligence*, 51, 95–143.
- Friswell, M. and Mottershead, J. (1995). *Finite Element Model Updating in Structural Dynamics*. Kluwer Academic Publishers.
- Grossman, R., Kamath, C., Kegelmeyer, P., Kumar, V., and Namburu, R., editors (2001). *Data Mining for Scientific and Engineering Applications*. Kluwer Academic Publishers.
- Hand, D., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining*. MIT Press.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer.

Lendasse, A., Lee, J., De Bodt, E., Wertz, V., and Verleysen, M. (2001). Dimension reduction of technical indicators for the prediction of financial time series - application to the bel20 market index. *European Journal of Economic and Social Systems*, 15(2), 31–48.

Lexin, L. and Hongzhe, L. (2004). Dimension reduction methods for microarrays with application to censored survival data. Technical report, Center for Bioinformatics & Molecular Biostatistics. <http://repositories.cdlib.org/cbmb/surv2>.

Liu, P. and Chian, C. (1997). Parametric identification of truss structures using static strains. *Journal of Structural Engineering*, 123(7), 927–933.

Ljung, L. (1999). *System Identification - Theory For the User*. Prentice Hall.

Nguyen, H. and Chan, C. (1999). Applications of data analysis techniques for oil production prediction. *Artificial Intelligence in Engineering*, 13, 257–272.

Packhama, I., Rafiqb, M., Borthwickb, M., and Denham, S. (2005). Interactive visualisation for

decision support and evaluation of robustness in theory and in practice. *Advanced Engineering Informatics*, 19, 263–280.

Pan, X., Ye, X., and Zhang, S. (2005). A hybrid method for robust car plate character recognition. *Engineering Applications of Artificial Intelligence*, 18(8), 963–972.

Picone, J. (1990). Duration in context clustering for speech recognition. *Speech Communications*, 9(2), 119–128.

Raphael, B. and Smith, I. (1998). Finding the right model for bridge diagnosis. *Artificial Intelligence in Structural Engineering*, 1454, 308–319.

Raphael, B. and Smith, I. (2003a). A direct stochastic algorithm for global search. *Journal of Applied Mathematics and Computation*, 146(2-3), 729–758.

Raphael, B. and Smith, I. (2003b). *Fundamentals of Computer-Aided Engineering*. Wiley.

Reich, G. W. and Park, K. C. (2001). A theory for strain-based structural system identification. *Journal of Applied Mechanics*, 68(4), 521–527.

Robert-Nicoud, Y. (2003). *Une methodologie mesures-modeles pour l'identification de systemes de genie civil*. PhD thesis, EPFL, Lausanne, Switzerland.

Robert-Nicoud, Y., Raphael, B., Burdet, O., and Smith, I. (2005a). Model identification of bridges using measurement data. *Computer-Aided Civil and Infrastructure Engineering*, 20(2), 118–131.

Robert-Nicoud, Y., Raphael, B., and Smith, I. (2000). Decision support through multiple models and probabilistic search. In *Proceedings of Construction Information Technology*. 765–779.

Robert-Nicoud, Y., Raphael, B., and Smith, I. (2004). Improving the reliability of system identification. *Next Generation Intelligent Systems in Engineering*, 4(199), 100–109.

Robert-Nicoud, Y., Raphael, B., and Smith, I. (2005b). Configuration of measurement systems using shannon's entropy function. *Computers and structures*, 83(8-9), 599–612.

Robert-Nicoud, Y., Raphael, B., and Smith, I. (2005c). System identification through model composition and stochastic search. *Journal of Computing in Civil Engineering*, 19(3), 239–247.

- Saitta, S., Raphael, B., and Smith, I. (2005). Data mining techniques for improving the reliability of system identification. *Advanced Engineering Informatics*, 19(4), 289–298.
- Salvo, A. (2006). Ponts de robert maillart. Technical report, EPFL-MCS, Lausanne, Switzerland.
- Sempere, J. and Lopez, D. (2003). Learning decision trees and tree automata for a syntactic pattern recognition task. *Pattern Recognition and Image Analysis*, 943–950.
- Shirazi Kia, S., Noroozi, S., Carse, B., and Vinney, J. (2005). Application of data mining techniques in predicting the behaviour of composite joints. In *The Eighth International Conference on the Application of Artificial Intelligence to Civil, Structural and Environmental Engineering*, Civil-Comp Press. CDROM.
- Smith, I. (2005). Sensors, models and videotape. In *Proceedings of the 2005 ASCE International Conference on Computing in Civil Engineering*, American Society of Civil Engineers. Reston VA. CDROM.
- Tan, P.-N., Steinbach, M., and Kumar, V. (2006). *Introduction to Data Mining*. Addison Wesley.

Washio, T. (2007). Applications eligible for data mining. *Advanced Engineering Informatics*, 21, 241–242.

Webb, A. (2002). *Statistical Pattern Recognition*. Wiley.

Wilking, D. and Roefer, T. (2004). Realtime object recognition using decision tree learning. In *RoboCup*, Springer. volume 3276 of *Lecture Notes in Artificial Intelligence*, 556–563.

Xu, L., Yan, Y., Cornwell, S., and Riley, G. (2005). Online fuel tracking by combining principal component analysis and neural network techniques. *IEEE Transactions on Instrumentation and Measurement*, 54(4), 1640–1645.

Yan, L., Fraser, M., Oliver, K., Elgamal, A., Conte, J., and Fountain, T. (2005). Traffic pattern recognition using an active learning neural network and principal components analysis. In *The Eighth International Conference on the Application of Artificial Intelligence to Civil, Structural and Environmental Engineering*, Civil-Comp Press. CDROM.

Yun, C.-B., Yi, J.-H., and Bahng, E. (2001). Joint damage assessment of framed structures using a neural networks technique. *Engineering Structures*, 23, 425–435.

Zhou, L., Shi, Y., Feng, J., and Sears, A. (2005). Data mining for detecting errors in dictation speech recognition. *IEEE Transactions on Speech and Audio Processing*, 13(5), 681–688.

Case	Assumptions		% change
	Hinged verticals	Fixed verticals	
Design Load	-5.8	-5.4	7%
Temperature ($\Delta T=20C^{\circ}$)	11.4	11.1	3%
Asymmetric load	-0.9	-1.0	10%

Table 1: Midspan deflection in *cm* (positive upwards) for three cases and two joint rigidity assumptions for the spandrel elements.

Case	Assumptions		% change
	Roller supports	Hinged supports	
Design Load	-5.4	-5.8	7%
Temperature ($\Delta T=20C^{\circ}$)	11.1	13.2	16%
Asymmetric load	-1.0	-0.8	10%

Table 2: Midspan deflection in *cm* for three cases and two support assumptions for the longitudinal beam elements that support the deck.

Model parameters	p_1	p_2	p_3	p_4	p_5
Connection numbers	1, 10, 11, 20	2, 9, 12, 19	3, 8, 13, 18	4, 7, 14, 17	5, 6, 15, 16
State 1	7.9×10^6	1.0×10^2	3.2×10^5	1.3×10^2	1.0×10^8
State 2	5.0×10^0	4.0×10^7	1.3×10^6	6.3×10^3	1.0×10^8
State 3	2.5×10^4	4.0×10^5	2.0×10^2	4.0×10^5	1.0×10^8

Table 3: Three cluster centers representing the three possible states of the structure. Each parameter is in Nm/rad . Connection numbers are given in Figure 4.

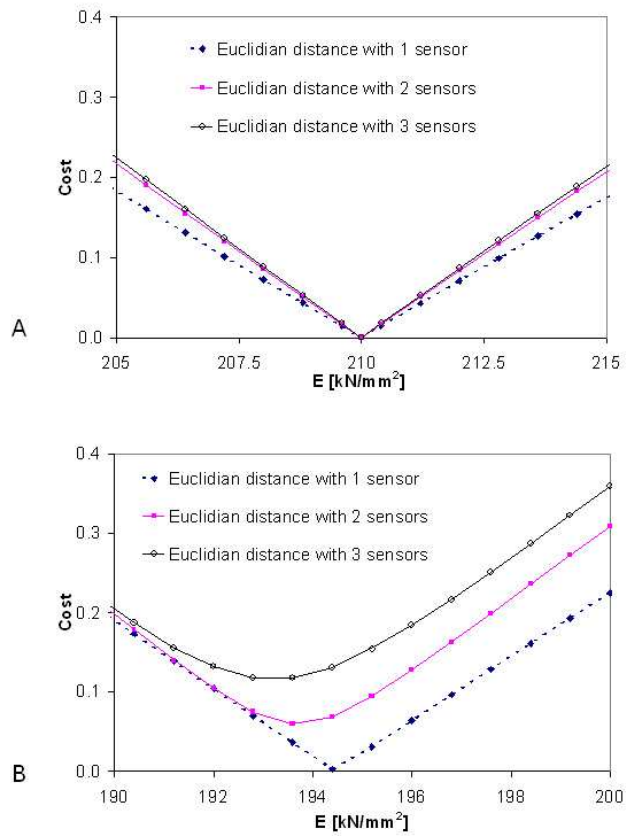


Figure 1: An example from (Robert-Nicoud et al. 2005c) of the effect of errors on system identification for the case of a steel beam in bending where the value of Young's modulus, E , needs to be found. In the absence of errors (Figure A), the lowest value of the objective function is independent of the number of sensors and the minimum is the correct value of E (210 kN/mm^2). In the presence of errors (Figure B), error compensation means that the lowest value is not the correct value and that this deviation increases with more sensors.

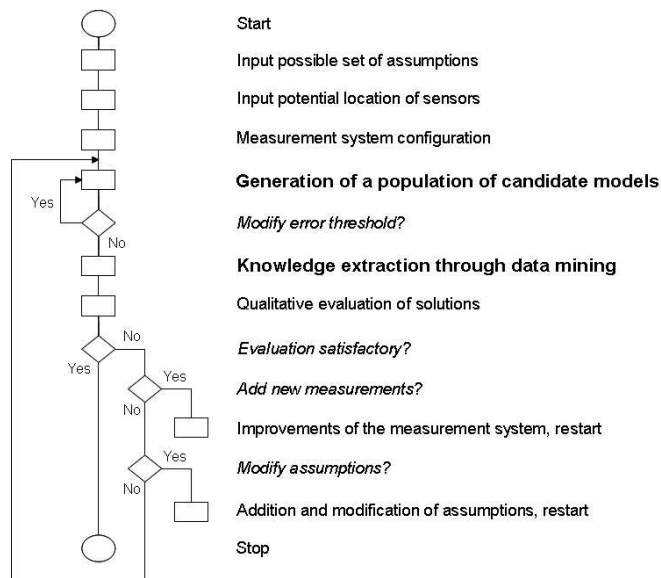


Figure 2: Methodology flowchart for multi-model diagnosis. This paper discusses steps marked in bold. Legends: ○ start/stop; □ processes; ◇ choices.

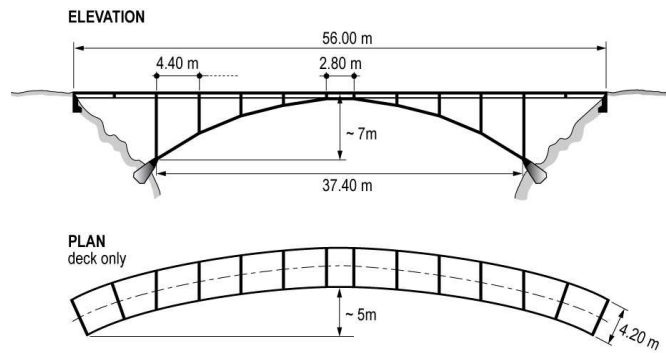


Figure 3: Schema of the Schwandbach Bridge used in the case study.

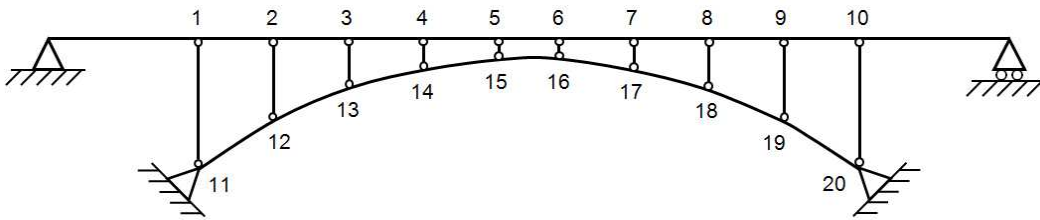


Figure 4: Schema of the bridge showing the set of twenty connections that have unknown stiffness values.

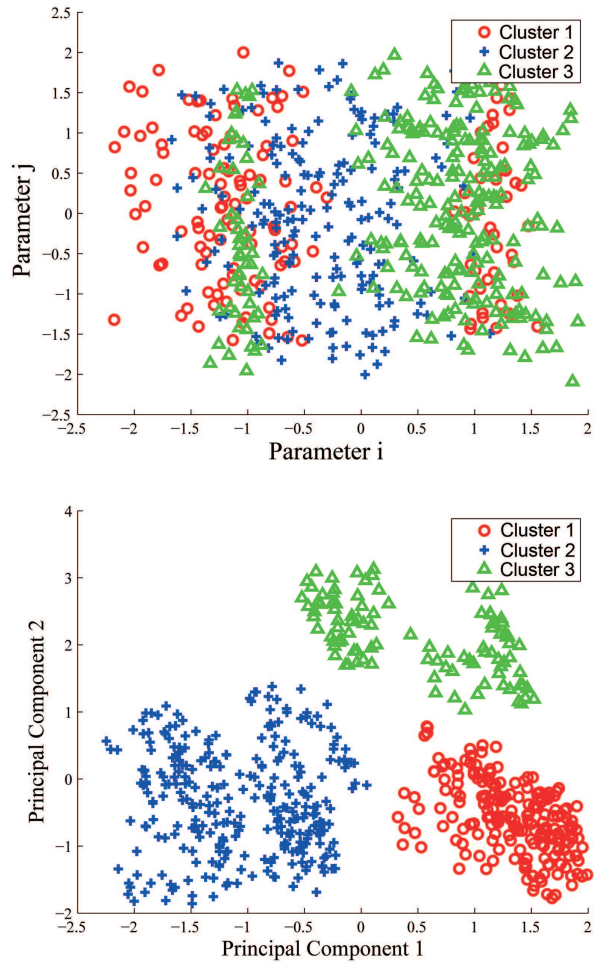


Figure 5: Comparison between standard k-means (top) and the proposed algorithm (bottom) output for the Schwandbach bridge. Six hundred models have been generated. Every point corresponds to a model. In the first case, two parameters are chosen to display the clusters. In the second, the two first principal components are used.

*

List of Tables

1	Midspan deflection in <i>cm</i> (positive upwards) for three cases and two joint rigidity assumptions for the spandrel elements.	40
2	Midspan deflection in <i>cm</i> for three cases and two support assumptions for the longitudinal beam elements that support the deck.	41
3	Three cluster centers representing the three possible states of the structure. Each parameter is in <i>Nm/rad</i> . Connection numbers are given in Figure 4.	42

*

List of Figures

- 1 An example from (Robert-Nicoud et al. 2005c) of the effect of errors on system identification for the case of a steel beam in bending where the value of Young's modulus, E , needs to be found. In the absence of errors (Figure A), the lowest value of the objective function is independent of the number of sensors and the minimum is the correct value of E ($210kN/mm^2$). In the presence of errors (Figure B), error compensation means that the lowest value is not the correct value and that this deviation increases with more sensors. 43
- 2 Methodology flowchart for multi-model diagnosis. This paper discusses steps marked in bold. Legends: ○ start/stop; □ processes; ◇ choices. 44
- 3 Schema of the Schwandbach Bridge used in the case study. 45
- 4 Schema of the bridge showing the set of twenty connections that have unknown stiffness values. 46

5 Comparison between standard k-means (top) and the proposed algorithm (bottom)

output for the Schwandbach bridge. Six hundred models have been generated.

Every point corresponds to a model. In the first case, two parameters are chosen to

display the clusters. In the second, the two first principal components are used. . . . 47