# Certification in Structural Health Monitoring Systems

C. M. S. KABBAN[1] and M. M. DERRISO[2]

## ABSTRACT

Despite advances in health monitoring systems, the certification process for automated SHM systems remain rooted in the probability of detection (POD). POD is a fundamental part of the certification process, however, it is not the only measure of system validity. The process to certify a system must contain measures of accuracy and reliability. Disclosing the accuracy and reliability at the time of system design allows the end user to weigh differences among various SHM system architectures. In this manner, end users may identify the architecture that meets their specific overall system requirements. Further, it creates a benchmark from which system redesigns may be assessed and compared in order to improve overall system performance.

In order to determine the appropriate measures of accuracy and reliability for the SHM system, first, the necessary output of such a system must be defined. Typically, the system is built in order to identify and assess structural damage. The appearance of damage is a *detection* problem whereas the extent of damage (crack length and location) is an *estimation* problem. Thus, the proper certification of a SHM system must incorporate measures of accuracy and reliability for both the detection and estimation of structural damage. These include probabilities related to detection such as the probabilities of true and false alarms as well as positive and negative predictive values. Confidence intervals must be defined as measures of the reliability of these probability estimates. Measures for the accuracy of the location and extent of structural damage include summary measures of the estimation process or model as well as the predicted location or crack length and its associated confidence interval.

Establishment of the certification process as related to SHM will lead to better quantification of system performance for both detection and estimation and ultimately better system designs. These criteria may then act as a springboard for certification in subsequent prediction of structural failure, time remaining until structural failure, and other general aspects of risk analysis related to structural health.

---

[1]Christine M Schubert Kabban, Ph.D., MBA, Assistant Professor, Department of Mathematics and Statistics, AFIT/ENC, Air Force Institute of Technology, 2950 Hobson Way, WPAFB, OH 45433-7765 (937) 255-3636 x4549, christine.schubertkabban@afit.edu

[2]Mark M. Derriso, Technical Advisor, Integrated Systems Branch, Air Force Research Laboratory, 2130 8th Street, WPAFB, OH 45433

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| | | |
|---|---|---|
| 1. REPORT DATE<br>**SEP 2011** | 2. REPORT TYPE<br>**N/A** | 3. DATES COVERED<br>**-** |
| 4. TITLE AND SUBTITLE<br>**Certification in Structural Health Monitoring Systems** | | 5a. CONTRACT NUMBER |
| | | 5b. GRANT NUMBER |
| | | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | | 5d. PROJECT NUMBER |
| | | 5e. TASK NUMBER |
| | | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**Department of Mathematics and Statistics, AFIT/ENC, Air Force Institute of Technology, 2950 Hobson Way, WPAFB, OH 45433-7765** | | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | 10. SPONSOR/MONITOR'S ACRONYM(S) |
| | | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT<br>**Approved for public release, distribution unlimited** | | |
| 13. SUPPLEMENTARY NOTES<br>**See also ADA580921. International Workshop on Structural Health Monitoring: From Condition-based Maintenance to Autonomous Structures. Held in Stanford, California on September 13-15, 2011 . U.S. Government or Federal Purpose Rights License.** | | |

14. ABSTRACT

**Despite advances in health monitoring systems, the certification process for automated SHM systems remain rooted in the probability of detection (POD). POD is a fundamental part of the certification process, however, it is not the only measure of system validity. The process to certify a system must contain measures of accuracy and reliability. Disclosing the accuracy and reliability at the time of system design allows the end user to weigh differences among various SHM system architectures. In this manner, end users may identify the architecture that meets their specific overall system requirements. Further, it creates a benchmark from which system redesigns may be assessed and compared in order to improve overall system performance. In order to determine the appropriate measures of accuracy and reliability for the SHM system, first, the necessary output of such a system must be defined. Typically, the system is built in order to identify and assess structural damage. The appearance of damage is a detection problem whereas the extent of damage (crack length and location) is an estimation problem. Thus, the proper certification of a SHM system must incorporate measures of accuracy and reliability for both the detection and estimation of structural damage. These include probabilities related to detection such as the probabilities of true and false alarms as well as positive and negative predictive values. Confidence intervals must be defined as measures of the reliability of these probability estimates. Measures for the accuracy of the location and extent of structural damage include summary measures of the estimation process or model as well as the predicted location or crack length and its associated confidence interval. Establishment of the certification process as related to SHM will lead to better quantification of system performance for both detection and estimation and ultimately better system designs. These criteria may then act as a springboard for certification in subsequent prediction of structural failure, time remaining until structural failure, and other general aspects of risk analysis related to structural health.**

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT **unclassified** | b. ABSTRACT **unclassified** | c. THIS PAGE **unclassified** | **SAR** | **8** | |

## INTRODUCTION

It is inherent that the process to certify a system must contain measures of both accuracy and reliability. Disclosing measures of accuracy and reliability at the time of system design will allow the end user to weight differences among various SHM system architectures and to build-in features at the design stage to assure a specific level of performance. In this manner, end users may identify the architecture that meets their specific overall system requirements and further, it creates a benchmark from which system redesigns may be assessed and compared in order to meet, or improve, overall system performance.

SHM systems for aircraft, as for other structures, are mounted as a permanent part of the structure. As such, these systems collect and process data in situ. The advantages of such systems are apparent through the real-time monitoring of structural integrity. To ignore or not make proper use of such real-time data streams is to negate the advantages and purpose of SHM technology.

The information in the data streams that such systems collect as to structural health, takes on several forms, but contains: 1) the presence of structural damage, 2) the extent of structural damage, and 3) the location of the structural damage. How the system captures information to estimate these measures of damage is not the focus of this paper, yet clearly the tools required to capture, process and predict presence and extent of damage must also embrace SHM technology and capitalize on continuous monitoring. To the extent that these tools must embrace SHM technology, so does the certification (validation) process of SHM systems. Thus, this paper seeks to establish a statistical framework with which statistical tools may be applied to assess how well the system captures this information, that is, the validity and reliability of the system output. An understanding of the nature of the measures comprising the framework for structural damage as well as how to statistically estimate them is of importance.

## DEVELOPMENT OF THE STATISTICAL FRAMEWORK

As previously stated, the statistical framework for the certification of a SHM system should contain measures that assess the system's accuracy and reliability with respect to the system's outcomes: 1) the presence, 2) the extent, and 3) the location of the structural damage. The appearance of structural damage, typically a crack, is a *detection* problem whereas the extent and location of the structural damage are *estimation* problems.

To create the framework, statistical tools assessing detection (for the presence of damage) and estimation (for the extent and location of damage) must be utilized to create a multi-dimensional assessment for the SHM system. Paramount to assessing the ability of the SHM system to detect and estimate, is the use and/or development of statistical methods that are adjusted to capitalize on the added enhancement of the SHM system's continuous data collection, that is, dependent measurements. To not consider such adjustments is to negate the use and added utility of the SHM system. First, methods for assessing the detection of damage will be examined, then methods for assessing the estimation of extent and location of damage will be examined.

**Detection of Damage**

There are several probabilities with which to assess the system as to its ability to detect damage. Indeed, the same probabilities that are of current use are the same probabilities that should continue to be used to assess the validity and reliability of damage detection. However, the method to compute these probabilities must take advantage of the added value of continuous monitoring, namely, the time-related, continuous assessment of damage occurrence. Because the data to detect damage for a SHM system is obtained in a manner different that current standard inspections, so should the estimation and validation process reflect and capitalize on these differences. Although the point of this paper is to develop the statistical framework for certification, a discussion of how SHM systems differ will be used to suggest how the methods to compute these probabilities should be enhanced to incorporate the continuous data collection.

In general, with any system that seeks to classify whether or not damage is present, there are four probabilities related to system outcomes: the probability that the system detects damage is present when it is not (false positive), the probability that the system detects damage is present when it is (true positive), the probability that the system detects no damage when it is (false negative) and the probability that the system detects no damage when it is not (true negative). Conjunctive equations stemming from set theory [1-2] demonstrate that the probabilities of true positive and false negative add to one as do the probabilities of false positive and true negative. Thus, not all four probabilities need be disclosed, but only one from each equation is necessary. The two probabilities typically used are the probability of true positive (or probability of detection) and the probability of false positive. In addition, these probabilities may also be used to incorporate how likely (prevalent) structural damage is to occur. These probabilities are discussed in turn.

PROBABILITY OF DETECTION (POD)

Current standards for detection center on the POD of the system also known as the sensitivity, hit rate, or more precisely as the probability of true positive. The POD is defined as the probability that the system detects structural damage when damage is truly present. Because current practice is to perform periodic assessment of structural damage, a "snap-shot" of structural integrity is obtained. The time of occurrence of damage that is present may not be known with certainty. However, the extent of the damage present enhances the probability that the damage will be detected, that is, enhances the POD. Thus, current standards advocate that the certification process of damage assessment should include a POD that is computed after adjustment for the target (damage) size. Therefore, experimental (or real) data is collected as to the detection of damage. Truth data is observed and a GLM is fit using a hypothesized relationship between the target and observed damage size. Such a relationship is assumed to adjust for the increased POD that accompanies increased structural damage.

To fit such a model, standard statistical techniques are employed. A special link function is specified that transforms the outcome ("damage" or "no damage") into a probability estimate with respect to the target sizes. Such a model is known typically as a logistic regression [3-5], or a form of a log-linear model [4] such as a logit or probit model, and its properties and solutions are well established [3-5] as are its confidence bounds. Once fit, the POD is estimated from this model (e.g. POD=0.90) and a 95% confidence bound at this point is used as the threshold for detection. If multiple inspectors (sensors) are used to estimate the POD, then the POD estimate contains an adjustment for multiple inspectors, that is, contains a dampening effect to adjust the estimated POD and its confidence interval for the variability in inspectors (e.g. sensors). Ideally, in future sampling as set by the schedule in the application's standards of practice, the target POD, or at least the lower bound of the confidence interval for the target POD is sought to assure continued validity of the system.

The POD as well as the confidence bound (as a measure of reliability of the POD estimate) is extremely important as a measure of validity of the SHM system and its ability to detect structural damage. Recall, by the conjunctive equations, once an estimate of the POD is ascertained, so is the estimate for the probability of false negative of the system.

## PROBABILITY OF FALSE POSITIVE (PFP)

Current standards also suggest that along with the target size associated with the POD, the disclosure of the false alarm rate (specifically at the POD=0.90 point) is also recommended. The false alarm rate is defined as the probability that the system detects structural damage when damage is truly not present, that is, it is an estimate of the PFP of the system. In application of current standards, the false alarm rate is adjusted for target size as it is computed from a model of POD that assumes so. The PFP is just as important as the probability of true positive (or POD) for a system. Indeed, if everything is labeled as damage, then surely, the POD = 1.0, however, the PFP is also 1.0. Such a misnomer is a costly waste of resources for structures with no damage. In fact, any false positive is a costly waste of resources as damage is mistakenly thought to be present. Thus, the estimate of the PFP needs to be disclosed for the system. In addition, just as the POD requires a confidence interval, so should the PFP. As all probability estimates for the system are based on data, the measure of variability associated with the estimate must be included. A confidence interval for the PFP will provide the reliability associated with this measure of system accuracy. Formulas for the confidence interval of the PFP (as well as POD) when independent of target size are well established [2].

## POSITIVE (PPV) AND NEGATIVE (NPV) PREDICTIVE VALUES

The positive (negative) predictive value refers to the probability that damage does (does not) exist when the system says that damage does (does not) exist. These probabilities are computed using Bayes rule [1,2]. In situ, these probabilities may be of great importance. If a sensor detects damage, then the probability that damage is really present given that the system detects damage, becomes quite important, especially for structure in which immediate decisions must be made based on this information. However, unlike the POD and PFP, the PPV and NPV are functions of the prevalence, or likelihood of damage being present. Prevalence may be unknown, though estimated, but if not estimated accurately, may mislead users. If accurate estimation of damage prevalence (likelihood) is possible, then these probabilities are useful to disclose. In a simplified example, if it is known that the likelihood of damage for an aircraft is related to the number of flight hours, and that at onset of the system, the chance of damage is 0.01, then the PPV and its confidence bound may be estimated across the possible range of flight hours. This information may be used to predict and schedule the necessary resources over time with respect to maintenance. In addition, it allows for an assessment of the likelihood of a false positive occurring. If the PPV is quite low, yet the system warns of the presence of structural damage, other measures or assessment tools may be used to determine if the warning is true or if a false positive is occurring. The use of estimation of the extent and location of structural damage, as well as the conditions during which the warning occurred may add to this profile.

## DAMAGE DETECTION PROBABILITIES AS RELATED TO SHM SYSTEMS

Differences exist for the automated SHM system. SHM systems are permanently mounted and continuously collect data in situ using fixed sensors (inspectors) surrounding a region of interest. Thus, measures to assess the accuracy of the system need not be adjusted for multiple inspectors as the same inspector(s) assess the system. Specifically, it is reasonable to assume that the probability of inspection is equal to 1.0. Further, there is no “snap shot” for assessment. Assessment is continual and virtually immediate. Thus there is little question when damage occurs as continual assessment may be used to pinpoint its occurrence. In addition, the system is capable of knowing what it just assumed about a region of interest (whether there was damage or not). Further, just as for current practices, the POD will increase over time because the probability of damage will increase over time. However, in an SHM system, the system’s ability to detect damage will rely almost entirely on the minimally detectable damage (MDD), e.g. minimally detectable crack length, and the variability associated with this MDD. Assumably, almost with probability one, will the system of sensors detect any target size outside the range of variability associated with the MDD. This is because the system continuously monitors this region and smart systems with feedback information may capitalize on previous observations. With this feedback, though, is the real possibility to inflate the false alarm rate should a detection error be propagated forward. Such feedback, though useful, must make adjustments and add the checks and balances necessary so as not to inflate the false alarm rate.

Finally, SHM systems may be adjusted to make use of the PPV and NPV estimates. Since the PPV and NPV rely heavily on the prevalence of structural damage, some understanding and knowledge of the occurrence of structural damage over time, especially under specific environmental conditions, may aid the determination of the presence of structural damage when damage is detected. For instance, at the onset of a new SHM system on a new structure, the PPV is quite low because the (expected) prevalence of structural damage is low. Thus, a sensor denoting damage, without the occurrence of other environmental factors that facilitate damage, may be more suspect (i.e. more likely to be a false alarm) than a sensor denoting damage after extended use and/or exposure to environmental factors. This ability is unique to SHM and is not readily incorporated in current inspection practice.

## Estimation

Unique to a SHM system, is the ability of the system to measure and track the development of structural damage. Therefore, in conjunction with assessment tools to evaluate the ability of the system to detect damage, are tools with which to assess how well the system estimates damage, both for the extent (crack length) and location of damage. Since crack length and location are continuous measurements, prediction models will most likely be used for estimation. Assessment measures in the statistical framework for such models as well as the estimates themselves should be disclosed. Thus a recommendation is made for several measures within the framework in order to build a more comprehensive picture of the accuracy of the system. A description of suggested measures follows.

### SUMMARY MEASURES FOR PREDICTION MODEL

The most common tool for estimating a continuous measurement is a form of a general linear model (GLM), which includes supervised learning algorithms, linear regression, time series, random effects models and neural networks, though the model need not be linear. The model's root mean square error (RMSE) is an important measure for model validity. Although many models are designed to minimize MSE, the MSE does not have much meaning in and of itself. In linear regression models, the RMSE does have special meaning. Even without the assumption of normality, about 95% of the observed (predicted) values should lie within 2 times the RMSE. That is, a model that fits well will have a 2 x RMSE value that is very small and that demonstrates a level of precision acceptable for any predicted damage, ideally, within the error of sensor measurement.

Secondly, the model should demonstrate no lack of fit (LOF). When available, data used to create the prediction model should contain measurements to estimate the pure error of the model [3]. These estimates may be used to statistically test the prediction model for LOF. Models exhibiting LOF indicate that for at least some range of data, the predicted values may be inaccurate. Thus a good system will have a prediction model that passes a LOF test so that estimation in situ will not be subject to errors in prediction related to model fit.

Finally, model errors should follow the standard assumptions required of that modeling technique. For instance, in simple or multivariate linear regression, prediction errors are typically normally distributed with mean zero, constant variance and are independent of each other.

### PREDICTED VALUE OF DAMAGE AND LOCATION

The estimates for crack length and location are of primary importance for any system designed to estimate structural damage. Thus, although a fitted model or estimation equation may possess properties that are not traditionally desirable (such as non-significant terms), these models may produce accurate predicted values and are thus meaningful. A good system will have an accurate predicted value for crack length and location with a small residual value. Not only the predicted value should be disclosed, but a confidence for the estimate must be disclosed. This confidence is a function of the model variability (mean square error). For crack length estimation this confidence should form a prediction interval, though for location which may be multi-dimensional it could form a prediction ellipse [3]. The prediction interval (or ellipse) should be constructed to meet the statistical requirements of being uniformly most accurate, that is, will possess minimum width [1,3] thus minimizing the probability of false coverage, and ideally lie within the range of error (noise) of estimation by the sensor.

Prediction intervals which meet the criteria of minimal width [1] are derived for well established modeling techniques such as forms of the GLM [3,6]. For other methods that are non-parametric in nature, alternate techniques may be necessary with which to establish an estimate of the variability of the predicted damage. Common methods to do so often use re-sampling techniques, the most commonly applied in many situations being bootstrap methods [7].

Of utmost importance, is to assess the model fit by cross-validation [3,8]. This may be accomplished by computing the sum of squares of pure error (SSPE) and its associated squared correlation [3,8]. To compute these values, a cross-validation sample must be established. In general, if the SSPE is high, the model does not predict well on independent data and thus will most likely produce poor estimation. The validity of the model will be overstated and will not perform as expected in situ. That is, the model does not translate beyond the data on which it was trained and is over-fit to the data used to create the model. Many modeling procedures, especially those trained on data, lend themselves to overfitting and model summary measures such as R-square and MSE mislead users to believing that the data is better than it is. A model that cannot translate beyond the data on which it was derived will not be useful in practice. There are a plethora of cross-validation methods, some of which are more useful for certain models than others [3,8]. When possible, a disclosure of the SSPE and its associated R-square value is recommended to assess the validity of estimation for the SHM system.

## ESTIMATION AS RELATED TO SHM SYSTEMS

Unique to the SHM system is its continuous data collection surrounding a region of interest. To capitalize on this feature, models developed to estimate damage should include the stream of time related measurements in the estimation process. Standard GLM forms exist for such data structures, and further, may incorporate multiple outcomes [3-6]. Such complex models that make use of time-related data streams are seemingly appropriate and may offer better estimation, i.e., tighter prediction intervals [3,6].

Summary measures for the prediction model are not directly translatable from simple prediction models to those that incorporate the time related data with the exception of the SSPE. This measure and the lack of fit for SHM prediction models may still be employed. Further, by virtue of continuous data collection, an extensive subset of data may be readily obtained from which to conduct cross-validation, and fine intervals of data cycles may be created with which to group data for formal lack of fit testing.

## CONCLUSION

The statistical framework suggested in this paper and summarized in Figure 1 includes statistical measures for systems whose purpose is detection and estimation. To a large

- **Detection (presence of damage):** Probability of detection (POD), Probability of false positive (PFP), other enhancers (positive predictive value, negative predictive value).
- **Estimation (extent and location of damage):** predicted value and associated prediction interval or ellipse, pure error sum of squares, p-value for lack of fit testing, assurance that model assumptions are met, prediction model RMSE.

Figure 1. Summary measures to include in the statistical framework for SHM system validation.

extent, these are the same measures as in current standard practice. However, the methods of current practice do not account for the added value of continuous monitoring in permanently mounted SHM systems.

Specific criteria for each measure in the framework have not been suggested. This was, for the most part, purposely avoided because each SHM system may be uniquely designed for the application of interest. Thus, a system that performs poorer on location estimation, but is highly accurate in detection may be a valid system in applications that do not require location estimation. Such systems may be preferred due to cost or design specifications as compared to a more complex system which estimates location well.

Not all systems need to include measures for detection and estimation of structural damage. Indeed the necessary measures to validate the system must come from the purpose and application of the system itself. In some applications, "hot spots" may be targeted and thus estimates for damage location are not necessary. For SHM certification in this application, measures of detection and crack length, but not location, provide system validation. In other applications, the SHM system may monitor a large area, and thus pinpointing the location of damage may impact greatly the response to such damage. Although not every SHM systems must include all measures as described in Figure 1, the statistical framework of the certification process must identify the appropriate statistical estimates with which to validate the SHM system for both detection and estimation.

Establishment of the certification process and subsequent estimation methods as related to SHM will lead to better quantification of system performance and ultimately better system designs. These criteria may then act as a springboard for certification in subsequent prediction of structural failure, time until structural failure, and other aspects of structural health; all which capitalize on the added value of SHM.

## REFERENCES


1. Casella, G. and R. L. Berger. 2002. *Statistical Inference*, 2nd ed. Duxbury Advanced Series.
2. Zhou, X-H., Obuchowski, N. A. and D. K. McClish. 2011. *Statistical Methods in Diagnostic Medicine*, 2nd ed. Wiley.
3. Kutner, M. H., Nachtsheim, C. J., Neter, J. and W. Li. 2005. *Applied Linear Statistical Models*, 5th ed. McGraw-Hill Irwin.
4. Christensen, R. 1990. *Log-Linear Models*. Springer-Verlag.
5. Molenberghs, G. and G. Berbeke. 2006. *Models for Discrete Longitudinal Data*. Springer Series in Statistics.
6. Liang, K-Y. and S. L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73: 13-22.
7. Efron B. and R. J. Tibshirani 1993. *An Introduction to the Bootstrap*. Chapman & Hall.
8. Hawkins, D. M., Basak, S. C. and D. Mills. 2003 "Assessing Model Fit by Cross-Validation," *J Chem. Inf. Comput. Sci*, 43: 579-586.