COST Action FP1101 Assessment, reinforcement and monitoring of timber structures

State of the Art Report

WG 1 / TG 2

COMBINE USE OF NDT/SDT METHODS FOR ASSESSMENT OF STRUCTURAL TIMBER MEMBERS

Editors:
José Saporiti Machado and Mariapaola Riggio

EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

Multi sensor approach combined with multivariate analysis for assessment of timber structures

Jakub Sandak, Anna Sandak, Mariapaola Riggio
CNR Ivalsa, S.Michele all'Adige (TN), Italy

Introduction

Currently blooming engineering research provides us with numerous methods to be used for improving (reengineering) existing structure assessment routines, including also continuous monitoring of the structure performance. The availability of novel statistical tools to handle many variables simultaneously is another stimulus for rapid changes within the field of measurement technology and the sensors domain. Current trend for using multiple sensors simultaneously is more favorable than a single sensor approach due to far better representation of the real-world cases: the world is multivariate.

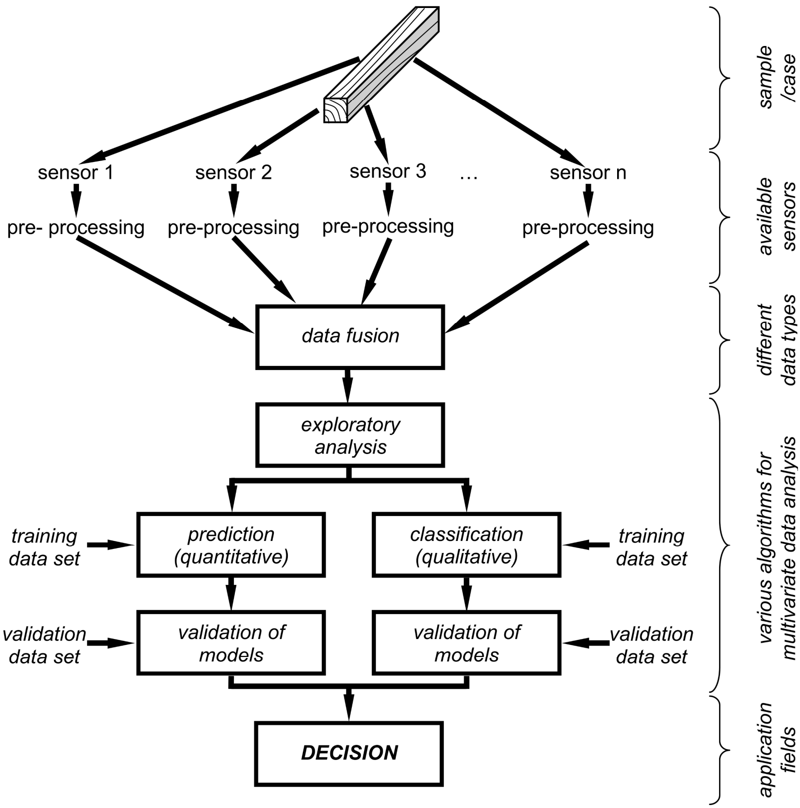
Multi-sensor monitoring generates new issues and challenges, where the fusion of different sources of information is fundamental. Data collected from different types of sensors are often based on diverse physical phenomena, therefore interpretations of results is complicated. A general flowchart of the multi-sensor approach in timber structures assessment is summarized on Figure below. It consists of several layers, including sample/object/case, sensors measuring member properties (through generating various types of data), and numerical models/tools to deal with such data in order to support expert in decision making.

Sample

Samples, cases or objects are the physical units on which the evaluation/measurements are performed. It can be a single wooden member or the whole structure, depending on the scope of evaluation and/or the goal of inspection. There are numerous sample(s) characteristics of interest when assessing timber structures, including:

- material properties
- degradation stage of wooden members due to biotic and a-biotic agents
- presence, position and incidence of strength-affecting defects in wooden members

- presence of damp areas and not uniform moisture distribution in wooden members
- mechanical damage in wooden members and connections
- geometrical alterations in the wooden members and assemblies
- overall performance of the structure
- and others.



Sensors

The visual assessment of timber structures on can be complemented by a series of instrumental techniques, giving information on unreachable object, about not-visible features, and on measurable/quantifiable parameters. The range of sensing techniques suitable for characterization of wood within structures is very wide and includes:

- vision systems in various spectral ranges; visible, infrared, thermo-vision, hyperspectral cameras
- measurement of electromagnetic radiations penetrating structure of wood; γ-rays, X-rays, microwaves, radar detectors
- analysis of mechanical/stress waves propagation; vibration; ultrasound sensors, accelerometers, microphones, laser vibration-meters
- semi-destructive methods; drilling/penetration/cutting equipment as well as screw withdrawal portable testers
- portable spectrophotometers in visible, near-infrared, mid-infrared, XRF spectral bands
- wood moisture meters of various types
- Environment condition monitoring systems including measurements of temperature, relative humidity, solar radiation, rain events and intensity, among others

Pre-processing

The pre-processing of raw signals is a routine task usually performed before any further data evaluation. Several treatments are available, including:

- electronic signal manipulation; amplification, filtering, compensation, etc.
- numeric signal manipulation; normalization, filtering, correction, derivative, integration, noise reduction, smoothing, interpolation, averaging, convolution, etc.
- numerical processing; compression, filtering, wavelet analysis, Fourier transform, etc.

The optimal selection of signal pre-processing is crucial for the overall performance of the multi-sensor system, as well as the presentation of the results/parameters/data. Even if all the existing sensors help in characterization, the human contribution to the structure assessment is indispensable.

Data types

The final output of the sensor (following the pre-processing procedures) is very diverse and depends on the sensor itself, the moment of data analysis, and nature of investigated object/structure, among others. Various sorts of data may be accessible, including:

- single/multiple variables such as scalars or vectors; value at a given time of measurement, change of this value (in the function of time, temperature, pressure, frequency, etc) measured with constant time laps or randomly
- waves, in the form of series of measurements, with defined starting point and distance between measurement points; such as stress-waves, vibrations, radar signals, microwaves
- images, or spatially resolved data, in the form of matrices (a rectangular array of numbers/variables arranged in rows and columns); gray, color, x-ray absorption, thermal images
- spectra, where series of data are representing frequency, wavelength, wavenumber resolved signals; UV-, Vis-, NIR-, IR-, XRF- spectra
- hyperspectral/multispectral cubes, what are hybrids of images and spectra where each pixel represents the full spectra in a given range

Data fusion

The data fusion strategies are different when combining data in real-time or when analysis can be performed after measurement on the archived data. In the first case, dedicated interfaces are indispensable and such data fusion systems are rather complex/case-dependent.

Whilst data evaluation can be performed in post-process mode, the data fusion is rather straightforward. It can be carried out with the help of different software tools (suitable for dealing with various signals/sensors) and accessing diverse databases. The most common result of the data fusion process is a spreadsheet containing series of parameters extracted from various sensors corresponding to single sample/case, collected at a given time or period of time. The size of the spreadsheet may vary depending on the system complexity, number of sensors utilized, quantity of samples/cases and/or duration of the measurement/monitoring. In general, the more complex (in terms of the number of variables and samples) collection of fused data the more reliable/generalized numerical models may be created.

It has to be stated that having huge data sets is not equal to possessing “information”. In fact, high number of data may be cause of disturbance, misinterpretations or confusion, especially when “conventional” data analysis techniques are applied. Multivariate analysis (MVA) techniques, which allow more variables to be analyzed at once, are thus an alternative. The MVA can be divided into three groups:

- exploratory data analysis (data mining) – attempts to find the hidden structures in large and complex data sets
- classification models – are useful when identification of unknown sample/object within one of previously established classes is required

- regression analysis and predictive models – are used for developing statistical models on the base of available reference data.

Number of software packages suitable for MVA and for non-linear systems is already available on the market including, among others, Unscrambler X, OPUS, SIMCA. There is also a high number of dedicated modules for various software development environments such as C++, R, Matab or LabView, in some cases offered as an open source code. A list of MVA techniques suitable for applications toward timber structure assessment is presented below.

Cluster analysis

Cluster analysis (CA) is a statistical method used for matching multivariate data into particular groups according to their similarities. CA divides similar samples into groups called classes or clusters. Clustering methods belongs to unsupervised statistical algorithms; therefore do not require previous information about the objects' memberships, which are obtained according to the data's intrinsic characteristics, or dissimilarities. The clustering can be displayed in the form of a dendrogram where the heterogeneity explains the similarity between the samples. The higher the heterogeneity, the higher is the difference between samples.

Principal Components Analysis

Principal Components Analysis (PCA) is a powerful statistical method for decorrelation of highly correlated data and to reduce the high dimensional data set to lower dimensions. PCA decomposes a linear combination of original variables into few PC (principal components or factors). Each PC explains part of the data set variability. The number of significant factors is case dependant, but as a rule it should be as low as possible. In analogy to cluster analysis, PCA searches for unique properties of samples and separates set of input data into groups of peculiar similarities allowing visualization of natural clustering of the data.

Identity test

Identity test is a method extending use of Principal Components Analysis for differentiation of sample/cases and classification of unknown samples within previously defined groups. The identification of the unknown sample/case is based on computation of principal components by using loadings corresponding to the model. The resulting components are then compared to each group within the model. The result of such comparison is the sample distance called hit quality. The better sample match with the model group, the smaller is the distance. The hit quality of each comparison is weighted against threshold corresponding to each modeled group/class. Three possibilities of unknown sample identification are possible as a result of the identity test:

- the sample is identified as one of the modeled classes (hit quality < threshold in case of only one class)
- the sample is identified as probably belonging to more than one modeled classes, therefore not unique identification is possible (hit quality < threshold in case of more than one class)
- the sample is identified as none of the modeled classes (hit quality > threshold in case of all classes)

SIMCA

Soft Independent Modeling of Class Analogy (SIMCA) is another classification/identification algorithm using Principal Components Analysis for differentiation of sample classes. In analogy to identity test, set of meaningful principal components is derived from the data set. The difference lays in modeling of classes, as in SIMCA each class is modeled separately (local models) and number of principal components may vary between classes. The prediction of a probable class membership for new samples/observations is performed by determination of best fitting to the respective class (local model).

Partial Least Squares

Partial Least Squares (PLS) is a statistical method considered as an expansion of Principal Components Analysis toward quantitative analyses. Basically, PLS finds a linear model describing some predicted variables in terms of other observable variables. The development of PLS model starts with computation of principal components on the base of calibration dataset. The obtained principal components are regressed in the next step against reference variables to be predicted. The PLS model has to be validated after calibration. The coefficient of determination, the root-mean square error of prediction and ratio of standard error of prediction to sample standard deviation are commonly accepted indicators of the PLS model quality.

Multiple linear regression

Multiple linear regression (MLR) is a multivariate analysis tool for modeling the relationship between two (or more) explanatory variables by fitting a linear equation to the reference data. MLR can also be used to estimate the linear association between the predictors and responses, in analogy to Partial Least Squares. Another use of multiple regression is to understand the functional relationships between the dependent and independent variables, by discovering the cause of the variation. The relationship between all predictors and a given response is summarized by the regression coefficients.

Expert systems

The most common perceptive of the expert systems is a rule-based programming. In this programming paradigm, rules are used to represent heuristics, which specify a set of actions to be performed (or decisions to be taken) for a given situation. A rule is composed of an if portion and a then portion. The if portion of a rule is a series of patterns which specify the facts (or data) which cause the rule to be applicable. The then portion of a rule is the set of actions to be executed when the rule is applicable.

Fuzzy logic

The easiest method to emphasize “knowledge” is to use not exact expressions, avoiding precise quantifications and classifying variables into rough values/sets. The scientific usage of such semantics is implemented within fuzzy logic expert systems. The value of each variable (obtained from one or more sensors) is “fuzzyfied” according to pre-defined classes. As a result, detailed numerical value of variable is replaced by the fuzzy value such as “low”, “medium” or “high”. The set of fuzzyfied variables is then propagated to the module where if-then rules are tested. The output of the if-then module is a fuzzy number. It has to be “defuzzyfied” in order to make it a numerical/predicted value and become compatible with the following steps of evaluation and/or decision making.

Neural networks

Neural networks (NN) are widely used for processing of very complex and multivariable data sets. NN are very functional and have a number of great advantages; they have a parallel computation nature, can be applied in various applications, can continuously learn and adopt themselves for changing circumstances. NN is alike a “black box” as the knowledge acquired by the NN is hidden in the neuron weights. The set of variables is propagated to the input of the NN. The properly trained NN will process the input vector and will generate an output – predicted value to be later used for decision making or other actions. Back propagation algorithm is the most popular method for NN training, even if other procedures (including non-supervised learning) are also available.

Decision

The overall purpose of characterizing wooden members with different sensors, as well as developing MVA solutions to deal with multivariate data, is to assist the inspector in making the correct decision regarding the structure assessment, safety and optimal maintenance. It is impossible to generalize the final reasoning process, and to even think to reduce the importance of the inspector in the final decision making, as each wooden structure is a unique case. It is clear, however, that the result of multivariate analysis has to be handy, reliable and intuitive for interpretation.

Conclusions

The proper timber structure assessment is of great importance to assure safe service of buildings as well as to preserve cultural heritage objects for future generations. A multi-sensor approach may be a very attractive alternative to conventional nondestructive method assessment and can provide supplementary data to be considered when inspector decision is measured. The problematic issue is, however, the high number of data/signals to be dealt as a result of measurement with several sensors. It is important, therefore, to assure proper pre-processing of the signals from sensors, appropriate data fusion and optimal data analysis.

The multivariate analysis systems can be implemented for various applications. Exploratory analysis are the most basic, but also most useful for preliminary data screening and identification of trends within data.

Very important task is to continuously upgrade the models. In a perfect case, the software system should acquire new knowledge automatically along the service.

It is assumed that, after additional developments, such methodologies can serve as assisting tools for non-destructive assessment of the wooden structures, service life prediction of structural elements and to support selection of optimal conservation process.