Structural Damage Detection Using Multivariate Time Series ARX Model

Su Chen*
Department of Mathematical Sciences, The University of Memphis, USA

Abstract
In this paper, a novel statistical method of structural health monitoring with multiple wireless sensors is presented. Damage sources in a structure are usually identified by multiple instead of single wireless sensors via the tracking of acceleration time histories stored under undamaged and damaged conditions. Firstly, this method selects a reference signal from the database of undamaged structures. A two-stage prediction model, a combination of multivariate time series auto-regressive and multivariate auto-regressive with exogenous (MARX) input model is constructed to analyze the measurements from reference and test signals respectively. The variance-covariance matrices of the residue errors generated from MARX modeling on test status (health condition to be determined) and the reference status signal data are obtained. Thereafter a log-likelihood ratio test is proposed to compare the two matrices and determine the health status of a structure. Furthermore, to demonstrate the reliability of the proposed multiple wireless sensors damage detection method in Structural Health Monitoring (SHM), we apply our method to two different types of experiments conducted by Los Alamos National Laboratory (LANL). Data are collected from 24 accelerometers placed on a three-story bookshelf. In contrast to prior damage detection methods which can only detect severe damage sources and is insensitive to the less severe damage source, our proposed novel multiple wireless sensors detection method is able to localize all types of damage source much more efficient and reliable in SHM.

INTRODUCTION
Structural health monitoring (SHM) system is a process of detecting the damage condition of a civil structure, providing spatial and quantitative information with respect to structure damage, or predicting the performance of the structure during its remaining operation life-cycle [1]. To seek a sensitive and measurable dynamic damage indicator is one of the important parts in SHM technology. Recent research studies on damage identification are based on modal parameters, such as notably resonant frequencies, natural frequencies, modal shaping and so on. However, these model parameters are insensitive to local damage for the reason that damage, a local phenomenon, cannot be accurately characterized by global parameters. Therefore, it is unwise for researchers to detect local damage in SHM through modal parameters [2].

Alternatively, time-domain and frequency domain damage measurement indicators based on time series analysis were provided to detect the local damage status in SHM. Unlike some mathematical methods such as finite element analysis, the statistical methods such as time series analysis [3] are far more efficient and easy to perform as well. Many researchers adopted the single-input single-output linear time invariant model to interpret the time history of structural response and the residual errors generated by an auto-regressive (AR) and auto-regressive with exogenous input (ARX) model to classify the current damage status of the structure via comparing the current state with the reference state [4]. A reference state is the standard undamaged state and is regarded as the benchmark. The current state is a health condition of a structure to be detected and is unknown [5]. Proposed a combined AR and ARX model to predict the healthy and damaged conditions of bridges. Mattson and Pandit [6] chose the sample standard deviation of residual series for vector auto-regressive (VAR) model as damage diagnosis indicators. Sohn and Farrar [7] and Qing, Zhi and Li [8] used the standard deviation of the residual of (AR-ARX) model as damage indicators. Nair, Kiremidjian and Law [9] proposed an ARM model whose first three (AR) coefficients are regarded as damage sensitive features.

However, above models can only incorporate one sensor in one statistical test. As we know, a structural damage expresses itself in various aspects. One sensor can only catch very limited aspects of damage signals. If multiple wireless sensors are gathered together in a node for online structural health monitoring, they would probably be more sensitive to the damage, even very

tiny one. The earlier the damage source is detected, the faster maintenance can be made to remedy the structure. Suppose two or more sensors are available in a given location, simple (AR-ARX) model would either perform the statistical test for each sensor separately, or choose one important sensor for testing. The potential problems are apparent. On one hand, different sensors may lead to distinctive test results, or even opposite health conditions. If the test is rejected based on the database from sensor 1, but accepted based on the database of sensor 2, it is very difficult to withdraw a persuasive conclusion. On the other hand, no reasonable criterion is reachable to select which sensor is the most important and efficient to detect the structure damage. Furthermore, failing to incorporate the information of all other sensors is a huge waste, and might result in avoidable error.

To solve the problems discussed above, this paper proposed a multivariate time series analysis that handles the database collected by multiple wireless sensors in a node for online structural health monitoring. The advantage of this method is that it utilizes all measurement data of multiple sensors simultaneously, incorporates more complete information and improves the stability of the result. Multiple wireless sensors measure the physical characteristics for both undamaged structures and tested unknown structures. Then these wireless sensors transmit the signals to a cluster head through radio frequency. A standard data normalization process will be performed before analyzing the multiple measurement data. Then we apply a multivariate time series auto-regression with exogenous input (MARX) model to both reference and test signal database. The variance-covariance of residue errors from both MARX models is compared and a log-likelihood ratio test is applied to identify the damage location and severity. The flowchart of this novel damage detection method is schematically depicted in Figure1.

At the end of this paper, a set of bookshelf measurement data is used to justify the feasibility of this multivariate time series (ARX) model. A univariate time series (ARX) model was utilized to demonstrate the advantage of this novel damage detection method as well. It is concluded from our test results that our proposed method improves the accuracy and efficiency in SHM.

This paper was organized as follows. Derivation of multivariate time series (ARX) models on reference signals as well as test signal database and analysis of covariance matrix for residue error are provided in Section 2. Experiment description and results were given in Section 3. To demonstrate the effectiveness of proposed multivariate method in contrast to univariate method, comparison studies are provided as well. Section 4 highlights the advantages of this novel method and summarizes this paper.

**MATERIALS AND METHODS**

Assume that \( k \) wireless sensors are gathered in one node to detect the structure damage status in a specific physical position. Let \( Z_t \) be a vector of \( k \) signal responses from the corresponding wireless sensors at time \( t \), i.e. \( Z_t = (Z_{1t}, Z_{2t}, \ldots, Z_{kt})^T \). To eliminate the errors brought by the different measure units of the signal responses, all time signals are standardized by

\[
Z_{it} = \frac{z_{it} - \mu_i}{\sigma_i},
\]

where \( z_{it} \) is the standardized signal response, \( \mu_i \) and \( \sigma_i \) are the mean and standard deviation of \( Z_{it} \), and \( i = 1, 2, \ldots, k \). Generally, the mean and standard deviation of \( Z_t \) are unknown. Thus, the sample mean \( \bar{z} \) and sample standard deviation \( \sigma_i \) can be used instead. This standardization procedure is applied to all signal responses in the following study. (For simplicity, \( Z_i \) is used to denote \( Z_{it} \) hereafter).

(a) Multivariate Auto-Regression (MAR) Model

Let \( Z_{iDB} \) be a vector of \( k \) signal responses for the \( k \) wireless sensors at the time \( t \) in the reference state, which are usually stored in the reference database. Assume that there are \( T \) consecutive samples in time, and then a multivariate \( AR_{(p)} \) model based on the reference signal responses \( Z_{iDB} \) can be written as:

\[
Z_{iDB} = \sum_{i=1}^{p} \alpha^i_{iDB} Z_{iDB,i-1} + r_{iDB},
\]

where \( \alpha^i_{iDB} = \left[ \alpha^i_{1j}, \alpha^i_{2j}, \ldots, \alpha^i_{kj} \right] \) is a \( k \times k \) coefficient matrix, with columns \( \alpha^i_j = \left[ \alpha^i_{1j}, \alpha^i_{2j}, \ldots, \alpha^i_{kj} \right]^T \) for \( i = 1, 2, \ldots, p; \ j = 1, 2, \ldots, k \). The \( k \) dimensional vector \( r_{iDB} = \left[ r_{1i}, r_{2i}, \ldots, r_{ki} \right]^T \) is the residual error of \( AR_{(p)} \) model for \( T = (p+1), (p+2), \ldots, T \). It is clear that in the \( MAR_{(p)} \) model the response vector \( Z_{iDB} \) of the \( k \) sensors is the weighted sum of \( p \) previous responses of the \( k \) sensors, plus the residual error \( r_{iDB} \).

(b) Multivariate Auto-Regression with Exogenous Input (MARX) Model

It is assumed that the residual error after fitting the \( MAR \) model is mainly influenced by some unknown external input \([10]\). Consequently, a second time-series model for inference database with exogenous inputs added to the multivariate \( AR \) model is constructed as:
Z_i DB = \sum_{i=1}^{p} A_i DB Z_{t-i} DB + \sum_{j=1}^{q} B_j DB r_{t-j} DB + \varepsilon_i DB  \tag{2}

Here $A_i DB = \begin{bmatrix} \alpha_1^i, \alpha_2^i, \ldots, \alpha_k^i \end{bmatrix}$ is a $k \times k$ coefficient matrix for signal responses in previous times with columns

$\alpha_m = \begin{bmatrix} \alpha_{1m}, \alpha_{2m}, \ldots, \alpha_{km} \end{bmatrix}^T$ for $i = 1, 2, \ldots, p$;

$m = 1, 2, \ldots, k$ and $B_j DB = \begin{bmatrix} b_1^j, b_2^j, \ldots, b_k^j \end{bmatrix}$, is a $k \times k$ coefficient matrix for exogenous inputs with columns $b_m^j = \begin{bmatrix} b_{1m}^j, b_{2m}^j, \ldots, b_{km}^j \end{bmatrix}$ for $j = 1, 2, \ldots, q$;

$m = 1, 2, \ldots, k$. The $k$-dimensional vector $\varepsilon_i DB = \begin{bmatrix} \varepsilon_{1u}, \varepsilon_{2u}, \ldots, \varepsilon_{ku} \end{bmatrix}^T$ is the residual error of MARX model, which is assumed to be normally distributed with mean and variance-covariance matrix $\Sigma DB$, for $t = (p + q + 1), (p + q + 2), \ldots, T$. Note that this MARX modeling is similar to a linear Auto-Regressive Moving-Average (ARMA) model presented in [11]. To make the coefficients estimable, $p$ and $q$ should satisfy that $p \geq q + 1, q \geq 1$ and $T \geq p + q + 1$.

In the MARX model, the response vector $Z_i DB$ of $k$ sensors consist of weighted sum of $p$ previous responses, weighted sum of $q$ exogenous inputs $r_{t-i} DB, r_{t-j} DB$ and residual error $\varepsilon_i DB$. Notice that, the exogenous inputs in (2) are residual errors from MAR models.

(c) Estimates of the Parameters in MAR and MARX Model

To estimate the coefficient in (1), define

$Z DB = \begin{bmatrix} Z_{p+1, DB}^{DB}, Z_{p+2, DB}^{DB}, \ldots, Z_{T, DB}^{DB} \end{bmatrix}$, $X DB = \begin{bmatrix} X_{p+1, DB}^{DB}, X_{p+2, DB}^{DB}, \ldots, X_{T, DB}^{DB} \end{bmatrix}^T$

with columns

$X_i DB = \begin{bmatrix} Z_{i-1, DB}, Z_{i-2, DB}, \ldots, Z_{i-p, DB} \end{bmatrix}$,

$\alpha_i DB = \begin{bmatrix} \alpha_1^i, \alpha_2^i, \ldots, \alpha_k^i \end{bmatrix}^T$

and $r_i DB = \begin{bmatrix} r_{p+1, DB}, r_{p+2, DB}, \ldots, r_{T, DB} \end{bmatrix}$, then (1) can be rewritten as:

$Z DB = X DB \alpha DB + r DB \tag{3}$

This is a generalized linear model, and thus Maximum Likelihood Estimate (MLE) of the coefficient matrix $\alpha DB$ is given by [12]:

$\hat{\alpha} DB = \left(X' DB X DB \right)^{-1} X' DB Z DB \tag{4}$

which is used as exogenous inputs to $MAR X_{(p)}$ model.

Similarly, to estimate the coefficient $A_i DB$ and $B_j DB$ in (2), define

$Z DB = \begin{bmatrix} Z_{p+q+1, DB}^{DB}, Z_{p+q+2, DB}^{DB}, \ldots, Z_{T, DB}^{DB} \end{bmatrix}^T$

$C DB = \begin{bmatrix} A_1 DB, A_2 DB, \ldots, A_p DB, B_1 DB, B_2 DB, \ldots, B_q DB \end{bmatrix}$.

$X DB = \begin{bmatrix} X_{p+q+1, DB}, X_{p+q+2, DB}, \ldots, X_{T, DB} \end{bmatrix}^T$, with columns

$X_i DB = \begin{bmatrix} Z_{i-1, DB}, Z_{i-2, DB}, \ldots, Z_{i-p, DB}, r_{i-1, DB}, r_{i-2, DB}, \ldots, r_{i-q, DB} \end{bmatrix}^T$

and then equation (2) can be rewritten in the matrix form as:

$Z DB = X DB C DB + \varepsilon DB$

The maximum likelihood estimator (MLE) of the coefficients of generalized linear model (4) is given by [12]:

$C = \left(X' DB X DB \right)^{-1} \left(X' DB Z DB \right)$

That is to say, coefficient matrix for signal responses and exogenous inputs is estimated by:

$\hat{C} DB = \left(X' DB X DB \right)^{-1} \left(X' DB Z DB \right)$

Meanwhile, the Least Square Estimate (LSE) of the variance-covariance matrix of MAR model’s residual error $\varepsilon DB$ for reference database is given by [12]:

$\Sigma = \frac{1}{T - p - pk - qk} (Z DB - X DB \hat{C}) (Z DB - X DB \hat{C})^T$

(d) Analysis of Covariance Matrix for Residue Errors

In previous subsection, $AR-ARX$ model was developed to estimate the variance-covariance matrix of residual error $\varepsilon DB$ for reference database $\Sigma$. Assume that the variance and covariance matrix for residual error of MARX model on current database is $\hat{\Sigma}$. Thus, repeating the procedure (a) (b) (c) in Section 2, it is not hard to estimate the covariance matrix of residual error for current database as:

$\Sigma = \frac{1}{T - p - pk - qk} (Z - X \hat{C}) (Z - X \hat{C})^T$

Where $\hat{C}$ is the MLE of coefficient matrix for signal responses and exogenous inputs for current database.

To detect whether current database is significantly different from the reference database, a statistical test is performed to compare the variance and covariance of residual error for the two cases. The null hypothesis of test is $H_0: \Sigma = \Sigma DB$ which
is equivalent to “no damage detected”, while the alternative hypothesis is \( H_\theta \neq \Sigma \), which is equivalent to “damage detected”. A likelihood ratio test is usually used to test the difference of two variance matrices. The likelihood ratio, which measures how much more or less the statistical evidence supports some other variance matrix \( \Sigma \) compared with the reference variance matrix \( \Sigma^{DB} \), is defined numerically as follows:

\[
\Lambda = \frac{\max L(\Sigma | H_\theta)}{\max L(\Sigma | H_\theta \cup H_\varnothing)}.
\]

Where \( L(\Sigma | H_\theta) \) represents the likelihood functions conditional on residual variance-covariance matrix when null hypothesis is true and \( L(\Sigma | H_\theta \cup H_\varnothing) \) represents the likelihood functions conditional on residual variance-covariance matrix when null hypothesis or alternative hypothesis is true. It has been proved that (5) can be simplified as in [13]:

\[
\Lambda = \frac{N^{4N/2}}{N^{4N/2}} \frac{|V_1|^{|N/2|} |V_2|^{|N/2|}}{\sum_{i=1}^{n_1} V_i^2 \sum_{i=1}^{n_2} V_i^2},
\]

Where \( V_1 = V_2 \). \( V_i \) and \( V_i \) are adjusted variance-covariance matrices of residual error for current database and reference database respectively, i.e. \( V_1 = \frac{\sum_{i=1}^{n_1} V_i^2}{N} \). \( N\ = N_x \cdot N_y \), \( N_1 = n_{x_1} \cdot 1 \), \( N_2 = n_{x_2} \cdot 1 \), and \( n_1 \) and \( n_2 \) are the sample size in MARX model for current database and reference database respectively, i.e. \( n_{x_1} = T_1 - p-q \), \( n_{x_2} = T_2 - p-q \). \( T_1 \) is the length of the measured detected data and \( T_2 \) is the length of the measured reference data. Fortunately, it is shown by [13] that \(-2\rho \log(\Lambda)\) is asymptotically distributed \( \chi^2(f) \) with degrees of freedom \( f = \frac{1}{2} k(k+1) \), where the factor \( \rho = 1 - \frac{2k^2 + 3k - 1}{6n(k+1)} \left( \frac{n_1}{n_1} - \frac{n_2}{n_2} - 1 \right) \).

**EXPERIMENTAL APPLICATION AND RESULTS**

An experiment of structural damage detection on a three-story bookshelf frame structure was established by Los Alamos National Laboratory (LANL). Details of the experimental program are available in the report by Fasel et al. (LA-UR-03-1468), and the experimental data can be downloaded from the LANL related website: http://institute.lanl.gov/ei/software-and-data.

The bookshelf in this experiment was instrumented with 24 piezoelectric single axis accelerometers. Eight accelerometers were mounted on each floor. A shaker was coupled to the structure by a 15 cm long, 9.5 mm diameter stinger connected to a tapped hole at the mid-height of the base plate. The shaker is attached at the corner D, so that both translational and torsion motions can be excited.

There are 24 channels (1-24) representing for the 24 accelerometers placed on the bookshelf. Channel 3 and Channel 4 represent for the pair of sensors across a joint. Each sensor position is marked with either a P (plate) or C (column) to indicate the position related to the joint. The 3rd floor column support bracket was completely removed and the input level to the shaker is 5 volts in the damage case that this paper studied on. Figure 2 shows a side view of this bookshelf and the placement of those eight accelerometers on one floor. 3AP and 3AC are the 3rd floor position A close to the plate and the column respectively. Similarly 1CP and 1CC are the 1st floor position C near the plate and the column respectively.

**(a) Multivariate Time Series ARX Damage Detection**

The two points 3AP and 3AC are very close to each other as shown in Figure 2. Hence the two sensors mounted at 3AP and 3AC are very close as well and thus it is reasonable to ignore the distance affect between them. In another word, these two sensors can be treated as they were measure data the same damage position merely two different types of damage. One type of damage is related to the column effect, the other one related to the plate. This damage detection test could be assumed that those two sensors in 3AP and 3AC can be bundled together at one node. Likewise 1CP and 1CC can be assumed to be bundle together at one node. 8-second time histories were sampled at a rate of 1024 Hz (1/1024 second), producing 8192 signal points. The first 160 vibration signal data of points 3AP, 3AC, 1CP and 1CC were chosen to analyze the damage status of this three story bookshelf. In this experiment study, these optimal model orders of model are set to be \( p=20 \) and \( q=10 \) based on a partial auto-correlation analysis described in [10].

The experiment results of the proposed log-likelihood ratio test on the covariance of residue error from MARX models are listed in Table 2. It is shown that there are damage at 3AP and 3AC, while no damage detected at position 1CP and 1CC. The test result fits for the actual damage status of bookshelf exactly.

**(b) Univariate Time Series ARXD image Detection**

A number of existing Structural Health Monitoring (SHM) models incorporate only one single sensor in the model to detect damage even in presence of multiple sensors. This paper proposes an innovative two-stage MARX model to detect damage by simultaneously utilizing all the measurements from multiple sensors. To illustrate the advantages of our proposed method, a univariate time series ARX model is applied to measurements of each single sensor for structural damage detection. The details of this method can be found in [8]. The univariate time series ARX model in [8] defines the ratio of the variance of the residual errors to that in the database as:

\[
y = \frac{\sigma^2(\varepsilon_i^2)}{\sigma^2(\varepsilon_i^{DB})}.
\]

Where \( \sigma^2(\varepsilon_i^2) \) the variance of residual errors in unknown state, and \( \sigma^2(\varepsilon_i^{DB}) \) is the variance of residual errors in undamaged state. The coefficient \( y \) in this method follows the F-distribution if the structure is undamaged. Inputting the measurement data in this univariate time series ARX method, those variances of residual errors and corresponding damage coefficients are listed below in Table 1.

By applying the time series univariate ARX method, the
Table 1: Analysis of the residual errors.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Positions</th>
<th>3AP</th>
<th>3AC</th>
<th>1CP</th>
<th>1CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damage Status</td>
<td>0.0036</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undamaged Status</td>
<td>0.0050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Damage Coefficient</td>
<td>0.7200</td>
<td>0.0035</td>
<td></td>
<td>0.0016</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

Table 2: Multiple sensors damage detection result analysis.

<table>
<thead>
<tr>
<th>Test Position</th>
<th>Statistics Value</th>
<th>p-value</th>
<th>Test Status</th>
<th>Actual Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>3AC, 3AP</td>
<td>9.4101</td>
<td>0.0243</td>
<td>Damage</td>
<td>Damage</td>
</tr>
<tr>
<td>1CC, 1CP</td>
<td>5.5344</td>
<td>0.1366</td>
<td>Undamaged</td>
<td>Undamaged</td>
</tr>
</tbody>
</table>

Note: 1. Given significance level: 
2. The degrees of freedom for this experiment is: $f=1/2×2×(2+1)=3^2$.

Experiment results which listed in Table 2 displayed that there are no damage at 3AP, 3AC, 1CP and 1CC. However, this test results didn’t exactly fit for the actual damage status of bookshelf. Actually, there are damage at 3AP and 3AC because the support bracket was completely removed at this position in the experiment.

(c) Comparison of Multiple Sensors and Single Sensor Techniques in Damage Detection

Table 2 and 3 provide the test results (including p-values and damage status) for multiple-sensor and single-sensor damage detection tests respectively.

**p-value** is an indicator of the strength of the evidence against the null hypothesis in favor of the alternative, which shown in fact that there is damage in the bookshelf structure in this experiment. The significance level of these above tests is set as $\alpha=0.05$. If the $p$-value shown in those tables above is less than $\alpha=0.05$ then there is damage at this detection point. Otherwise there is no damage there. In Table 2, $p$-value of 1CC and 1CP is less than $\alpha=0.05$ which predict that there are no damage at all four detected positions, but actually there are damage at positions 3AC and 3AP. To summarize, **MARX** method is much better and more sensitive than the single sensor **ARX** method.

CONCLUSION

In this paper, a multiple wireless sensor structural health monitoring system was discussed. To detect the damage source in the structure efficiently, a novel damage detection method based on multivariate auto-regressive (MAR) and multivariate auto-regression with exogenous input (MARX) modeling was proposed to take advantage of multiple-sensor measurements. A log-likelihood ratio test was constructed to determine the structural health condition through a ratio of the variance and
covariance matrices of residual errors from MARX modeling on the current database and reference database. This novel damage detection method was then applied to detect the health status of a three-story bookshelf. To illustrate the advantage of the proposed MARX method in health structure monitoring, a AR-ARX model was also performed to each single sensor in the three-story bookshelf database for comparison purpose. It had been verified that this MARX model correctly detected the damage when the sensors were close to the damage source. While the sensors are very far from the damage source, no significant damage signal is detected which satisfies the common sense. In contrast, simple AR-ARX model cannot detect the existence of damage through analyzing any of single sensors alone at the significant level as the MARX model, no matter whether detected sensors are close to or far from the damage source. This comparison infers that this MAR-MARX modeling method, as expected, is more sensitive to damage source because of incorporation of multidimensional information of damage detection.

Another advantage of this novel damage detection method is that it is not only efficient, but also easy to perform in the real life. By inheriting the benefit of low computation load from AR-ARX model, MAR-MARX model does not require any complicated mathematical models and time-consuming computation as finite element modeling method did. Consequently, it is possible to process data at sensor node location through embedded algorithms since this damage detection method is particularly suitable for wireless sensor network analysis. Last but not least, this novel damage detection method could also be widely applied to multiple type of sensors detection, as well as, multiple wireless sensors gather together so close to each other that distance affect can be ignored.

REFERENCES