Paper ID: IA-ICCAEMILN-26088-9345

IASTEM
INTERNATIONAL ACADEMY OF SCIENCE, TECHNOLOGY, ENGINEERING AND MANAGEMENT

International Conference on Civil and Architectural Engineering

Certificate

This is to certify that Ali Behkamal has presented a paper entitled "An Automatic Method for Identification of Time Series Models in Vibration-Based Applications" at the International Conference on Civil and Architectural Engineering (ICCAE) held in Milan, Italy on 29th-30th August 2018.

Chairman
International Academy of Science, Technology, Engineering and Management

AN AUTOMATIC METHOD FOR IDENTIFICATION OF TIME SERIES MODELS IN VIBRATION-BASED APPLICATIONS

[1]ALI BEHKAMAL, [2]ALIREZA ENTEZAMI, [3]HASSAN SARMADI, [4]HASHEM SHARIATMADAR

[1]M.sc. student in Construction Management and Engineering, Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[2]Ph.D student in Structural Engineering, Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[3]M.Eng. graduate in Structural Engineering, Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[4]Associate Professor, Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
E-mail: [4]shariatmadar@um.ac.ir

Abstract -
Time series modelling is an influential and successful method for vibration-based applications under data-driven approaches. Because it is a parametric statistical method, one needs to define more details and parameters compare with non-parametric techniques. Time series modelling is generally based on fitting a time series representation to raw vibration measurements and using its statistical characteristics. In vibration-based applications, these characteristics are used for some problems such as system identification, modal analysis, damage detection, etc. The primary step of time series modelling is to identify an appropriate time series representation is such a way that is should be compatible with the nature of time series data. Although the graphical techniques such as Box-Jenkins methodology are often the initial choices, the model identification via such approaches may be difficult and time-consuming along with some limitations. Therefore, this study proposes an automatic model identification approach by incorporating the statistical and engineering aspects when vibration time-domain measurements are linear and stationary. In the first step of the proposed approach, it is necessary to perform some data analyses to recognize the nature of vibration time-domain measurements. For the process of model identification, the proposed method relies on numerical evidence based on some information criteria including Akaike's final prediction error (FPE) and Bayesian information criterion (BIC). The measured vibration responses of an experimental four-story steel structure under ambient excitations are utilized to demonstrate the capability of the proposed method. Results will show that the proposed automatic approach succeeds in identifying the best time series model for linear and stationary time series data and facilitates the process of model identification compared with the Box-Jenkins methodology.

Keywords - Vibration; Time Series Modelling; Data Analysis; Model Identification

I. INTRODUCTION

Vibration-based applications such as system identification, modal analysis, and damage detection recently focus on data-driven approaches. The main characteristic of these approaches is to use raw vibration measurements acquired from dynamic tests. In contrast, another approach to vibration-based application relies upon constructing an accurate finite element (FE) or analytical model of the real or tested structure and applying the inherent structural properties such as mass, stiffness, and damping matrices [1-5]. In reality, it is impossible to build an analytical or numerical model of a real structure due to some modelling errors. In such case, it is indispensable to implement the model updating procedure [6]. Under this limitation and the other difficulties such as constructing the FE model and transforming raw vibration measurements into the frequency and modal domains, one can realize that data-driven approaches are more useful and appropriate ways for vibration-based applications.
In most cases, the measured vibration data is in time domain such as acceleration time histories acquired from accelerometers. An important fact is that it is very difficult or impossible to directly utilize the raw vibration measurements for most of the vibration-based applications. On this basis, time series modelling provides a powerful statistical tool for analysing and modelling of the vibration time-domain data in an effort to some important vibration-based applications such as modal identification [7-9], long-term condition assessment [10]. Apart from the data analysis and modelling, it makes an efficient and influential approach to extracting meaningful and important patterns or features from the time series data for damage detection under structural health monitoring [11, 12].
One of the significant parts of time series modelling is to select an appropriate time series representation in such a way that it should be compatible with the nature of time series data [13]. In most cases, the selection of a time series model needs an expertise because most of the available model identification techniques are graphical and visual-based tools. The inaccurate choice of a time series model may also lead to a time-consuming process with redundant details. The mentioned issues are related to the statistical aspects of time series modelling, while the engineering aspects play important roles in the use of time series modelling in the vibration-based applications. Therefore, this study is intended to focus on the process of model identification based on the statistical and engineering aspects by proposing an automatic approach. The central idea behind this method is to

utilize numerical evidence under the theory of information criteria that are often applied to determine the orders of time series models. Experimental datasets of vibration measurements of a four-story steel structure under ambient excitation are used to validate the accuracy and performance of the proposed automatic model identification method.

II. TIME SERIES MODELLING

Time series modelling is a statistical tool for fitting a mathematical model to time-domain samples [13]. This is a parametric approach because one needs to define and characterize the main details of time series models. In contrast, a non-parametric does not require describing the details of the model and often relies upon a limited number of parameters. On the other hand, time series modelling depends strongly on the type and nature of time series data. There are a broad range of time-domain data including stationary vs. non-stationary, linear vs. nonlinear, Gaussian vs. non-Gaussian, univariate vs. multivariate, seasonal vs. non-seasonal etc. [13]. As such, one can find a large number of time series representations suitable for each of the time series datasets. In the time series modelling, time-invariant linear representations are widely used models in vibration-based applications. This is because of the use of a linear model with stationarity behaviour is considerably simpler than a nonlinear or non-stationary representation. Moreover, time-invariant linear models have a few details in comparison with the other types of representations and decrease the computational costs.

In general, most of the time-invariant linear models consists of output, input, and error terms. These terms are equivalent to Autoregressive (AR), eXogenous (X), and Moving Average (MA). Using these terms, one can construct different linear and stationary time series models. Assume that $x(t)$ and $y(t)$ denote the excitation and vibration response at the specific time t. The general formulation of a time-invariant representation belongs to Autoregressive Moving Average with eXogenous input (ARMAX), which includes all of the above-mentioned terms. The ARMAX model is defined as (see "armax" MATLAB function):

$$y(t)=\sum_{\mathrm{i}=1}^{p}\theta_{\mathrm{i}}y(t-\mathrm{i})+\sum_{\mathrm{j}=1}^{r}\varphi_{\mathrm{i}}x(t-\mathrm{j})+\sum_{\mathrm{k}=1}^{q}\psi_{\mathrm{i}}e(t-\mathrm{k})+e(t) \quad (1)$$

where p, r, and q represents the orders of the output (AR), input (X), and error (MA) terms, respectively. Hence, $\mathbf{\Theta}=[\theta_1\ldots\theta_p]$, $\mathbf{\Phi}=[\varphi_1\ldots\varphi_r]$, and $\mathbf{\Psi}=[\psi_1\ldots\psi_q]$ denotes the coefficients of these terms. Moreover, $e(t)$ is the error or discrepancy between the measured time series data and predicted one via the model. It is possible to define the other kinds of time-invariant linear models by removing each of the X and MA terms. It should be mentioned that the main aim of using time series modelling in vibration-based applications is to model the structural responses or the outputs of the dynamic systems. Therefore, it is essentially needed to preserve the AR or output term of the linear and stationary models. By eliminating the error term, one can define Autoregressive with eXogenous (ARX) model as follows (see "arx" MATLAB function):

$$y(t)=\sum_{\mathrm{i}=1}^{p}\theta_{\mathrm{i}}y(t-\mathrm{i})+\sum_{\mathrm{j}=1}^{r}\varphi_{\mathrm{i}}x(t-\mathrm{j})+e(t) \quad (2)$$

If the input term is removed from Eq. (1), it can be constructed Autoregressive Moving Average (ARMA) representation in the following form:

$$y(t)=\sum_{\mathrm{i}=1}^{p}\theta_{\mathrm{i}}y(t-\mathrm{i})+\sum_{\mathrm{k}=1}^{q}\psi_{\mathrm{i}}e(t-\mathrm{k})+e(t) \quad (3)$$

Eventually, the last time-invariant linear model is AR that only requires the output term (see "ar" MATLAB function):

$$y(t)=\sum_{\mathrm{i}=1}^{p}\theta_{\mathrm{i}}y(t-\mathrm{i})+e(t) \quad (4)$$

Unlike the non-parametric representations (e.g. auto-covariance function, cross-covariance function, cross spectral density, auto power spectral density, etc. [11]), it is apparent that the parametric models require some prominent ingredients such as identifying an appropriate time series representation compatible with the type and nature of time series data (model identification), determining adequate orders (model order determination), and estimating the model coefficients (parameter estimation), and validating the accuracy and adequacy of the model (model diagnostic checking). The main objective of this study is to pay attention to the model identification.

III. MODEL IDENTIFICATION BASED ON ENGINEERING ASPECTS

The measurement of vibration data can be implemented by diverse dynamic tests through sensing and data acquisition equipment [14]. Generally, there are two types of excitation methods including forced and ambient vibration loads [15]. When some excitation devices such as … are available, one can apply controlled forces to structures and measure both the input (excitation) and output (structural responses) datasets. However, this approach is not always practical, particularly in civil engineering structures due to the requirement of a large force for exciting large and complex civil structures and occurrence of probable damage [15]. Due to the advanced development of sensing and data acquisition systems, the use of ambient vibration for the excitation of civil structures has received more attention. The main limitation of using this excitation source is that it is difficult or impossible to measure the unknown and unpredictable ambient loads. Therefore, one can realize that in addition to the nature of time series data, the application of time series modelling also depends on the type of data acquisition.

With these descriptions, the model identification based on engineering aspects can be carried out by the input-output and output-only time series model classes. In this regard, the ARX and ARMAX models require the input or eXogenous term and fall into the input-output model class. On the contrary, the AR and ARMA representations do not need to assign the input term in their formulations. Therefore, these fall into the output-only model classes.

IV. MODEL IDENTIFICATION BASED ON STATISTICAL ASPECTS

From a statistical viewpoint, the time-invariant linear representations are constructed from one or all of the output, input, and error terms, which are equivalent to the AR, X, and MA. Hence, the model identification based on a statistical approach can be implemented by approaches that are able to recognize these terms based on the nature of time series data.

4.1 Conventional graphical approach

The Box-Jenkins methodology is a conventional graphical approach to identifying a time-invariant linear representation [13]. This methodology relies on the autocorrelation function (ACF) and partial auto-correlation function (PACF). The ACF is the correlation between any two values in a time series with a specific time shift called lag (see "autocorr" MATLAB function). The PACF is the correlation between any two samples with a specific lag (see "parcorr" MATLAB function), where the linear effects of the samples are removed [16]. Once the type of model class (i.e. input-only or output-only) has been recognized, the plots of ACF and PACF on the univariate time series data enable the researcher, engineer, and analyst to graphically identify a model. The strategy for model identification by using these function is presented in Table 1.

Function	Model class		
	AR(*p*)	MA(*q*)	ARMA(*p*,*q*)
ACF	Tails off	Cut off after lag *q*	Tails off
PACF	Cut off after lag *p*	Tails off	Tails off

Table 1. Graphical model identification by the Box-Jenkins methodology

4.2 Proposed automatic approach

Although the Box-Jenkins methodology is a well-known graphical approach to the model identification, some limitations may lead to difficulties and erroneous selections. The first limitation is that the use of this approach needs high expertise for recognizing when the samples of ACF or PACFs tail off or cut off. In other words, it is not a trivial process to graphically choose a model. As the other limitation, the plots of ACF and PACF need to define a lag number. Small and large lag numbers may cause an erroneous model identification. Moreover, the Box-Jenkins methodology is not automatic, which means that it is necessary to perform the model identification on each time series data leading to a time-consuming process. To deal with these limitations, the proposed automatic approach utilizes numerical evidence for the best model identification. The fundamental principle of this approach relies on the use of information criteria under the selection of different sample orders. The process of model identification via the proposed automatic approach is similar and equivalent to the Box-Jenkins methodology with a difference that instead of applying the plots of ACF and PACF, it uses the information criteria as the numerical evidence for choosing the best AR, MA, ARMA models. At the first step of this approach, it is necessary to carry out initial data analyses in an effort to recognize the nature of vibration time-domain measurements. The first information criterion used in the proposed automatic model identification method is Akaike's final prediction error (FPE) [17], which is expressed as:

$$FPE = \det\left(\frac{1}{n}\sum_{t=1}^{n}\mathbf{e}(\mathrm{t},\mathbf{\Omega})\mathbf{e}(\mathrm{t},\mathbf{\Omega})^{\mathrm{T}}\right)\left(\frac{1+\frac{d}{n}}{1-\frac{d}{n}}\right) \quad (5)$$

where $\mathbf{e}(t,\mathbf{\Omega})$ is the vector of model residuals at time t by using the vector of model coefficients $\mathbf{\Omega}$, which can be one of the coefficient vectors of output ($\mathbf{\Theta}$), input ($\mathbf{\Phi}$), and error ($\mathbf{\Psi}$) terms. In Eq. (5), n denotes the number of data samples and d represents the number of estimated coefficients (i.e. d=p for AR, d=q for M, and d=p+q for ARMA). Furthermore, "*det*" refers to the determination operator. Bayesian information criterion (BIC) is another tool used in the automatic model identification approach. The formula of BIC is given by [13]:

$$\mathrm{BIC} = -2\ln\left(L_{\max}\right) + d\ln\left(n\right) \quad (6)$$

where $L_{\max}$ represents the maximum value of likelihood function. In statistics, the information criteria are useful techniques for order determination. For this purpose, one initially needs to examine different orders (e.g. 1-100) and then compute the amounts of information criteria. A number that has the smallest quantity of information criteria is chosen as the model order. However, in this article, the same procedure is adopted to identify the best time-invariant linear models by using the FPE and BIC.

V. AN EXPERIMENTAL FOUR-STORY STEEL STRUCTURE

The capability and performance of the proposed automatic approach to model identification are validated by the measured vibration responses of an experimental four-story steel structure under ambient excitations. This structure is a well-known benchmark problem for vibration-based applications, which belongs to the American Civil Engineer Society (ASCE) [18]. The structure was constructed from 2-bay-by-2-bay steel frame in scale-model with 2.5×2.5m in plan and 3.6m in tall as shown in Fig. 1(a). It was subjected to ambient excitations including excitations present from the environment due to the wind, pedestrians, and traffic. The vibration time-domain signals were measured by 15 accelerometers with 5 Volts/g sensitivity distributed on the four stories and the base of the structure as shown in Fig. 1 (b).

(a)

(b)

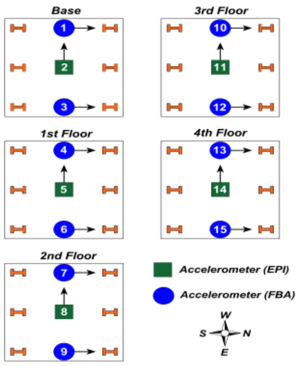

Figure 1. (a) The four-story ASCE steel structure, (b) sensor locations

In order to demonstrate the ability of the proposed methods to identify a time series model, the measured vibration responses at the sensors 4, 7, 10, and 15 of the undamaged condition are utilized. Note that the undamaged state used in this article is equivalent to the test configuration 1 in [18]. Since the measured vibration responses were acquired from the ambient excitation sources, one can use the output-only model classes. At first, it is necessary to analyse the vibration responses and recognize their nature. As a sample, Fig. 2 illustrates the vibration measurements at the sensors 4, 7, 10, and 15. The visual inspection indicates that the time series data at these sensors are linear.

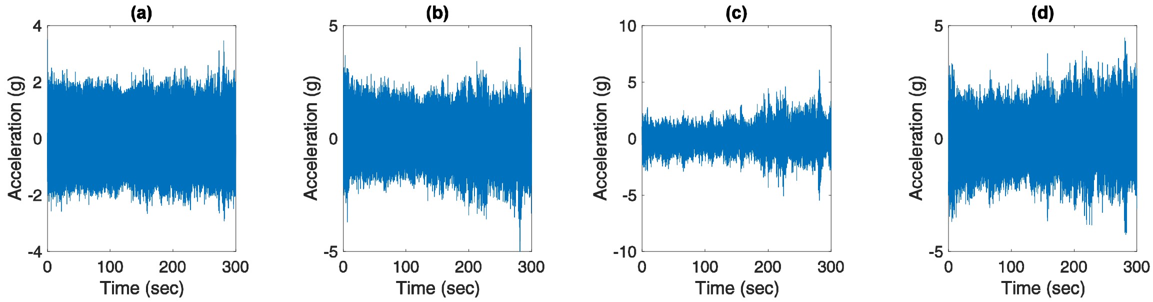

Figure 2. The measured vibration responses: (a) Sensor 4, (b) Sensor 7, (c) Sensor 10, (d) Sensor 15

In order to examine the stationarity or non-stationarity of vibration measurements, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) hypothesis test [19] is utilized that assesses the null hypothesis of a univariate time series is trend stationary. Similar to most of the statistical hypothesis test, it gives some important outputs such as the test statistics (Q_{KPSS}) and a critical value (c-value) under a significance level (see "kpsstest" MATLAB function). Based on the KPSS test, if the value of Q_{KPSS} is larger than the c-value, the test rejects the null hypothesis in the sense that the time series data is non-stationary. Fig. 3(a) demonstrates the amounts of Q_{KPSS} by using 5% significance level, which leads to the c-value (the dashed red line) corresponds to 0.1460.

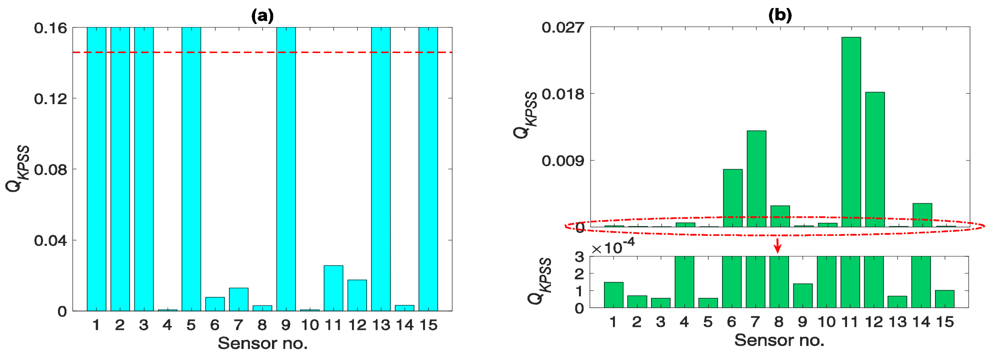

Figure 3. Stationarity assessment of the vibration responses at all sensors by the KPSS test: (a) before differencing, (b) after differencing

It is apparent from Fig. 3(a) that the acceleration time histories at the sensors 1-3, 5, 9, 13, and 15 are non-stationary since their KPSS test statistics exceed the *c*-value. Under such circumstances, it is not accurate to fit the time-invariant linear models to the non-stationary data. An efficient way for dealing with this problem is to use the differencing of time series data [13]. Performing the first-order differencing on the non-stationary time-domain responses (i.e. "diff" MATLAB function), Fig. 3(b) illustrate the new amounts of Q_{KPSS} after the data differencing. As can be seen, all values of Q_{KPSS} are smaller than the *c*-value, which mean that the non-stationary time-domain responses become stationary. At this moment, one can utilize the graphical and automatic model identification techniques based on the linearity and stationarity of the vibration responses. Fig. 4 shows the results of model identification by the Box-Jenkins methodology at the sensor 15. In this figure, the plots of ACF and PACF are depicted by using 100 lags. It can be discerned that the samples of both ACF and PACF tail off and do not cut off at specific lags. Based on Table 1, one can state that the most appropriate time series model for the acceleration time histories at the sensor 15 is most likely the ARMA representation. As can be seen, it is difficult to certainly select a suitable model because it needs a professional expertise and knowledge about the variations of the ACF and PACF samples. On the other hand, if the predefined lag becomes more than 100, it is probable that the selection of a model between the AR and ARMA to be more difficult. This is because the samples of PACF tend to zero or cut off at a lag, which implies the selection of AR model. As a result, although the Box-Jenkins methodology is known as the useful graphical tool for the model identification, it does not guarantee an appropriate and accurate selection of a time series model.

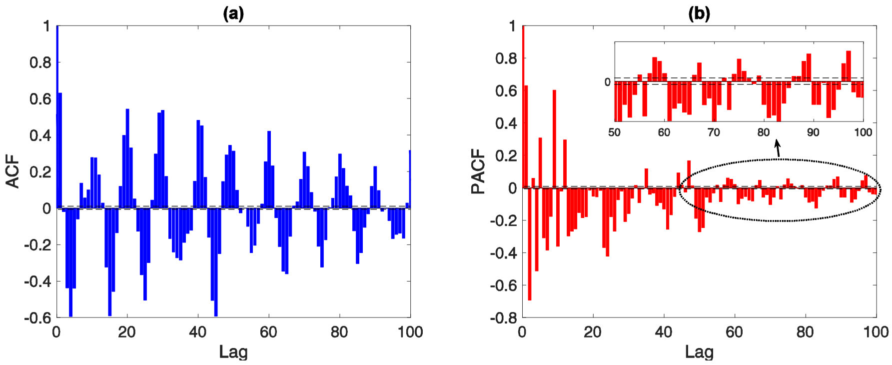

Figure 4. Graphical model identification via the conventional Box-Jenkins methodology at the sensor 15: (a) ACF, (b) PACF

In order to begin the automatic model identification, it is necessary to compute the amounts of these FPE and BIC in different sample orders. The process is based on fitting AR, MA, and ARMA models with diverse orders (i.e. p for AR, q for Ma, and both p and q for ARMA), estimating the coefficients of the mentioned representations, and calculating the FPE and BIC values. The sample order for AR and MA are 10, 20, 30, 40, 50, 60, 80, and 100.

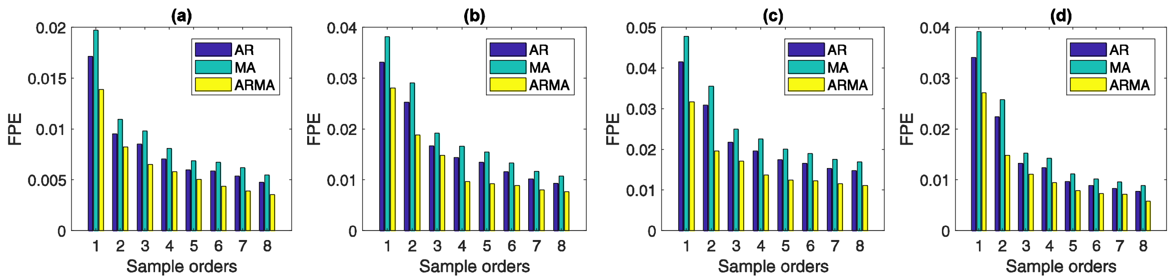

Figure 5. Automatic model identification via the proposed approach based on the FPE: (a) Sensor 4, (b) Sensor 7, (c) Sensor 10, (d) Sensor 15

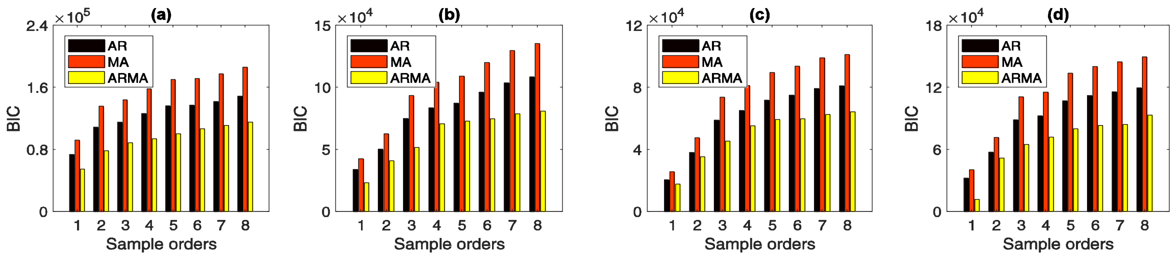

Figure 6. Automatic model identification via the proposed approach based on the BIC: (a) Sensor 4, (b) Sensor 7, (c) Sensor 10, (d) Sensor 15

Because the ARMA contains the two types of orders, the half of the mentioned quantities are allocated to the ARMA model; that is, ARMA(5,5), ARMA(10,10),…,ARMA(50,50). The results of the automatic model identification based on the FPE and BIC under the eight sample orders at the sensors 4, 7, 10, and 15 are shown in Figs. 5 and 6, respectively. The observations in these figures reveal that the smallest BIC and FPE quantities in all eight sample orders belong to the ARMA, while the MA has the largest amounts. Therefore, it can be concluded that the ARMA model is the most appropriate time series representation for modelling.

VI. CONCLUSION

This article proposed an automatic model identification approach based on numerical evidence through the FPE and BIC based on the statistical and engineering aspects. The proposed approach is equivalent to the Box-Jenkins methodology as the conventional graphical approach to the model identification. The central idea of the automatic method is to fit AR, MA, and ARMA representations with diverse sample orders and choose one of them that has the smallest FPE and BIC values. Using the measured vibration datasets of the ASCE benchmark problem under ambient vibration, it was seen that the applications of both the automatic and graphical model identification techniques essentially need an initial data analysis for understanding the nature of time series data in terms of linearity and stationarity. Furthermore, the results showed that the selection of a suitable model by the conventional Bo-Jenkins methodology may be difficult and erroneous along with a professional expertise. In contrast, it was observed that the proposed automatic approach not only addresses the limitations of the graphical technique but also facilitates the process of model identification based on its automatic algorithm.

REFERENCES

[1] Entezami, H. Shariatmadar, and H. Sarmadi, "Structural damage detection by a new iterative regularization method and an improved sensitivity function," Journal of Sound and Vibration, vol. 399, pp. 285-307, 2017/07/07/ 2017.

[2] Entezami and H. Shariatmadar, "Damage localization in shear buildings by direct updating of physical properties," International Journal of Advanced Structural Engineering (IJASE), vol. 6, p. 4, 2014.

[3] Entezami and H. Shariatmadar, "New sensitivity-based methods for structural damage diagnosis by least square minimal residual techniques," Iranian Journal of Science and Technology-Transactions of Civil Engineering, vol. 39, 2015.

[4] Entezami and H. Shariatmadar, "Damage detection in structural systems by improved sensitivity of modal strain energy and Tikhonov regularization method," International Journal of Dynamics and Control, vol. 2, pp. 509-520, 2014.

[5] Entezami, H. Shariatmadar, and M. Ghalehnovi, "Damage detection by updating structural models based on linear objective functions," Journal of Civil Structural Health Monitoring, vol. 4, pp. 165-176, 2014.

[6] H. Sarmadi, A. Karamodin, and A. Entezami, "A new iterative model updating technique based on least squares minimal residual method using measured modal data," Applied Mathematical Modelling, vol. 40, pp. 10323-10341, 2016.

[7] M. Rezaiee-Pajand, A. Entezami, and H. Shariatmadar, "An iterative order determination method for time-series modeling in structural health monitoring," Advances in Structural Engineering, vol. 0, 2017.

[8] Entezami and H. Shariatmadar, "An unsupervised learning approach by novel damage indices in structural health monitoring for damage localization and quantification," Structural Health Monitoring, vol. 0, 2017.

[9] Bao, H. Hao, and Z. Li, "Vibration-based structural health monitoring of offshore pipelines: numerical and experimental study," Structural Control and Health Monitoring, vol. 20, pp. 769-788, 2013.

[10] P. A. Samara, J. S. Sakellariou, G. N. Fouskitakis, J. D. Hios, and S. D. Fassois, "Aircraft virtual sensor design via a Time-dependent Functional Pooling NARX methodology," Aerospace Science and Technology, vol. 29, pp. 114-124, 8// 2013.

[11] S. D. Fassois and J. S. Sakellariou, "Statistical time series methods for structural health monitoring," in Encyclopedia of Structural Health Monitoring, ed: John Wiley & Sons, Ltd, 2009, pp. 443-472.

[12] R. Yao and S. N. Pakzad, "Autoregressive statistical pattern recognition algorithms for damage detection in civil structures," Mechanical Systems and Signal Processing, vol. 31, pp. 355-368, 2012.

[13] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time series analysis: forecasting and control, Fifth ed.: John Wiley & Sons, 2015.

[14] M. L. Wang, J. P. Lynch, and H. Sohn, Sensor Technologies for Civil Infrastructures: Applications in Structural Health Monitoring: Woodhead, 2017.

[15] Cunha, E. Caetano, F. Magalhães, and C. Moutinho, "Recent perspectives in dynamic testing and monitoring of bridges," Structural Control and Health Monitoring, vol. 20, pp. 853-877, 2013.

[16] M. Bögl, W. Aigner, P. Filzmoser, T. Lammarsch, S. Miksch, and A. Rind, "Visual analytics for model selection in time series analysis," IEEE transactions on visualization and computer graphics, vol. 19, pp. 2237-2246, 2013.

[17] L. Ljung, System Identification: Theory for the User, Second Edition. Upper Saddle River, NJ: Prentice-Hall, 1999.

[18] S. J. Dyke, D. Bernal, J. Beck, and C. Ventura, "Experimental phase II of the structural health monitoring benchmark problem," in Proceedings of the 16th ASCE engineering mechanics conference, 2003.

[19] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root," Journal of Econometrics, vol. 54, pp. 159-178, 1992.

★★★