

## Subspace system identification of support-excited structures—part I: theory and black-box system identification

Junhee Kim and Jerome P. Lynch<sup>\*,†</sup>

*Department of Civil and Environmental Engineering, University of Michigan, 2380 G. G. Brown Building, Ann Arbor, MI USA*

### SUMMARY

This paper reviews the theoretical principles of subspace system identification as applied to the problem of estimating black-box state-space models of support-excited structures (e.g., structures exposed to earthquakes). The work distinguishes itself from past studies by providing readers with a powerful geometric interpretation of subspace operations that relates directly to theoretical structural dynamics. To validate the performance of subspace system identification, a series of experiments are conducted on a multistory steel frame structure exposed to moderate seismic ground motions; structural response data is used off-line to estimate black-box state-space models. Ground motions and structural response measurements are used by the subspace system identification method to derive a complete input–output state-space model of the steel frame system. The modal parameters of the structure are extracted from the estimated input–output state-space model. With the use of only structural response data, output-only state-space models of the system are also estimated by subspace system identification. The paper concludes with a comparison study of the modal parameters extracted from the input–output and output-only state-space models in order to quantify the uncertainties present in modal parameters extracted from output-only models. Copyright © 2012 John Wiley & Sons, Ltd.

Received 23 March 2011; Revised 14 February 2012; Accepted 21 February 2012

**KEY WORDS:** subspace system identification; support-excited structure; structural dynamics; data-driven model; system identification

### INTRODUCTION

Output-only system identification using ambient vibrations is a popular practice in the civil engineering field. The use of ambient excitations is convenient because of the technical difficulties associated with exciting large civil engineering structures in a controlled and repeatable manner [1–5]. Although output-only identification is a popular choice among civil engineers, the approach has a number of notable limitations. Foremost among those limitations is that a mathematical evaluation of the accuracy of the system models resulting from output-only methods is difficult to perform because of the absence of a measured input. Hence, output-only system identification must be treated as operational modal analysis as opposed to experimental modal analysis. In light of these technical challenges, output-only methods might be insufficient to completely attain the goal of system identification, which is the accurate estimation of the physical quantities and output prediction of the system extracted from a mathematical model created using measurement data [6]. Hence, input–output system identification is strongly favored over output-only methods if the excitation (i.e., system input) can be accurately measured.

<sup>\*</sup>Correspondence to: Jerome P. Lynch, Department of Civil and Environmental Engineering, University of Michigan, 2380 G. G. Brown Building, Ann Arbor, MI 48109-2125, USA, Tel: (734) 615-5290.

<sup>†</sup>E-mail: jerlynch@umich.edu

The vibrations of foundation systems is one of the most widely studied problems in the civil engineering field [7]. Foundation excitations originating from earthquakes, rotating machinery, vehicular loads, and construction equipment may be harmful to civil engineering structures. Paradoxically, the structural vibrations derived from support excitations can also be used to examine the performance of the structure [8]. For example, structural characteristics extracted from vibration measurements can be correlated to the structural integrity of the system. Structural characteristics observed after the application of strong ground motion could be used to assess the health of the structure by comparing structural characteristics to a set of baseline characteristics extracted before the application of ground motion. Unlike other structural dynamic problems where the system input corresponds to time-varying forcing functions applied to the system degrees-of-freedom, the system input for support-excited structures is based on the motion of the structure boundary (i.e., support motion). The input motion of the foundation is typically as easy to measure as any other system degree-of-freedom; this allows input–output system identification methods to be applied to support-excited structures.

Since the introduction of stochastic subspace identification (SSI) by Peeters and Roeck [9], subspace methods have become popular system identification tools used by the civil and mechanical engineering communities. The SSI algorithm is just one member of the general subspace state-space system identification (4SID) family [10] and is recognized as a significant achievement of the theoretical dynamics and control communities [11]. Since 4SID was introduced to the civil engineering community as a set of output-only system identification methods, a rigorous mathematical mapping of state-space model parameters to the physical parameters of the system has yet to be undertaken. The lack of a mathematical mapping has limited 4SID as a purely black-box data-driven tool whose results are difficult to interpret by engineers. The overarching goal of this paper is to revisit subspace system identification for the modeling of support-excited civil engineering structures. To accomplish this goal, a detailed description of the subspace system identification method is provided with an emphasis placed on the physical meaning of subspace operations using geometric interpretations intuitive to researchers in the field of structural dynamics. This paper is intended to serve as the theoretical foundation of a companion paper [12] whose focus is on the extension of subspace system identification to realize a physically meaningful model derived from experimental data. Central to this study is the use of a six-story steel frame test structure experimentally excited with moderate seismic base motions. Both input–output and output-only state-space models are extracted by subspace system identification methods. Uncertainties associated with output-only system identification are quantified by comparing the modal parameters extracted from the output-only models with those extracted from the input–output models.

## 2. THEORY OF SUBSPACE SYSTEM IDENTIFICATION

### 2.1. Problem statement: system identification of a state-space model

Consider a combined deterministic–stochastic discrete-time state-space model (Figure 1(a)) as

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k \quad (1)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{v}_k \quad (2)$$

where  $\mathbf{u}_k \in \mathbf{R}^m$  is a vector of  $m$  measured inputs at time step  $k$ ,  $\mathbf{y}_k \in \mathbf{R}^l$  is a vector of  $l$  measured outputs at time step  $k$ , and  $\mathbf{x}_k \in \mathbf{R}^n$  is an  $n$ -dimensional unknown discrete state vector. The model considers two additional stochastic processes: process,  $\mathbf{w}_k \in \mathbf{R}^n$ , and measurement noise,  $\mathbf{v}_k \in \mathbf{R}^l$ . Assuming that  $\mathbf{w}_k$  and  $\mathbf{v}_k$  are uncorrelated Gaussian zero-mean white noise processes, their covariance matrices are defined simply as

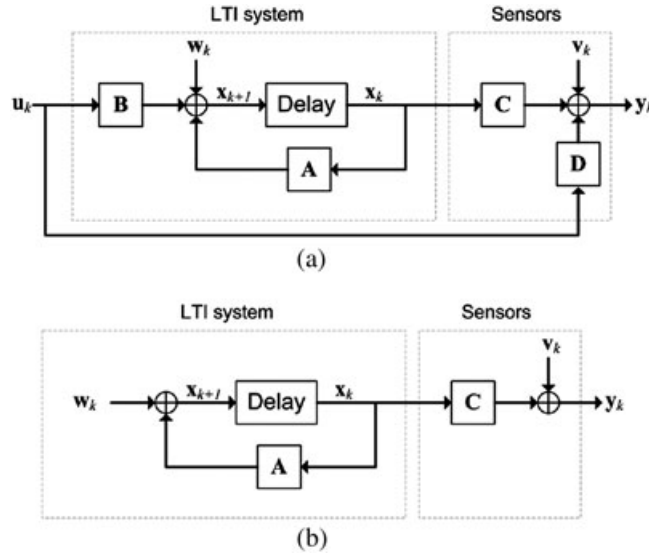


Figure 1. (a) Complete input–output state-space model for system identification; (b) state-space model for stochastic system identification.

$$E \left[ \begin{pmatrix} w_k \\ v_k \end{pmatrix} \begin{pmatrix} w_q^T & v_q^T \end{pmatrix} \right] := \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix} \delta_{kq} \tag{3}$$

where  $\mathbf{Q} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{S} \in \mathbb{R}^{n \times l}$ , and  $\mathbf{R} \in \mathbb{R}^{l \times l}$ . By assuming a linear system, the states and outputs in the model (Equations (1) and (2)) are split into deterministic and stochastic components as follows:

$$\mathbf{x}_k = \mathbf{x}_k^d + \mathbf{x}_k^s; \mathbf{y}_k = \mathbf{y}_k^d + \mathbf{y}_k^s \tag{4}$$

By considering Equations (1), (2), and (4), the deterministic subsystem is formulated as

$$\mathbf{x}_{k+1}^d = \mathbf{A} \mathbf{x}_k^d + \mathbf{B} \mathbf{u}_k; \mathbf{y}_k^d = \mathbf{C} \mathbf{x}_k^d + \mathbf{D} \mathbf{u}_k \tag{5}$$

Similarly, the stochastic subsystem is written as

$$\mathbf{x}_{k+1}^s = \mathbf{A} \mathbf{x}_k^s + \mathbf{w}_k; \mathbf{y}_k^s = \mathbf{C} \mathbf{x}_k^s + \mathbf{v}_k \tag{6}$$

If the stochastic state process,  $\mathbf{x}_k^s$ , is stationary, then the state covariance matrix,  $\Sigma$ , can be defined as

$$\Sigma := E \left[ \mathbf{x}_k^s \left( \mathbf{x}_k^s \right)^T \right] = \mathbf{A} \Sigma \mathbf{A}^T + \mathbf{Q} \in \mathbb{R}^{n \times n} \tag{7}$$

Now, the problem statement for the identification of a linear time-invariant (LTI) system can be stated as the optimal estimation of  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ ,  $\mathbf{D}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$ ,  $\mathbf{S}$  given the measured input sequence (i.e.,  $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{N-1}$ ) and output sequence (i.e.,  $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{N-1}$ ) as  $N \rightarrow \infty$  [13, 14].

### 2.2. Subspace state-space system identification family

The methods associated with 4SID are generally categorized into two groups: realization-based and direct 4SID methods [10]. Realization-based 4SID methods find their origins in the seminal work of Ho and Kalman [15] and offer a means of extracting state-space models from the extended observability matrix. At the core of the realization-based 4SID methods is the need for a reliable estimate of system impulse responses, often termed Markov parameters (MPs); the extended observability matrix is

estimated directly from the MPs. In contrast, direct 4SID methods, also referred to as data-driven subspace identification in the civil engineering community [16], strive to estimate a state-space model directly from an arbitrary set of input and output sequences (i.e., without requiring the estimation of system impulse response functions). Extensive research in the 1970s and 1980s (e.g., stochastic realization [17]) led to the establishment of numerous direct 4SID numerical algorithms such as the multivariable output-error state-space (MOESP) [14] and the numerical algorithms for subspace state-space system identification (N4SID) methods [13]. Generally, direct 4SID methods are simply referred to as subspace methods or subspace system identifications.

Figure 2(a) illustrates the geometric interpretation of the direct 4SID methods from which the name ‘subspace methods’ is derived. Given the measurements of the system input,  $\mathbf{U}$ , and output,  $\mathbf{Y}$ , two subspaces spanned by the input  $\mathbf{U}$  and the colored noise output  $\mathbf{Y}^s$  and one subspace spanned by the joint null space of the input and the colored noise output  $\{\mathbf{U}^\perp, \mathbf{Y}^{s\perp}\}$  are defined. The output  $\mathbf{Y}$  is placed at a specific point in the space. The colored noise output  $\mathbf{Y}^s$  is the system output from the unmeasured colored noise input. The measured system response comprises the forced and free vibration response plus a noise component. In this context,  $\mathbf{Y}$  can be divided into three orthogonal matrices ( $\mathbf{H}\mathbf{U}$ ,  $\mathbf{O}\mathbf{X}$ , and  $\mathbf{Y}^s$ ), which are parallel to the spans of  $\{\mathbf{U}\}$ ,  $\{\mathbf{U}^\perp, \mathbf{Y}^{s\perp}\}$ , and  $\{\mathbf{Y}^s\}$ , respectively. Here,  $\mathbf{O}$  is the extended observability matrix of the system, and  $\mathbf{H}$  is the lower block triangular Toeplitz matrix composed of deterministic MPs. Because the free vibration component of the system response is only dependent on the system dynamic properties, 4SID estimates the terms of  $\mathbf{O}\mathbf{X}$  (i.e., the product of the extended observability matrix and the state sequence) along the span of  $\{\mathbf{U}^\perp, \mathbf{Y}^{s\perp}\}$  (as denoted with the question mark in Figure 2(a)) from which the system state-space matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , and  $\mathbf{D}$  are extracted.

Subspace operations that estimate the term  $\mathbf{O}\mathbf{X}$  from the measured output are illustrated in Figure 2(b) and (c) for the MOESP and N4SID algorithms, respectively. In the MOESP method, the orthogonal projection of  $\mathbf{Y}$  onto the null space of the input is conducted first in order to remove the input dependence. Then, the coloredness of the system output is eliminated through the adoption of instrumental variables. In contrast, the two-step MOESP operation is combined to a one-step oblique projection in the N4SID method (Figure 2(c)). Essentially, the two subspace operations are identical (i.e., MOESP and N4SID) except for different weighting terms used in the two methods; hence, the product of the extended observability matrix and the state sequence ( $\mathbf{O}\mathbf{X}$ ) obtained in both methods are similar. The estimated product of  $\mathbf{O}$  and  $\mathbf{X}$  can be further simplified to extract a state-space model of the system. Specifically, singular value decomposition (SVD) of the estimated  $\mathbf{O}\mathbf{X}$  matrix can be performed. Once the appropriate system order is established by the singular values, a truncated extended observability matrix can be determined. The main difference between the MOESP and N4SID methods can be found in the numerical procedures used to estimate the

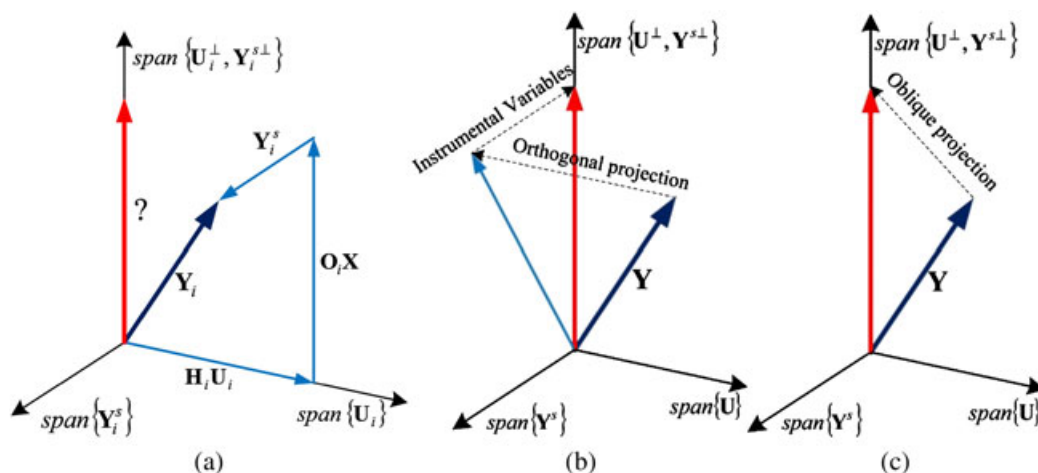


Figure 2. Geometric interpretation of subspace system identification: (a) definition of the three subspaces defined by the input–output measurements; (b) subspace operations in the MOESP algorithm; (c) subspace operations in the N4SID algorithm.

state-space system matrices (**A**, **B**, **C**, and **D**). In the MOESP method, system matrices **A** and **C** are first calculated from the estimated observability matrix. Then, the input system matrices **B** and **D** are calculated by linear regression from the estimated system matrices (**A**, **C**) and measured input–output data. In contrast, the N4SID method adopts a one-step estimation of the complete state-space model by solving a least square problem. To ensure the problem is over-determined, the system state sequence needs to be estimated first. Van Overschee and De Moor argue that the intermediate system state sequence is the state sequence of the Kalman filter [13]. Even though the two methods in the 4SID family are posed differently, the accuracy of both methods are reported to be comparable [18].

During the development of the aforementioned N4SID method, the stochastic identification problem (Figure 1(b)) played an important role because it established the concept of the Kalman filter state sequence and provided a least square solution for the determination of system matrices [19]. The stochastic identification problem can be stated as the estimation of **A**, **C**, **Q**, **R**, and **S** given the measured output sequence (i.e.,  $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{N-1}$  as  $N \rightarrow \infty$ ). Because the stochastic system identification problem can also be interpreted as an output-only system identification method under the assumption of an unmeasured white noise input, it has been widely applied to the identification of civil engineering structures excited by ambient vibrations [9]. After Peeters and Roeck's introduction of data-driven SSI to the civil engineering community, numerous researchers have adopted SSI for output-only modal analysis of structural system, thereby attaining greater accuracy compared with classical frequency domain methods [5, 20–22]. In this paper, the N4SID method will be exclusively used for the system identification of support-excited structures.

### 3. NUMERICAL ALGORITHMS FOR SUBSPACE STATE-SPACE SYSTEM IDENTIFICATION

Numerical algorithms for 4SID is considered for input–output system identification (i.e., combined deterministic–stochastic subspace identification) [23] of support-excited structures. This general case can be further specialized for output-only system identification (i.e., SSI) by ignoring the input system matrices **B** and **D** that represent the deterministic subsystem. A convenience of using N4SID is the availability of a MATLAB function (i.e., `n4sid`) for its execution [24]. A more detailed description of the N4SID algorithm can be found from the works of Van Overschee and De Moor [13, 23].

#### 3.1. Least square solution for state-space model estimation

An input–output discrete-time state-space model is considered with unknown state, **X**, measurable input, **U**, and output, **Y**, sequences as

$$\mathbf{X}_{i+1} = \mathbf{A}\mathbf{X}_i + \mathbf{B}\mathbf{U}_{i|i} + \mathbf{W} \quad (8)$$

$$\mathbf{Y}_{i|i} = \mathbf{C}\mathbf{X}_i + \mathbf{D}\mathbf{U}_{i|i} + \mathbf{V} \quad (9)$$

where the state sequence consist of  $j$  sequential state vectors

$$\mathbf{X}_i = [\mathbf{x}_i \quad \mathbf{x}_{i+1} \quad \mathbf{x}_{i+2} \quad \cdots \quad \mathbf{x}_{i+j-1}] \quad (10)$$

Similarly, the input and output sequences are also composed from the system measurements

$$\mathbf{U}_{i|i} = [\mathbf{u}_i \quad \mathbf{u}_{i+1} \quad \mathbf{u}_{i+2} \quad \cdots \quad \mathbf{u}_{i+j-1}] \quad (11)$$

$$\mathbf{Y}_{i|i} = [\mathbf{y}_i \quad \mathbf{y}_{i+1} \quad \mathbf{y}_{i+2} \quad \cdots \quad \mathbf{y}_{i+j-1}] \quad (12)$$

**W** and **V** are zero-mean Gaussian random sequences corresponding to the process noise and measurement noise, respectively. Combination of Equations (8) and (9) leads to

$$\begin{bmatrix} \mathbf{X}_{i+1} \\ \mathbf{Y}_{i|i} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} \mathbf{X}_i + \begin{bmatrix} \mathbf{B} \\ \mathbf{D} \end{bmatrix} \mathbf{U}_{i|i} + \begin{bmatrix} \mathbf{W} \\ \mathbf{V} \end{bmatrix} \tag{13}$$

If the two state sequence estimates  $(\hat{\mathbf{X}}_{i+1}, \hat{\mathbf{X}}_i)$  are given, then the system matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , and  $\mathbf{D}$  are produced by the least square solution as

$$\begin{bmatrix} \hat{\mathbf{A}} & \hat{\mathbf{B}} \\ \hat{\mathbf{C}} & \hat{\mathbf{D}} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{X}}_{i+1} \\ \mathbf{Y}_{i|i} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{X}}_i \\ \mathbf{U}_{i|i} \end{bmatrix}^\dagger \tag{14}$$

where ‘ $\dagger$ ’ is notation for the pseudo-inverse. As a result, subspace system identification algorithms seek to first estimate the state sequence using the measured system input and output sequences without requiring *a priori* knowledge of the system matrices.

### 3.2. Oblique projection by LQ decomposition

Because subspace system identification deals with batch data (i.e., sequential data), the Hankel matrix is introduced. A block Hankel matrix is constructed from  $2i + j - 1$  sequences of the system output and partitioned as the past,  $\mathbf{Y}_p$ , and future,  $\mathbf{Y}_f$ , output as follows:

$$\mathbf{Y}_{0|2i-1} = \begin{bmatrix} \mathbf{y}_0 & \mathbf{y}_1 & \cdots & \mathbf{y}_{j-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{i-1} & \mathbf{y}_i & \cdots & \mathbf{y}_{i+j-2} \\ \mathbf{y}_i & \mathbf{y}_{i+1} & \cdots & \mathbf{y}_{i+j-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{2i-1} & \mathbf{y}_{2i} & \cdots & \mathbf{y}_{2i+j-2} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_{0|i-1} \\ \mathbf{Y}_{i|2i-1} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_p \\ \mathbf{Y}_f \end{bmatrix} \tag{15}$$

Analogously, the past and future input block Hankel matrices are defined as

$$\mathbf{U}_{0|2i-1} = \begin{bmatrix} \mathbf{U}_{0|i-1} \\ \mathbf{U}_{i|2i-1} \end{bmatrix} := \begin{bmatrix} \mathbf{U}_p \\ \mathbf{U}_f \end{bmatrix} \tag{16}$$

The past and future deterministic state sequences are also defined as

$$\mathbf{X}_p^d := \mathbf{X}_0^d = [\mathbf{x}_0^d \quad \mathbf{x}_1^d \quad \mathbf{x}_2^d \quad \cdots \quad \mathbf{x}_{j-1}^d] \tag{17}$$

$$\mathbf{X}_f^d := \mathbf{X}_i^d = [\mathbf{x}_i^d \quad \mathbf{x}_{i+1}^d \quad \mathbf{x}_{i+2}^d \quad \cdots \quad \mathbf{x}_{i+j-1}^d] \tag{18}$$

The oblique projection in Figure 2(c) can be numerically implemented using LQ decomposition. Householder transformations are preferred among the numerous LQ decomposition algorithms [25]. The LQ decomposition of the system input,  $\mathbf{U}$ , and output,  $\mathbf{Y}$ , is

$$\begin{bmatrix} \mathbf{U}_{0|2i-1} \\ \mathbf{Y}_{0|2i-1} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_p \\ \mathbf{U}_f \\ \mathbf{Y}_p \\ \mathbf{Y}_f \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{0|i-1} \\ \mathbf{U}_{i|i} \\ \mathbf{U}_{i+1|2i-1} \\ \mathbf{Y}_{0|i-1} \\ \mathbf{Y}_{i|i} \\ \mathbf{Y}_{i+1|2i-1} \end{bmatrix} \xrightarrow{\text{LQ}} \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{L}_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{L}_{31} & \mathbf{L}_{32} & \mathbf{L}_{33} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{L}_{41} & \mathbf{L}_{42} & \mathbf{L}_{43} & \mathbf{L}_{44} & \mathbf{0} & \mathbf{0} \\ \mathbf{L}_{51} & \mathbf{L}_{52} & \mathbf{L}_{53} & \mathbf{L}_{54} & \mathbf{L}_{55} & \mathbf{0} \\ \mathbf{L}_{61} & \mathbf{L}_{62} & \mathbf{L}_{63} & \mathbf{L}_{64} & \mathbf{L}_{65} & \mathbf{L}_{66} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \\ \mathbf{Q}_3^T \\ \mathbf{Q}_4^T \\ \mathbf{Q}_5^T \\ \mathbf{Q}_6^T \end{bmatrix} \tag{19}$$

where  $\mathbf{L}_{11}, \dots, \mathbf{L}_{66}$  are lower triangular matrices and  $\mathbf{Q}_1, \dots, \mathbf{Q}_6$  are orthogonal matrices. Because the null space of the colored noise output  $\mathbf{Y}^{s\perp}$  can be determined from the span of ‘past’ input and output

$\{U_p, Y_p\}$  (this is also related with the concept of instrumental variables), the oblique projection of  $Y$  onto the span of the joint null space  $\{U^\perp, Y^{s\perp}\}$  is considered as a projection of the future output  $Y_f$  onto the span of the past input and output  $\{U_p, Y_p\}$  and the null space of the future input. This projection is

$$P_i := Y_f / U_f^\perp \begin{pmatrix} U_p \\ Y_p \end{pmatrix} = L_{U_p} L_{11} Q_1^T + L_{Y_p} [L_{41} \ L_{42} \ L_{43} \ L_{44}] [Q_1 \ Q_2 \ Q_3 \ Q_4]^T \quad (20)$$

where  $L_{U_p}$  and  $L_{Y_p}$  are subspace weighting matrices corresponding to past input and past output, respectively, and determined by

$$\begin{bmatrix} L_{U_p} & L_{U_f} & L_{Y_p} \end{bmatrix} = \begin{bmatrix} L_{51} & L_{52} & L_{53} & L_{54} \\ L_{61} & L_{62} & L_{63} & L_{64} \end{bmatrix} \begin{bmatrix} L_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ L_{21} & L_{22} & \mathbf{0} & \mathbf{0} \\ L_{31} & L_{32} & L_{33} & \mathbf{0} \\ L_{41} & L_{42} & L_{43} & L_{44} \end{bmatrix}^{-1} \quad (21)$$

Another oblique projection  $P_{i-1}$  can be defined similarly as

$$P_{i-1} := Y_f^- / (U_f^-)^\perp \begin{pmatrix} U_p^+ \\ Y_p^+ \end{pmatrix} \quad (22)$$

where one block row has been added (superscript +) or deleted (superscript -) from the input and output matrices as defined respectively as

$$\begin{bmatrix} U_p^+ \\ U_f^- \end{bmatrix} := \begin{bmatrix} U_{0i} \\ U_{i+1|2i-1} \end{bmatrix} \quad (23)$$

$$\begin{bmatrix} Y_p^+ \\ Y_f^- \end{bmatrix} := \begin{bmatrix} Y_{0i} \\ Y_{i+1|2i-1} \end{bmatrix} \quad (24)$$

Thus, the oblique projection can be calculated as

$$P_{i-1} = L_{U_p^+} \begin{bmatrix} L_{11} & \mathbf{0} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} + L_{Y_p^+} \begin{bmatrix} L_{41} & L_{42} & L_{43} & L_{44} & \mathbf{0} \\ L_{51} & L_{52} & L_{53} & L_{54} & L_{55} \end{bmatrix} [Q_1 \ Q_2 \ Q_3 \ Q_4 \ Q_5]^T \quad (25)$$

where both subspace weighting matrices are calculated as

$$\begin{bmatrix} L_{U_p^+} & L_{U_f^-} & L_{Y_p^+} \end{bmatrix} = \begin{bmatrix} L_{61} & L_{62} & L_{63} & L_{64} & L_{65} \end{bmatrix} \begin{bmatrix} L_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ L_{21} & L_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ L_{31} & L_{32} & L_{33} & \mathbf{0} & \mathbf{0} \\ L_{41} & L_{42} & L_{43} & L_{44} & \mathbf{0} \\ L_{51} & L_{52} & L_{53} & L_{54} & L_{55} \end{bmatrix}^{-1} \quad (26)$$

### 3.3. Span of the Kalman filter state sequence

Before exploring the oblique projections,  $P_i$  and  $P_{i-1}$ , the nonsteady state Kalman filter state estimate (i.e., the optimal state estimate based on the measured input and output data) is first reviewed [26]. Given the initial state estimate ( $\hat{x}_0$ ), initial stochastic state covariance ( $\Sigma_0$ ) noise covariances ( $Q, R$ ,



and  $\mathbf{S}$ ), input sequence  $(\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{k-1})$ , and output sequence,  $(\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{k-1})$ , along with a model of an LTI system  $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ , then the nonsteady state Kalman filter state estimate at time step  $k$ ,  $\hat{\mathbf{x}}_k$ , is calculated as

$$\hat{\mathbf{x}}_k = \mathbf{A} \hat{\mathbf{x}}_{k-1} + \mathbf{B} \mathbf{u}_{k-1} + \mathbf{K}_{k-1} (\mathbf{y}_{k-1} - \mathbf{C} \hat{\mathbf{x}}_{k-1} - \mathbf{D} \mathbf{u}_{k-1}) \tag{27}$$

where the Kalman gain,  $\mathbf{K}_{k-1}$ , is determined based on the stochastic state covariance,  $\Sigma_{k-1}$ . The Kalman filter state estimate can be also written in matrix form as

$$\hat{\mathbf{x}}_k = [\mathbf{A} - \mathbf{K}_{k-1} \mathbf{C} \mid \mathbf{B} - \mathbf{K}_{k-1} \mathbf{D} \mid \mathbf{K}_{k-1}] [\hat{\mathbf{x}}_{k-1}^T \mid \mathbf{u}_{k-1}^T \mid \mathbf{y}_{k-1}^T]^T \tag{28}$$

An alternative recursive form of Equation (28) is written from the initial time step 0 to step  $k$  as

$$\hat{\mathbf{x}}_k = [\mathbf{A}^k - \Omega_k \mathbf{O}_k \mid \Delta_k^d - \Omega_k \mathbf{H}_k^d \mid \Omega_k] [\hat{\mathbf{x}}_0^T \mid \mathbf{u}_0^T \cdots \mathbf{u}_{k-1}^T \mid \mathbf{y}_0^T \cdots \mathbf{y}_{k-1}^T]^T \tag{29}$$

where the modified Kalman gain  $\Omega_k := (\Delta_k^s + \mathbf{A}^k \Sigma_0 \mathbf{O}_k^T) (\mathbf{L}_k^s + \mathbf{O}_k \Sigma_0 \mathbf{O}_k^T)^{-1}$ , the extended observability matrix  $\mathbf{O}_k := [\mathbf{C}^T \ (\mathbf{C}\mathbf{A})^T \ (\mathbf{C}\mathbf{A}^2)^T \ \cdots \ (\mathbf{C}\mathbf{A}^{k-1})^T]^T$ , the reversed extended controllability matrix  $\Delta_k^d := [\mathbf{A}^{k-1} \mathbf{B} \ \cdots \ \mathbf{A} \mathbf{B} \ \mathbf{B}]$ , the reversed extended stochastic controllability matrix  $\Delta_k^s := [\mathbf{A}^{k-1} \mathbf{G} \ \cdots \ \mathbf{A} \mathbf{G} \ \mathbf{G}]$ , the stationary stochastic state and output covariance matrix  $\mathbf{G} := E[\mathbf{x}_k^s (\mathbf{y}_k^s)^T]$  and the limit for the past stochastic output  $\mathbf{L}_k^s := \lim_{j \rightarrow \infty} 1/j \mathbf{y}_p^s (\mathbf{y}_p^s)^T$  are all included. A sequence of the future Kalman filter state estimate vectors can be expressed as

$$\hat{\mathbf{X}}_f := [\hat{\mathbf{x}}_i \ \hat{\mathbf{x}}_{i+1} \ \hat{\mathbf{x}}_{i+2} \ \cdots \ \hat{\mathbf{x}}_{i+j-1}] \tag{30}$$

Substituting Equation (29) into Equation (30) leads to

$$\hat{\mathbf{X}}_f = [\mathbf{A}^i - \Omega_i \mathbf{O}_i \mid \Delta_i^d - \Omega_i \mathbf{H}_i^d \mid \Omega_i] \begin{bmatrix} \hat{\mathbf{X}}_0 \\ \mathbf{U}_{0|i-1} \\ \mathbf{Y}_{0|i-1} \end{bmatrix} = [\mathbf{A}^i - \Omega_i \mathbf{O}_i \mid \Delta_i^d - \Omega_i \mathbf{H}_i^d \mid \Omega_i] \begin{bmatrix} \hat{\mathbf{X}}_p \\ \mathbf{U}_p \\ \mathbf{Y}_p \end{bmatrix} \tag{31}$$

This expression for the Kalman filter state estimate sequences is significant. It implies that the future Kalman filter state estimate sequence is a linear combination of the past state estimate sequence and the past input and output sequences. Namely, it is located in the subspace spanned by the past input and output. This means that the span of the Kalman filter state estimate sequence is identified.

### 3.4. Estimation of state sequences

Because the span of the Kalman filter state sequence is identified, the Kalman filter state sequence is geometrically calculated by projections. First, a projection of the future output sequence to the subspace of past input and output  $\{\mathbf{U}_p, \mathbf{Y}_p\}$  eliminates the stochastic term,  $\mathbf{Y}_f^s$ :

$$\mathbf{P}'_i = \mathbf{O}_i \hat{\mathbf{X}}_i + \mathbf{H}_i^d \mathbf{U}_f \tag{32}$$

Then, a projection of  $\mathbf{P}'_i$  to the null space of the future input  $\mathbf{U}_f$  (i.e., Equation (20)) results in

$$\mathbf{P}_i = \mathbf{O}_i \hat{\mathbf{X}}_i \tag{33}$$

SVD can be applied to factorize Equation (33) as



$$\mathbf{P}_i \cong [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} = \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^T \quad (34)$$

Thus, the Kalman filter state estimate sequence can be calculated as

$$\hat{\mathbf{X}}_i = \mathbf{S}_1^{1/2} \mathbf{V}_1^T \quad (35)$$

Now, the Kalman filter state sequences are estimated directly from the input and output data sequences without *a priori* knowledge of system matrices by the geometrical interpretation of the batch implementation of Kalman filtering.

Similarly, the oblique projection  $\mathbf{P}_{i-1}$  is calculated with one-step shifted data sequences as

$$\mathbf{P}_{i-1} = \mathbf{O}_{i-1} \hat{\mathbf{X}}_{i+1} \quad (36)$$

As a result, the one-step shifted Kalman filter state estimate can be calculated as

$$\hat{\mathbf{X}}_{i+1} = \mathbf{O}_{i-1}^\dagger \mathbf{P}_{i-1} \quad (37)$$

where  $\mathbf{O}_{i-1}$  is the notation of  $\mathbf{O}_i$  (in Equation (33)) without the last block row. Based on the two estimated state sequences,  $\hat{\mathbf{X}}_i$  and  $\hat{\mathbf{X}}_{i+1}$ , the state-space system matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , and  $\mathbf{D}$  can be estimated by a least square solution in Equation (14). Finally, the noise covariance matrices are estimated from the residuals of Equation (13):

$$\begin{bmatrix} \hat{\mathbf{Q}} & \hat{\mathbf{S}} \\ \hat{\mathbf{S}}^T & \hat{\mathbf{R}} \end{bmatrix} = \frac{1}{j} \begin{bmatrix} \boldsymbol{\rho}_w \\ \boldsymbol{\rho}_v \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho}_w^T & \boldsymbol{\rho}_v^T \end{bmatrix} \quad (38)$$

The estimated system matrices represent a black-box state-space model for the system by input–output identification. Now, with an algorithmic foundation for the N4SID algorithm, black-box state-space models will be estimated for support-excited (i.e., seismically excited) structures.

## 4. SYSTEM IDENTIFICATION OF SUPPORT-EXCITED STRUCTURES

### 4.1. Testbed structure and support excitations

A single-bay, steel frame structure (Figure 3) is constructed on a large shaking table at the National Center for Research in Earthquake Engineering (NCREE) at National Taiwan University (NTU). The structure is a partial-scale six-story single-bay steel frame building with an interstory height of 1 m and a total building height of 6 m. Each story consists of four steel columns supporting a heavy steel plate floor (1 m by 1.5 m). Because of the rectangular cross section of the columns, the structure has two orthogonal axes: a flexurally weak axis (termed the *x*-axis) and a flexurally strong axis (termed the *y*-axis). Because system identification of the structure focuses on the estimation of a black-box state-space model, a more detailed physical description of the structure is not necessary.

To study the behavior of the test structure under seismic base motion, the shaking table is used to excite the structure in two lateral directions. In total, 20 accelerometers are installed (Figure 3(b)). Each floor is instrumented with three accelerometers: two oriented in the *x*-axis (denoted as *xa* and *xb*) and one oriented in the *y*-axis (denoted as *y*). Additionally, two accelerometers are installed in two orthogonal directions at the ground level to measure the support motion in two orthogonal directions (denoted as *x* base and *y* base). Two moderate base excitations are applied to both axes of the shaking table. First, the 40 s-long El Centro (1940) earthquake motion is applied to the two table axes simultaneously. The peak ground accelerations are scaled to 0.053 *g* and 0.044 *g* for the *x*-axis and the *y*-axis, respectively. Second, a 120 s-long white noise base motion is applied with

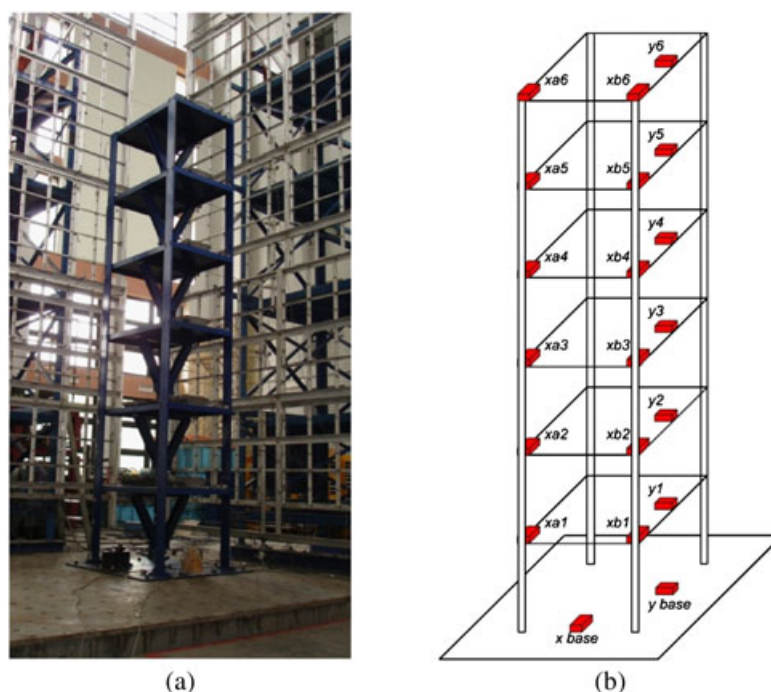


Figure 3. Large-scale six-story steel frame building structure: (a) perspective view of test structure on the NCREE shake table; (b) schematic of the sensor installation (a total of 20 accelerometers installed).

variances of  $0.061\text{ g}$  and  $0.0745\text{ g}$  in the  $x$ -axis and the  $y$ -axis, respectively. During the tests, the lateral acceleration of the building is measured at each floor with a 100-Hz sampling rate. A fourth-order Butterworth anti-aliasing filter with a 25-Hz cut-off is utilized on each data channel.

#### 4.2. Input–output black-box model estimation and evaluation

Subspace system identification by the N4SID method is conducted off-line using 30-s segments of the input (i.e., base acceleration) and output (i.e., floor acceleration) for the El Centro and white noise excitation tests. Implementation of the N4SID algorithm entails *a priori* determination of two user-defined parameters: (1) the number of block rows of the past input or output data in the block Hankel matrix (i.e., the size of the instrumental variables or  $i$  in Equation (15)) and (2) the model order (i.e., the dimension of the Kalman filter state sequence or the number of significant singular values in Equation (34)). Theoretically, the number of block rows of the past input or output data should be the same or larger than the model order [23]. The default setting of  $i$  is 2 times the model order in the `n4sid` MATLAB function [24]. Hence, the selection of the model order will be the most significant factor in determining the quality of the final system identification results. In the control theory community, the model order is determined using a residual error analysis of model predictions [6]. However, the civil engineering community commonly uses stabilization diagram to determine the model order. In this study, both approaches are adopted to determine the correct model order. Input–output N4SID analyses, followed by modal parameter estimation are conducted with the white noise test data sets for eight different model orders ranging from 18 to 60 with a model order increment of 6. Then, structural modes with small modal damping ratios (i.e., less than 3%) are selected, and results are plotted in Figure 4(a, b). After a model order of 36, 12 modes consistently appear with reasonable and stable damping ratios. Thus, the final model order of 36 is determined based on the stabilization diagram results of Figure 4.

To validate the estimated black-box model, the model's ability to reproduce the output signal of the system for a given input signal is checked. Figure 5 displays the predicted system output (i.e., the floor accelerations) from the black-box model compared to that measured. A very close match is discovered for both the white noise and El Centro tests with the deterministic predicted output signal in strong

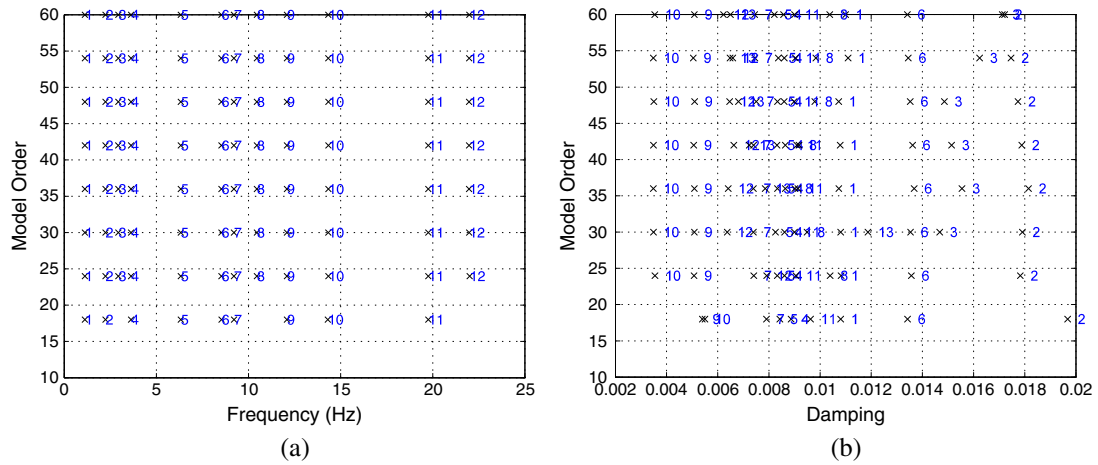


Figure 4. Stabilization plots of the input-output system identification study using N4SID on the white noise excited test data: (a) frequency stabilization plot; (b) damping ratio stabilization plot. The numbers in the plots denote modes tabulated in Table II.

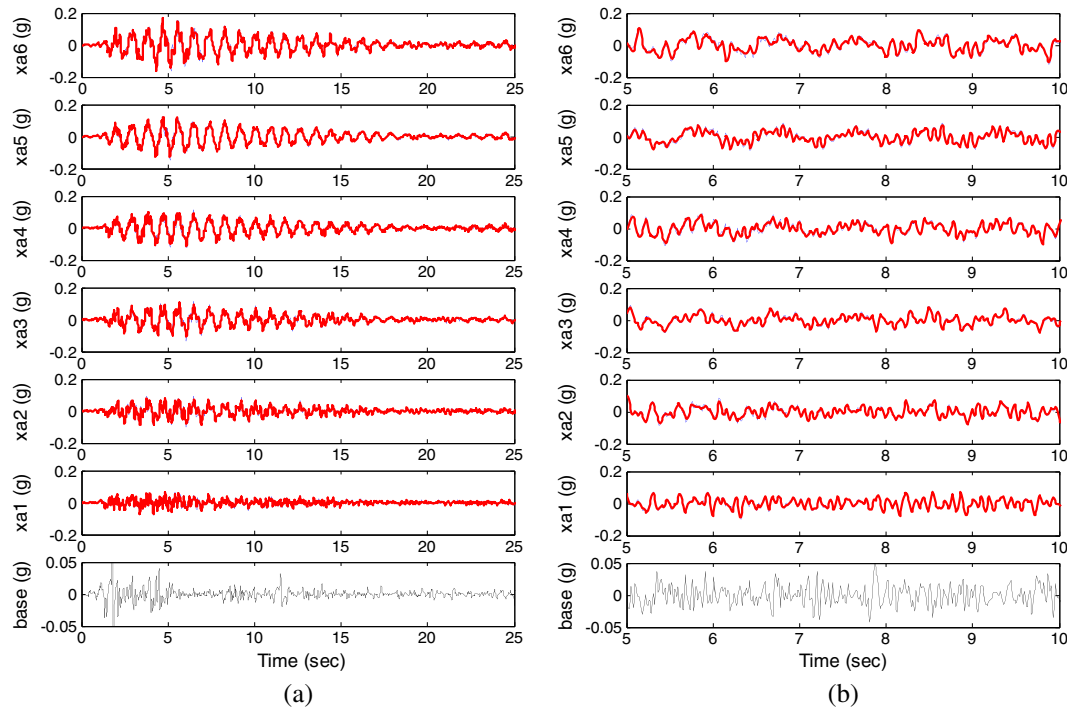


Figure 5. Comparison plots of the measured (blue thin) versus predicted (red thick) floor accelerations in the  $x$ -direction: (a) El Centro base motion; (b) white noise base motion. For completeness, the base acceleration is also shown for each test.

agreement with the measured signal. The small differences encountered between the predicted and measured accelerations seem to be random along the time axis and are considered as innovation processes or random residuals of the prediction. To confirm this belief, stochastic signal analysis is conducted on the prediction residual. Ideally, the residual is independent of the system input and is a Gaussian process with zero mean. This ideal condition can be checked by using the autocorrelation function of the residual and the cross-correlation function between each degree-of-freedom residual and the system input. Figure 6 plots the autocorrelation and cross-correlation functions for the El

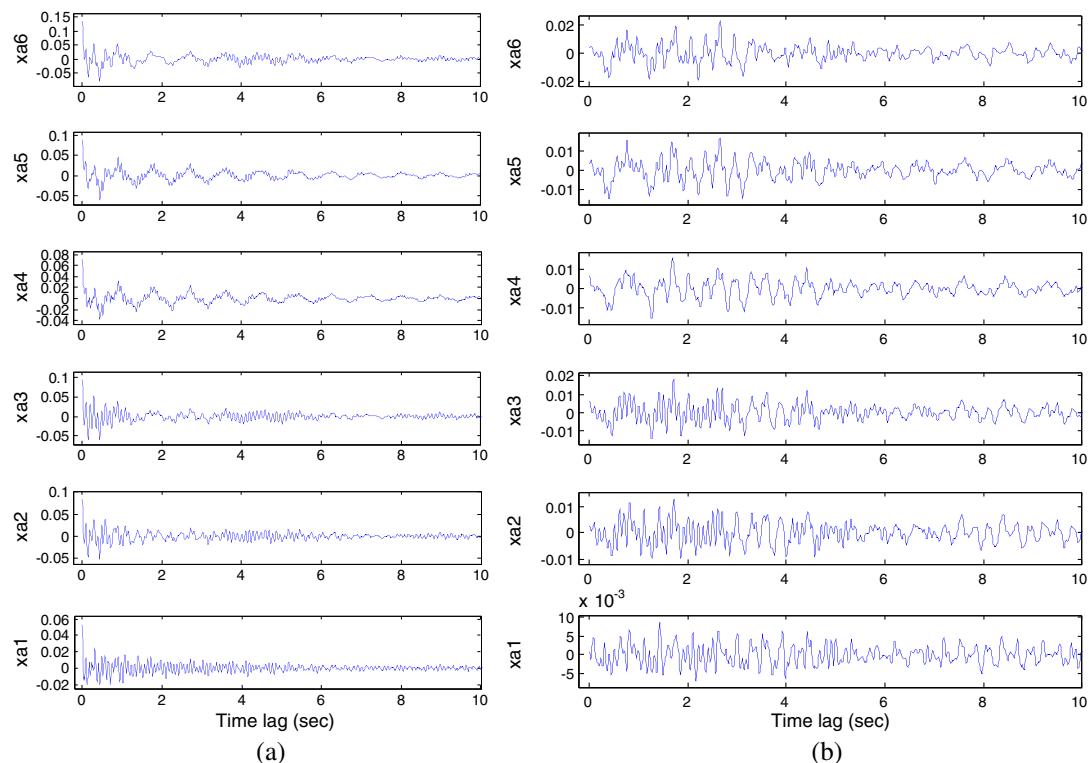


Figure 6. Stochastic residual analysis for the El Centro base motion: (a) autocorrelations of the model residuals; (b) cross-correlations of the model output residual and the applied base excitation (in the same direction as the output).

Centro test. In the case of the autocorrelation function (Figure 6(a)), the consistent high peak at the zero time lag implies strong whiteness of the model residual at each degree of freedom. After the zero time lag, a small-value bounded signal reveals a slight degree of coloredness to the model residual. Furthermore, a small arbitrary signal along the time lag axis is found in the cross-correlation function (Figure 6(b)), confirming that the residual is not correlated to the input signal.

#### 4.3. Output-only black-box model estimation

The deterministic subsystem (Equation (5)) is ignored to apply SSI to the system identification problem. The projection operation by LQ decomposition with the output data block Hankel matrix (i.e., Equation (15)) and its SVD lead to estimates of the Kalman filter sequences. Then, a black-box model consisting of the system matrices  $\mathbf{A}$  and  $\mathbf{C}$  is calculated by the least square method

$$\begin{bmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{C}} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{X}}_{i+1} \\ \mathbf{Y}_{i|i} \end{bmatrix} \hat{\mathbf{X}}_i^\dagger \quad (39)$$

Being similar to input–output subspace identification, SSI by the N4SID is conducted off-line again using 30-s segments of the measured floor accelerations from both base excitations (i.e., white noise and El Centro). The appropriate model order is again assessed using stabilization diagrams. As shown in Figure 7(a), the stabilization plot of system natural frequencies suggests model orders of 30 or larger to fully capture the system dynamics. However, unlike the stabilization plot of damping ratio for the input–output analysis (Figure 4(b)), the plot is less stable even for cases of high model orders. In addition, the actual damping ratios for each mode do not compare well to those found for the input–output model. With higher model orders not resolving the lack of stability in the damping ratio, a final model order of 36 is selected in order to be consistent with the models formulated for the

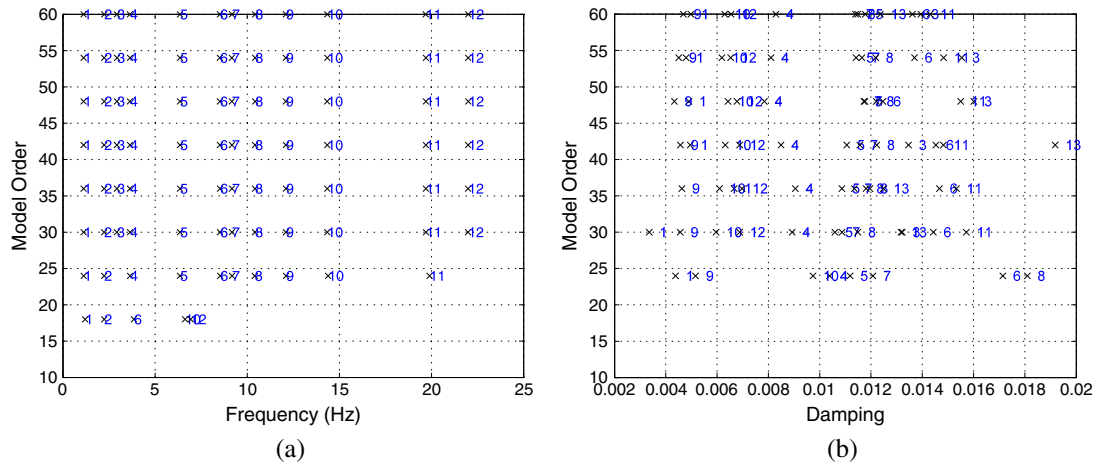


Figure 7. Stabilization plots of the output-only system identification study using N4SID on the white noise excited test data: (a) frequency stabilization plot; (b) damping ratio stabilization plot. The numbers in figures denote modes tabulated in Table II.

input–output case. The lack of consistency in the estimated damping parameters has been reported in the literature for covariance-based stochastic subspace identification (cov-SSI). For example, Reynders *et al.* [27] methodically explored variations in estimated system modal parameters based on perturbations to the output covariance. The uncertainty bound (with a 95% reliability) on the modal parameters from the covariance-based output-only system identification were calculated under the assumption of an ideal white noise input. Unlike the uncertainty bound of modal frequency, that of damping ratio was found to be 360% for the output-only analysis of a steel transmitter mast excited under wind loads.

5. COMPARISON OF INPUT–OUTPUT AND OUTPUT-ONLY BLACK-BOX MODELS

The black-box model generated by the output-only identification analysis (i.e., SSI) is difficult to be verified by prediction methods largely because of the absence of a deterministic system input. Hence, a comparison study between the output-only and input–output models is conducted by comparing the modal parameters derived from each model. Henceforth, the input–output model will be utilized as a baseline to which the output-only model will be compared. Towards this end, modal parameters (e.g., mode shapes, natural frequencies, and damping ratios) will be estimated from both the input–output and output-only black-box models. The modal parameters are extracted from the estimated system matrices,  $\hat{A}$  and  $\hat{C}$ . The estimated system matrix  $\hat{A}$  can be decomposed by eigen-decomposition as  $\hat{A} = \Psi \Lambda \Psi^{-1}$ , where  $\Psi \in C^{2n \times 2n}$  is the eigenvector matrix and  $\Lambda = \text{diag}(\lambda_{di}) \in C^{2n \times 2n}$  ( $i = 1, \dots, 2n$ ) is the diagonal eigenvalue matrix. The matrix  $\Lambda$  contains the  $n$  discrete-time eigenvalues  $\lambda_{di}$  of which the complex conjugated pairs are directly related to the frequency and damping properties of the structure. The discrete-time eigenvalues are first converted to continuous-time eigenvalues  $\lambda_{ci}$  as  $\lambda_{ci} = \ln(\lambda_{di}) / \Delta t$ , where  $\Delta t$  is the time step of the digital data acquisition system. The natural frequencies  $\omega_{ni}$  and damping ratios  $\zeta_i$  can then be easily calculated from the conjugate pair of complex-valued eigenvalues:  $\lambda_{ci}, \lambda_{ci}^* = -\zeta_i \omega_{ni} \pm j \omega_{ni} \sqrt{1 - \zeta_i^2}$ . The mode shape vector for the  $i$ th mode  $\Phi_i \in C^n$  can be calculated as  $[\Phi_1 \ \dots \ \Phi_i \ \dots \ \Phi_n] = \hat{C} \Psi$ . A comparison of the modal parameters extracted from the input–output and output-only models is tabulated in Tables I and II for the El Centro and white noise tests, respectively. Mode shapes for the El Centro test (both input–output and output-only models) are depicted in Figure 8. In total, 12 modes are reliably identified for both the models derived from the El Centro and the white noise excitations. Modes 1, 4, 5, 7, 9, and 10 are clearly flexural modes in the  $x$ -direction, whereas Modes 2, 6, and 11 are flexural modes in the  $y$ -direction. Modes 3, 8, and 12 are pure torsional modes.

Table I. Comparisons of estimated modal parameters for the El Centro test from input–output identification and output-only identification with 30 s-long data segment.

Mode number	Natural frequencies (Hz)			Damping ratio			MAC value
	Input–output	Output only	Error (%)	Input–output	Output only	Error (%)	
Mode 1	1.115	1.125	0.880	0.013	0.006	57.477	0.999
Mode 2	2.206	2.192	0.643	0.019	0.008	57.384	0.037
Mode 3	3.056	3.198	4.650	0.062	0.051	16.855	0.927
Mode 4	3.626	3.630	0.123	0.010	0.010	0.558	0.990
Mode 5	6.324	6.310	0.227	0.009	0.017	85.510	0.998
Mode 6	8.454	8.333	1.432	0.015	0.026	70.177	0.996
Mode 7	9.225	9.254	0.323	0.009	0.011	19.914	0.989
Mode 8	10.270	10.240	0.286	0.011	0.015	30.130	0.997
Mode 9	12.122	12.055	0.556	0.004	0.009	152.256	0.861
Mode 10	14.329	14.337	0.052	0.004	0.004	3.457	0.963
Mode 11	19.544	19.583	0.202	0.010	0.023	121.802	0.832
Mode 12	21.870	21.812	0.262	0.016	0.016	3.332	0.971
Mean			0.803	0.015	0.016	51.571	0.880

Numbers in italics indicate anomalous results.  
MAC, modal assurance criteria.

Table II. Comparisons of estimated modal parameters for the white noise test from input–output identification and output-only identification with 30 s-long data segment.

Mode number	Natural frequencies (Hz)			Damping ratio			MAC value
	Input–output	Output only	Error (%)	Input–output	Output only	Error (%)	
Mode 1	1.123	1.121	0.011	0.011	0.005	52.094	1.000
Mode 2	2.243	2.232	0.018	0.018	0.030	67.774	0.999
Mode 3	2.937	2.931	0.013	0.013	0.013	3.464	0.228
Mode 4	3.627	3.639	0.009	0.009	0.009	0.894	1.000
Mode 5	6.327	6.352	0.008	0.008	0.010	23.874	1.000
Mode 6	8.535	8.525	0.013	0.013	0.015	9.703	0.999
Mode 7	9.208	9.169	0.007	0.007	0.011	50.648	0.998
Mode 8	10.452	10.421	0.009	0.009	0.012	27.985	0.981
Mode 9	12.090	12.092	0.005	0.005	0.004	14.979	0.954
Mode 10	14.335	14.343	0.003	0.003	0.006	69.438	0.990
Mode 11	19.758	19.701	0.009	0.009	0.015	60.688	0.999
Mode 12	21.975	21.987	0.006	0.006	0.007	13.804	0.996
Mean			0.009	0.009	0.011	32.946	0.929

Numbers in italics indicate anomalous results.

The output-only SSI algorithm might generate some estimation outliers because it is based on a stochastic analysis. In this study, anomalous outliers are identified when comparing the natural frequencies, damping ratios, and mode shapes (more specifically the modal assurance criteria also termed MAC) of the output-only models to those from the input–output models. Namely, modal parameters deemed as anomalously big or small appears in italics in Tables I and II. The output-only model derived from the white noise test data reveals stronger agreement (i.e., fewer outliers) with the input–output model. In contrast, the output-only models derived from the El Centro test data appear to have more anomalous results. This discovery is not surprising due to the narrow-band colored attributes of the El Centro excitation in the frequency domain. If it is assumed that the input–output model estimated from the white noise test data is the most accurate model, this input–output model can then serve as the baseline model to which all the other models can be compared. Therefore, it can then be concluded that the quality of the output-only models depends on the input quality (i.e., degree of coloredness) because the fundamental assumption of output-only identification is the excitation is white noise. However, it should be emphasized that even for colored narrow-band excitations (e.g., seismic excitations such as El Centro), the output-only model is still an excellent model capturing the behavior of the structure.



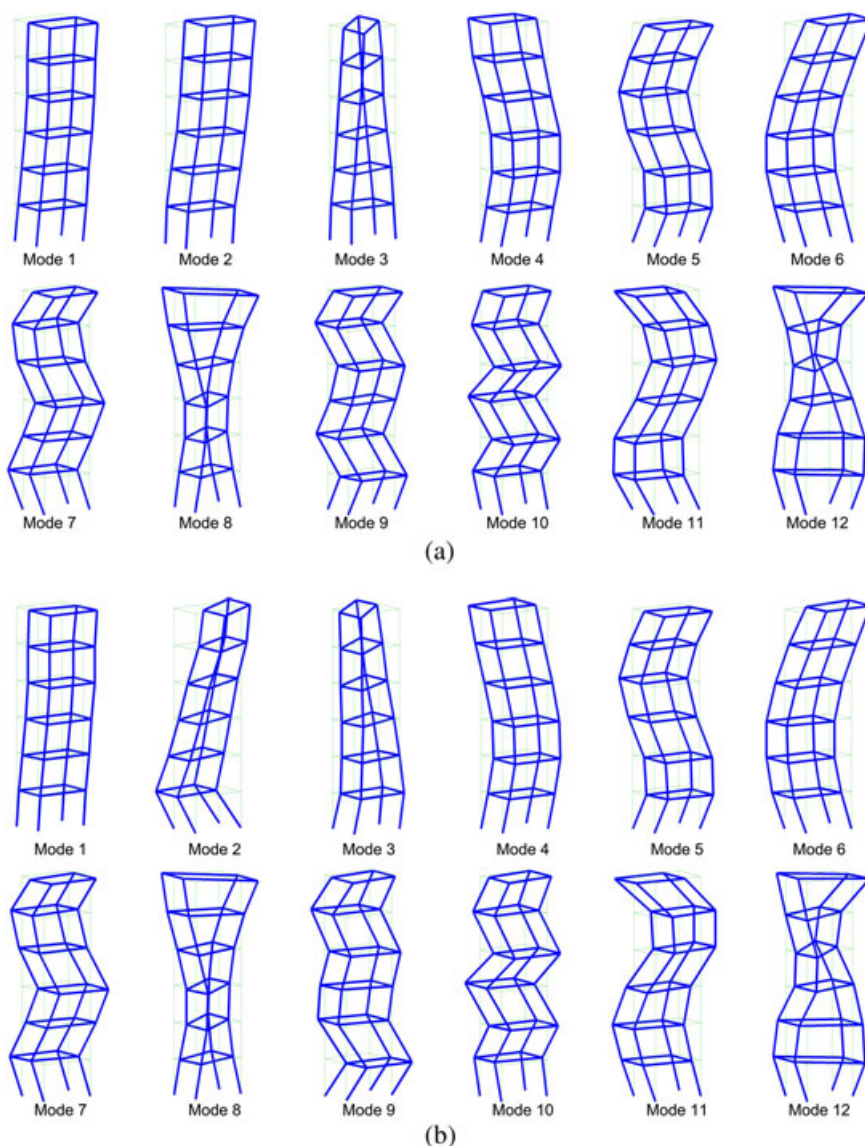


Figure 8. Mode shapes extracted from the system matrices  $\mathbf{A}$  and  $\mathbf{C}$  estimated from the El Centro test data: (a) input-output analysis; (b) output-only analysis.

## 6. CONCLUSIONS

This paper provides a detailed explanation of the N4SID subspace system identification algorithm. N4SID is a powerful estimation tool for formulating black-box state-space models using measurement data collected from a base excited structure. One of the most significant concepts in N4SID is the estimation of the nonstationary Kalman filter state sequence without *a priori* knowledge of the system matrices. To provide civil engineers with a deeper appreciation for this concept, a geometric explanation has been provided. In addition to the theoretical exploration of N4SID, application of the method to classical identification problems in structural engineering (i.e., system identification of support-excited structures) has been undertaken. Toward this end, a partial-scale six-story steel frame structure mounted on a shaking table (NCREE, National Taiwan University, Taiwan) was excited at its base using bidirectional seismic and white noise excitations. The accuracy of the estimated model using input-output data was evaluated deterministically (i.e., comparison of the estimated model's prediction and the measured response) and stochastically (i.e., correlation analysis of the residual error of the prediction); outstanding predictive capabilities were found with the model yielding strongly Gaussian residuals. The



accuracy of the output-only model was checked by comparing modal parameters of the output-only model to those of the input–output model derived from the white noise base excitation. It was confirmed that the estimated model by output-only identification using white noise test data provided the most accurate output-only model. The modal parameters of these models were in excellent agreement with those of the input–output model.

The subspace system identification methods presented in this study yield black-box models that can be used for accurate modeling of support-excited structures. However, one drawback of the models is that they are black-box models. This will prevent these models from being fully utilized for structural health monitoring (SHM). The companion paper to this paper will convert the N4SID models into ‘gray’-box models from which structural parameters can be extracted for SHM purposes [12].

#### ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the generous support offered by the National Science Foundation under grant CMMI-0726812 (program manager: Dr. S.C. Liu). The authors would also like to thank Prof. C. H. Loh of National Taiwan University and the National Center for Research in Earthquake Engineering (NCREE) technical staff for their support and guidance during experimental testing at the NCREE shake table facility.

#### REFERENCES

1. Abdelghani M, *et al.* Comparison study of subspace identification methods applied to flexible structures. *Mechanical Systems and Signal Processing* 1998; **12**(5):679–692.
2. Hermans L, Van Der Auweraer H. Modal testing and analysis of structures under operational conditions: industrial applications. *Mechanical Systems and Signal Processing* 1999; **13**(2):193–216.
3. Brownjohn JMW. Ambient vibration studies for system identification of tall buildings. *Earthquake Engineering and Structural Dynamics* 2003; **32**(1):71–95.
4. Yi JH, Yun CB. Comparative study on modal identification methods using output-only information. *Structural Engineering and Mechanics* 2004; **17**(3–4):445–466.
5. Weng JH, *et al.* Output-only modal identification of a cable-stayed bridge using wireless monitoring systems. *Engineering Structures* 2008; **30**: 1820–1830.
6. Ljung L. *System Identification: Theory for the User*. Prentice Hall: Upper Saddle River, NJ, 1999.
7. Richart FE, Hall JR, Wood RD. *Vibrations of Soils and Foundations*. Prentice Hall: Englewood Cliffs, NJ, 1970.
8. Chaudhary MTA, *et al.* System identification of two based-isolated bridges using seismic records. *Journal of Structural Engineering, ASCE* 2000; **126**(10):1187–1195.
9. Peeters B, Roeck GD. Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing* 1999; **13**(6):855–878.
10. Viberg M. Subspace-based methods for the identification of linear time-invariant systems. *Automatica* 1995; **31**(12):1835–1851.
11. Gevers M. A personal view on the development of system identification. In *13th Federation of Automatic Control Symposium on System Identification (IFAC SYSID)*. Rotterdam, Netherlands, 2003.
12. Kim J, Lynch JP. Subspace system identification of support excited structures—part ii: gray-box interpretations and damage detection. *Earthquake Engineering and Structural Dynamics* 2012. DOI: 10.1002/eqe.2185.
13. Van Overschee P, De Moor B. N4SID: subspace algorithms for the identification of combined deterministic–stochastic systems. *Automatica* 1994; **30**(1):75–93.
14. Verhaegen M. Identification of the deterministic part of MIMO state space models given in innovations form from input–output data. *Automatica* 1994; **30**(1):61–74.
15. Ho BL, Kalman RE. Effective construction of linear state-variable models from input–output functions. *Regelungstechnik* 1965; **12**:545–548.
16. Peeters B, Ventura CE. Comparative study of modal analysis techniques for bridge dynamic characteristics. *Mechanical Systems and Signal Processing* 2003; **17**(5):965–988.
17. Akaike H. Stochastic theory of minimal realization. *IEEE Transactions on Automatic Control* 1974; **19**(6):667–674.
18. Viberg M, *et al.* Performance of subspace-based system identification methods. In *12th Federation of Automatic Control Symposium on System Identification (IFAC SYSID)*. Sydney, Australia, 1993.
19. Van Overschee P, De Moor B. Subspace algorithm for the stochastic identification problem. *Automatica* 1993; **29**(3):649–660.
20. Arici Y, Mosalam KM. Modal identification of bridge systems using state-space methods. *Structural Control and Health Monitoring* 2005; **12**:381–404.
21. Yan AM, Golinval JC. Null subspace-based damage detection of structures using vibration measurements. *Mechanical Systems and Signal Processing* 2006; **20**(3):611–626.
22. Kim J, Lynch JP. Comparison study of output-only subspace and frequency-domain methods for system identification of base excited civil engineering structures. In *29th International Modal Analysis Conference (IMAC-XXIX)*. Jacksonville, FL, 2011.

23. Van Overschee P, De Moor B. *Subspace Identification for Linear Systems* Kluwer Academic Publishers: Dordrecht, Netherlands, 1996.
24. Ljung L. *System Identification Toolbox 7*. The Mathworks, 2009.
25. Golub GH, Van Loan CF. *Matrix Computations* The Johns Hopkins University Press: Baltimore, Maryland, 1996.
26. Kalman RE. New results in linear filtering and prediction problems. *Journal of Basic Engineering, ASME* 1960; **82**(Series D):95–108.
27. Reynders E, Pintelon R, De Roeck G. Uncertainty bounds on modal parameters obtained from stochastic subspace identification. *Mechanical Systems and Signal Processing* 2008; **22**:948–969.