Title
Discussion of user-defined parameters for recursive subspace identification: Application to seismic response of building structures

Journal
Earthquake Engineering & Structural Dynamics, 49(15)

ISSN
0098-8847

Authors
Huang, Shieh-Kung
Chen, Jun-Da
Loh, Kenneth J
et al.

Publication Date
2020-12-01

DOI
10.1002/eqe.3327

Peer reviewed

DISCUSSION OF USER-DEFINED PARAMETERS FOR RECURSIVE SUBSPACE IDENTIFICATION: APPLICATION TO SEISMIC RESPONSE OF BUILDING STRUCTURES

Shieh-Kung Huang[1], Jun-Da Chen[2], Kenneth J Loh[3], Chin-Hsiung Loh[2,3]

[1]National Center for Research on Earthquake Engineering, NARL, Taipei, Taiwan.

[2]Department of Civil Engineering, national Taiwan University, Taipei, Taiwan.

Department of Structural Engineering, University of California, San Diego, CA USA

Structural damage assessment under external loading, such as earthquake excitation, is an important issue in structural safety evaluation. In this regard, an appropriate data analysis and system identification technique is required to interpret the measured data and to identify the state of the structure. Generally, the recursive system identification algorithm is used. In this study, the recursive subspace identification algorithm based on the matrix inversion lemma algorithm with oblique projection technique (RSI-Inversion-Oblique) is applied to investigate the time-varying dynamic characteristics. The user-defined parameters used in the RSI-Inversion-Oblique technique are carefully discussed, which include the size of the data Hankel matrix (i), model order to extract the physical modes, and forgetting factor (FF) to detect the time-varying system modal frequencies. Response data from the Northridge earthquake from the Sherman Oaks building (CSMIP) is used as an example to examine a systematic method to determine the suitable user-defined parameters in RSI. It is concluded that the number of rows in the data Hankel matrix significantly influences the identification of the time-varying fundamental modal frequency of the structure. An algorithmic model order selection method using the eigenvalue distribution of RSI-Inversion can detect the system modal frequencies at each appending data window without causing any abnormality.

KEYWORDS

State space model, Recursive subspace identification, forgetting factor, building seismic response, time-varying modal frequencies.

INTRODUCTION

During the last two decades, state space subspace system identification (4SID) methods have attracted substantial interest in the system identification and control communities as the method can identify the system matrices of the state space model directly from the input and output data. Several well-known subspace identification algorithms including CVA, N4SID, MOESP, and IV-4SID [1-5] have been developed. Since the traditional subspace identification method is not suitable for online computations due to the computational complexity of singular value decomposition (SVD) and an inability to detect time-varying systems, several recursive subspace identification (RSI) algorithms have been proposed that involve the determination of a mathematical model representing the underlying time-varying dynamics. Kameyama et al. [6] proposed a recursive 4SID-based identification algorithm with fixed input-output data size and developed the RSI algorithm to avoid the use of repetitive LQ decomposition. The recursive update algorithm PI-MOESP (Past-Input/Past Output Multivariable Output-Error State sPace) was proposed by Tamaoki et al. [7] to determine the order and parameters of a time-varying system. The RSI-BonaFide algorithm was derived from LQ decomposition applied in PO-MOESP to estimate system matrices and modal parameters recursively at each time step. One of the applications of the RSI algorithm in the civil engineering field is to identify the dynamic characteristics of a structure during strong earthquake excitation and provide online tracking of the modal parameters of structures. Some preliminary studies have reported using the RSI algorithm to track structural modal parameters from a building's seismic response [8, 9].

For the application of subspace identification to building structural dynamic characteristics, it is necessary to use stabilization diagrams implemented with some criteria to remove spurious modes. These criteria include the mode shape assurance criterion, phase collinearity, etc. It will be time-consuming if the stabilization diagram and spurious modes removal criteria need to be applied at each recursive time step. Therefore, a further analysis on the development of the Hankel matrix as well as its projection analysis can be implemented in advance to remove the spurious modes. In this study, the following three user-defined parameters for RSI will be discussed: size of the Hankel matrix (i), model order selection, and forgetting factor. Through the building seismic response data, a detailed discussion on the determination of these three user-defined parameters in RSI will be presented.

BASIC THEORY OF RECURSIVE SUBSPACE IDENTIFICATION (rsi)

During the past decade, several recursive subspace identification algorithms have been studied. In the present study, a brief description is given of RSI using the matrix inversion lemma renewing method on orthogonal projection (RSI-Inversion-Orthogonal) as well as oblique projection (RSI-Inversion-Oblique). To derive matrix-input-output equations for subspace identification, input and output measurements should be arranged in the form of a "data Hankel matrix", i.e. $\mathbf{U}_p \in \mathbb{R}^{mi\times j}$ is defined as the "past input data Hankel matrix" and $\mathbf{U}_f \in \mathbb{R}^{mi\times j}$ is the "future input data Hankel matrix". Similarly, $\mathbf{Y}_p \in \mathbb{R}^{li\times j}$ and $\mathbf{Y}_f \in \mathbb{R}^{li\times j}$ are the "past output data Hankel matrix" and "future output data Hankel matrix", respectively. Here, m is the number of inputs, l is the number of outputs, n is the modal order of system, j is the number of columns, and i is the number of data points in each block row for each record in the data Hankel matrix, which is a user-defined parameter. Additionally, $WL = 2i + j - 1$ is the number of available samples in a Hankel matrix defined as the window length (WL). It is important to define an instrumental variable (IV) matrix $\mathbf{\Xi}_p$, which is crucial to subspace identification, consisting of both past input and past output data Hankel matrices:

$$\mathbf{\Xi}_p = \begin{bmatrix} \mathbf{U}_p \\ \mathbf{Y}_p \end{bmatrix} \in \mathbb{R}^{(m+l)i\times j}. \tag{1}$$

The original state-space model can be transformed into "Matrix Input-Output Equations" as [10]

$$\begin{aligned} \mathbf{Y}_p &= \mathbf{\Gamma}_i \cdot \mathbf{X}_p + \mathbf{H}_i \cdot \mathbf{U}_p + \mathbf{G}_i \cdot \mathbf{W}_p + \mathbf{V}_p \\ \mathbf{Y}_f &= \mathbf{\Gamma}_i \cdot \mathbf{X}_f + \mathbf{H}_i \cdot \mathbf{U}_f + \mathbf{G}_i \cdot \mathbf{W}_f + \mathbf{V}_f \\ \mathbf{X}_f &= \mathbf{A}_d^i \cdot \mathbf{X}_p + \mathbf{\Delta}_i \cdot \mathbf{U}_p \end{aligned} \tag{2}$$

with

$$\boldsymbol{\Delta}_i \equiv \begin{bmatrix} \boldsymbol{A}_d^{i-1}\boldsymbol{B}_d & \boldsymbol{A}_d^{i-2}\boldsymbol{B}_d & \cdots & \boldsymbol{A}_d\boldsymbol{B}_d & \boldsymbol{B}_d \end{bmatrix} \in \mathbb{R}^{2n\times mi}$$

$$\boldsymbol{\Gamma}_i \equiv \begin{bmatrix} \boldsymbol{C}_c \\ \boldsymbol{C}_c\boldsymbol{A}_d \\ \boldsymbol{C}_c\boldsymbol{A}_d^2 \\ \vdots \\ \boldsymbol{C}_c\boldsymbol{A}_d^{i-1} \end{bmatrix} \in \mathbb{R}^{li\times 2n}, \quad \boldsymbol{H}_i \equiv \begin{bmatrix} \boldsymbol{D} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{C}_c\boldsymbol{B}_d & \boldsymbol{D} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{C}_c\boldsymbol{A}_d\boldsymbol{B}_d & \boldsymbol{C}_c\boldsymbol{B}_d & \boldsymbol{D} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C}_c\boldsymbol{A}_d^{i-2}\boldsymbol{B}_d & \boldsymbol{C}_c\boldsymbol{A}_d^{i-3}\boldsymbol{B}_d & \boldsymbol{C}_c\boldsymbol{A}_d^{i-4}\boldsymbol{B}_d & \cdots & \boldsymbol{D} \end{bmatrix} \in \mathbb{R}^{li\times mi}$$

$$\boldsymbol{G}_i \equiv \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{C}_c & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{C}_c\boldsymbol{A}_d & \boldsymbol{C}_c & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C}_c\boldsymbol{A}_d^{i-2} & \boldsymbol{C}_c\boldsymbol{A}_d^{i-3} & \boldsymbol{C}_c\boldsymbol{A}_d^{i-4} & \cdots & \boldsymbol{0} \end{bmatrix} \in \mathbb{R}^{li\times 2ni} \tag{2a}$$

where $\mathbf{\Gamma}_i$ is the extended observability matrix that contains information of system matrices (i.e. $\mathbf{A}_d$ and $\mathbf{C}_c$) and is the primary outcome of the subspace identification. $\mathbf{X}_p$ and $\mathbf{X}_f$ is past and future state sequences. Two different geometric projection methods can be implemented in subspace identification and will be briefly introduced in this section: (a) orthogonal projection, and (b) oblique projection. One can choose either orthogonal or oblique projection for extracting the extended observability matrix. Based on the format of the projection matrix one can use SVD (Orthogonal projection) or ED (Oblique projection) to decide model order and the dimension of the extended observability matrix of the target system, from which the extended observability matrix $\mathbf{\Gamma}_i$ can be obtained.

Besides the offline subspace identification, the online system identification is also needed to detect the time-varying structural system. For the transformation from offline SI toward online RSI, two projection categories (orthogonal projection and oblique projection) can be used for conducting data-driven subspace identification.

RSI using the Matrix Inversion Lemma renewing method on Orthogonal Projection (RSI-Inversion-Orthogonal)

RSI-Inversion-Orthogonal is a projection matrix renewing algorithm mathematically based on matrix inversion lemma numerically expanding the inverse operation. The transformation from offline SI toward RSI is first achieved by computing the projection. Considering the matrix input-output equation $\mathbf{Y}_f$ (shown in Eq. (2)), the future input Hankel matrix can be eliminated by projecting the whole equation onto the orthogonal component of the future input matrix. The orthogonal projection matrix, $\mathbf{O}_{(k)}^{Orthogonal}$, is defined as

$$\mathbf{O}_{(k)}^{Orthogonal} = \mathbf{Y}_{f(k)} \mathbf{\Pi}_{\mathbf{U}_{f(k)}}^{\perp} \mathbf{\Xi}_{p(k)}^{T} . \tag{3}$$

where the subscript (k) indicated the k-th time step in recursive identification $\mathbf{\Pi}_{\mathbf{U}_{f(k)}}^{\perp}$ is a geometric operator that projects the row space of a matrix onto the orthogonal complement of the row space of the matrix $\mathbf{U}_{f(k)}$. Then, the resultant projection matrix can be derived from the above-mentioned procedure, which contains the target extended observability matrix.

$$\begin{aligned} \mathbf{O}_{(k)}^{Orthogonal} &= \mathbf{Y}_{f(k)} \mathbf{\Pi}_{\mathbf{U}_{f(k)}}^{\perp} \mathbf{\Xi}_{p(k)}^{T} \approx \mathbf{\Gamma}_i \mathbf{X}_{f(k)} \mathbf{\Pi}_{\mathbf{U}_{f(k)}}^{\perp} \mathbf{\Xi}_{p(k)}^{T} \\ &= \mathbf{Y}_{f(k)} \mathbf{\Xi}_{p(k)}^{T} - \mathbf{Y}_{f(k)} \mathbf{U}_{f(k)}^{T} \cdot (\mathbf{U}_{f(k)} \mathbf{U}_{f(k)}^{T})^{-1} \cdot \mathbf{U}_{f(k)} \mathbf{\Xi}_{p(k)}^{T} \end{aligned} \tag{3}$$

$$\begin{aligned} &= \left[\sum_{g=k-j+1}^{k} \mathbf{y}_{f(g)} \mathbf{\xi}_{p(g)}^{T} \right] - \left[\sum_{g=k-j+1}^{k} \mathbf{y}_{f(g)} \mathbf{u}_{f(g)}^{T} \right] \\ &\quad \times \left[\sum_{g=k-j+1}^{k} \mathbf{u}_{f(g)} \mathbf{u}_{f(g)}^{T} \right]^{-1} \times \left[\sum_{g=k-j+1}^{k} \mathbf{u}_{f(g)} \mathbf{\xi}_{p(g)}^{T} \right] \end{aligned} \tag{4}$$

where $\mathbf{y}_{f(g)}$, $\mathbf{u}_{f(g)}$, and $\mathbf{\xi}_{p(g)}$, are g-th column of the matrices $\mathbf{Y}_f$, $\mathbf{U}_f$, and $\mathbf{\Xi}_p$. The column space of $\mathbf{\Gamma}_i$ can be estimated by the column space of the projected matrix $\mathbf{O}_{(k)}^{Orthogonal}$, which can be obtained by SVD.

Then the enlarged window is used in RSI-Inversion-Orthogonal, which keeps appending new data points rather than using a fixed-length moving window; therefore, a forgetting factor can be introduced in this algorithm to eliminate the influence of previous measurements to identify the latest state of the system. An exponential forgetting factor λ (smaller than one) is added in the expanding projection matrix operation as mentioned above, which is multiplied in front of the data vector product at each time step.

$$\begin{aligned} \mathbf{O}_{(k)}^{Orthogonal} &= \left[\sum_{g=k-j+1}^{k} \lambda^{k-g} \cdot \mathbf{y}_{f(g)} \mathbf{\xi}_{p(g)}^{T} \right] - \left[\sum_{g=k-j+1}^{k} \lambda^{k-g} \cdot \mathbf{y}_{f(g)} \mathbf{u}_{f(g)}^{T} \right] \\ &\quad \times \left[\sum_{g=k-j+1}^{k} \lambda^{k-g} \cdot \mathbf{u}_{f(g)} \mathbf{u}_{f(g)}^{T} \right]^{-1} \times \left[\sum_{g=k-j+1}^{k} \lambda^{k-g} \cdot \mathbf{u}_{f(g)} \mathbf{\xi}_{p(g)}^{T} \right] \end{aligned} \tag{5}$$

For

$$\mathbf{R}_{(k+1)} = \left[\mathbf{U}_{f(k+1)} \mathbf{U}_{f(k+1)}^{T} \right]^{-1} = \left[\mathbf{U}_{f(k)} \mathbf{U}_{f(k)}^{T} + \mathbf{u}_{f(k+1)} \mathbf{u}_{f(k+1)}^{T} \right]^{-1} = \left[\mathbf{R}_{(k)}^{-1} + \mathbf{u}_{f(k+1)} \mathbf{u}_{f(k+1)}^{T} \right]^{-1}, \tag{6}$$

The matrix inversion lemma [11] can be implemented as

$$\begin{aligned} \left[\mathbf{R}_{(k)}^{-1} + \mathbf{u}_{f(k+1)} \mathbf{u}_{f(k+1)}^{T} \right]^{-1} &= \mathbf{R}_{(k)} - (1 + \mathbf{u}_{f(k+1)}^{T} \mathbf{R}_{(k)} \mathbf{u}_{f(k+1)})^{-1} \cdot \mathbf{R}_{(k)} \mathbf{u}_{f(k+1)} \mathbf{u}_{f(k+1)}^{T} \mathbf{R}_{(k)} \\ &= \mathbf{R}_{(k)} - \alpha_{(k+1)} \cdot \mathbf{R}_{(k)} \mathbf{u}_{f(k+1)} \mathbf{u}_{f(k+1)}^{T} \mathbf{R}_{(k)} \end{aligned} . \tag{6a}$$

Now given the new incoming data vectors $\mathbf{u}_{f(k+1)}$, $\mathbf{y}_{f(k+1)}$, and $\boldsymbol{\xi}_{p(k+1)}$, the following recursive computation elements can be constructed:

$$\begin{aligned}\alpha_{(k+1)} &= (\lambda + \mathbf{u}^T_{f(k+1)}\mathbf{R}_{(k)}\mathbf{u}_{f(k+1)})^{-1} \\ \boldsymbol{\beta}_{(k+1)} &= \mathbf{y}_{f(k+1)} - \mathbf{Y}_{f(k)}\mathbf{U}^T_{f(k)}\mathbf{R}_{(k)}\mathbf{u}_{f(k+1)} \\ \boldsymbol{\gamma}_{(k+1)} &= \boldsymbol{\Xi}_{p(k)}\mathbf{U}^T_{f(k)}\mathbf{R}_{(k)}\mathbf{u}_{f(k+1)} - \boldsymbol{\xi}_{p(k+1)}\end{aligned} \quad . \tag{7}$$

The orthogonal projection matrix $\mathbf{O}^{Orthogonal}_{(k)}$ is then updated to $\mathbf{O}^{Orthogonal}_{(k+1)}$, following Eq. (4), given by

$$\begin{aligned}\mathbf{O}^{Orthogonal}_{(k+1)} &= \mathbf{Y}_{f(k+1)}\boldsymbol{\Pi}^{\perp}_{\mathbf{U}_{f(k+1)}}\boldsymbol{\Xi}^T_{p(k+1)} \\ &= \lambda\cdot\mathbf{Y}_{f(k)}\boldsymbol{\Pi}^{\perp}_{\mathbf{U}_{f(k)}}\boldsymbol{\Xi}^T_{p(k)} - \lambda\cdot\alpha_{(k+1)}\cdot\boldsymbol{\beta}_{(k+1)}\cdot\boldsymbol{\gamma}^T_{(k+1)} \\ &= \lambda\cdot\mathbf{O}^{Orthogonal}_{(k)} - \lambda\cdot\alpha_{(k+1)}\cdot\boldsymbol{\beta}_{(k+1)}\cdot\boldsymbol{\gamma}^T_{(k+1)}\end{aligned} \quad . \tag{8}$$

Only the initial data window is necessary to be constructed and computed following the definition of orthogonal projection, then the projection matrix at each time step can be computed by the recursive formulae when a set of new data points are appended. Singular value decomposition (SVD) is applied on each newly updated projection matrix to estimate the dominant subspace from the distribution of singular values to obtain the extended observability matrix. RSI-Inversion-Orthogonal is categorized as a projection updated method. It is noted that the time-consuming process of explicitly and repeatedly calculating an orthogonal projection with different data Hankel matrices with large dimensions is avoided; however, SVD at every time step is required in this algorithm for modal parameter identification. A basic derivation of the RSI-Inversion-Orthogonal can be found in [7].

RSI using the Matrix Inversion Lemma renewing method on Oblique Projection (RSI-Inversion-Oblique)

RSI-Inversion-Oblique is a projection matrix renewing algorithm that is similar to RSI-Inversion-Orthogonal; however, this algorithm carries out a more sophisticated oblique projection instead of using orthogonal projection as its identification strategy. The projection matrix implemented here is the product of the original projection matrix $\mathbf{O}^{Oblique}_{(k)}$ and its transpose, or

$$\begin{aligned}\mathbf{O}^{Oblique}_{(k)} &= \left[(\mathbf{Y}_{f(k)}/_{\mathbf{U}_{f(k)}}\boldsymbol{\Xi}_{p(k)})/\mathbf{U}^{\perp}_{f(k)}\right]\cdot\left[(\mathbf{Y}_{f(k)}/_{\mathbf{U}_{f(k)}}\boldsymbol{\Xi}_{p(k)})/\mathbf{U}^{\perp}_{f(k)}\right]^T \\ &= (\mathbf{Y}_{f(k)}\boldsymbol{\Pi}^{\perp}_{\mathbf{U}_{f(k)}}\boldsymbol{\Xi}^T_{p(k)})\cdot\boldsymbol{\Psi}_{(k)}\cdot(\mathbf{Y}_{f(k)}\boldsymbol{\Pi}^{\perp}_{\mathbf{U}_{f(k)}}\boldsymbol{\Xi}^T_{p(k)})^T \\ &= \mathbf{O}^{Orthogonal}_{(k)}\cdot\boldsymbol{\Psi}_{(k)}\cdot(\mathbf{O}^{Orthogonal}_{(k)})^T\end{aligned} \tag{9}$$

with the recursive computation elements $\mathbf{R}_{(k)}$, $\mathbf{Y}_{f(k)}\mathbf{U}^T_{f(k)}$, $\boldsymbol{\Xi}_{p(k)}\mathbf{U}^T_{f(k)}$, $\boldsymbol{\Psi}_{(k)}$ defined as

$$\mathbf{R}_{(k)} = \left[\mathbf{U}_{f(k)}\mathbf{U}^T_{f(k)}\right]^{-1} = \left[\sum_{g=k-j+1}^{k}\lambda^{k-g}\cdot\mathbf{u}_{f(g)}\mathbf{u}^T_{f(g)}\right]^{-1} \quad , \quad \mathbf{Y}_{f(k)}\mathbf{U}^T_{f(k)} = \sum_{g=k-j+1}^{k}\lambda^{k-g}\cdot\mathbf{y}_{f(g)}\mathbf{u}^T_{f(g)}$$

$$\boldsymbol{\Xi}_{p(k)}\mathbf{U}^T_{f(k)} = \sum_{g=k-j+1}^{k}\lambda^{k-g}\cdot\boldsymbol{\xi}_{p(g)}\mathbf{u}^T_{f(g)}$$

$$\begin{aligned}\boldsymbol{\Psi}_{(k)} &= \left[\boldsymbol{\Xi}_{p(k)}\boldsymbol{\Pi}^{\perp}_{\mathbf{U}_f(k)}\boldsymbol{\Xi}^T_{p(k)}\right]^{-1} \\ &= \left\{\boldsymbol{\Xi}_{p(k)}\cdot\left[\mathbf{I}-\mathbf{U}^T_{f(k)}(\mathbf{U}_{f(k)}\mathbf{U}^T_{f(k)})^{-1}\mathbf{U}_{f(k)}\right]\cdot\boldsymbol{\Xi}^T_{p(k)}\right\}^{-1} \\ &= \left\{\begin{aligned}&\left[\sum_{g=k-j+1}^{k}\lambda^{k-g}\cdot\boldsymbol{\xi}_{p(g)}\boldsymbol{\xi}^T_{p(g)}\right] - \left[\sum_{g=k-j+1}^{k}\lambda^{k-g}\cdot\boldsymbol{\xi}_{p(g)}\mathbf{u}^T_{f(g)}\right] \\ &\times\left[\sum_{g=k-j+1}^{k}\lambda^{k-g}\cdot\mathbf{u}_{f(g)}\mathbf{u}^T_{f(g)}\right]^{-1}\times\left[\sum_{g=k-j+1}^{k}\lambda^{k-g}\cdot\mathbf{u}_{f(g)}\boldsymbol{\xi}^T_{p(g)}\right]\end{aligned}\right\}^{-1}\end{aligned} \quad . \tag{10}$$

where $\mathbf{Y}_{f(k)} /_{\mathbf{U}_{f(k)}} \mathbf{\Xi}_{p(k)}$ denotes the operator that projects the row space of a matrix $\mathbf{Y}_{f(k)}$ along the row space of matrix $\mathbf{U}_{f(k)}$ onto the row space of the matrix $\mathbf{\Xi}_{p(k)}$.

In this way, as shown in Eq. (9), SVD is now replaced by an eigenvalue decomposition (ED) that is applied on the re-formulated oblique projection matrix $\mathbf{O}_{(k)}^{Oblique}$ to estimate the dominant subspace required for obtaining the extended observability matrix. RSI-Inversion-Oblique shares some procedures with RSI-Inversion-Orthogonal, since the projection matrix $\mathbf{O}_{(k)}^{Orthogonal}$ also appears in the formulation for $\mathbf{O}_{(k)}^{Oblique}$. Therefore, the updating technique will be divided into two parts: updating the orthogonal projection matrix $\mathbf{O}_{(k)}^{Orthogonal}$ as per RSI-Inversion-Orthogonal, and updating the matrix $\mathbf{\Psi}_{(k)}$ when a set of new input/output data vectors are appended. Then, the product of these two renewed terms will derive the target projection matrix $\mathbf{O}_{(k)}^{Oblique}$ of the latest state for each time step. Based on this update technique, a series of recursive formulae can be built, which is similar to the RSI-Inversion-Orthogonal algorithm. By combining an exponential forgetting factor added at each time step due to its enlarging time window, and applying the oblique projection matrix $\mathbf{O}_{(k)}^{Oblique}$ at each time step, one can recursively obtain the dynamic properties of the system during the excitation process. **Figure 1** shows the flowchart of the RSI-Inversion-Oblique computation. Detailed derivation of RSI-Inversion-Oblique can be found in [12, 13].

Through applying the on-line RSI algorithm to the seismic response of buildings, the identification may provide ambiguous system dynamic characteristics due to the selection of different user-defined parameters for the RSI method. These user-defined parameters in RSI include the size of the data Hankel matrix, the determination of model order (similar to the criteria to remove spurious modes) and the forgetting factor. To have a consistent estimation of system dynamic characteristics, discussions on the determination of user-defined parameters for RSI are important. In the following section, three important user-defined parameters are discussed.

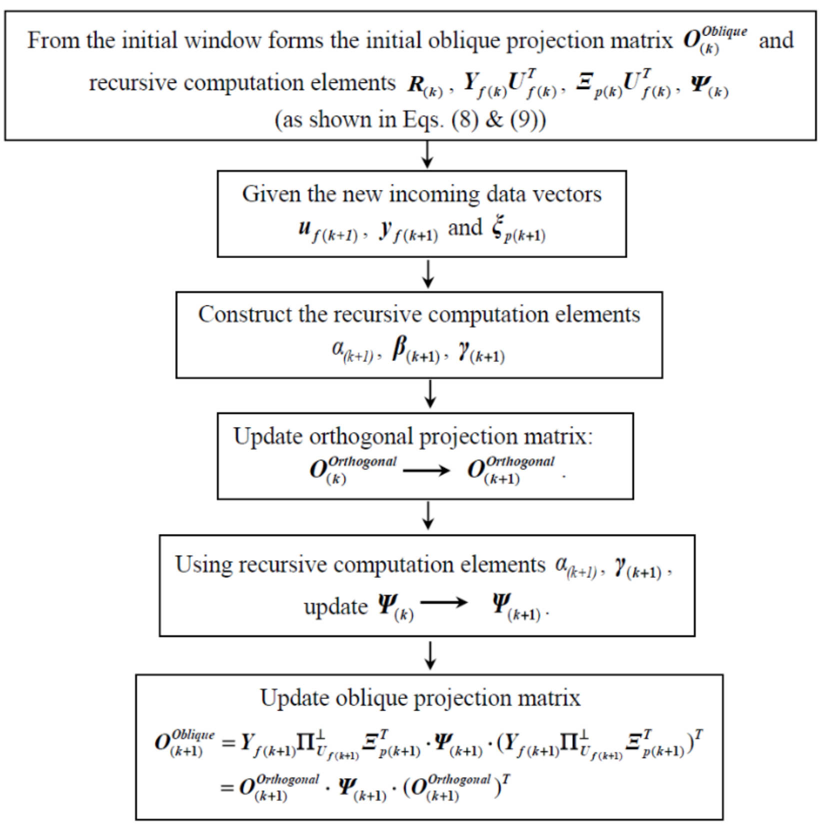


Figure 1: Flowchart on the computation of RSI-Inversion-Oblique.

DISCUSSIONS ON THE USER-DEFINED PARAMETERS IN RSI

Seismic response data collected from the Sherman Oaks building is used as an example to examine the user-defined parameters on the identification of system modal frequencies using the subspace identification technique. In 1977, the building was instrumented with fifteen accelerometers on five levels (under CMSIP). This building was retrofitted with friction dampers after the 1994 Northridge Earthquake**. Figure 2** shows the instrument layout in the building. In the present study, data collected from this building during the Northridge earthquake is used for the experimental study. Since the seismic response data collected from the structure may contain the pre-event memory data before the trigger level, to enhance the accuracy of RSI the ambient vibration signal needs to be removed from the recorded data. Therefore, the method to select the duration of seismic response data for SI (or RSI) must be defined. The initial time from the recorded seismic response data can be determined from the concept of P-wave picker through [14] through the following equation

$$AIC(t) = t\log(\mathrm{var}(a[1:t])) + (N - t + 1)\log(\mathrm{var}(a[t+1:N])), \tag{11}$$

where *AIC* is Akaike information criterion, t is the time moving window length and N is total number of data points, and var(a[1:t]) is the variance of the recorded data $a(t)$ from the first data point to time t. Then the initial starting P-wave arrival time can be determined from the slope of the information criteria *AIC*, which begins to change dramatically. To determine the end point of the recorded data, one can plot the normalized Arias Intensity and select the data point at a time of 99.5%. Two earthquake event datasets from the Chino Hills and Northridge earthquakes as recorded from the Sherman Oak building are used as examples to determine the strong motion duration for RSI, as shown in **Figure 3**. Almost all of the recorded data from the Northridge earthquake will be used while the Chino Hill earthquake response data has a long pre-event memory data that needs to be removed before the implementation of RSI. Based on the proposed criteria, the entire recorded dataset from the Northridge earthquake will be used for RSI.

To identify the system modal frequencies using the subspace identification technique, besides the determination of the record duration as well as the selection of criteria to remove the spurious modes (such as using mode shape assurance criteria, phase collinearity among different model order, etc.) [15], the user-define parameters to generate the system A-matrix from using either the SI or RSI algorithm need to be carefully selected. In the following sections, the determination of the three user-defined parameters in RSI are discussed: (1) the number of rows "*i*" in the block data Hankel matrix, (2) the method for model order estimation, and (c) the forgetting factor.

Figure 2: Photo and layout of strong motion instrumentation of the Sherman Oaks 13-story building (CSMIP).

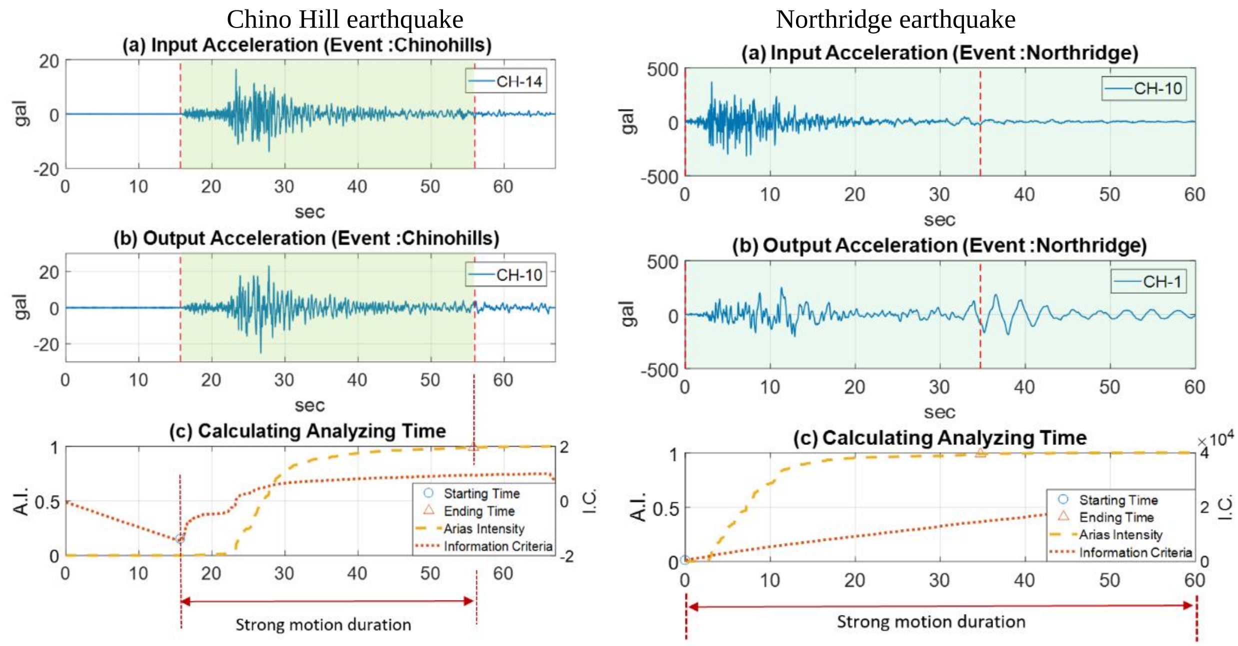


Figure 3: Plot the AIC and normalized Arias intensity to identify the strong motion duration from two different seismic event data recorded from Sherman Oaks building (the initial and the ended data point of strong motion duration are identified).

Determining the number of rows "i" in the block data Hankel matrix

The number of rows and columns of the block data Hankel matrix must be determined first. Caicedo suggested the number of columns of the block data Hankel matrix shall be four times the number of expected modes [16]. To assure the complete waveform of the longest period of the structure can be included in each column of the data Hankel matrix, in this study, the number of block rows in the data Hankel matrix (consider only single output record) should be larger than the sampling frequency divided by two times the fundamental frequency of the structure, i.e.

$$i = \frac{\text{Sampling Rate (Hz)}}{2 \times \text{Fundamental freqeuncy (Hz)}} . \qquad (12)$$

The most economic number of columns "*j*" in the block Hankel matrix is chosen so that the Hankel matrix remains square, therefore $j = 2i \times (m+l)$, and sometimes for convenience of computing *WL* in integers, *j* is assigned as $j = 2i \times (m+l) + 1$, where *m* is the number of input records and *l* is the number of output records. Once *i* and *j* are determined, the initial number of data points arranged in a Hankel matrix can be calculated from $WL = 2i + j - 1$, i.e. the initial window length for the moving time window using RSI is decided. It is obvious that a larger *i* will create a longer initial window length, which may remove the ability to detect the dynamic characteristics of the time-varying system.

On the contrary, if a large number of response measurements are installed in the structure, then the initial window length will be very long, which may also obscure the identification of time-varying dynamic characteristics of the structure in the initial time window. Therefore, for selecting the number of rows in the block Hankel matrix, one can at least choose the data length with a duration that covers half of the fundamental dominant period of the structure. It should be noted that a smaller *i* may obscure the identification of lower modes while it may increase the ability to detect the higher modes of the structure (which will be demonstrated below).

Determining the Model Order for RSI

The model order decides the dimension of subspace $\boldsymbol{U_1}$ for estimating the extended observability matrix after **U** is obtained from SVD or ED of the projection matrix, as shown by

$$\mathbf{H} = \mathbf{USV}^T = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_1^T \end{bmatrix}. \tag{13}$$

The singular values obtained from SVD or eigenvalues from ED of the projection matrix will be arranged in descending order as shown by

$$diag([\mathbf{S}]) = \begin{bmatrix} s_1 & s_2 & \cdots & s_j \end{bmatrix} = [\mathbf{SV}] \tag{13a}$$

The number shown in Eq. (13) can be defined from the percentage of the singular value to be removed (i.e. percentage of noise to be excluded: $100 \times \mathrm{diag}(\mathbf{S}_2) / [\mathrm{diag}(\mathbf{S}_1) + \mathrm{diag}(\mathbf{S}_2)]$). Generally, a fixed percentage of noise will be pre-assigned, from which one can decide the model order to extract subspace $\mathbf{U}_1$ from SVD. For each impending new dataset, the fixed noise percentages remain constant through the recursive analysis. The estimation of the subspace dimension will be taken at the position corresponding to singular values (or eigen-value) starting from the largest one to the pre-assigned value of C_{svd} percentage multiplied by the maximum singular value, as shown by

$$s_N > C_{svd} \cdot s_1 > s_{N+1} \,. \tag{14}$$

This method lacks a theoretical definition, particularly, on how to determine the percentage of noise or the value of C_{svd}. Therefore, different approaches to determine the model order will be discussed in the following section.

A simple way to calculate the change of singular value (s_j), from Eq. (13a), by calculating the difference between each pair of singular values is given by

$$ND_i = \frac{s_i - s_{i+1}}{s_i} \qquad i = 1, 2, \cdots, j \qquad . \tag{15}$$

where ND is the normalized difference of singular values. Then the sum of ND_i can be calculated and defined as A_k, or

$$A_k = \frac{1}{k} \sum_{i=1}^{k} ND_i \quad k = 1, 2, \ldots, j \,. \tag{16}$$

where k is the searching step. Through the recursive subspace identification technique, one can determine the steep descent of A_k to estimate the model order at the location of steep descent. To use this method for model order estimation in each time window, the level of noise in the recorded data needs to be identified. Then, the signal to noise (S/N) ratio of the recorded data can be determined, from which the maximum reference S/N ratio is given, and thus the model order can be determined (**Method-1**). **Figure 4** presents a flowchart for the procedure to estimate the model order.

The first approach to determine the model order is proposed and describe below. An example to demonstrate the estimation of model order is shown in **Figure 5** for time t = 19.98 s and at time t = 39.98 s. First, the difference between two consecutive eigenvalues is plotted with respect to the sequence number of eigenvalues. Then, A_k as defined in Eq. (16) is plotted, from which a model order of 14 (t = 19.98 s) and 8 (t = 39.98 s) can be identified, as the point at which A_k begins to decrease can be selected as the model order.

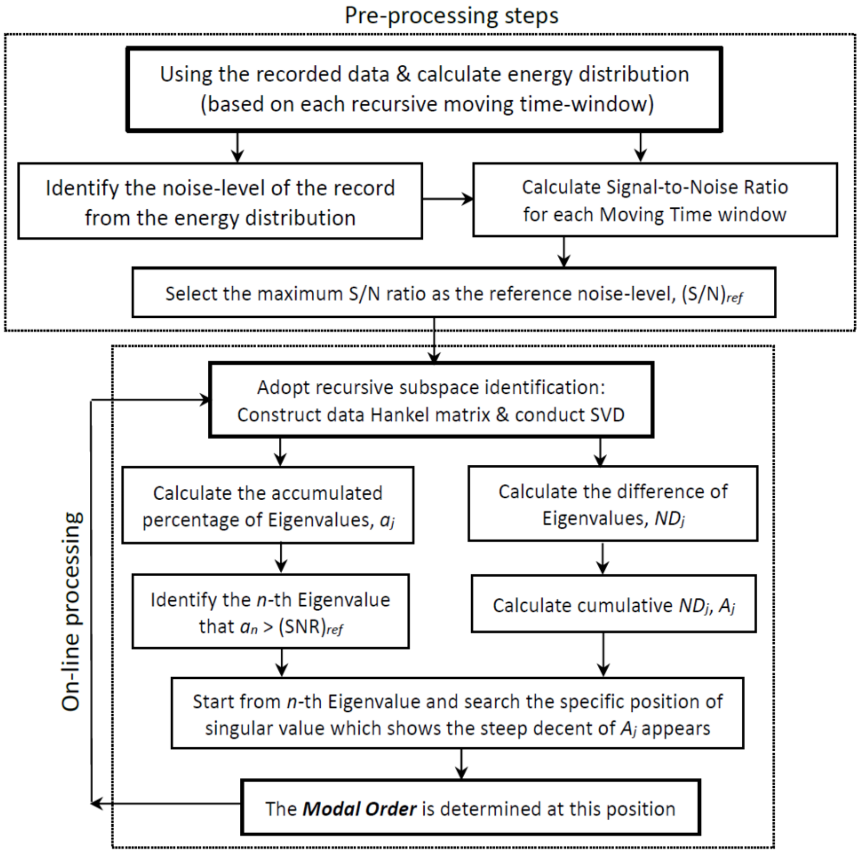
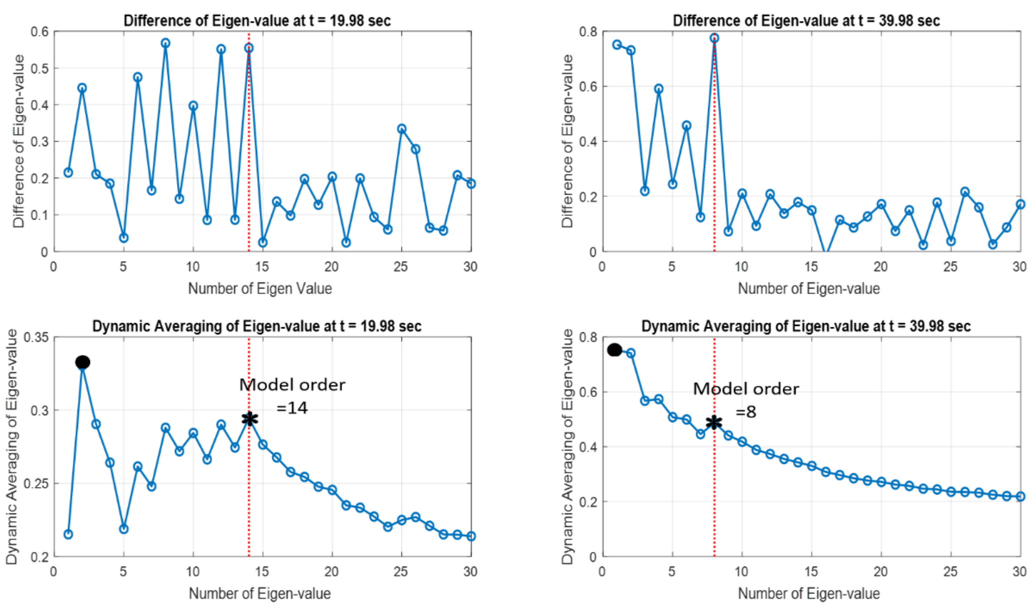


Figure 4: Flowchart for estimating the model order using the difference of eigenvalue (Method-1).



(a) at time t=19.98 sec. (b) at time t=39.98 sec

Figure 5: Plot the difference of eigenvalues and cumulative of ND_i (Eq.(15)) at time t=19.98 sec and t=39.98 sec, respectively, from which the model order can be estimated (used λ=0.99, i =45).

Instead of using the normal scale of the singular value to determine the model order, the logarithms of the singular values from Eq. (13a) can also be used, as shown by

$$\begin{aligned} \langle LogSV \rangle &= < log_{10}(s_1), log_{10}(s_2), \cdots, log_{10}(s_j) > \\ &= \langle lsv_1\ ,\ lsv_2,\ \cdots\ \ ,\ lsv_j \rangle \end{aligned} \quad (17)$$

This method was originally proposed by [7] to determine the model order. Since the singular value (s_i) shown in Eq. (13a) will be arranged in descending order, therefore, lsv_i shown in Eq. (17) will also in descending order. To avoid the negative value in the element of $\langle LogSV \rangle$, the element in the Eq. (17) is subtracted by the smallest element, among singular values, defined as $lsv_{\min}$, then the equation can be revised to

$$\begin{aligned} \langle LogSV_p \rangle &= \langle lsv_1 - lsv_{min} + \eta,\ \ lsv_2 - lsv_{min} + \eta\ ,\ \cdots\ \ ,\ lsv_j - lsv_{min} + \eta \rangle \\ &= \langle\ lsv_{p1}\ \ lsv_{p2}\ \ \ \cdots\ \ \ lsv_{pj}\ \ \rangle \end{aligned} \quad (18)$$

where η is an arbitrary positive constant and all element in $\langle LogSV_p \rangle$ are positive.

To determine the significant difference between two consecutive singular values in the element of $\langle LogSV_p \rangle$, the sequential difference between two consecutive values of $lsv_{ps(j-1)}$ and lsv_{psj} is calculated and given as

$$\begin{aligned} < LSV_{ps} > &= < \frac{(lsv_{p1} - lsv_{p2})}{lsv_{p1}}\ \ \frac{(lsv_{p2} - lsv_{p3})}{lsv_{p2}}\ \ \cdots\ \ \frac{(lsv_{p(j-1)} - lsv_{pj})}{lsv_{p(j-1)}} > \\ &= <\ lsv_{ps1}\ \ lsv_{ps2}\ \ \cdots\ \ lsv_{ps(j-1)} > \end{aligned} \quad (19)$$

Furthermore, all the element in $< LSV_{ps} >$ can be normalized with the largest value of lsv_{psj} (defined as $lsv_{ps\_\max}$), and defined as:

$$\begin{aligned} < NLSV_{ps} > &= < \frac{lsv_{ps1}}{lsv_{ps\_max}}\ \ \frac{lsv_{ps2}}{lsv_{ps\_max}}\ \ \cdots\ \ \frac{lsv_{ps(j-1)}}{lsv_{ps\_max}} > \\ &= <\ nlsv_{ps1}\ \ nlsv_{ps2}\ \ \cdots\ \ nlsv_{ps(j-1)}\ \ > \end{aligned} \quad (20)$$

The other approach to estimate the model order is based on Eq. (20). Instead of directly using the eigenvalue differences, from$< NLSV_{ps} >$, the position with a value in the element of $< NLSV_{ps} >$ equal to 1 is the model order. Since by using $nlsv_{psj}$ =1 might still be too conservative, therefore, it is accepted that a higher model order is better for some data windows that can include more modes in the analysis. Therefore, the order at the second largest value from the element of $< NLSV_{ps} >$ can also be selected as the decided model order.

At some specific time-window, if the value at the second largest singular value is chosen as the model order, one may include too many higher modes, which may also obscure the fundamental modes. Therefore, to prevent this phenomenon from occurring on a time step, a constraint needs to be added. Since the major elements in the normalized [LSV_{ps}] are preferred for determining the model order, one can use the normalized [LSV_{ps}] data as a signal and calculate the cumulative signal energy of $nlsv_{psj}$ values. 90% of the cumulative signal energy values of $nlsv_{psj}$ can be used as a threshold to avoid selecting too many trivial $nlsv_{psj}$ and thus including too many higher modes. Then, the model order can be determined from the minimum location between the number of the second largest singular value and the index of 90% of the cumulative value of $nlsv_{psj}$ (**Method-2**):

$$Min.\left[\text{No. of the } 2^{\text{nd}} \text{ largest singular value, 90\% of the signal energy value of } nlsv_{psj}\right]$$

As an example, consider time t = 20 s and t = 50 s. **Figure 6** shows the distribution of singular values at these two discrete times. The normalized difference in the singular value lsv_{psj} as well as the percentage of the cumulative signal energy values is also plotted. Based on the proposed Method-2, at t = 20 s, the 2nd largest singular value lsv_{psj} is located at 14, while 90% of the cumulative signal energy values of $nlsv_{psj}$ is located at 28, therefore, for t = 20 s the model order is determined as 14. Alternatively, at t = 40 s the 90% of the cumulative signal energy of $nlsv_{psj}$ is located at 28 and the second largest $nlsv_{psj}$ is located at 8, and therefore, at t = 40 s the model order is decided as 8.

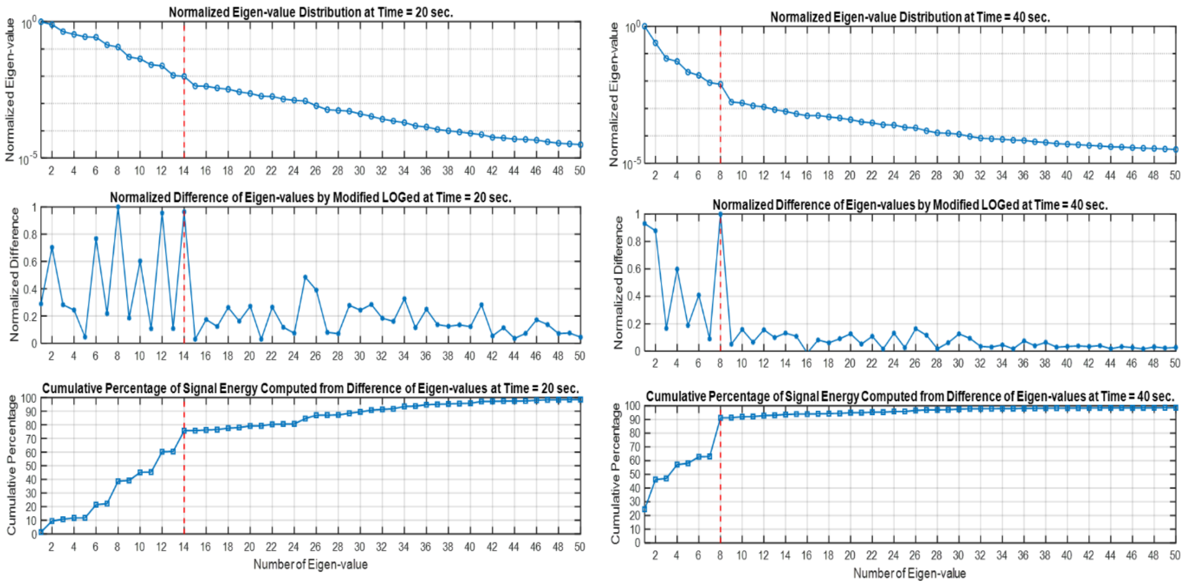


Figure 6: Model order estimation using the minimum location of the 2nd largest singular value and the location of 90% of the cumulative value of $nlsv_{psi}$; (a) at time *t*=20 sec, (b) at time *t*=40. Sec.

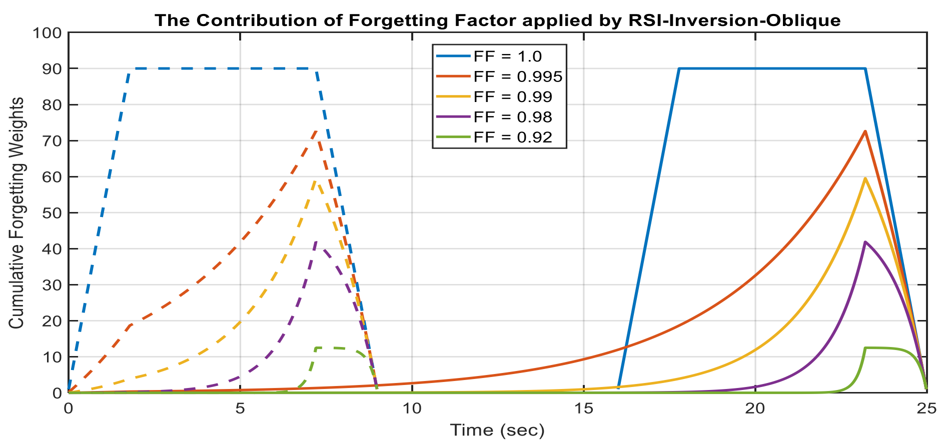


Figure 7: The weighting factor on data (from time *t*=0.0 to 9.0 sec and from time *t*=0.0 to 25.0 sec) using different forgetting factor (initial window length is 9.0 sec).

Influence of the forgetting factor in RSI

The purpose of using the forgetting factor (FF) is to reduce the weight of data (or fading memory) that are far away from the current time step on the modal parameter estimation. The weighting factor of the fading memory data is changed by using a different forgetting factor. If a smaller forgetting factor is used then more weight is given to the current time step, as shown in **Figure 7**. In Figure 7, two time steps are considered, at *t* = 9.0 s and at *t* = 25.0 s. To estimate the modal parameters at *t* = 9.0 s the weighting for the time window from 0.0 s to 9.0 s is used and to estimate the modal parameters at *t* = 25.0 s the weighting for the time window data is from 0.0 s to 25.0 s is used. If a lower FF is used then the fading memory becomes more significant for data away from the estimated time step. Therefore, for estimating the abrupt change of modal parameters (such as the abrupt reduction of floor stiffness) one can select a lower FF, as the fading memory on the past data will be more significant then when using a larger FF. For the reinforced concrete structure, due to the inelastic hysteretic behavior of the restoring force, which is different from an abrupt change of stiffness, the FF will be set near 0.99 as the fading memory cannot be so significant. Detailed experimental study results will be presented in the following section.

EXAMINE THE USER-DEFINED PARAMETERS IN RSI USING BUILDING SEISMIC RESPONSE DATA

It has been pointed out that by using different user-defined parameters in RSI will create different results on tracking the structural dynamic modal properties. An example is used to implement the proposed algorithm to building seismic response data. A discussion on using different combinations of *i* and the model order selection method on using RSI to identify the time-varying modal parameters of the Sherman Oaks building during the Northridge earthquake is presented below. The inputs and outputs in transverse, longitudinal, and vertical directions can be used to identify the modes simultaneously; however, the number of inputs and outputs will increase together, increasing the windows length dramatically and delaying the online system identification eventually. Considering that the torsion effect in the examination is less significant, the identification of transverse and longitudinal data is performed separately. Furthermore, a sensitivity analysis on using different combinations of *i* value to construct data Hankel matrix (*i* = 20 or 45) and different methods on model order estimation (Method-1 and Method-2) will be presented.

Using single input single output versus single input multiple output on system fundamental frequency estimation

For the seismic response data of the Sherman Oaks building, the data sampling rate is 50 Hz and base on the Fourier amplitude spectrum of roof response the target fundamental frequency of the building is approximately 0.325 Hz; therefore, based on Eq. (12) the suggested the *i* value is 75. First, if only the roof response data is considered for RSI (single-input and single-output case) and using $i = 75$ with initial window length of 9.0sec, **Figure 8b** shows the identified time-varying fundamental frequency of the building in the transverse direction. Second, instead of using the roof response data only as the output, but all the floor response measurements are used as output (multiple output) in RSI. In order to use the same initial time window in RSI using single-output measurement (as shown in Figure 8b), the *i* = 45 was selected to construct the data Hankel matrix for multiple output measurement. **Figure 8c** shows the identified modal frequency using all the measurement data with *i*=45. From the comparison between these two results (SISO vs SIMO), the identified fundamental modal frequency is similar. Since *i* = 45 (data length = 0.02 s x 2*i* = 1.8 s) does not cover the structural fundamental period (~2.9 s), but with the consideration of multiple output measurements, there is enough energy contribution from all measurement that can contribute to the fundamental mode. Thus, by using multiple outputs one can select a smaller *i* while maintaining strong identification on the fundamental frequency of the structure. Therefore, it is summarized that to have a better identification on time-varying modal parameters, the *i* (number of rows), the number of measurements and the initial time window length need to be considered.

Besides, the forgetting factor also needs to be discussed in advance. **Figure 9** shows the identified modal frequency by using two forgetting factors (0.99 and 0.995) and two *i* values (*i* = 20 and *i* = 45). Figure 9 demonstrates that for smaller FF (=0.99) the fading memory applied to each data point is more significant than that using FF (=0.995); therefore, larger FF value will have less fluctuation on the identified modal frequency because of less fading memory. If one needs to detect the time-varying system, the FF needs to have a smaller value so as to detect the system natural frequency change with respect to time. In the following analysis on the model order estimation method, the FF is 0.99.

Comparison of RSI result using different model order estimation technique

As described above, *i* = 45 will be used in the following RSI and the initial time window will need $2i+2i(m+l)-1$ = 2×45+2×45×(1+3)–1 = 449 data points, which is 8.98 s. In such a case, due to the longer initial window length, the tracking ability of time-varying modal properties for data in the initial time window may be lost by using RSI. Thus, a lower *i* is also considered. If $i = 20$ and the square data Hankel matrix is also used, the number of column of the Hankel matrix is $j = 2ix(m+l) = 2x20x(1+3)$ = 160 data points. Then the initial window length is $WL= 2i+2i(m+l)-1$ = 2×20+2×20×(1+3)–1= *199 points* = 3.98 s, which is shorter for initial time window than using *i* = 45 (*WL* = 8.98 s). Therefore, a comparison of the RSI analysis results from using two different values of *i* (*i* = 20 and *i* = 45) is made.

First, consider the method from [7] for model order estimation with two different *i* (20 and 45). Based on the normalized difference of a maximum value of lsv_{ps} (i.e. $lsv_{ps\_\max}$), as shown in Eq. (20), the position with the value equal to one (or $nlsv_{psj} = 1$) is selected as the model order [7]. **Figure 10** compares the result of identification from data of both longitudinal and transverse directions. It is observed that by using smaller values of *i*, the initial time window is shorter (cannot include the period of the 1st mode), therefore, the component energy from higher modes will be enhanced relatively. As shown in Figure 10a, the fundamental frequency of the structure cannot be identified from the data between 0.0 sec and 11.0 sec).

However, for i = 45 the initial window length is 8.98 sec, the identification of higher modes is poor than using i =20. In summary, if the position of $nlsv_{psj}$ = 1 is used to select the model order, no matter high or low i value be used it, the method on identifying time-varying modal frequencies have its drawback. Of course, the result may change when using a different model order selection method. A discussion on using different model selection methods for RSI is given in the following section.

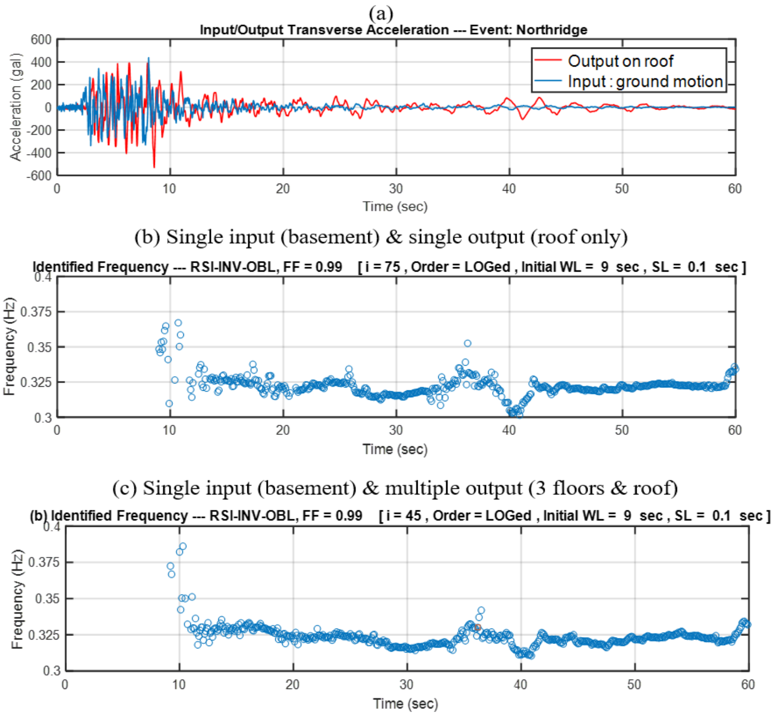
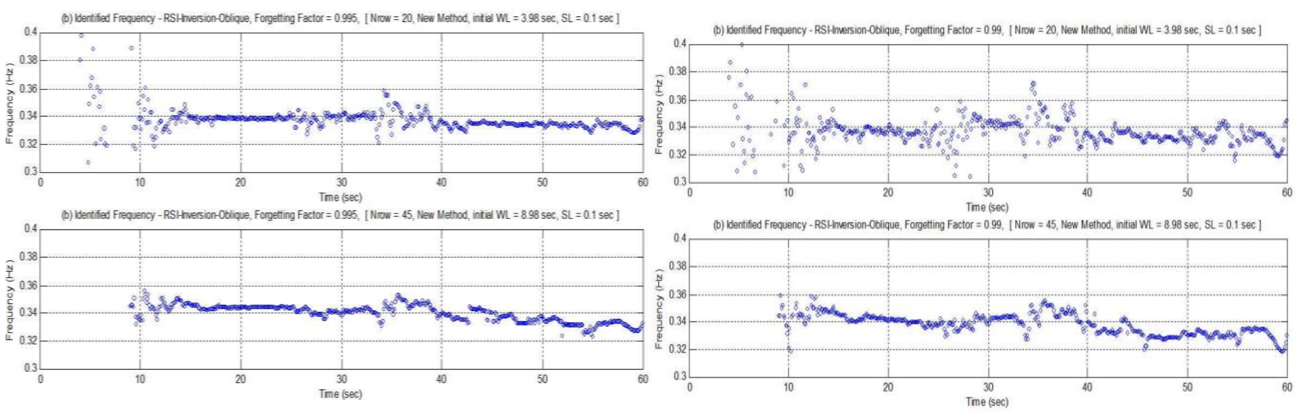
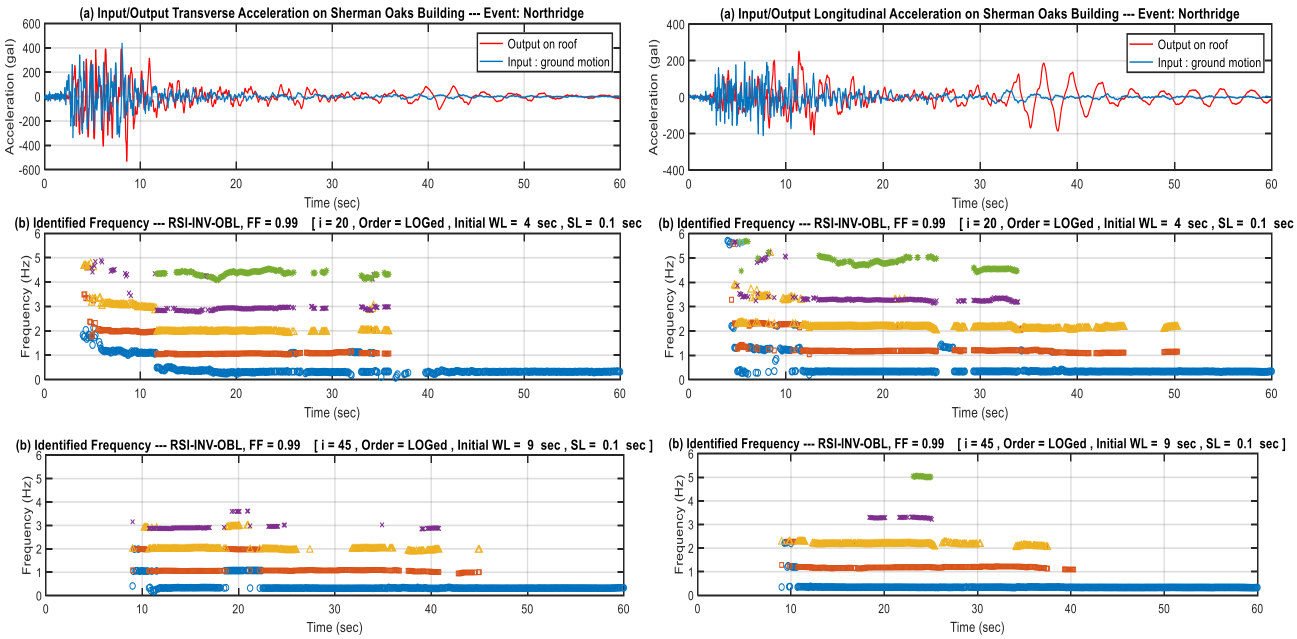


Figure 8:
(a) Recorded seismic response data (basement and roof),
(b) Using *i*=75 for case of RSI with single input (basement data) and single output (roof response data),
(c) Using *i*=45 for case of RSI with single input (basement data) and multiple (all floor response data).



(a) Forgetting factor λ=0.995 (b) Forgetting factor λ=0.99

Figure 9: Plot of time-varying modal frequency using two different *i* values (*i*=20 and *i*=45);
(a) with forgetting factor λ=0.995, (b) with forgetting factor λ=0.99.

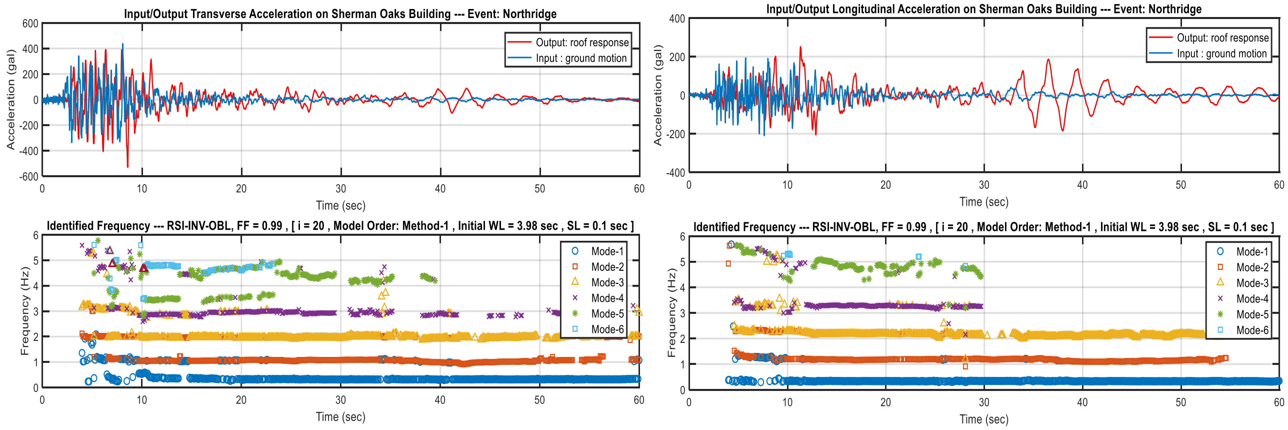


(a) Transverse Direction (b) Longitudinal Direction

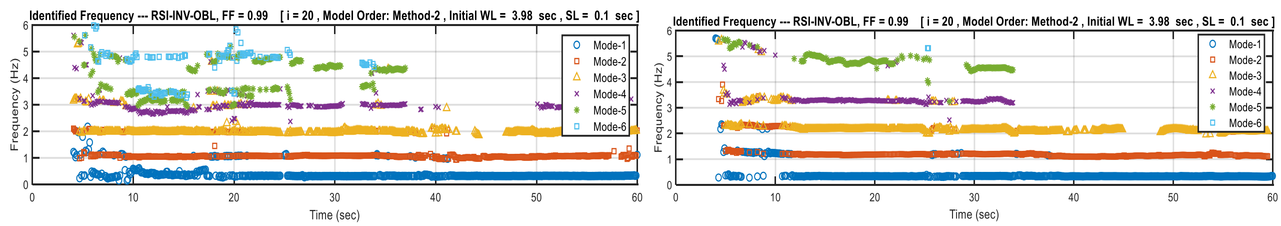
Figure 10: Identified time-varying model frequencies of Sherman Oaks building during Northridge earthquake excitation using model order from Ref.7 and with *i*=20 and *i*=45; (a) Longitudinal direction, (b) Transverse direction.

Different from using the traditional method [7] to select the model order, both ***Method-1*** and ***Method-2*** for model order selection are used to identify the time-varying modal frequencies of the system. **Figures 11a** and **11b** show the comparison on the identified time-varying modal frequencies using two different model order selection methods using i = 20. By using smaller i value, the energy contribution from higher modes is stronger than the contribution from fundamental mode, therefore, more higher modes can be identified. Besides, the initial window length is also short, one can detect the modal frequencies earlier than using longer initial time window (as shown in **Figures 11c** and **11d** for i = 45). Using shorter initial time window one can have a better adaptability on the identification of system nonlinearity. A comparison between the identification results using Method-1 and Method-2 for model selection is also shown in **Figure 11.** It is challenging to judge which method is better for identifying the time-varying modal frequencies. **Figure 12** shows the comparison on the estimated model order using two different methods. In general, Method-2 provides a lower number of model order estimation than Method-1, but the difference is not so significant. If only the fundamental modal frequency is concerned, selection a suitable row number of data Hankel matrix is much important than using different method of model order estimation. **Figure 13** shows that the difference on the estimation of fundamental frequency is minor by using two different model order estimation if i = 45 is used. The estimated modal damping of the building is also shown in **Figure 14**, which shows that both methods can have a similar result.

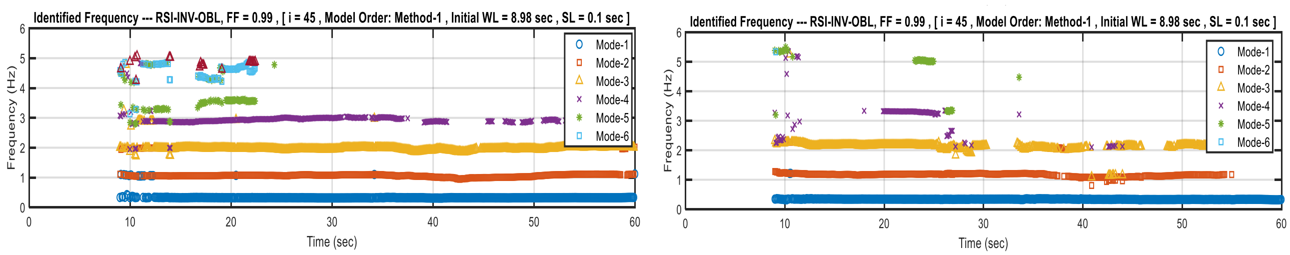
(I) Transverse direction (II) Longitudinal direction



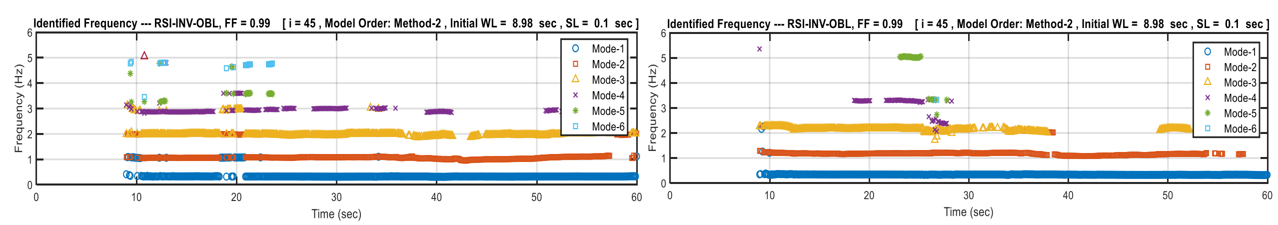
(a) result of RSI using *i*=20 and method-1 for model order estimation



(b) result of RSI using *i*=20 and method-2 for model order estimation

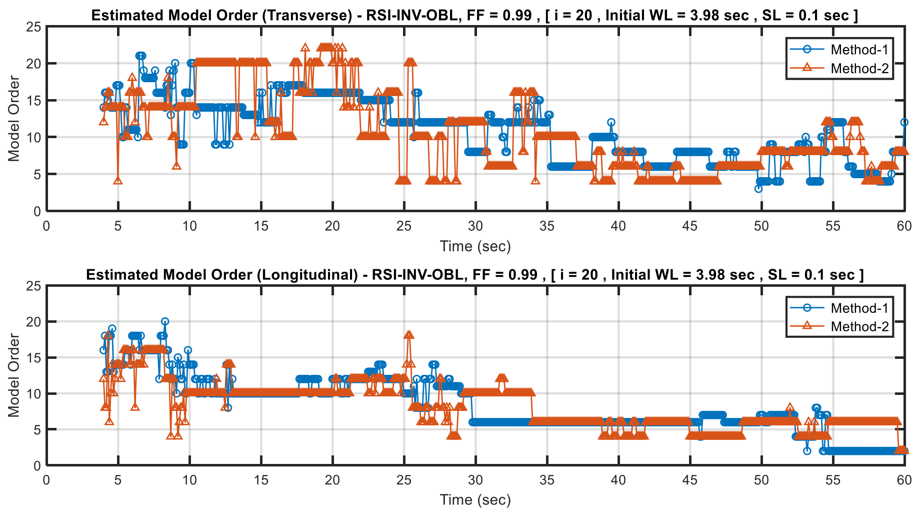


(c) result of RSI using *i*=45 and method-1 for model order estimation

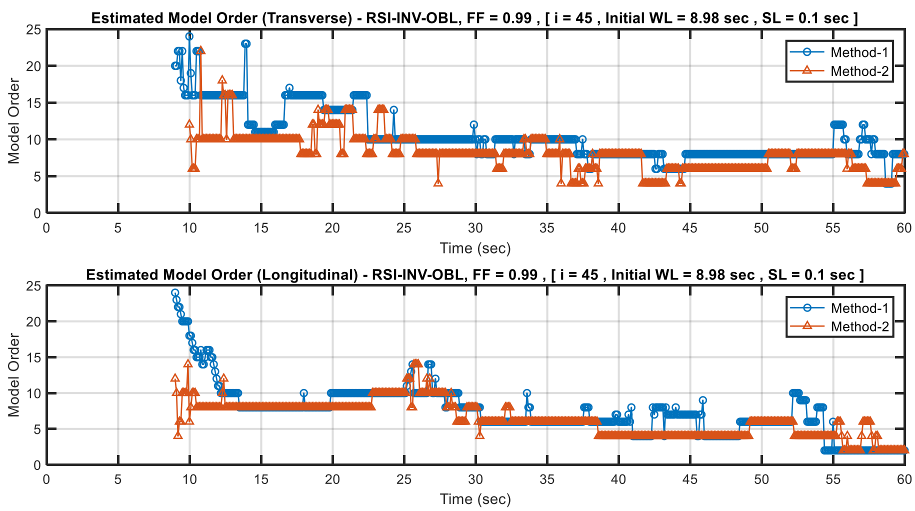


(d) result of RSI using *i*=45 and method-2 for model order estimation

Figure 11: Identified time-varying model frequencies of Sherman Oaks building during Northridge earthquake excitation using:
(a) method-1 for model order selection and with *i*=20
(d) method-2 for model order selection and with *i*=20
(c) method-1 for model order selection and with *i*=45
(d) method-2 for model order selection and with *i*=45

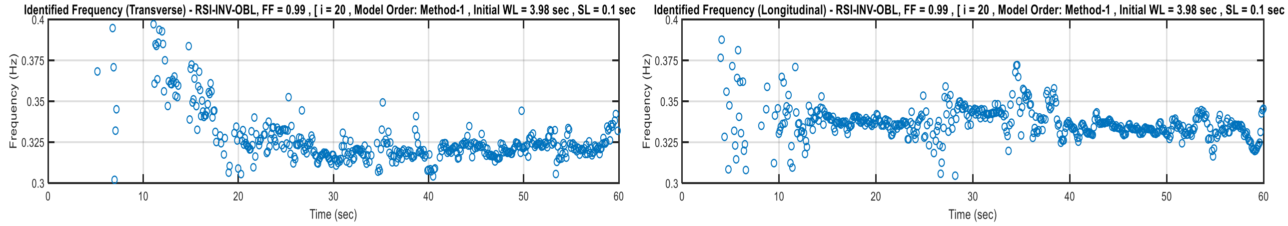


(a) Estimated model order using *i*=20

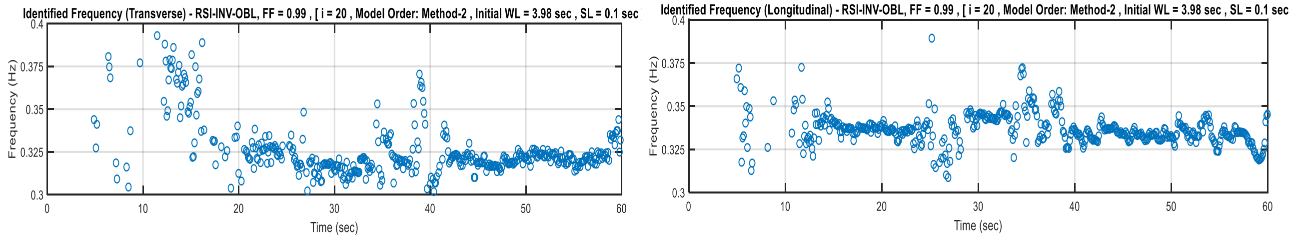


(b) Estimated model order using *i*=45

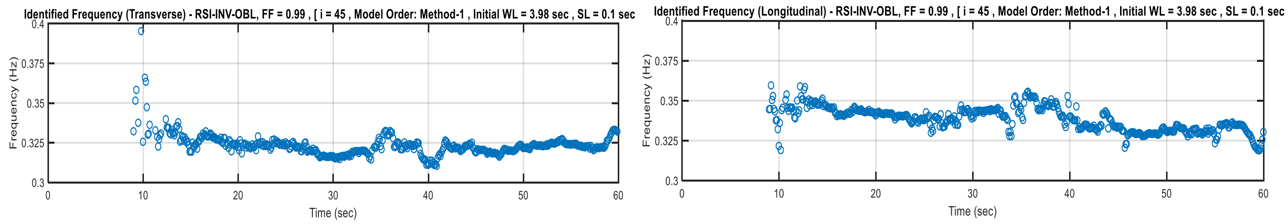
Figure 12: Comparison on the estimated model order using two different method and two different *i* values.



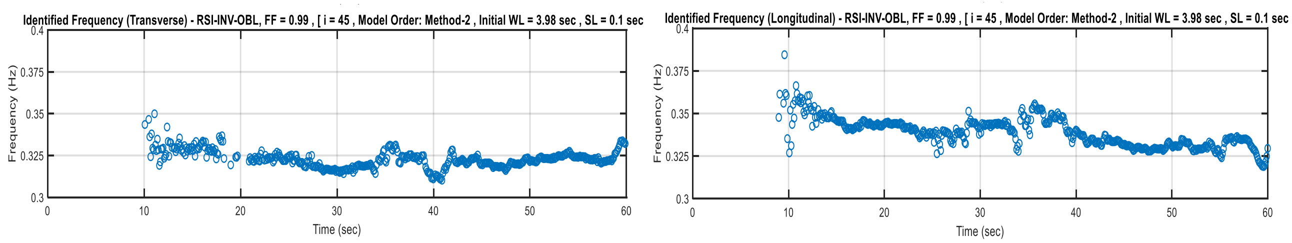
(a) Result from RSI using method-1 for model order estimtion and *i*=20



(b) Result from RSI using method-2 for model order estimtion and *i*=20



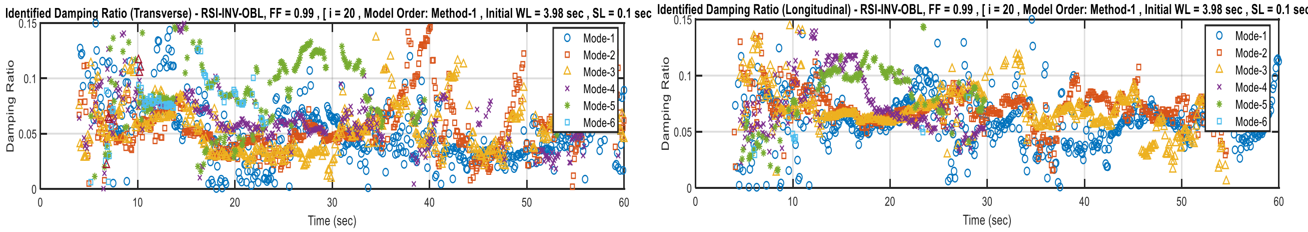
(c) Result from RSI using method-1 for model order estimtion and *i*=45



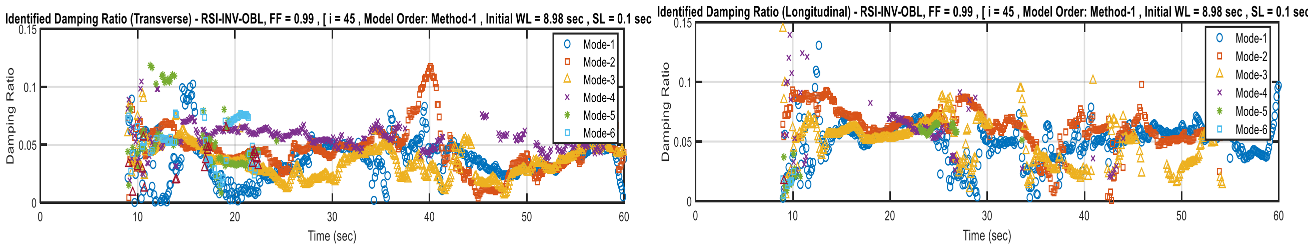
(d) Result from RSI using method-2 for model order estimtion and *i*=45

Figure 13: Comparison on the identified fundamental modal frequency with two different *i* values (*i*=20 & *i*=45) using method-1 and method-2 for model order estimationand.

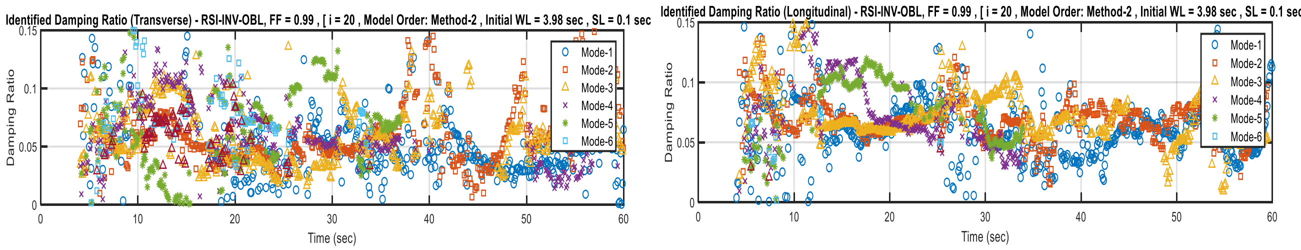
(I) Transverse direction (II) Longitudinal direction



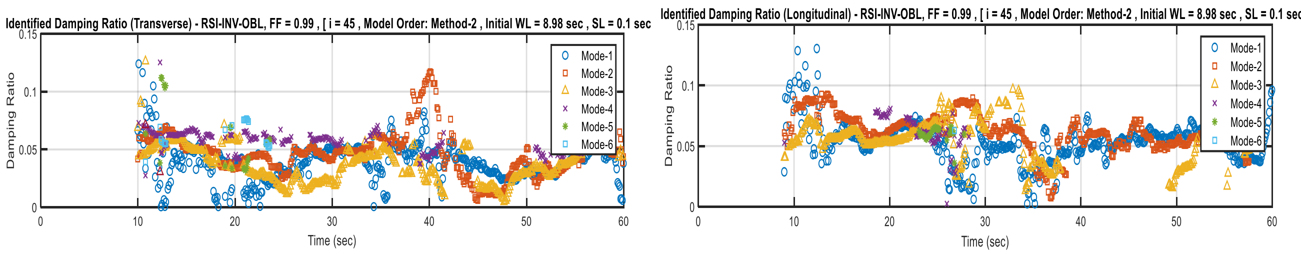
(a) Result from RSI using method-1 for model order estimation and *i*=20



(b) Result from RSI using method-1 for model order estimation and *i*=45.



(c) Result from RSI using method-2 for model order estimation and *i*=20.



(d) Result from RSI using method-2 for model order estimation and *i*=45.

Figure 14: Identified time-varying model damping ratio; (a) using *method-1* for model order selection and with *i*=20, (b) using *method-1* for model order selection and with *i*=45, (c) using *method-2* for model order selection and with *i*=20, (d) using *method-2* for model order selection and with *i*=45.

CONCLUSIONS

Recursive subspace identification (RSI) is an efficient on-line method to identify the time-varying modal frequencies of a structure during earthquake excitation. For application of RSI, the user-defined parameters, such as the size of data Hankel matrix and model order selection, et al., need to be carefully selected otherwise some wrong message may be interpreted after identification. Among these user-defined parameters, in this paper the number of rows in the data Hankel matrix as well as the selection of model order, based on the distribution of eigenvalues of the oblique projection matrix ($\mathbf{O}_{(k)}^{Oblique}$), are carefully discussed for RSI analysis of seismic response of building structure. From the study of seismic response data using RSI the following conclusions are made:

1. The number of rows ($2i$) to construct the data Hankel matrix needs to have enough length to cover the fundamental period of the structure ($2 \times i \times \Delta t \geq T_1(1^{st}\ period\ of\ a\ structure)$). Since RSI considers the measurement from all sensing nodes (multi-output measurement), then the number of rows in data Hankel matrix will be $2 \times i \times (m + l)$.
2. Once the number of rows in the data Hankel matrix is decided ($2i$), if the square data Hankel matrix is used for RSI, then the initial time window for identification is found ($WL = 2i + j - 1$). A larger i–value will create a longer initial time window, which may not provide the recursive identification of time-varying modal frequencies within the initial time window. This situation is even worse if multiple input and multiple output measurement are included in the analysis ($2i+2i(m+l)-1$). To enhance the applicability of recursive identification in the initial window, reduce the number of rows in data Hankel matrix is acceptable.
3. To determine the model order to extract the structural modes, either Method-1 or Method-2 can provide a good estimation on the lower structural modes. To use Method-1 for model order estimation, the S/N ratio of the recorded data must be determined. Therefore, data from the pre-event memory is required to determine the S/N value. As for Method-2, the minimum value of the number of eigenvalues between the second largest normalized eigenvalue difference is selected and 90% of the cumulative signal energy is used (treating the normalized eigenvalue difference value as a signal).
4. From the RSI analysis on the earthquake response data (Northridge earthquake) of the Sherman Oaks Building, i = 45 is selected which is less than the fundamental period of the structure (should be i=75 to cover the fundamental period). Since multiple input and output measurements are included in RSI, for system identification, there are enough energy to enhance the fundamental frequency of the structure. Without loss of generosity, smaller i value can be used for RSI.
5. In the case study on seismic response data of the Sherman Oaks (Northridge earthquake), the initial window length is 8.98 sec, within that window length the RSI can not be applied (too long for initial window). As described before, one can reduce the initial time window by selecting smaller number of i. To compensate for this un-identified time window at the beginning of the earthquake excitation, the RSI-BonaFide algorithm was used. The RSI-BonaFide algorithm is derived from LQ decomposition applied in PO-MOESP to estimate system matrices and modal parameters recursively at each time instant step [6]. It is a fixed time window RSI method. The data contribution for each time window is shown in Figure 7 (trapezoid shape). Considering a time window of 5.0 s without a forgetting factor and with a sliding window shift length of 0.1 s, **Figure 15** shows the result of RSI using the BonaFide algorithm (for response data from 0.0 s to 20.0 s). A comparison of the identified first fundamental frequency of the structure in the transverse direction using RSI-BonaFide [8] and RSI-Inversion-Oblique is also shown in **Figure 15**. The result of identification from both methods for data after 9.0 s is very similar. But using RSI-Inversion-Oblique the estimation can only begin at 9.0 s, while with the implementation of the RSI-BonaFide algorithm, the identification can begin from 5.0 s (see shaded area of Figure 15). The trend on the identified fundamental frequency between 5.0 s and 9.0 s looks lower than the trend of the identified frequency after 9.0 s because of the significant response of the structure.

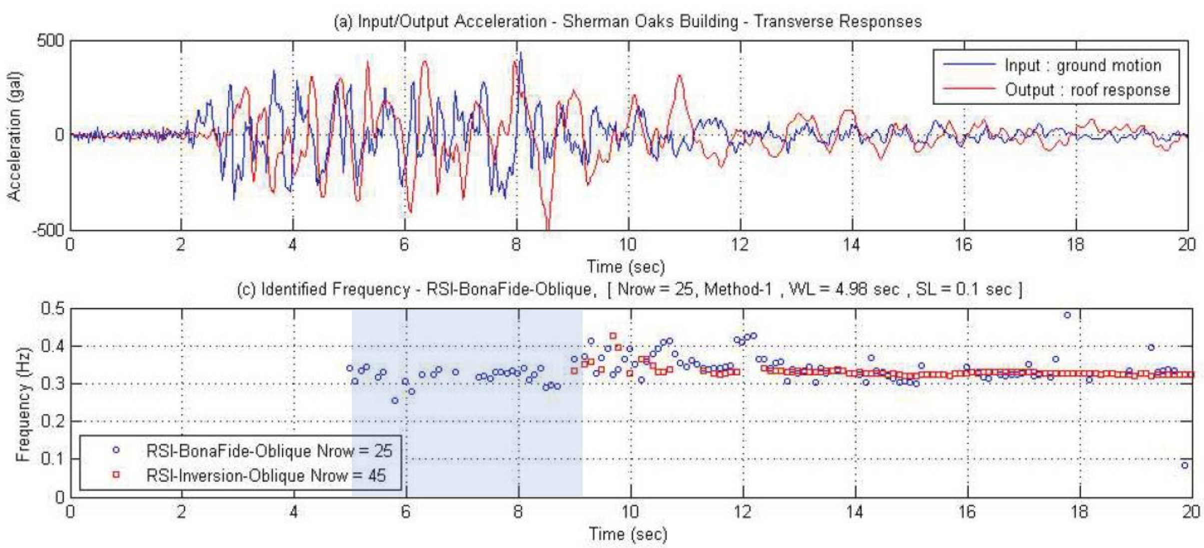


Figure 15: Comparison on the identified 1st modal frequency of Sherman Oak Building during Northridge earthquake (0.0 sec ~ 20 sec) using RSI-Inversion-Oblique and BonaFide algorithms.

Data Availability

The authors acknowledge accessing strong-motion data through the Center for Engineering Strong Motion Data (CESMD). The networks or agencies providing the data used in this paper are the California Strong Motion Instrumentation Program (CSMIP) and the USGS National Strong Motion Project (NSMP).

Acknowledgements

This research was supported by the U.S. Department of Conservation for the California Strong Motion Instrumentation Program (CSMIP) data interpretation project no. 1018-567. Some suggestions from the National Center for Research on Earthquake Engineering (NCREE), Taiwan, is also acknowledged.

References

[1] Viberg, M. (1994), *Subspace Methods in System Identification*, Proc. of SYSID'94, July, Copenhagen, Denmark, p1-12.
[2] Van Overschee, P. & De Moor, B. (1996), *Subspace identification for linear systems: Theory-Implementation-Applications.* Dordrecht, Netherlands: Kluwer Academic Publishers.
[3] Larimore, W. E. (1994), *The Optimality of Canonical Variate Identification by Example*, Proc. of SYSID '94, 4-6 July, Copenhagen, Denmark, 2, p151-156.
[4] Van Overschee, P. & De Moor, B. (1994), *N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems*, Automatica, 30(1), p75-93.
[5] Verhaegen, M. (1994), *Identification of the deterministic part of MIMO state space models given in innovations from input-output data*, Automatica, 30(1), p61-74.
[6] Kameyama, K. and A. Ohsumi, Y. Matsuura and K. Sawada. (2005), *Recursive 4SID-based identification algorithm with fixed input-output data size*, Int. J. of Innovative Computing, Information and Control, Vol.1, No 1, p17-33.
[7] Tamaoki, M., K. Akizuki and K. Oura. (2006), *Order and parameter estimation of time-varying system by subspace method*, Electrical Engineering in Japan, Vol, 157, No.2, p57-64.
[8] Chin-Hsiung Loh & Jun-Da Chen, "Tracking Modal Parameters from Building Seismic Response Data Using Recursive Subspace Identification Algorithm," *Published in Int. J. Earthquake Engineering & Structural Dynamics, DOI: 10.1002/eqe.2900,* 46:2163–2183 *(2017)*
[9] Jun-Da Chen & Chin-Hsiung Loh, "Two-stage damage detection algorithms of structure using modal parameters identified from recursive subspace identification," *Int. J. Earthquake Engineering & Structural Dynamics,* DOI: 10.1002/eqe.2980, Vol. 47 (3), pp. 573-593 March 2018.
[10] De Moor, B., Van Overschee, P., & Favoreel, W, "Algorithms for subspace state-space system identification: an overview," In Applied and Computational Control, Signals, and Circuits (pp. 247-311), Birkhäuser, Boston, MA. (1999)
[11] Boyd, S. and L. Vandenberghe, Convex Optimization (Appendix C.4.3), Cambridge University Press, 2004
[12] Hiroshi Oku, Hidenori Kimura, "A RECURSIVE 4SID FROM THE INPUT-OUTPUT POINT OF VIEW", Asian Journal of Control, Vol.1, pp258-269. (1999)
[13] H. Oku, G. Nijsse, M. Verhaegen, V. Verdult, "Change detection in the dynamics with recursive subspace identification", Proceedings of the 40th IEEE CDC, Orlando, Florida, pp2297-2302. (2001)
[14] Zhang, H.J., C. Thurber, and C. Rowe (2003), "Automatic P-Wave Arrival Detection and Picking with Multiscale Wavelet Analysis for Single-Component Recordings," *Bulletin of the Seismological Society of America* 93(5), 1904–1912.
[15] Reynders, E., Houbrechts, J., and De Roeck, G., "*Fully automated (operational) modal analysis," Mechanical Systems and Signal Processing*,'' 29, 228-250, 2012.
[16] Caicedo, J. M., "Practical guidelines for the natural excitation technique (NExT) and the eigensystem realization algorithm (ERA) for modal identification using ambient vibration,'' Experimental Techniques, 35(4), 52-58. (2011)