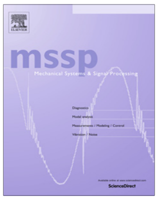
Mechanical Systems and Signal Processing 100 (2018) 501–523

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp



On the selection of user-defined parameters in data-driven stochastic subspace identification

C. Priori [a,*], M. De Angelis [a], R. Betti [b]

[a] Department of Structural and Geotechnical Engineering, Sapienza University of Rome, Via Eudossiana, 18, 00100 Rome, Italy
[b] Department of Civil Engineering and Engineering Mechanics, Columbia University, 640 SW Mudd, 4709, New York, USA

ARTICLE INFO

Article history:
Received 9 January 2017
Received in revised form 1 June 2017
Accepted 22 July 2017
Available online 5 August 2017

Keywords:
Output-only identification
Data-driven stochastic subspace identification
Laboratory test
In-situ test
Ambient vibration

ABSTRACT

The paper focuses on the time domain output-only technique called Data-Driven Stochastic Subspace Identification (DD-SSI); in order to identify modal models (frequencies, damping ratios and mode shapes), the role of its user-defined parameters is studied, and rules to determine their minimum values are proposed. Such investigation is carried out using, first, the time histories of structural responses to stationary excitations, with a large number of samples, satisfying the hypothesis on the input imposed by DD-SSI. Then, the case of non-stationary seismic excitations with a reduced number of samples is considered. In this paper, partitions of the data matrix different from the one proposed in the SSI literature are investigated, together with the influence of different choices of the weighting matrices.

The study is carried out considering two different applications: (1) data obtained from vibration tests on a scaled structure and (2) in-situ tests on a reinforced concrete building. Referring to the former, the identification of a steel frame structure tested on a shaking table is performed using its responses in terms of absolute accelerations to a stationary (white noise) base excitation and to non-stationary seismic excitations of low intensity. Black-box and modal models are identified in both cases and the results are compared with those from an input-output subspace technique. With regards to the latter, the identification of a complex hospital building is conducted using data obtained from ambient vibration tests.

1. Introduction

Structural Identification (St-Id) for civil infrastructure systems is the process which provides structural parameters and mathematical models of buildings, bridges, etc., from experimental measurements [1,2]. Traditional St-Id methods rely on the measurements of both the applied input and the corresponding response output and, for this reason, are referred to as input-output (I/O) St-Id methods. Several applications of such methods can be found in the literature [3–6].

On the contrary, output-only (O/O) St-Id methods use only the measured structural response and, because of the fact that they do not require measurements of the input, are perfectly suited for civil engineering applications. For buildings and bridges, since they must be kept operational during monitoring, the most commonly used measurements are those of the structural response induced by wind and/or traffic, excitations that are very difficult (if not impossible) to measure. For the O/O St-Id algorithms to work, only general information about the input (*e.g.* stationarity, white noise, etc.) are needed.

---

[*] Corresponding author.
E-mail addresses: ing.priori@gmail.com (C. Priori), maurizio.deangelis@uniroma1.it (M. De Angelis), rb68@columbia.edu (R. Betti).

http://dx.doi.org/10.1016/j.ymssp.2017.07.045
0888-3270/© 2017 Elsevier Ltd. All rights reserved.

Hence, the possibility of performing a structural identification without the need to measure the applied input excitation makes these O/O St-Id methods particularly suited for many applications within Structural Health Monitoring, especially when dealing with large structural systems. However, not having measurements of the input excitation will undoubtedly introduce some limitations in the analyses that can be performed.

O/O identification techniques can be classified in parametric and non-parametric, the first generally in time domain, the second in frequency domain [7–9]. Working in the frequency domain, the most common identification technique is the Peak Picking approach [10]: it takes its name from the fact that the frequencies of the structure under consideration are selected from the peaks of the corresponding output response spectra. On the contrary, the Frequency Domain Decomposition algorithm (FDD) and its enhanced version (EFDD) [11] rely on the Singular Value Decomposition (SVD) for the decomposition of the power spectral density matrices into multiple single-degree-of-freedom systems; although frequently used, these approach may encounter some difficulties in the identification of closely spaced modes. With regard to time domain techniques, the Polyreference Least Square Complex Exponential (PLSCE) algorithm and the Eigensystem Realization Algorithm (ERA) [12] extract the modal parameters of a structure from its impulse response. Another widely used time-domain approach is the Auto-Regressive Moving Average approach which, however, requires large computations due to a nonlinear optimization step in the algorithm. Other techniques involve the introduction of an observer to avoid the inversion of the large convolution matrix; among them, the Observer Kalman filter Identification (OKID) algorithm, often coupled with the ERA [13,14].

A widely used time domain class of algorithms is the one that follows the approach of Stochastic Subspace Identification (SSI). Among them, those procedures that directly use the output measurements to identify a system's model without the explicit computation of the covariance between outputs are called direct or data-driven methods [15]. Such Data-Driven SSI (DD-SSI) rely on the use of well-established linear algebra/geometry tools, such as orthogonal projection, singular value decomposition, QR factorization, etc., and on a certain number of user-defined parameters to extract first-order models of a system. A distinctive classification of such methods can be done by looking at the form of the weighting matrices that are involved in the computation of the observability range space [16]: the effects of these different choices are investigated in [17,18]. Since such algorithms are based on the assumptions of Linear Time-Invariant (LTI) systems and of the input representation as a stationary stochastic process, they are well suited in the analysis of systems whose measurements represent the ambient vibration response: however, few examples of DD-SSI involving the analysis of the structural response to non-stationary excitations (e.g. seismic excitation [19]) are also available in the literature.

As mentioned above, some parameters in DD-SSI are chosen by the user and they influence the effectiveness of the algorithm, leading to unreliable identifications if they are not properly selected. These parameters are linked to the dimensions of the Hankel matrix built with the measurements of the output response and of its subpartitions (namely, the past output submatrix and the future output submatrix). In particular, we are referring to:

- The number of block rows of the Hankel matrix, i;
- The number of block rows, g, of the past output subpartition $\mathbf{Y}_p$
- The number of block rows, h, of future output subpartition $\mathbf{Y}_f$, with:

$$g + h = i \tag{1}$$

- The number of columns of the Hankel matrix, j.

Usually, the parameters g and h are chosen equal to half the number of rows of the Hankel matrix i (*e.g.* $g = h = i/2$); in this case, named as symmetric partition, the two subpartitions of the Hankel matrix are equal in dimensions.

A rule on how to select the minimum value of i is proposed in [20] and in [21], while Pridham and Wilson [22] suggest a convenient value of j to consider. In [23], a study on the variability of the modal parameter estimates using DD-SSI is carried out, but only considering fixed values of the user-defined parameters and symmetric partition. A sensitivity analysis on the effect of i on the modal model identification is conducted in [24] using values of g and h typically used in the current literature. Thus, to the best of these authors' knowledge, there is no study/application that considers "asymmetric" partitions of the output Hankel matrix (*e.g.* $h < 0.5i$ or $h > 0.5i$) as well as there are no rules to determine the minimum values for each of the user-defined parameters. Since the quality of the identification results is strongly affected by the values of such parameters, it is important to address this issue and try to fill this gap.

The scope of this paper is the investigation of the role of the parameters that appear in DD-SSI and the definition of rules to guide the user to choose the most-appropriate values for such parameters. In addition, the influence of different choices of the weighting matrices and the performance of DD-SSI algorithms using output measurements of the structural response to non-stationary excitations are also investigated. DD-SSI algorithms are tested to identify first-order, state-space models as well as modal models (*e.g.* frequencies and damping ratios) of a steel frame laboratory structure and of a reinforced concrete hospital building. The identification of the steel frame scaled structure, tested on a shaking table, is performed using its response to a white noise excitation as well as to moderate seismic excitation so to test the SS-DDI algorithms' performance in the presence of non-stationary excitations. The results of such identifications are then compared with those obtained using an input-output subspace identification technique. In the case of the hospital building, the identification is carried out using the measurements of the building response to ambient excitation. The recommendations and conclusions of this