# Model Order Determination for Modal Parameter Estimation: Application to a Composite Fan Stage

Corentin Jorajuria, Claude Gibert, Cécile Esteves, Fabrice Thouverez

# Model order determination for modal parameter estimation: application to a composite fan stage

Corentin Jorajuria[1*], Claude Gibert[1], Cécile Esteves[2] and Fabrice Thouverez[1]

[1] Ecole Centrale de Lyon, CNRS, ENTPE, Laboratoire de Tribologie et Dynamique des Systèmes, UMR5513, 69130 Ecully, France
[2] Safran Aircraft Engines, 77550 Moissy-Cramayel, France

**Abstract**

*Experimental methods are necessary to quantify structural damping. Among these methods, modal testing of bladed disks are of particular interest because it provides an efficient experimental characterization of these structures which are critical parts of aeronautic turbojet engines. This work presents techniques to determine the optimal order of subspace state-space identification methods for modal parameter estimation of a realistic fan stage. Indeed, the order of the identification model is unknown during modal analysis requiring techniques to determine which model order should be used for parameter estimation. To investigate these techniques, several methods to estimate the optimal model order are reviewed. A specific focus is proposed on the determination of model order through statistical tests as these techniques have the benefits of providing indicators along with a threshold built over hypothesis testing and a probability of rejection. In order to further analyze model order determination techniques, some of these techniques are evaluated by means of a numerical model of a realistic fan stage, hence making possible to assess the performances of estimation methods close to experimental conditions. In particular, estimation in a context of high modal density as well as model order determination through "M-test" are addressed using this model. After assessing the implemented estimation method over the numerical model, this method is applied over experimental data obtained by performing modal tests over a full-scale composite rotating fan in vacuum condition.*

**Keywords**: modal analysis, subspace identification, bladed disks, aeronautic turbojet engines.

## Introduction

Experimental characterizations of structures are crucial to validate simulations as well as understanding and quantifying dissipation phenomena. On the other hand, fans bring the major part of thrust of aeronautic civilian turbojet engines, thus making the investigation of these structures an important part of the design of new turbojet engines. As modal analysis provides an efficient characterization of the dynamics of systems, modal analysis of aeronautic civilian fans is of particular interest to aeronautic turbojet designers. However, effects of energy dissipation, gyroscopic phenomena or mistuning can raise significant issues for the modal characterization of such systems. In order to perform efficient experimental analysis of these structures, there is a need for modal parameter estimation methods which overcome these challenges. One of the key challenges consists in providing an appropriate model to explain the observed dynamics. More precisely, for a given class of models, the choice of the optimal order of model is an important step to successfully estimate the modal parameters of the structure. Consequently, several techniques have been proposed to determine the optimal order of identification model concerning maximum of likelihood based estimation methods [1]. On the other hand, state-space subspace identification methods have the advantage of processing efficiently vector-valued time series, thus presenting advantages for the estimation of modal parameters of system exhibiting high modal density [2]. They also provide efficient data processing as well as numerical stable techniques for system identification [3]. This article proposes to focus on model order determination for modal parameter estimation methods applied to composite fan stage using state-space subspace identification methods.

The first section presents the implemented subspace state-space identification method used in this article. The second section describes different techniques to determine the number of states for the presented subspace state-space identification method. The third section deals with the assessment of estimation performances of the investigated techniques over a numerical model. Finally, the fourth section presents an application of these techniques over modal tests of an actual full-scale composite rotating fan in laboratory vacuum conditions.

## 1 Subspace identification methods

Subspace identification methods propose to estimate a system using a set of signals by identifying subspaces generated by these signals. Considering a discrete time signal $s_t \in \mathbb{R}^{n\times 1}$, the subspace generated from this signal can be defined as the set of linear combinations of $s_{t-k}$ where $k$ is a discrete time shift along with a condition to obtain a complete metric subspace.

$$\mathcal{H} = \left\{ \zeta_t = \sum_{k=-\infty}^{+\infty} G_k s_{t-k} \;\middle|\; G_k \in \mathbb{R}^{n\times n},\ \sum_{k=-\infty}^{+\infty} \|G_k\|_{\mathrm{F}}^2 < +\infty \right\} \quad (1)$$

The vector-valued series $\zeta_t$ form the vector elements of $\mathcal{H}$. They are built from finite linear combinations of $s_{t-k}$ with

*Corresponding author: corentin.jorajuria@ec-lyon.fr
**Received:** April 18, 2024, **Published:** August 28, 2024

matrix coefficients $G_k \in \mathbb{R}^{n\times n}$. The condition $\sum_k \|G_k\|_F^2 < +\infty$ ensures that $\mathcal{H}$ is a complete metric subspace with respect to a norm based on the variance of the signal $s_t$ which is assumed weakly stationary. At time $t$, the restriction of $\mathcal{H}$ to past samples is associated to the past dynamics and denoted $\mathcal{S}_{p,t}$ while the restriction to future samples is associated to the possible future dynamics and denoted $\mathcal{S}_{f,t}$.

$$\mathcal{S}_{p,t} = \left\{ \zeta_t = \sum_{k=0}^{+\infty} G_k s_{t-k} \;\middle|\; G_k \in \mathbb{R}^{n\times n},\ \sum_{k=0}^{+\infty} \|G_k\|_F^2 < +\infty \right\} \quad (2a)$$

$$\mathcal{S}_{f,t} = \left\{ \zeta_t = \sum_{k=1}^{+\infty} G_k s_{t+k} \;\middle|\; G_k \in \mathbb{R}^{n\times n},\ \sum_{k=1}^{+\infty} \|G_k\|_F^2 < +\infty \right\} \quad (2b)$$

More details over subspaces generated by signals can be consulted in a previous article [4]. Subspace identification methods model system dynamics through a state-space representation. In the following, the investigated systems are assumed linear, time invariant and with finite order. The order of the system, denoted $n_x$, is assumed to be a known parameter in this section. The system dynamics can be modeled by the following set of equations where $x_t \in \mathbb{R}^{n_x\times 1}$ is the state vector and $(w_t, v_t)$ are zero mean perturbation processes.

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + w_t \\ y_t &= Cx_t + Du_t + Sw_t + v_t \end{aligned} \quad (3)$$

$$\text{with} \quad \mathbb{E}\left[w_t w_s^T\right] = Q\delta_{ts}, \quad \mathbb{E}\left[v_t v_s^T\right] = R\delta_{ts} \quad \text{and} \quad \mathbb{E}\left[w_t v_s^T\right] = 0$$

In the above, $y_t \in \mathbb{R}^{n_y\times 1}$ and $u_t \in \mathbb{R}^{n_u\times 1}$ denote respectively the observation vector and the excitation vector at time sample $t$. The notations $A, B, C$ and $D$ denote the transition, control, observation and feedforward matrices. The stochastic process $w_t$ models the perturbations over the state transition equation while $v_t$ depicts perturbations specific to the observation equation. These processes are assumed independent of $x_t$ and $u_t$, uncorrelated in time and normally distributed with covariance matrices $Q$ and $R$. The term $Sw_t$ accounts for additional perturbations of the observation equation depending on the state perturbations [5]. In order to identify the state matrices of eq. (3), the state-space representation is considered over a finite time recursion $h$, which defines the horizon of the subspace identification method. This leads to stacked equations as follows.

$$y_{t|h} = \mathcal{O}_h x_t + \Psi_h u_{t|h} + F_h w_{t|h} + v_{t|h} \quad (4)$$

$$\text{with,} \quad \mathcal{O}_h = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{h-1} \end{bmatrix} \qquad y_{t|h} = \begin{bmatrix} y_t \\ y_{t+1} \\ \vdots \\ y_{t+h-1} \end{bmatrix}$$

$$\Psi_h = \begin{bmatrix} D & 0 & \dots & 0 \\ CB & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{h-2}B & CA^{h-3}B & \dots & D \end{bmatrix} \quad F_h = \begin{bmatrix} S & 0 & \dots & 0 \\ C & S & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{h-2} & CA^{h-3} & \dots & S \end{bmatrix}$$

The above equation introduces the extended observability matrix $\mathcal{O}_h$ which is linked to the observability concept in control theory [6]. The notations $u_{t|h}$, $v_{t|h}$, $w_{t|h}$ results of the same operations to obtain $y_{t|h}$ from $y_t$. Moreover, subspace identification methods build structured matrices from the signals of the state-space representation (eq. 3). The structured matrix built from the signal $y_t$ is denoted $Y_{t|h,N}$ and defined in the equation below.

$$Y_{t|h,N} = \begin{bmatrix} y_t & y_{t+1} & \dots & y_{t+N-1} \\ y_{t+1} & y_{t+2} & \dots & y_{t+N} \\ \vdots & \vdots & \ddots & \vdots \\ y_{t+h-1} & y_{t+h} & \dots & y_{t+h+N-2} \end{bmatrix} \in \mathbb{R}^{\dim(y)h\times N} \quad (5)$$

The order of time shift between columns, denoted $N$, is the number of time samples used by the subspace identification methods to estimate the ensemble average through the ergodicity hypothesis. Generally, this order verifies $N \gg \dim(y)h$ leading to consider wide rectangular matrices. One can notice that the linear combinations of rows of $Y_{t|h,N}$ span the subspace generated from $N$ samples of the signal $y_t$ over a finite horizon $h$. The same operations to define $Y_{t|h,N}$ from $y_t$ are used to define $U_{t|h,N}$, $V_{t|h,N}$ and $W_{t|h,N}$ from $u_t$, $v_t$ and $w_t$ respectively. As the investigated systems are assumed to have a finite order, these structured matrices are particularly suited to study subspaces generated by the signals of the state-space representation. Considering eq. (4) over $N$ time samples leads to the following equation between the structured matrices of the state-space representation.

$$Y_{t|h,N} = \mathcal{O}_h X_{t|1,N} + \Psi_h U_{t|h,N} + F_h W_{t|h,N} + V_{t|h,N} \quad (6)$$

$$\text{with,} \quad X_{t|1,N} = \begin{bmatrix} x_t & x_{t+1} & \dots & x_{t+N-1} \end{bmatrix}$$

In order to study the subspaces of eq. (2a-b), the above equation is split into past and future dynamics both over a horizon $h$. The structured matrices of eq. (6) are considered to belong to past dynamics and are denoted more briefly by $Y_p$, $U_p$, $V_p$, $W_p$ and $X_p$. Matrices of future dynamics are obtained by considering matrices of eq. (6) at time sample $t+h$ leading to structured matrices $X_f = X_{t+h|1,N}$, $Y_f = Y_{t+h|h,N}$ and similarly for $U_f$, $V_f$ and $W_f$. The state-space representation gives the following equations between these structured matrices.

$$X_f = A^h X_p + \mathcal{C}_h' U_p + \mathcal{F}_h' W_p \quad (7a)$$

$$Y_f = \mathcal{O}_h X_f + \Psi_h U_f + \underbrace{F_h W_f + V_f}_{\widehat{V}_f} \quad (7b)$$

$$\text{with} \quad \mathcal{C}_h' = [A^{h-1}B \ \dots \ AB \ B] \quad \text{and} \quad \mathcal{F}_h' = [A^{h-1} \ \dots \ A \ \mathrm{I}]$$

The equations (7a-b) give the transition from the $N$ sampled past dynamics of states $X_p$ to the future dynamics of states $X_f$ as well as an expression of the structured matrix $Y_f$ using the state-space representation. The symbol $\widehat{V}_f$ denotes the purely stochastic part in the decomposition of $Y_f$ of eq. (7b). As signals $w_t$ and $v_t$ are assumed independent of $x_t$ and $u_t$, one obtains in the limit $N$ tends to $+\infty$: $\frac{1}{N} X_f \widehat{V}_f^T = 0$ and $\frac{1}{N} U_f \widehat{V}_f^T = 0$ [7].

This article is concerned with model order determination issues of Numerical algorithms for SubSpace State-Space IDentification (`N4SID`) [7, 8] in order to estimate modal parameters of dynamical systems. These algorithms use a LQ decomposition of a specific matrix, denoted $H$, which is defined with $Y_f$, $U_f$ and $Z_p = \begin{bmatrix} U_p \\ Y_p \end{bmatrix}$.

$$H = \begin{bmatrix} U_f \\ Z_p \\ Y_f \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} \tag{8}$$

The above decomposition gives an efficient data compression as well as a useful expression of the structured matrix $Y_f$. Indeed, `N4SID` algorithms use this decomposition with the identification hypotheses (H1-3) to identify the state-space subspaces of the system.

$$\text{rank}(X_p) = n_x \tag{H1}$$
$$\text{rank}(U_p) = \text{rank}(U_f) = hn_u \quad \text{where } hn_u \geq n_x \tag{H2}$$
$$\text{span}'(X_p) \cap \text{span}'(U_p) = \text{span}'(X_f) \cap \text{span}'(U_f) = \{0\} \tag{H3}$$

In the above, $\text{span}'(\bullet)$ denotes the subspace generated from linear combinations of rows of the matrix in argument. These hypotheses can be used to obtain an expression of $Y_f$ as a decomposition of $Z_p$, $U_f$ and $Q_3$.

$$Y_f = L_{32}L_{22}^{\dagger}Z_p + \left(L_{31} - L_{32}L_{22}^{\dagger}L_{21}\right)L_{11}^{-1}U_f + L_{33}Q_3 \tag{9}$$

More details over the above result as well as the identification hypotheses (H1-3) can be consulted in a previous article [4]. Furthermore, the term $L_{33}Q_3$ in eq. (9) verifies $\frac{1}{N}U_f(L_{33}Q_3)^T = 0$ and $\frac{1}{N}Z_p(L_{33}Q_3)^T = 0$ thanks to the orthogonality properties between $Q_1$, $Q_2$ and $Q_3$ of the LQ decomposition. These relations make possible to identify the purely stochastic part of $Y_f$ between eq. (7b) and (9).

$$\widehat{V}_f = L_{33}Q_3$$

Moreover, it is possible to establish the identification term by term between eq. (7b) and (9) [2, p. 276] leading to an estimation of $\zeta \overset{\text{def}}{=} O_h X_f$, the state contribution in the decomposition of $Y_f$ in eq. (7b).

$$\zeta \overset{\text{def}}{=} O_h X_f = L_{32}L_{22}^{\dagger}Z_p \tag{10}$$

Equation (10) is described as the *optimal prediction* of the state contribution $\zeta$. Indeed, the right-hand side of this equation gives an estimation of $\zeta$ using the LQ decomposition and past structured matrices $U_p$ and $Y_p$. In order to estimate $O_h$ and $X_f$, `N4SID` algorithms propose to split this estimation into a product of two matrices using the singular value decomposition (SVD). The optimal prediction of eq. (10) can be moderated in order to be compliant with a complexity reduction principle [8]. This moderation can be realized through weighting matrices $W_1$, $W_2$ and the truncation of the SVD to the order $n_x$ which is assumed to be known.

$$W_1 \zeta W_2 = U\Sigma V^T \approx U_x \Sigma_x V_x^T \tag{11}$$

Equation (11) describes the *complexity reduction* step and leads to truncated SVD matrices $\Sigma_x \in \mathbb{R}^{n_x \times n_x}$, $U_x \in \mathbb{R}^{hn_y \times n_x}$ and $V_x \in \mathbb{R}^{N \times n_x}$. The weighting matrices $W_1, W_2$ can be used to obtain algorithms equivalent to other subspace identification methods; especially the Multi Output Error State sPace (MOESP) method [9] and the Canonical Variate Analysis (CVA) method [10]. In the following, the weighting matrices are chosen as $W_1 = W_2 = \text{I}$ which is equivalent to the original `N4SID` subspace identification method [7]. The observability matrix $O_h$ and the future state dynamics $X_f$ are estimated with a balance realization of the right-hand side of eq. (11).

$$O_h X_f \approx \left(U_x \Sigma_x^{1/2}\right)\left(\Sigma_x^{1/2} V_x^T\right) \tag{12}$$
$$\widehat{O}_h = U_x \Sigma_x^{1/2} \tag{12a}$$
$$\widehat{X}_f = \Sigma_x^{1/2} V_x^T \tag{12b}$$

The estimation of the state-space matrices $\widehat{A}$, $\widehat{B}$, $\widehat{C}$, $\widehat{D}$ can be obtained from the estimation of the observability matrix eq. (12a) or the state dynamics eq. (12b). Concerning the implemented estimation method used in this article, the estimation of the state-space matrices are obtained from $\widehat{X}_f$. More precisely, the state-space matrices are estimated by solving eq. (13).

$$\begin{bmatrix} X_1 \\ Y_0 \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_0 \\ U_0 \end{bmatrix} = \epsilon \tag{13}$$
$$\underset{A,B,C,D}{\text{argmin}} \ \|\epsilon\|_F^2$$
$$\text{with} \quad X_0 = \widehat{X}_{t+h|1,N-1} \qquad U_0 = U_{t+h|1,N-1}$$
$$X_1 = \widehat{X}_{t+h+1|1,N-1} \qquad Y_0 = Y_{t+h|1,N-1}$$

The matrices $\widehat{A}$, $\widehat{B}$, $\widehat{C}$, $\widehat{D}$ solutions of eq. (13) are as follows.

$$\begin{bmatrix} \widehat{A} & \widehat{B} \\ \widehat{C} & \widehat{D} \end{bmatrix} = \begin{bmatrix} X_1 \\ Y_0 \end{bmatrix} \begin{bmatrix} X_0^T & U_0^T \end{bmatrix} \left( \begin{bmatrix} X_0 \\ U_0 \end{bmatrix} \begin{bmatrix} X_0^T & U_0^T \end{bmatrix} \right)^{-1}$$

The estimated matrices $\widehat{Q}$, $\widehat{R}$ and $\widehat{S}$ are also obtained from $X_0$, $X_1$, $U_0$ and $Y_0$ [10].

The following section presents techniques to determine the order of the state-space representation which has been assumed to be a known parameter in this section.

## 2 Model order determination

The identification of the state-space representation with the estimation method presented in section 1 requires to determine the optimal order of truncation for the complexity reduction step. Indeed, the order of truncation, denoted $n_x$, should be equal to the order of the "true" state-space representation of the system dynamics which is denoted $\bar{n}_x$. The determination of this order can be obtained from the observation of the mean square error between the measurements and the synthesized response. In the dedicated

literature, this technique is sometimes referred as Final Prediction Error (FPE) [11]. The optimal order is then obtained as the order after which the mean square error is constant or does not decrease significantly. A similar order determination technique consists in observing the evolution of the singular values of $\zeta \overset{\text{def}}{=} O_h X_f$ [2, p. 166]. However, noise in the data set can jeopardize the detection of $\bar{n}_x$ with this technique. Indeed, an increase of the order of the identification method exceeding $\bar{n}_x$ will fit the noise dynamics leading to a decrease of the mean square error after $\bar{n}_x$ which can be similar to the decrease before $\bar{n}_x$, and thus preventing the detection of $\bar{n}_x$. When a subspace identification method uses too many states to identify the system, the identification is qualified as overfitting. In such cases, the modes fitting the noise dynamics are called numerical modes to mark the difference between modes fitting the system dynamics which are qualified as physical modes. Another kind of model order determination technique proposes to obtain the physical modes through stabilization charts. These techniques make the hypothesis that the physical modes are independent of the order of identification whereas the numerical modes are dependent of this order. Then, the physical modes are identified as the modes which are constant while increasing the order of the identification method. For such techniques, the optimal order is determined using a complexity reduction principle as the minimum order after which all physical modes are stabilized [12]. Both determination of model order through error minimization techniques and stabilization charts require a choice of threshold to determine if the error or the poles are stabilized. These choices are often made according to heuristic rules meaning that an adequate threshold is taken without showing that it is the optimal threshold.

Furthermore, other model order determination techniques propose to find the optimal order of the system by using an information criterion. These techniques have the advantage to determine the optimal order by minimizing a cost function based on a chosen information criterion. Several information criteria have been proposed. A selection of these model order determination techniques can be consulted in references [cf. 1]. These criteria include the AIC criterion which can be interpreted as a compromise between mean square error and the size of the model [13], the corrected AIC (AICc) which proposes a correction to prevent statistical bias when analyzing small sized samples [14] and the Minimum Description Length (MDL) based on information theory [15]. Moreover, the Bayesian Information Criterion (BIC) proposes to determine the optimal order based on the Bayesian estimation framework [16].

Another group of techniques uses statistical tests to determine the optimal model order. These techniques have the advantages to propose a threshold based on a probability of rejection of an investigated hypothesis. The "F-test", named by G. W. Snedecor after R. Fisher, makes possible to compare the variance of two random variables to detect

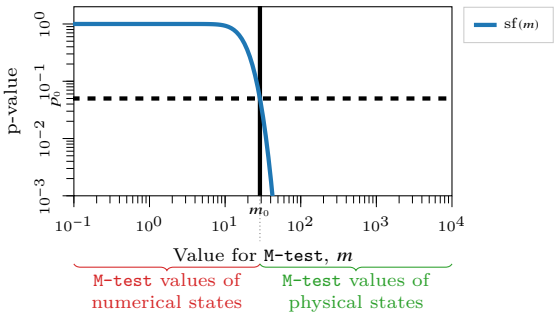



**Fig. 1:** *Example of $m_0$ threshold computation for* `M-test`.

changes [17]. This test can be used over identification results of increasing order to detect when the variance of the error does not change significantly making possible to determine $\bar{n}_x$ as the minimum order after which the variance of the error is considered to remain unchanged. M. Aoki and al. [18] have proposed another statistical test based on theoretical results over the distribution of correlation coefficients and often mentioned as "C-test" in the dedicated literature. This article focuses more precisely over a similar test, denoted "M-test" [19]. The `C-test` and `M-test` have the advantages of investigating the statistical significance of each estimated states independently. Compare to the `C-test`, the `M-test` has the advantage of focusing over an indicator asymptotically distributed according to the $\chi^2$ law without correction factor. It proposes to investigate each column of the observability matrix by testing a statistical hypothesis defined as follows.

$H_0^{(j)}$: "column $j$ of $O_h$ is associated to a numerical mode"

The `M-test` uses an indicator $M_j$ built as a specific scalar product of the column $j$ of the observability matrix with itself. The column $j$ of the observability matrix is denoted $O_h^{(j)}$.

$$M_j = O_h^{(j)^T} \operatorname{cov}\left[O_h^{(j)}\right]^{-1} O_h^{(j)}$$

More precisely, the column $j$ of the observability matrix can be expressed with the column vector $e_j$ having all coefficients to zero expect in position $j$ where it takes value 1 (cf. eq. 14). Moreover, if the hypothesis $H_0^{(j)}$ is verified, then $j > \bar{n}_x$. Indeed, using the property of strict nesting [20] coming from the orthogonality properties of the SVD, it is possible to split the state vector $x_t$ at time sample $t$ into ${x^{[x]}}_t \in \mathbb{R}^{\bar{n}_x \times 1}$ distributed according to the dynamics of physical states and ${x^{[w]}}_t \in \mathbb{R}^{(n_x - \bar{n}_x)\times 1}$ distributed according to the dynamics of numerical states.

$$O_h^{(j)} = \begin{bmatrix} Ce_j \\ CAe_j \\ \vdots \\ CA^{h-1}e_j \end{bmatrix} \qquad x_t = \begin{bmatrix} {x^{[x]}}_t \\ {x^{[w]}}_t \end{bmatrix} \tag{14}$$

Using all the identification hypotheses and assuming $H_0^{(j)}$ is satisfied, the coefficients of $O_h^{(j)}$ are zero-mean, normally distributed and independent each one to another. Indeed,

with these hypotheses $O_h^{(j)}$ as expressed in eq. (14) is the free response with specific initial conditions for the dynamics of $x^{[w]}{}_t$ which is distributed as $w_t$. Then, $M_j$ is distributed according to the $\chi^2{}_{(hn_y)}$ law (assuming $H_0^{(j)}$ is true). As we know the cumulative distribution of $\chi^2$ law, it is possible to get, given a probability of rejection $p_0$, the threshold $m_0$ that the indicator $M_j$ should not overcome if the hypothesis $H_0^{(j)}$ is true. The computation of this threshold makes use of a survival function (cf. eq. 20). Figure 1 shows an example of $m_0$ threshold computation choosing $p_0 = 0.05$ and $hn_y = 18$. One of the difficulty to apply the `M-test` consists in getting an estimation of $\mathrm{cov}\left[O_h^{(j)}\right]^{-1}$. This article investigates the possibility to use a similar estimation of $\mathrm{cov}\left[O_h^{(j)}\right]^{-1}$ for `N4SID` compare to the subspace identification method used by Dorfman and al. [19] which performs estimation without measurements of the excitations.

$$\mathrm{cov}\left[\mathrm{vec}\left(O_h\right)\right]^{-1} = \left(\mathrm{cov}\left[\widehat{V}_f\right]^{-1}\right) \otimes \left(\widehat{X}_f \widehat{X}_f^T\right)$$

More details to obtain this estimator are proposed in appendix A. Using the above results, the indicator of `M-test` for state $j$ is estimated as follows.

$$M_j = e_j^T \left(\widehat{X}_f \widehat{X}_f^T\right) e_j \; O_h^{(j)\,T} \left(\mathrm{cov}\left[\widehat{V}_f\right]^{-1}\right) O_h^{(j)} \qquad (15)$$
$$\mathrm{cov}\left[\widehat{V}_f\right] = F_h \mathrm{diag}_h(\widehat{Q}) F_h^T + \mathrm{diag}_h(\widehat{R})$$

Equation (15) assumes that $v_t$ and $w_t$ are white noise and independent one to another. Finally, if the indicator $M_j$ takes significant value, *i.e.* superior to $m_0$, the hypothesis $H_0^{(j)}$ is rejected leading to select state $j$ as a physical state.

# 3 Evaluation with a numerical model

This section proposes to assess the estimation performances of the `N4SID` algorithm presented in section 1 combined with a determination of the optimal model order through `M-test`. A first subsection describes the numerical model used to assess estimation performances while the second subsection shows estimation results obtained over the numerical model.

## 3.1 Experimental Modal Model

The proposed numerical model uses experimental results of a modal analysis of a full-scale rotating fan in vacuum condition. Measurements of the system are obtained from strain gauges over blades excited by a forward travelling wave with steps in frequency. More details concerning these modal tests are given in section 4.1. The modal parameters defining the modal model have been estimated using a method based on Least Square Rational Functions (`LSRF`) and described in a previous article [21]. Figure 2 shows

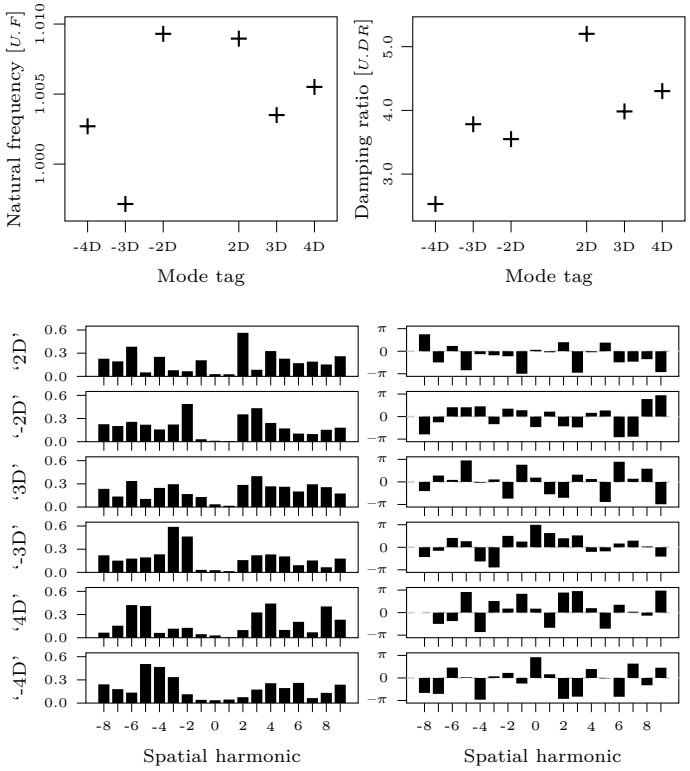



**Fig. 2:** *Definition of the experimental modal model*

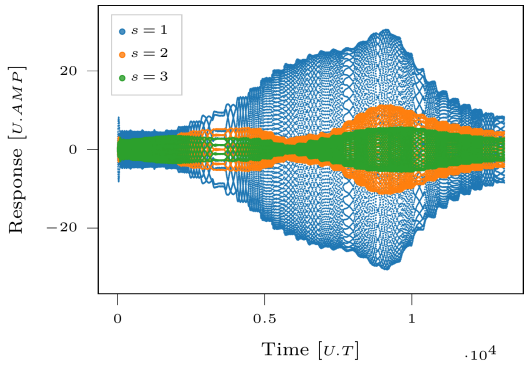
the poles of the experimental modal model by representing the natural frequencies and damping ratios of the $n_m = 6$ modes of the model. Notations $U.F$ and $U.DR$ stand for the normalized unit of frequency and damping ratio used in this article to preserve industrial confidentiality. Figure 2 also illustrates the spatial analysis of the associated mode shapes which take into account mistuning phenomena as observed in the experimental results. Each mode is assigned a mode tag reflecting the dominant spatial harmonic of its estimated mode shape. The modal model is expressed with the modal coordinates $q \in \mathbb{C}^{2n_m \times 1}$.

$$\begin{aligned} \dot{q}(t) &= \Lambda q(t) + \widehat{\Theta}^T \widehat{g} + w(t) \\ y(t) &= \Phi q(t) + v(t) \end{aligned} \qquad (16)$$

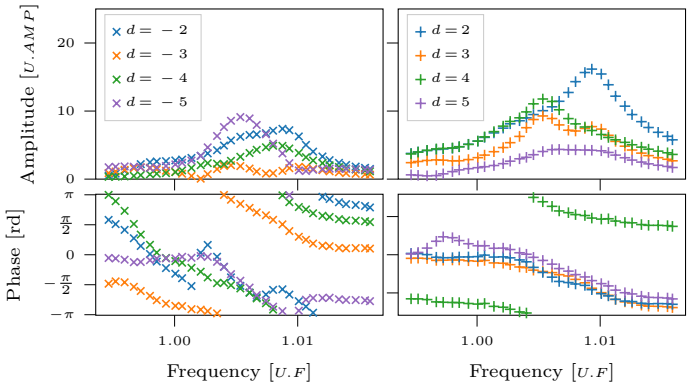
In eq. (16), the matrix of eigenvalues $\Lambda \in \mathbb{C}^{2n_m \times 2n_m}$, the matrix of mode shapes $\Phi \in \mathbb{C}^{n_y \times 2n_m}$, the augmented signal of forces applied to the system $\hat{g}(t) \in \mathbb{C}^{2n_g \times 1}$ and the augmented matrix of left mode shapes $\widehat{\Theta} \in \mathbb{C}^{2n_g \times 2n_m}$ have been introduced. They verify the following relationships.

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_1^* & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_{nm} & 0 \\ 0 & 0 & \dots & 0 & \lambda_{nm}^* \end{bmatrix} \qquad \Phi = \begin{bmatrix} \varphi_1 & \varphi_1^* & \dots & \varphi_{n_m} & \varphi_{n_m}^* \end{bmatrix}$$
$$\widehat{\Phi} = \begin{bmatrix} \Phi\Lambda \\ \Phi \end{bmatrix} \qquad \widehat{\Theta} \underset{(17a)}{=} \begin{bmatrix} -\Theta\Lambda \\ \Theta \end{bmatrix} \qquad \widehat{g} = \begin{bmatrix} g \\ 0 \end{bmatrix} \qquad (17)$$
$$\widehat{\Theta}^T \widehat{g} = -\Lambda \Theta^T g \underset{(17b)}{=} -\Lambda \Phi^H g$$

Equation (17) uses superscript $.^H$ to denote the transpose conjugate of a matrix and $g(t)$ the signal of forces applied to




**(a)** *Temporal response for three consecutive sectors.*




**(b)** *Wave transform of the response for* $d=\pm 2,3,4$ *and* 5.

**Fig. 3:** *Numerical model response using a travelling wave excitation having one spatial harmonic at* $d=2$.

the system. Also, the equal signs denoted by (17a-b) assume a symmetric stiffness matrix as well as the hypothesis $\Theta=\Phi^*$ which is verified by the mode shapes of a cyclic conservative system. Equation (16) is integrated into the real domain as follows.

$$\begin{aligned} \dot{s}(t) &= \Lambda^{\mathbb{R}} s(t) + B^{\mathbb{R}} g(t) + w^{\mathbb{R}}(t) \\ y(t) &= \Phi^{\mathbb{R}} s(t) + v(t) \end{aligned} \tag{18}$$

$$\text{with, } \Lambda^{\mathbb{R}} = P^H \Lambda P \quad B^{\mathbb{R}} = -\Lambda^{\mathbb{R}} \Phi^{\mathbb{R}H} \quad P = \text{diag}(P^{(1)}, \ldots, P^{(1)})$$
$$\Phi^{\mathbb{R}} = \Phi P \qquad s = P^H q \qquad P^{(1)} = \frac{1}{2}\begin{bmatrix} 1-j & 1+j \\ 1+j & 1-j \end{bmatrix}$$

In the above, the perturbation process $w^{\mathbb{R}}$ verifies a similar relationship as the state vector of the real modal form: $w^{\mathbb{R}} = P^H w$. The numerical model simulates the response $y(t)$ from an excitation $g(t)$. Both of these signals have been taken with size $n_s \times 1$ where $n_s$ denotes the number of sectors. The integration is performed with a Runge-Kutta 45 (RK45) scheme [22] making possible to control the accuracy of the integration error at order 4 and the time step at order 5. The relative and absolute tolerance have been verified by comparison with an analytical computation taking mode shapes with a unique spatial harmonic and no perturbation processes ($v=w=0$). Figure 3a illustrates the numerical responses obtained with this model for a force signal $g(t)$ as a sine with steps in frequency and taking a forward travelling wave pattern with two nodal diameters. Figure 3b is

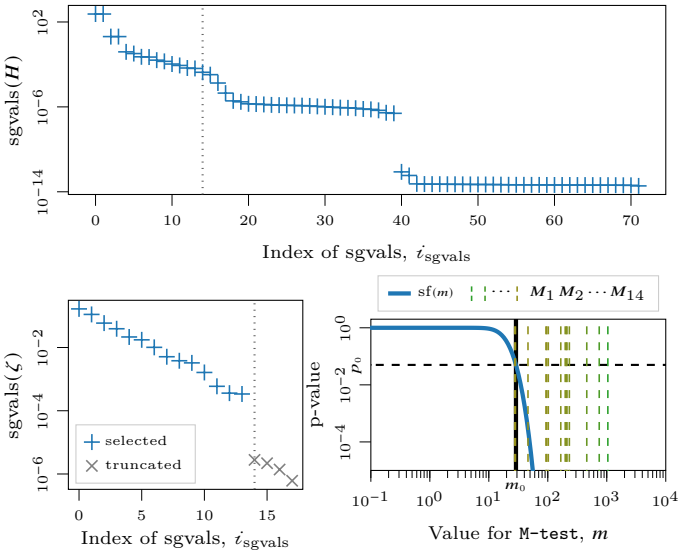



**Fig. 4:** *Model order determination by* `M-test` *over signals from the experimental modal model.*

the spatial transform of the estimated frequency response functions obtained by computing the discrete Fourier transform of the response over $n_{\text{bloc}}$ periods at the end of each frequency step divided by the Fourier transform over the same samples for the excitation signal over the reference sector. This figure shows that the harmonic component $d=2$ is higher than the other harmonic components of the response. However, the other harmonic components take significant values as the mode tag '2D' has a shape with significant values for harmonics different from $d=2$ and the other modes are also excited as they have significant values at spatial harmonic $d=2$. This test case has been studied with a level of noise (eq. 19) $\sqrt{\texttt{NSR}} \approx 2{\cdot}10^{-3}$ which can be considered relatively low compared to expected noise perturbation observed by comparing the level of background noise with the level of measured response.

$$\texttt{NSR} = \frac{\mathrm{P}_\varepsilon}{\mathrm{P}_z} \tag{19}$$
$$\text{with, } \mathrm{P}_m = \sum_{t=0}^{N-1} m_t^2 \quad \text{for } m = \varepsilon, z$$

In the above, $\varepsilon$ denotes the noise perturbation, and $z$ the signal. The NSR is taken similar for the perturbation over the observation equation $(\varepsilon, z) = (v, y)$ and the state transition equation $(\varepsilon, z) = (w^{\mathbb{R}}, \Lambda^{\mathbb{R}} s)$.

## 3.2 Estimation Performances Assessment

The estimation method presented in section 1 and the model order determination through `M-test` has been applied over the signals generated by the numerical model. Figure 4 shows model order determination results obtained with a horizon chosen at $h=1$. The top figure illustrates the decrease of the 72 singular values of the data matrix $H$

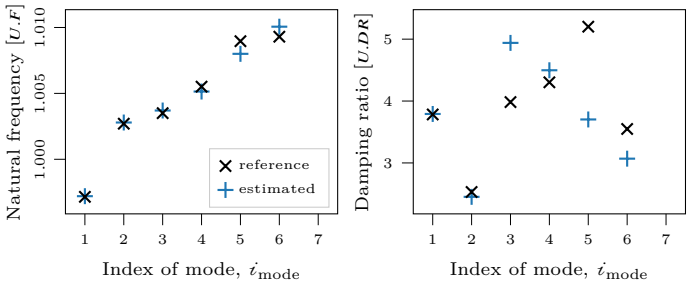



**Fig. 5:** *Estimation of modal parameters of the numerical model with* `N4SID` *using model order determination through* `M-test`.

($n_y = n_u = 18$). This figure shows several changes of the evolution of the singular values with respect to $i_{\text{sgvals}}$, their index in the SVD decomposition. One can observe that the determination of the optimal model order from this analysis is not straight forward. To go further into the optimal model order determination, a first `M-test` is performed over all the possible $hn_y = 18$ states, meaning that no truncation is performed at the complexity reduction step (eq. 11). This first statistical test indicates that only $\bar{n}_{x,\text{m-test}\,0} = 14$ states were physical states. The bottom figures (fig. 4) show the model order determination results for an estimation selecting 14 states in the complexity reduction step. More specifically, the bottom left figure shows singular values evolution of $\zeta$, the contribution of states in the decomposition of $Y_f$. It exhibits a specific change of singular values evolution after the selected states $i_{\text{sgvals}} = 0..13$ giving indication to confirm that the truncated states should not be selected. The bottom right figure gives a more objective confirmation over the determination of the optimal order by representing the `M-test` values $M_j$, for $j$ indexing all the 14 selected states. This figure also represents the survival function of the $\chi^2_{(hn_y)}$ law for $hn_y = 18$. The survival function is denoted $\text{sf}(x)$, and defined as $\text{sf}(x) = 1 - \text{cdf}(x)$ where "cdf" is the cumulative distribution function of the $\chi^2_{(hn_y)}$ law. The survival function gives the probability of having the indicator $M$ superior to a value $m$.

$$\text{sf}(m) = \Pr(M > m) \tag{20}$$

In the above, $\Pr(\bullet)$ denotes the probability of the event in argument. In this article, the rejection probability of the `M-test` is fixed at $p_0 = 0.05$. The intersection of the horizontal line at $p_0$ with the $\chi^2$ survival function gives the threshold over `M-test` values, denoted $m_0$ and verifying $\Pr(M > m_0) = p_0$. If the hypothesis $H_0^{(j)}$ is true for a state $j$, the `M-test` value $M_j$ should be low. More precisely, the probability of having $M_j$ higher than $m_0$ is $p_0 = 0.05$. Figure 4 shows that, for all the selected states, the indicator $M_j$ takes values higher than $m_0$. These values are too high to confirm hypothesis $H_0^{(j)}$ for $j$ indexing the selected states which leads to reject this hypothesis for each of these states. Hence, all the selected states are considered as physical states, and the determination of the optimal order through `M-test` leads to $\bar{n}_{x,\text{m-test}} = 14$. As the theoretical optimal order is $\bar{n}_x = 12$, the order determination through `M-test` is considered close to the expected model order determination. This model order determination has required two `M-tests`. It can be described through an iterative procedure defined by the below list of instructions.

Estimate $\mathcal{O}_h$ and $\mathcal{X}_f$ from eq. (12) selecting all states ($n_x = hn_y$)
  *output:* $\widehat{\mathcal{O}}_h, \widehat{\mathcal{X}}_f$
Initialize loop index: $i = 0$
**repeat**
  Perform `M-test` over the $n_x$ selected states
    *output:* $\bar{n}_{x,\text{m-test}\,i}$
  ▷ $\bar{n}_{x,\text{m-test}\,i}$ *number of physical states verifies* $\bar{n}_{x,\text{m-test}\,i} \leq n_x$
  **if** $\bar{n}_{x,\text{m-test}\,i} < n_x$ **then**
    converged = `False`
    $n_x \leftarrow \bar{n}_{x,\text{m-test}\,i}$
    Update $\widehat{\mathcal{O}}_h, \widehat{\mathcal{X}}_f$ selecting $n_x$ states in truncated SVD eq. (11-12)
    Increment loop index: $i \leftarrow i + 1$
  **else**
    converged = `True` ▷ *all selected states are physical*
**until** converged

**Alg. 1:** *Iterative* `M-test` *for order determination.*

After the determination of the optimal model order, the estimated modal parameters have been compared to the reference modal parameters defining the numerical model. Figure 5 shows the estimation results for natural frequencies and damping ratios. Reference and estimated modes are sorted by increasing values of natural frequencies with a mode index denoted $i_{\text{mode}}$. Estimation results of natural frequencies are close to reference natural frequencies from $i_{\text{mode}} = 1$ to 6. Mode $i_{\text{mode}} = 7$ takes a natural frequency value much higher than excited frequency band leading to consider this mode has a numerical mode despite the fact that the `M-test` detects it as a physical mode. The natural frequency of modes $i_{\text{mode}} = 1..6$ is estimated with a low relative error $|\Delta f/f| \leq 0.1\,\%$. The damping ratio estimation of modes corresponding to $i_{\text{mode}} = 1, 2, 4$ are close to reference values with a relative error $|\Delta\xi/\xi| \leq 5\,\%$ whereas modes corresponding to $i_{\text{mode}} = 3, 5, 6$ are less close to reference values exhibiting a relative error between 14 and 29 %. Estimated natural frequencies and damping ratios are considered close to the reference values despite few modes which can exhibit estimation errors over damping ratios.

# 4 Application to a composite rotating fan

The method presented in previous sections is applied over experimental data from a modal test of a composite rotating fan in vacuum condition. The first subsection describes the experimental procedures used to obtain the experimental data. The second subsection presents the modal parameter estimation results.

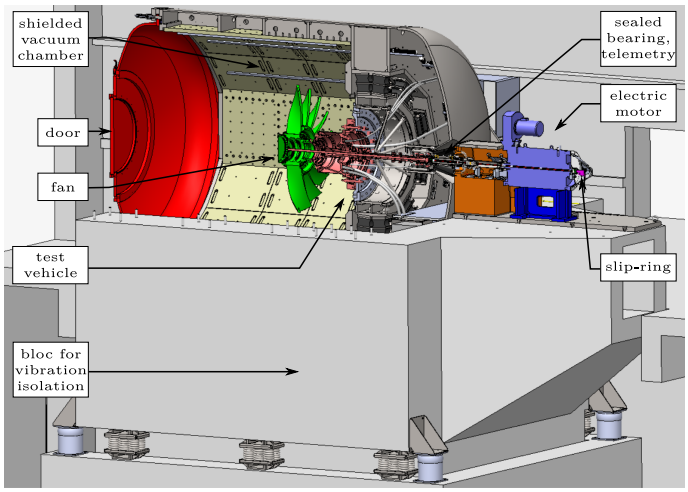



**Fig. 6:** *Cross-sectional view of test rig PHARE#1.*

## 4.1 Experimental facility and protocol

Experimental data are collected using a test rig dedicated to the measurements of full-scale machine dynamics in rotation and vacuum conditions. A cross-sectional view of the experimental facility, called PHARE#1, is presented in fig. 6. The fan specimen, mounted on the shaft of the test vehicle, is placed inside the vacuum chamber. This chamber has shielded walls with a useful diameter of 3.0 m and 3.6 m along the axis. It is strongly fixed to a seismic concrete bloc mounted on suspension springs and dampers in order to isolate the test rig against environmental vibrations and to prevent any transmissions of critical vibration levels to the building in case of heavy accidental loads. The shaft is driven by a 700 kW electric motor which can go up to the operating speed of the fan specimen and allows sufficiently fast ramp up and down in vacuum condition for experimental and security requirements. The driving shaft passes through the wall of the chamber with a special dynamic sealing which uses a circumferential carbon ring seal. A rotary vane vacuum pump coupled with a Roots pump maintain the pressure of the vacuum chamber at 0.1 mbar.

Each blade of the investigated fan specimen is instrumented with a set of interconnected piezoelectric patches composed of Lead Zirconate Titanate (PZT-5H) as illustrated in fig. 7. This figure also provides a picture of the fan specimen installed for tests. Blades are numbered according to the rotation direction from $p=0$ to $n_s-1$ with $n_s$ denoting the number of sectors. The excitation system generates forward travelling wave by controlling the phase difference between the piezoelectric patches of each of the blades [23]. Moreover, the excitation signal is chosen with steps in frequency leading to the following expression for the excitation signal of blade indexed by $p$.

$$u_p(t) = a \sin\left(2\pi f(t) t - \frac{2\pi}{n_s} dp\right)$$

In the above, $a$ denotes the amplitude of the excitation signal, $d$ denotes the spatial harmonic and $f(t)$ is a stepped function from $f_{\text{ini}}$ to $f_{\text{end}}$ by steps of $\delta f$ and performing $n_{\text{exci}}$ oscillations for each frequency step. More details over the experimental rules used to choose these parameters are provided in a previous article [21]. Excitation signals are amplified, passed into the rotating frame using an assembly of 18 circuits slip ring and passed through the vacuum chamber wall using a set of wires and connector stages placed inside the hollow shaft. Each of the blades is also instrumented with a strain gauge close to the leading edge (fig. 7). Wires route the strain measurements through the hollow shaft to the embedded telemetry unit which sends measurement signals to the fixed telemetry unit through radio waves. Strain measurements are simultaneously sampled at $f_s = 2{\cdot}10^3$ Hz. The experimental protocol aims at measuring frequency responses of the fan at constant conditions. The overall behavior of the test rig is monitored by numerous dynamic, displacement and thermal sensors. Moreover, the modal tests have been performed at stabilized rotation speed.

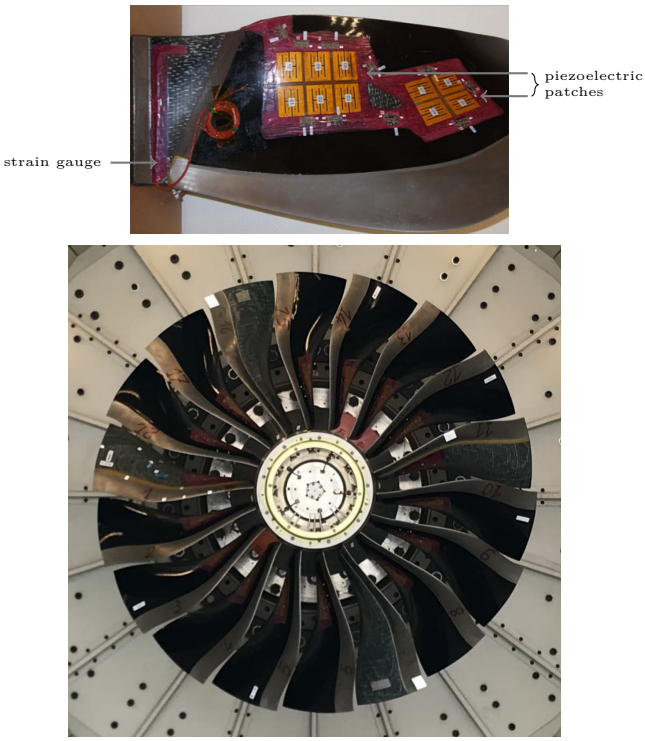
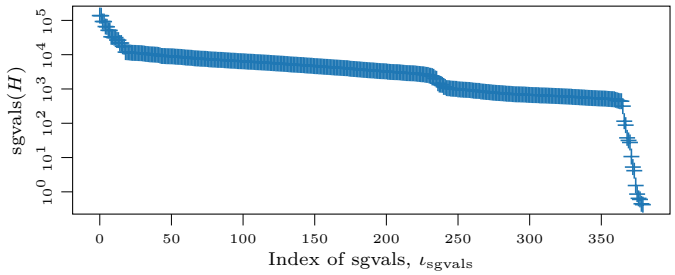


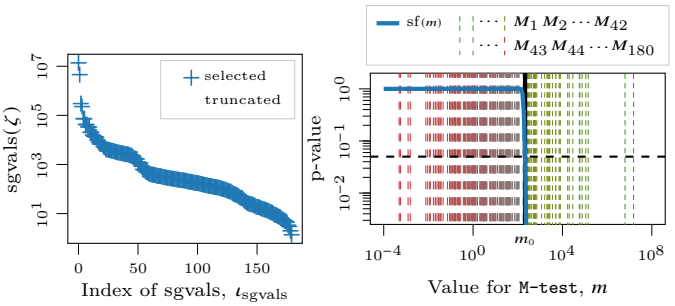

**Fig. 7:** *Blade and fan instrumentation.*

## 4.2 Estimation Results

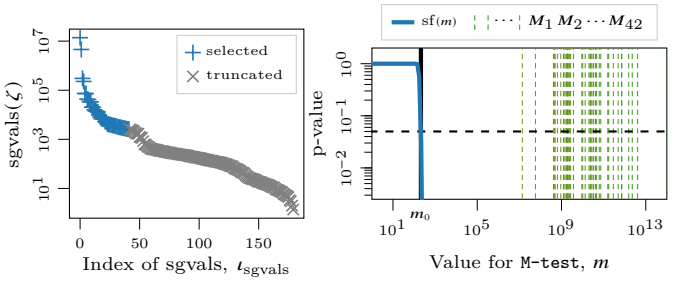
The experimental protocol described in the previous subsection 4.1 is particularly adapted to measure frequency responses of a rotating fan in vacuum condition. This article investigates the possibility to use subspace identification methods to estimate the modal parameters of the fan specimen with these measurements. To this end, the modal parameter estimation procedure using `N4SID` and a model order determination through `M-test` has been applied to




**(a)** *Singular value of data matrix.*




**(b)** *Iteration 0.*




**(c)** *Iteration 1.*

**Fig. 8:** *Model order determination by* `M-test` *over a modal test of a full-scale composite rotating fan.*

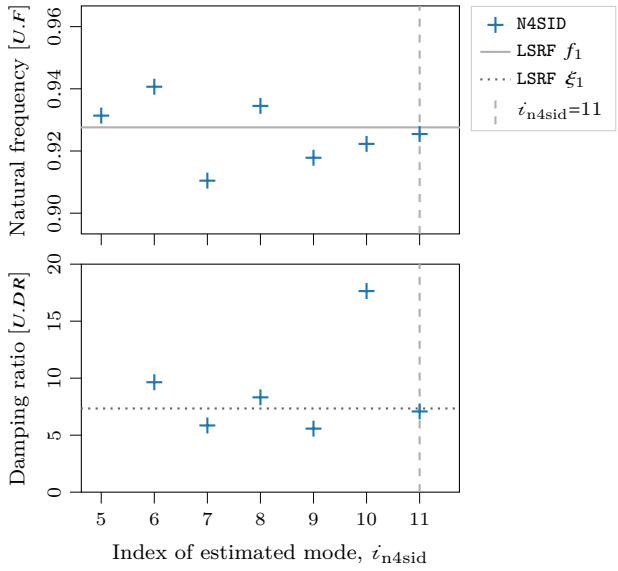



**Fig. 9:** *Estimation results with* `N4SID` *and model order determination through* `M-test`.

modal tests and compared to estimation results obtained over the same modal tests with a method based on Least Square Rational Function (LSRF) [21]. This section describes more precisely the modal parameter estimation results obtained over a modal test targeting the first bending mode at $50\%\Omega_n$, where $\Omega_n$ denotes the nominal rotation speed of the fan. The excitation has been chosen with a spatial pattern as a forward travelling wave with two nodal diameters. Experimental signals are composed of the strain measurements close to leading edge for each of the $n_s = 18$ sectors as well as the excitation signal sent to the piezoelectric patches of the reference sector. The horizon of the subspace identification method has been chosen as $h = 10$. In order to reduce the computational cost of the subspace estimation method, the experimental signals have been resampled at $f_s' = 2.7\ U.F$ after applying an ideal numerical filter. Figure 8a illustrates the singular values of the data matrix $H$ (eq. 8) possessing $h(2n_y + 2n_u) = 380$ singular values ($n_y = 18$ and $n_u = 1$). Figures 8b-c show more specifically the model order determination procedure through iterative `M-test` (algorithm 1). Figure 8b shows model order determination results of a first `M-test` applied on estimation results using `N4ISD` without any truncation at the complexity reduction step (eq. 11), so selecting $n_x = hn_y = 180$ states. The right figure of fig. 8b shows `M-test` values obtained for these selected states as well as the threshold $m_0$ computed from the survival function of $\chi^2_{(hn_y)}$ and rejection probability $p_0 = 0.05$ leading to determine $\bar{n}_{x,\text{m-test}\,0} = 42$ states as physical states. Figure 8c brings additional model order determination results over fig. 8b analysis by representing the `M-test` values for `N4SID` estimations with a truncation to $n_x = \bar{n}_{x,\text{m-test}\,0} = 42$ states. This figure shows that $\bar{n}_{x,\text{m-test}\,1} = 42$, as for each of the 42 states the associated `M-test` value is higher than the threshold $m_0$. Hence, all the selected states used to obtain these estimation results are associated with physical states which concludes the model order determination procedure. Modal parameter estimation results with `N4SID` and the above described model order determination procedure have been compared with estimations obtained using another method which is based on `LSRF` applied over the measured frequency response functions of the same modal test. Moreover, among all the estimated modes, 7 modes have their natural frequency in the interval $[f_{\text{ini}}, f_{\text{end}}]$. Figure 9 represents the natural frequency and damping ratio of these modes with respect to their index of estimation denoted $\imath_{\text{n4sid}}$. Estimation of the natural frequency and the damping ratio of the mode responding the most among the estimated modes with `LSRF` are also reported on this figure. These values are respectively denoted $f_1$ and $\xi_1$. Among the modes estimated with `N4SID`, mode indexed $\imath_{\text{n4sid}} = 11$ takes a natural frequency close to $f_1$ and a damping ratio also close to $\xi_1$. Indeed, the relative difference between these values is $|\Delta f/f| = 0.2\,\%$ and $|\Delta\xi/\xi| = 5\,\%$. The same estimation procedure applied over a second bending mode has also led to `N4SID` estimated modes containing

a mode which has close natural frequency and damping ratio values compare to the LSRF estimated mode which responds the most; however with a larger relative difference for damping ratio: $|\Delta f/f| = 0.1\,\%$ and $|\Delta\xi/\xi| = 33\,\%$. These results indicate that subspace identification method N4SID combined with a model determination through M-test can also provide estimation of natural frequency and damping ratio to characterize the dynamics of a full-scale rotating fan in vacuum condition excited with a forward travelling wave. Compare to other estimation methods, the investigated subspace identification method has the advantage of proposing an objective model order determination procedure based on statistical tests over the columns of the observability matrix.

## Conclusion

In this article, an estimation procedure based on N4SID algorithms and model order determination through iterative M-test is proposed. In a first step, N4SID algorithms are explained and several model order determination techniques are reviewed. The determination of the truncation order through iterative M-test has the advantage of proposing an optimal order using a threshold based over a rejection probability. This estimation procedure is assessed using a numerical model defined from experimental modal analysis of a full-scale composite fan in rotation and vacuum conditions. Natural frequencies estimated over the numerical model are close to the reference values with a relative error inferior to 0.1 %. Application over experimental data shows that estimated modes contain a mode having natural frequency and damping ratio which can be considered close to the most responding mode estimated with another estimation method.

This study provides promising results for modal parameter estimation of rotating bladed disks with subspace identification methods. Indeed, these methods have several advantages to face the estimation challenges imposed by the modal analysis of these systems. In particular, the efficient processing of vector-valued time series makes these methods interesting candidates to study system dynamics exhibiting high modal density. In addition, the experimental analysis of weakly nonlinear phenomena for which linearization assumption holds or time varying effects, can benefit from the efficient model order determination techniques proposed by these methods. Finally, it would be interesting to pursue the assessment of these estimation methods for experimental characterizations of stages of rotating machine exhibiting time varying effects. Indeed, the direct use of temporal measurement data can be an important asset to study these phenomena.

## Nomenclature

| | |
|---|---|
| SVD | Singular Value Decomposition |
| N4SID | Numerical algorithms for SubSpace State-Space IDentification |
| LSRF | Least Square Rational Function |
| $U.F, U.T$ | unit of frequency and time ($U.T = 1/U.F$) |
| $U.DR$ | unit of damping ratio |
| $x^*$ | complex conjugate of $x$ |
| $A^T, A^H$ | transpose matrix and Hermitian conjugate of $A$ |
| I | identity matrix |
| $\|A\|_F$ | Frobenius norm of matrix $A$ |
| sgvals$(A)$ | singular values of matrix $A$ |
| $A^\dashv$ | pseudoinverse of matrix $A$ |
| $\mathbb{E}[a]$ | statistical average of $a$ |
| cov $[a]$ | covariance matrix of $a$ |
| Pr $(\mathcal{A})$ | probability of event $\mathcal{A}$ |
| $\mathrm{cdf}_X(x)$ | cumulative distribution function |
| $\mathrm{sf}_X(x)$ | survival function $\mathrm{sf}_X(x) = 1 - \mathrm{cdf}_X(x) = \Pr(X > x)$ |
| $\otimes$ | tensor product |

## Acknowledgments

The authors are grateful to "Agence National de la Recherche" (France) for supporting PHARE#1 test rig through PIA EQUIPEX PHARE project (ANR-10-EQPX-0043), to Carnot Institute I@L and Safran Aircraft Engines for providing additional financial support as well as for giving permission to publish this work.

## References

[1] L. Ljung. *System Identification. Theory for the User (2nd Edition)*. Prentice Hall PTR, 1998, p. 609. ISBN: 9780136566953.

[2] T. Katayama. *Subspace methods for system identification*. Springer, 2005, p. 392. ISBN: 1852339810.

[3] P. Overschee. *Subspace Identification for Linear Systems Theory - Implementation - Applications. Theory - Implementation - Applications*. Springer US, 1996, p. 272. ISBN: 9781461380610.

[4] C. Jorajuria et al. "Comparison Of Subspace Identification Methods For Modal Estimation Of Bladed Disks". In: *Journal of Engineering for Gas Turbines and Power* 146.5 (Dec. 2023). ISSN: 1528-8919. DOI: 10.1115/1.4063634.

[5] W. E. Larimore. "Statistical optimality and canonical variate analysis system identification". In: *Signal Processing* 52.2 (July 1996), pp. 131–144. ISSN: 0165-1684. DOI: 10.1016/0165-1684(96)00049-7.

[6] T. Kailath. *Linear systems*. Prentice Hall, 1980.

[7] P. Van Overschee and B. De Moor. "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems". In: *Automatica* 30.1 (Jan. 1994), pp. 75–93. ISSN: 0005-1098. DOI: 10.1016/0005-1098(94)90230-5.

[8] P. Van Overschee and B. De Moor. "Choice of state-space basis in combined deterministic-stochastic subspace identification". In: *Automatica* 31.12 (Dec. 1995), pp. 1877–1883. ISSN: 0005-1098. DOI: 10.1016/0005-1098(95)00071-9.

[9] M. Verhaegen. "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data". In: *Automatica* 30.1 (Jan. 1994), pp. 61–74. ISSN: 0005-1098. DOI: 10.1016/0005-1098(94)90229-1.

[10] W. E. Larimore. "Canonical variate analysis in identification, filtering, and adaptive control". In: *29th IEEE Conference on Decision and Control*. IEEE, 1990. DOI: 10.1109/cdc.1990.203665.

[11] H. Akaike. "Statistical predictor identification". In: *Annals of the Institute of Statistical Mathematics* 22.1 (Dec. 1970), pp. 203–217. ISSN: 1572-9052. DOI: 10.1007/bf02506337.

[12] P. Guillaume et al. "A poly-reference implementation of the least-squares complex frequency-domain estimator". In: *Proceedings of IMAC*. Vol. 21. 2003, pp. 183–192.

[13] H. Akaike. "Information theory as an extension of the maximum likelihood principle". In: *Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 276-281*. 1973.

[14] N. Sugiura. "Further analysis of the data by Akaike's information criterion and the finite corrections". In: *Communications in Statistics - Theory and Methods* 7.1 (Jan. 1978), pp. 13–26. DOI: 10.1080/03610927808827599.

[15] J. Rissanen. "Modeling by shortest data description". In: *Automatica* 14.5 (Sept. 1978), pp. 465–471. DOI: 10.1016/0005-1098(78)90005-5.

[16] G. Schwarz. "Estimating the Dimension of a Model". In: *The Annals of Statistics* 6.2 (Mar. 1978). DOI: 10.1214/aos/1176344136.

[17] G. W. Snedecor and W. G. Cochran. *Statistical methods*. Eighth edition. 1996, pp. 71–82.

[18] M. Aoki and A. Havenner. "A method for approximate representation of vector-valued time series and its relation to two alternatives". In: *Journal of Econometrics* 42.2 (Oct. 1989). ISSN: 0304-4076. DOI: 10.1016/0304-4076(89)90002-x.

[19] J. H. Dorfman and A. Havenner. "Model specification tests for balanced representation state space models". In: *Communications in Statistics - Theory and Methods* 24.1 (Jan. 1995), pp. 97–119. ISSN: 1532-415X. DOI: 10.1080/03610929508831477.

[20] J. H. Dorfman and A. M. Havenner. "A Bayesian approach to state space multivariate time series modeling". In: *Journal of Econometrics* 52.3 (June 1992), pp. 315–346. ISSN: 0304-4076. DOI: 10.1016/0304-4076(92)90015-j.

[21] C. Jorajuria et al. "Experimental Modal Analysis of a Full-Scale Rotating Fan". In: *Volume 8B: Structures and Dynamics — Probabilistic Methods; Rotordynamics; Structural Mechanics and Vibration*. GT2022. American Society of Mechanical Engineers, June 2022. DOI: 10.1115/gt2022-82540.

[22] L. F. Shampine. "Some practical Runge-Kutta formulas". In: *Mathematics of Computation* 46.173 (1986), pp. 135–150. ISSN: 1088-6842. DOI: 10.1090/s0025-5718-1986-0815836-3.

[23] A. Mabilia et al. "Modal Testing of a Full-Scale Rotating Woven Composite Fan Using Piezoelectric Excitation. 2018". In: *Proceedings of the 10th International Conference on Rotor Dynamics – IFToMM*. Vol. 3. Mechanisms and machine science. Springer, Aug. 2018, pp. 291–305. ISBN: 9783319992709. DOI: 10.1007/978-3-319-99270-9_21.

# A Estimation of $M_j$

This appendix proposes more details over the estimation of the inverse of the covariance of the column $j$ of the observability matrix, which is denoted $\mathrm{cov}[O_h^{(j)}]^{-1}$. Indeed, this quantity is required to compute the `M-test` statistics and this article investigates the possibility to apply this statistical test to subspace identification methods with measurements of the excitation signals. To estimate this covariance, one can use the following set of equations which can be obtained using eq. (12) and (7b).

$$\widehat{O}_h = Y_f \widehat{X}_f^{\dagger} \tag{21}$$

$$\widehat{O}_h - O_h = \Psi_h U_f \widehat{X}_f^{\dagger} + \widehat{V}_f \widehat{X}_f^{\dagger} \tag{22}$$

The covariance of the observability matrix can be expressed in tensor notations as follows.

$$\mathrm{cov}\,[O_h]^{\alpha}{}_{\beta i}{}^{j} = \mathbb{E}\left[\left(\widehat{O}_h - O_h\right)^{\alpha}{}_{i}\left(\widehat{O}_h - O_h\right)_{\beta}{}^{j}\right]$$

As $U_f$ is considered deterministic, the associated term in eq. (22) does not contribute to the variance of the observability matrix. Denoting $\mathtt{X}$ the tensor associated with matrix $X_f^{\dagger}$ and $\mathtt{V}$ the tensor associated with matrix $\widehat{V}_f$, it is possible to obtain the following estimation of the covariance of the observability matrix.

$$\begin{aligned}\mathrm{cov}\,[O_h]^{\alpha}{}_{\beta i}{}^{j} &= \mathbb{E}\left[\mathtt{V}^{\alpha}{}_{n}\,\mathtt{X}^{n}{}_{i}\,\mathtt{V}_{\beta}{}^{m}\,\mathtt{X}_{m}{}^{j}\right] \\ &= \mathrm{cov}\,[\mathtt{V}]^{\alpha}{}_{\beta}\;\mathtt{X}^{n}{}_{i}\,\mathtt{X}_{n}{}^{j}\end{aligned}$$

The above estimator makes use of the hypotheses $x_t$ is independent of $(v_t, w_t)$, the perturbations $(v_t, w_t)$ are white noise processes and the ergodicity assumption to compute $\mathbb{E}\left[\mathtt{X}^{n}{}_{i}\mathtt{X}_{m}{}^{j}\right]$. With the operator of vectorization, denoted "vec", which consists in stacking column of a matrix to transform it into a vector, the estimation of covariance of the observability matrix and its inverse are expressed as follows [20].

$$\mathrm{cov}\left[\mathrm{vec}\,(O_h)\right] = \left(\mathrm{cov}\left[\widehat{V}_f\right]\right) \otimes \left(\left(\widehat{X}_f^{\dagger}\right)^{T}\widehat{X}_f^{\dagger}\right)$$

$$\mathrm{cov}\left[\mathrm{vec}\,(O_h)\right]^{-1} = \left(\mathrm{cov}\left[\widehat{V}_f\right]^{-1}\right) \otimes \left(\widehat{X}_f \widehat{X}_f^{T}\right)$$