Identification of Stochastic Systems Under Multiple Operating Conditions: The Vector Dependent FP–ARX Parametrization

Fotis P. Kopsaftopoulos and Spilios D. Fassois†

***Abstract*— The problem of identifying stochastic systems under multiple operating conditions, by using excitation – response signals obtained from each condition, is addressed. Each operating condition is characterized by several measurable variables forming a vector operating parameter. The problem is tackled within a novel framework consisting of postulated Vector dependent Functionally Pooled ARX (VFP–ARX) models, proper data pooling techniques, and statistical parameter estimation. Least Squares (LS) and Maximum Likelihood (ML) estimation methods are developed. Their strong consistency is established, and their performance characteristics are assessed via a Monte Carlo study.**

I. INTRODUCTION

In conventional system identification a mathematical model representing a system at a specific operating condition is identified based upon a single data record of excitation – response signals. Yet, in many applications, a system may operate under different operating conditions in different intervals of time, maintaining one such condition in each interval. These operating conditions affect the system characteristics, and thus its dynamics. Typical examples include physiological systems under different environmental conditions, mechanical systems under different load or lubrication conditions, systems under different configurations, hydraulic systems operating under different temperatures or fluid pressures, material and structures (civil–mechanical–aerospace) under different environmental (such as temperature and humidity) conditions, and so on.

In such cases it is of interest to identify a *"global"* model describing the system under *any* operating condition, based upon excitation – response data records available from each condition.

It could be, perhaps, argued that this may be handled by using conventional mathematical models and customary identification techniques that could artificially split the problem into a number of seemingly unrelated subproblems and derive a model based upon a single data record at a time. Nevertheless, such a solution would be both awkward and statistically suboptimal. Awkwardness has to do with the fact that a potentially large number of seemingly unrelated models (one per operating condition) would be obtained. Statistical suboptimality has to do with the fact that the set of identified models would be of suboptimal accuracy. This is due to two reasons. The first is the violation of the principle of statistical parsimony (model economy) as a large number of models would be used for representing the system. This would result in a large number of estimated parameters, and thus reduced accuracy. The second is the ineffective use of the information available in the totality of the data records. Indeed, not all available information would be extracted, as the *interrelations* among the different records would be ignored as a result of separating the problem into seemingly unrelated subproblems.

This work *aims* at the postulation of a proper framework and methods for effectively tackling the problem of identifying stochastic systems under *multiple operating conditions*. This is to be based upon three important entities:

(a) A novel, *Functionally Pooled (FP)*, stochastic model structure that explicitly allows for system modelling under *multiple* operating conditions via a single mathematical representation. This representation uses parameters that *functionally* depend upon the operating condition. It also uses a stochastic structure that accounts for the *statistical dependencies* among the different data records.

(b) *Data pooling* techniques (see [1]) for combining and optimally treating (as one entity) the data obtained from the various experiments.

(c) Statistical techniques for model estimation.

The resulting framework is referred to as a statistical *Functional Pooling framework*, and the corresponding models as stochastic *Functionally Pooled (FP) models*. A schematic representation is provided in Fig. 1.

The only essential practical condition for using this framework and identifying "global" system models is that each operating condition corresponds to a specific value of a measurable variable, henceforth referred to as the *operating parameter*. The case of a *scalar* operating parameter (for instance operating temperature) is treated in a companion paper [2]. The present paper focuses on the case of a *vector* operating parameter (consisting of two or more scalars, for instance operating temperature and humidity).

It should be also noted that early versions of the Functional Pooling framework, including certain simple models and estimation methods, have been already applied to practical fault diagnosis problems with very promising results. The interested reader is referred to [3], [4] for details.

F.P. Kopsaftopoulos and S.D. Fassois are with the Stochastic Mechanical Systems & Automation (SMSA) Group, Department of Mechanical & Aeronautical Engineering, University of Patras, GR 26500, Patras, Greece. Tel/fax: +30 2610 997 405. E-mail: {fkopsaf,fassois}@mech.upatras.gr, Internet: http://www.mech.upatras.gr/∼sms

†Corresponding author.

14th Mediterranean Conference on Control and Automation (MED 2006) Ancona, Italy, June 2006

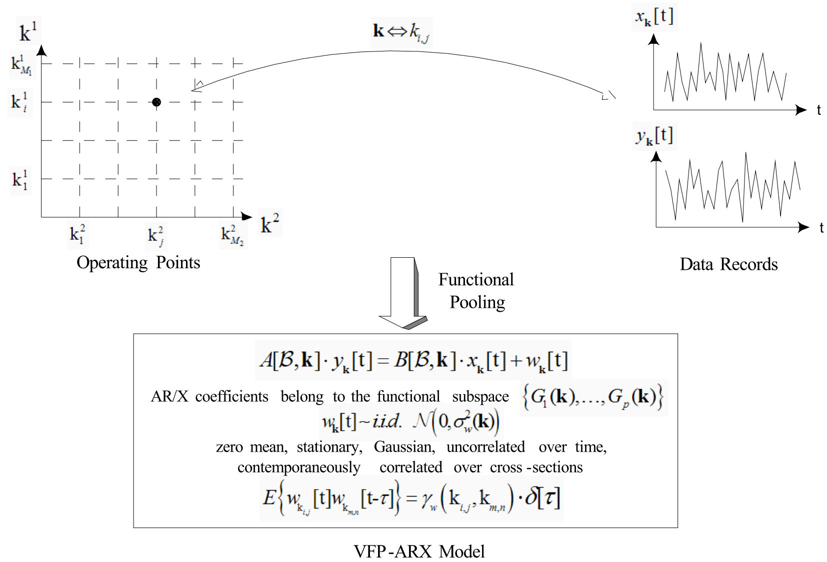

Fig. 1. Schematic representation of the problem showing the operating points on the (k^1, k^2) plane, an excitation–response data set corresponding to one particular point, and the VFP-ARX model structure.

II. THE DATA SET

Excitation – response data records from different operating points corresponding to various values of the (vector) operating parameter are used[1]:

$$Z^{NM_1M_2} \triangleq \{x_{\boldsymbol{k}}[t], y_{\boldsymbol{k}}[t] \mid \boldsymbol{k} \triangleq [k^1\ k^2]^T, \text{ with } t = 1, \ldots, N, \\ k^1 \in \{k^1_1, \ldots, k^1_{M_1}\},\ k^2 \in \{k^2_1, \ldots, k^2_{M_2}\}\}$$

In this expression t designates normalized discrete time (the corresponding analog time being $t \cdot T$ with T standing for the sampling period), $\boldsymbol{k} \triangleq [k^1\ \ k^2]^T$ the operating parameter (without loss of generality assumed to be two-dimensional), and $x_{\boldsymbol{k}}[t]$, $y_{\boldsymbol{k}}[t]$ the excitation and response signals corresponding to $\boldsymbol{k}$. N stands for the signal length (in samples) corresponding to each single experiment (each $\boldsymbol{k}$).

A total of $M_1 \times M_2$ experiments (one for each element of $\boldsymbol{k}$) are performed, with the complete series covering the required range of each scalar parameter, say $[k^1_{min},\ k^1_{max}]$ and $[k^2_{min},\ k^2_{max}]$, via the discretizations $k^1 = k^1_1, k^1_2, \ldots, k^1_{M_1}$ and $k^2 = k^2_1, k^2_2, \ldots, k^2_{M_2}$. Hence each experiment is characterized by a specific value of $\boldsymbol{k}$, say $\boldsymbol{k} = [k^1_i, k^2_j]$. This vector is, for simplicity of notation, also designated as the duplet $k_{i,j} = (k^1_i\ , k^2_j)$ (the first subscript of $k_{i,j}$ designating the value of k^1 and the second that of k^2).

III. THE VFP-ARX MODEL STRUCTURE

The Vector dependent Functionally Pooled AutoRegressive with eXogenous excitation (VFP-ARX) model structure postulated for treating this problem is of the form:

$$y_{\boldsymbol{k}}[t] + \sum_{i=1}^{na} a_i(\boldsymbol{k}) \cdot y_{\boldsymbol{k}}[t-i] = \sum_{i=0}^{nb} b_i(\boldsymbol{k}) \cdot x_{\boldsymbol{k}}[t-i] + w_{\boldsymbol{k}}[t] \quad (1)$$

[1]Lower case/capital bold face symbols designate vector/matrix quantities, respectively.

$$w_{\boldsymbol{k}}[t] \sim \text{iid}\,\mathcal{N}\big(0, \sigma^2_w(\boldsymbol{k})\big) \quad \boldsymbol{k} \in \mathbb{R}^2 \quad (2)$$

$$a_i(\boldsymbol{k}) \triangleq \sum_{j=1}^{p} a_{i,j} \cdot G_j(\boldsymbol{k}), \quad b_i(\boldsymbol{k}) \triangleq \sum_{j=1}^{p} b_{i,j} \cdot G_j(\boldsymbol{k}) \quad (3)$$

$$E\{w_{k_{i,j}}[t] \cdot w_{k_{m,n}}[t-\tau]\} = \gamma_w[k_{i,j}, k_{m,n}] \cdot \delta[\tau] \quad (4)$$

with na, nb designating the AutoRegressive (AR) and eXogenous (X) orders, respectively, $x_{\boldsymbol{k}}[t]$, $y_{\boldsymbol{k}}[t]$ the excitation and response signals, respectively, and $w_{\boldsymbol{k}}[t]$ the disturbance (innovations) signal that is a white (serially uncorrelated) zero-mean with variance $\sigma^2_w(\boldsymbol{k})$ and potentially cross-correlated with its counterparts corresponding to different experiments. The symbol $E\{\cdot\}$ designates statistical expectation, $\delta[\tau]$ the Kronecker delta (equal to unity for $\tau = 0$ and equal to zero for $\tau \neq 0$), $\mathcal{N}(\cdot,\cdot)$ Gaussian distribution with the indicated mean and variance, and iid stands for identically independently distributed.

As (3) indicates, the AR and X parameters $\alpha_i(\boldsymbol{k})$, $b_i(\boldsymbol{k})$ are modelled as explicit functions of the vector $\boldsymbol{k}$ belonging to a p-dimensional functional subspace spanned by the (mutually independent) functions $G_1(\boldsymbol{k}), G_2(\boldsymbol{k}), \ldots, G_p(\boldsymbol{k})$ (*functional basis*). The functional basis consists of polynomials of two variables (vector polynomials) obtained as cross-products from univariate polynomials (of the Chebyshev, Legendre, Jacobi and other families [5]). The constants $a_{i,j}$, $b_{i,j}$ designate the AR and X, respectively, coefficients of projection. Defining:

$$A[\mathcal{B}, \boldsymbol{k}] \triangleq 1 + \sum_{i=1}^{na} a_i(\boldsymbol{k})\mathcal{B}^i, \quad B[\mathcal{B}, \boldsymbol{k}] \triangleq \sum_{i=0}^{nb} b_i(\boldsymbol{k})\mathcal{B}^i$$

where $A[\mathcal{B}, \boldsymbol{k}], B[\mathcal{B}, \boldsymbol{k}]$ are the AutoRegressive (AR) and eXogenous (X) polynomials in the backshift operator $\mathcal{B}\big(\mathcal{B}^j \cdot u_{\boldsymbol{k}}[t] \triangleq u_{\boldsymbol{k}}[t-j]\big)$, the VFP-ARX representation of (1) is rewritten as:

$$A[\mathcal{B},\boldsymbol{k}]\cdot y_{\boldsymbol{k}}[t]=B[\mathcal{B},\boldsymbol{k}]\cdot x_{\boldsymbol{k}}[t]+w_{\boldsymbol{k}}[t]. \qquad (5)$$

As already mentioned, the innovations sequences $w_{\boldsymbol{k}}[t]$ corresponding to different operating conditions may be *contemporaneously correlated*, that is $E\{w_{k_{i,j}}[t]w_{k_{i,j}}[t]\} = \sigma_w^2([k_{i,j}]$ and $E\{w_{k_{i,j}}[t]w_{k_{m,n}}[t]\} = \gamma_w[k_{i,j},k_{m,n}]$. Defining the VFP-ARX model's *cross-section innovations vector* as:

$$\boldsymbol{w}[t]\triangleq\big[w_{k_{1,1}}[t]\,w_{k_{1,2}}[t]\,\ldots\,w_{k_{1,M_2}}[t]\,\ldots\,w_{k_{M_1,M_2}}[t]\big]^T \qquad (6)$$

with covariance matrix:

$$\begin{aligned}\boldsymbol{\Gamma}_{\boldsymbol{w}[t]} &= E\{\boldsymbol{w}[t]\boldsymbol{w}^T[t]\} \qquad (7)\\ &= \begin{bmatrix}\sigma_w^2[k_{1,1}] & \ldots & \gamma_w[k_{1,1},k_{M_1,M_2}]\\ \vdots & \ddots & \vdots\\ \gamma_w[k_{M_1,M_2},k_{1,1}] & \ldots & \sigma_w^2[k_{M_1,M_2}]\end{bmatrix}\end{aligned}$$

then the covariance matrix corresponding to the time instants $t=1,\ldots,N$ is given by:

$$\boldsymbol{\Gamma}_{\boldsymbol{w}}=\boldsymbol{\Gamma}_{\boldsymbol{w}[t]}\otimes\boldsymbol{I}_N \qquad (8)$$

with $\otimes$ designating Kronecker product [6, chap. 7].

In the case of *cross-sectionally uncorrelated* innovations sequences with *different variances* ($\sigma_w^2[k_{1,1}]\neq\sigma_w^2[k_{1,2}]\neq\ldots\neq\sigma_w^2[k_{M_1,M_2}]$, *groupwise heteroscedasticity*), the covariance matrix is given by:

$$\boldsymbol{\Gamma}_{\boldsymbol{w}}=\begin{bmatrix}\sigma_w^2[k_{1,1}]\boldsymbol{I}_N & \ldots & 0\\ \vdots & \ddots & \vdots\\ 0 & \ldots & \sigma_w^2[k_{M_1,M_2}]\boldsymbol{I}_N\end{bmatrix} \qquad (9)$$

In the simpler case of *cross-sectionally uncorrelated* innovations sequences with *equal variances* ($\sigma_w^2[k_{1,1}] = \sigma_w^2[k_{1,2}] = \ldots = \sigma_w^2[k_{M_1,M_2}] = \sigma_w^2$, *groupwise homoscedasticity*), the covariance matrix is given by $\boldsymbol{\Gamma}_{\boldsymbol{w}} = \sigma_w^2\boldsymbol{I}_{NM_1M_2}$ with $\boldsymbol{I}_{NM_1M_2}$ indicating the unity matrix.

The representation of equations (1) – (4) is referred to as a VFP-ARX model of orders (n_a,n_b) and functional subspace dimensionality p, or in short a VFP-ARX$(n_a,n_b)_p$ model. It is parameterized in terms of the parameter vector:

$$\bar{\boldsymbol{\theta}}\triangleq[\,\alpha_{i,j}\,\vdots\,b_{i,j}\,\vdots\gamma_w[k_{i,j},k_{m,n}]\,]^T \quad \forall\, i,j,m,n \qquad (10)$$

with $\gamma_w[k_{i,j},k_{i,j}]=\sigma_w^2[k_{i,j}]$.

The VFP-ARX representation is assumed to satisfy the following conditions:

A1. *Stability condition.* The poles of the AR polynomial (see (5)) lie inside the unit circle for all operating parameters $\boldsymbol{k}$.
A2. *Irreducibility condition* The polynomials $A[\mathcal{B},\boldsymbol{k}]$, $B[\mathcal{B},\boldsymbol{k}]$ are coprime (have no common factors) $\forall$ $\boldsymbol{k}$.
A3. The input signal $x_{\boldsymbol{k}}[t]$ is stationary, ergodic and persistently exciting with $E\{x_{k_{i,j}}[t]w_{k_{m,n}}[t]\}=0 \quad \forall\, i,j,m,n.$

IV. Model Estimation

A VFP-ARX model corresponding to the true system of (1) – (4) may be expressed as:

$$y_{\boldsymbol{k}}[t]+\sum_{i=1}^{na}a_i(\boldsymbol{k})\cdot y_{\boldsymbol{k}}[t-i]=\sum_{i=0}^{nb}b_i(\boldsymbol{k})\cdot x_{\boldsymbol{k}}[t-i]+e_{\boldsymbol{k}}[t] \qquad (11)$$

$$e_{\boldsymbol{k}}[t]\sim\ \text{iid}\,\mathcal{N}\big(0,\sigma_e^2(\boldsymbol{k})\big) \quad \boldsymbol{k}\in\mathbb{R}^2 \qquad (12)$$

$$a_i(\boldsymbol{k})\triangleq\sum_{j=1}^{p}a_{i,j}\cdot G_j(\boldsymbol{k}),\quad b_i(\boldsymbol{k})\triangleq\sum_{j=1}^{p}b_{i,j}\cdot G_j(\boldsymbol{k}) \qquad (13)$$

$$E\{e_{k_{i,j}}[t]\cdot e_{k_{m,n}}[t-\tau]\}=\gamma_e[k_{i,j},k_{m,n}]\cdot\delta[\tau] \qquad (14)$$

with $e_{\boldsymbol{k}}[t]$ designating the model's one-step-ahead prediction error or residual (corresponding to $w_{\boldsymbol{k}}[t]$) with variance $\sigma_e^2(\boldsymbol{k})$.

In the general case the model's one-step-ahead prediction error (residual) sequences $e_{\boldsymbol{k}}[t]$ may be contemporaneously correlated, that is $E\{e_{k_{i,j}}[t]e_{k_{i,j}}[t]\} = \sigma_e^2[k_{i,j}]$ and $E\{e_{k_{i,j}}[t]e_{k_{m,n}}[t]\} = \gamma_e[k_{i,j},k_{m,n}]$, with the model residual cross-section vector defined as $\boldsymbol{e}[t] \triangleq \big[e_{k_{1,1}}[t]\,\ldots\,e_{k_{M_1,M_2}}[t]\big]^T$ The cross-section vector covariance then is:

$$\begin{aligned}\boldsymbol{\Gamma}_{\boldsymbol{e}[t]} &= E\{\boldsymbol{e}[t]\boldsymbol{e}^T[t]\}\\ &= \begin{bmatrix}\sigma_e^2[k_{1,1}] & \ldots & \gamma_e[k_{1,1},k_{M_1,M_2}]\\ \vdots & \ddots & \vdots\\ \gamma_e[k_{M_1,M_2},k_{1,1}] & \ldots & \sigma_e^2[k_{M_1,M_2}]\end{bmatrix}\end{aligned}$$

and the covariance matrix for the time instants $t=1,\ldots,N$ is given as:

$$\boldsymbol{\Gamma}_{\boldsymbol{e}}=\boldsymbol{\Gamma}_{\boldsymbol{e}[t]}\otimes\boldsymbol{I}_N.$$

The VFP-ARX model estimation problem may then be stated as follows: "Given the excitation–response data records select an element $\mathcal{M}(\bar{\boldsymbol{\theta}})$ from the VFP-ARX model set:

$$\mathcal{M}\triangleq\{\mathcal{M}(\bar{\boldsymbol{\theta}}):A[\mathcal{B},\boldsymbol{k},\bar{\boldsymbol{\theta}}]\cdot y_{\boldsymbol{k}}[t]=B[\mathcal{B},\boldsymbol{k},\bar{\boldsymbol{\theta}}]\cdot x_{\boldsymbol{k}}[t]+e_{\boldsymbol{k}}[t,\bar{\boldsymbol{\theta}}]|$$
$$\gamma_w[k_{i,j},k_{m,n}]=E\{e_{k_{i,j}}[t,\bar{\boldsymbol{\theta}}]e_{k_{m,n}}[t,\bar{\boldsymbol{\theta}}]\},\ \forall\, i,j,m,n\ \}$$

that best fits the measured data."

The model identification problem is usually distinguished into two subproblems: the parameter estimation subproblem and the model structure selection subproblem. The present paper focuses on the parameter estimation part, while the model structure selection subproblem is treated in a forthcoming paper [7].

A. A Functionally Pooled Linear Regression Framework

The VFP-ARX model (11) may be rewritten as:

$$y_{\boldsymbol{k}}[t]=\big[\boldsymbol{\varphi}_{\boldsymbol{k}}^T[t]\otimes\boldsymbol{g}^T(\boldsymbol{k})\big]\cdot\boldsymbol{\theta}+e_{\boldsymbol{k}}[t]=\boldsymbol{\phi}_{\boldsymbol{k}}^T[t]\cdot\boldsymbol{\theta}+e_{\boldsymbol{k}}[t] \qquad (15)$$

with:

$$\begin{aligned}\boldsymbol{\varphi}_{\boldsymbol{k}}[t] &\triangleq \big[-y_{\boldsymbol{k}}[t-1]\,\ldots\,-y_{\boldsymbol{k}}[t-na]\,\vdots\,x_{\boldsymbol{k}}[t]\,\ldots\,x_{\boldsymbol{k}}[t-nb]\big]^T\\ \boldsymbol{g}(\boldsymbol{k}) &\triangleq \big[G_1(\boldsymbol{k})\,\ldots\,G_p(\boldsymbol{k})\big]^T\\ \boldsymbol{\theta} &\triangleq \big[a_{1,1}\,\ldots\,a_{na,p}\,\vdots\ \ b_{0,1}\,\ldots\,b_{nb,p}\big]^T.\end{aligned}$$

Pooling together the expressions of the VFP-ARX model [Equation (15)] corresponding to all operating parameters $\boldsymbol{k}$ $(k_{1,1}, k_{1,2}, \ldots, k_{M_1,M_2})$ considered in the experiments (cross-sectional pooling) yields:

$$\begin{bmatrix} y_{k_{1,1}}[t] \\ \vdots \\ y_{k_{M_1,M_2}}[t] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}^T_{k_{1,1}}[t] \\ \vdots \\ \boldsymbol{\phi}^T_{k_{M_1,M_2}}[t] \end{bmatrix} \cdot \boldsymbol{\theta} + \begin{bmatrix} e_{k_{1,1}}[t] \\ \vdots \\ e_{k_{M_1,M_2}}[t] \end{bmatrix} \Longrightarrow$$

$$\boldsymbol{y}[t] = \boldsymbol{\Phi}[t] \cdot \boldsymbol{\theta} + \boldsymbol{e}[t].$$

Then, following substitution of the data for $t = 1, \ldots, N$ the following expression is obtained:

$$\boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{\theta} + \boldsymbol{e} \tag{16}$$

with:

$$\boldsymbol{y} \triangleq \begin{bmatrix} \boldsymbol{y}[1] \\ \vdots \\ \boldsymbol{y}[N] \end{bmatrix}, \quad \boldsymbol{\Phi} \triangleq \begin{bmatrix} \boldsymbol{\Phi}[1] \\ \vdots \\ \boldsymbol{\Phi}[N] \end{bmatrix}, \quad \boldsymbol{e} \triangleq \begin{bmatrix} \boldsymbol{e}[1] \\ \vdots \\ \boldsymbol{e}[N] \end{bmatrix}.$$

B. Least Squares (LS) Based Estimation Methods

Uning the above linear regression framework the simplest possible approach to estimate the projection coefficients vector $\boldsymbol{\theta}$ is based upon minimization of the Ordinary Least Squares criterion:

$$J^{\text{OLS}}(\boldsymbol{\theta}, Z^{NM_1M_2}) \triangleq \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{e}^T[t]\boldsymbol{e}[t]$$

which leads to the *Ordinary Least Squares (OLS)* estimator:

$$\hat{\boldsymbol{\theta}}^{\text{OLS}} = \big[\boldsymbol{\Phi}^T\boldsymbol{\Phi}\big]^{-1}\big[\boldsymbol{\Phi}^T\boldsymbol{y}\big]. \tag{17}$$

A more appropriate criterion for the *contemporaneously correlated residual* case is (in view of the Gauss-Markov theorem [8]) the Weighted Least Squares (WLS) criterion:

$$J^{\text{WLS}}(\boldsymbol{\theta}, Z^{NM_1M_2}) \triangleq \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{e}^T[t]\boldsymbol{\Gamma}^{-1}_{\boldsymbol{w}[t]}\boldsymbol{e}[t] = \frac{1}{N}\boldsymbol{e}^T\boldsymbol{\Gamma}^{-1}_{\boldsymbol{w}}\boldsymbol{e}$$

with $\boldsymbol{\Gamma}_{\boldsymbol{w}[t]}, \boldsymbol{\Gamma}_{\boldsymbol{w}}$ given by (7) and (8), respectively. This leads to the *Weighted Least Squares (WLS)* estimator:

$$\hat{\boldsymbol{\theta}}^{\text{WLS}} = \big[\boldsymbol{\Phi}^T\boldsymbol{\Gamma}^{-1}_{\boldsymbol{w}}\boldsymbol{\Phi}\big]^{-1}\big[\boldsymbol{\Phi}^T\boldsymbol{\Gamma}^{-1}_{\boldsymbol{w}}\boldsymbol{y}\big]. \tag{18}$$

As the covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{w}}$ is practically unavailable, it may be consistently estimated by using the Ordinary Least Squares (OLS) estimator, thus:

$$\widehat{\boldsymbol{\Gamma}}^{\text{OLS}}_{\boldsymbol{w}[t]} = \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{e}[t, \hat{\boldsymbol{\theta}}^{\text{OLS}}]\boldsymbol{e}^T[t, \hat{\boldsymbol{\theta}}^{\text{OLS}}]$$

with $\boldsymbol{e}[t, \hat{\boldsymbol{\theta}}^{\text{OLS}}]$ designating the residuals $\boldsymbol{e}[t]$ for $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{\text{OLS}}$. Then:

$$\widehat{\boldsymbol{\Gamma}}^{\text{OLS}}_{\boldsymbol{w}} = \widehat{\boldsymbol{\Gamma}}^{\text{OLS}}_{\boldsymbol{w}[t]} \otimes \boldsymbol{I}_N.$$

The estimator in (18) is then expressed as:

$$\hat{\boldsymbol{\theta}}^{\text{WLS}} = \big[\boldsymbol{\Phi}^T(\widehat{\boldsymbol{\Gamma}}^{\text{OLS}}_{\boldsymbol{w}})^{-1}\boldsymbol{\Phi}\big]^{-1}\big[\boldsymbol{\Phi}^T(\widehat{\boldsymbol{\Gamma}}^{\text{OLS}}_{\boldsymbol{w}})^{-1}\boldsymbol{y}\big]. \tag{19}$$

while the final residual covariance matrix is estimated as:

$$\widehat{\boldsymbol{\Gamma}}^{\text{WLS}}_{\boldsymbol{w}[t]} = \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{e}[t, \hat{\boldsymbol{\theta}}^{\text{WLS}}]\boldsymbol{e}^T[t, \hat{\boldsymbol{\theta}}^{\text{WLS}}].$$

In the case of *cross-sectionally uncorrelated* residual sequences with *different variances* $(\sigma^2_e[k_{1,1}] \neq \sigma^2_e[k_{1,2}] \neq \ldots \neq \sigma^2_e[k_{M_1,M_2}]$, *groupwise heteroscedasticity*) the residual covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{w}}$ for all $\boldsymbol{k}$ has the same form as (9). As the variances are practically unavailable, they may be consistently estimated as [9]:

$$\hat{\sigma}^2_e(\boldsymbol{k}, \hat{\boldsymbol{\theta}}^{OLS}) = \frac{1}{N} \sum_{t=1}^{N} e^2_{\boldsymbol{k}}[t, \hat{\boldsymbol{\theta}}^{OLS}] \tag{20}$$

for all $\boldsymbol{k}$, with $e^2_{\boldsymbol{k}}[t, \hat{\boldsymbol{\theta}}^{\text{OLS}}]$ designating the residual sequences obtained by applying OLS. The $\hat{\boldsymbol{\theta}}^{\text{WLS}}$ estimator is then given by (19). The final residual variance is estimated as:

$$\hat{\sigma}^2_w(\boldsymbol{k}) = \hat{\sigma}^2_e(\boldsymbol{k}, \hat{\boldsymbol{\theta}}^{\text{WLS}}) = \frac{1}{N} \sum_{t=1}^{N} e^2_{\boldsymbol{k}}[t, \hat{\boldsymbol{\theta}}^{\text{WLS}}]. \tag{21}$$

In the simpler case of *cross-sectionally uncorrelated* residual sequences with *equal variances* $(\sigma^2_e[k_{1,1}] = \sigma^2_e[k_{1,2}] = \ldots = \sigma^2_e[k_{M_1,M_2}] = \sigma^2_e$, *groupwise homoscedasticity*) the covariance matrix is $\boldsymbol{\Gamma}_{\boldsymbol{w}} = \sigma^2_w \boldsymbol{I}_{NM_1M_2}$ with $\boldsymbol{I}_{NM_1M_2}$ designating the unit matrix. In this case the WLS estimator coincides with its OLS counterpart. The residual variance is estimated by (21).

C. The Maximum Likelihood (ML) Estimation Method

The complete parameter vector $\bar{\boldsymbol{\theta}}$ is estimated as:

$$\hat{\bar{\boldsymbol{\theta}}}^{\text{ML}} \triangleq \arg\max_{\bar{\boldsymbol{\theta}}} L(\boldsymbol{\theta}, \boldsymbol{\Gamma}_{\boldsymbol{w}[t]}/\boldsymbol{e})$$

with $L(\cdot)$ the natural logarithm of the conditional likelihood function [10], [11]. In the general case of *normally distributed* and *contemporaneously correlated* residuals $e_{\boldsymbol{k}}[t]$ $\forall$ $\boldsymbol{k}$ [10, p. 198] we have:

$$L(\boldsymbol{\theta}, \boldsymbol{\Gamma}_{\boldsymbol{w}[t]}/\boldsymbol{e}[t_1], \ldots, \boldsymbol{e}[t_N]) = \ln \prod_{t=1}^{N} p(\boldsymbol{e}[t]/\boldsymbol{\theta}, \boldsymbol{\Gamma}_{\boldsymbol{w}[t]})$$

$$= -\frac{1}{2} \sum_{t=1}^{N} \boldsymbol{e}^T[t]\boldsymbol{\Gamma}^{-1}_{\boldsymbol{w}[t]}\boldsymbol{e}[t] - \frac{NM_1M_2}{2}\ln 2\pi - \frac{N}{2}\ln det\{\boldsymbol{\Gamma}_{\boldsymbol{w}[t]}\} \tag{22}$$

with $p(\cdot)$ designating the Gaussian probability density function. By setting:

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}) \triangleq \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{e}[t, \boldsymbol{\theta}]\boldsymbol{e}^T[t, \boldsymbol{\theta}] \tag{23}$$

(22) becomes:

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\Gamma}_{\boldsymbol{w}[t]}/\boldsymbol{e}) = & - \frac{N}{2} Tr \boldsymbol{\Lambda}(\boldsymbol{\theta})\boldsymbol{\Gamma}^{-1}_{\boldsymbol{w}[t]} - \frac{N}{2}\ln det\{\boldsymbol{\Gamma}_{\boldsymbol{w}[t]}\} \\ & - \frac{NM_1M_2}{2}\ln 2\pi \end{aligned} \tag{24}$$

The first derivative of (24) with respect to $\mathbf{\Gamma}_{\boldsymbol{w}[t]}$ leads to:

$$\frac{\partial L(\boldsymbol{\theta}, \mathbf{\Gamma}_{\boldsymbol{w}[t]}/\boldsymbol{e})}{\partial \mathbf{\Gamma}_{\boldsymbol{w}[t]}} = \frac{N}{2}\mathbf{\Gamma}^{-1}_{\boldsymbol{w}[t]}\mathbf{\Lambda}(\boldsymbol{\theta})\mathbf{\Gamma}^{-1}_{\boldsymbol{w}[t]} - \frac{N}{2}\mathbf{\Gamma}^{-1}_{\boldsymbol{w}[t]}$$

and equating it to zero yields $\mathbf{\Gamma}_{\boldsymbol{w}[t]} = \mathbf{\Lambda}(\boldsymbol{\theta})$.

It is proven [10] that $L(\boldsymbol{\theta}, \mathbf{\Gamma}_{\boldsymbol{w}[t]}/\boldsymbol{e})$ is maximized with respect to $\mathbf{\Gamma}_{\boldsymbol{w}[t]}$ for $\mathbf{\Gamma}_{\boldsymbol{w}[t]} = \mathbf{\Lambda}(\boldsymbol{\theta})$ and the maximum likelihood estimate of $\mathbf{\Lambda}(\boldsymbol{\theta})$ is given by (23) for the optimum value of $\boldsymbol{\theta}$ that has to be determined. By replacing $\mathbf{\Gamma}_{\boldsymbol{w}[t]}$ with $\mathbf{\Lambda}(\boldsymbol{\theta})$ in (24) yields:

$$L(\boldsymbol{\theta}/\boldsymbol{e}) = -\frac{NM_1M_2}{2}(\ln 2\pi + 1) - \frac{N}{2}\ln det\{\mathbf{\Lambda}(\boldsymbol{\theta})\}. \quad (25)$$

Maximizing equation (25) with respect to $\boldsymbol{\theta}$ leads to the *ML estimator*:

$$\hat{\boldsymbol{\theta}}^{\mathrm{ML}} \triangleq \arg\min_{\boldsymbol{\theta}} det\{\mathbf{\Lambda}(\boldsymbol{\theta})\} \quad (26)$$

and $\widehat{\mathbf{\Gamma}}_{\boldsymbol{w}[t]} = \mathbf{\Lambda}(\hat{\boldsymbol{\theta}}^{\mathrm{ML}}) = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{e}[t, \hat{\boldsymbol{\theta}}^{\mathrm{ML}}]\boldsymbol{e}^T[t, \hat{\boldsymbol{\theta}}^{\mathrm{ML}}]$. Notice that obtaining $\hat{\boldsymbol{\theta}}^{\mathrm{ML}}$ requires the use of iterative optimization techniques [10].

In the *heteroscedastic* case we have:

$$\begin{aligned} \ln det\{\mathbf{\Lambda}(\boldsymbol{\theta})\} &= \ln\bigl(\sigma_e^2[k_{1,1}, \boldsymbol{\theta}] \cdot \ldots \cdot \sigma_e^2[k_{M_1,M_2}, \boldsymbol{\theta}]\bigr) \\ &= \ln\sigma_e^2[k_{1,1}, \boldsymbol{\theta}] + \ldots + \ln\sigma_e^2[k_{M_1,M_2}, \boldsymbol{\theta}] \\ &= \sum_{k^1=k^1_1}^{k^1_{M_1}}\sum_{k^2=k^2_1}^{k^2_{M_2}} \ln\sigma_e^2(\boldsymbol{k}, \boldsymbol{\theta}). \end{aligned} \quad (27)$$

Maximizing (25) with respect to $\boldsymbol{\theta}$ leads to the optimal value of $\boldsymbol{\theta}$ (as in (26)) and:

$$\hat{\sigma}_w^2(\boldsymbol{k}) = \hat{\sigma}_e^2(\boldsymbol{k}, \hat{\boldsymbol{\theta}}^{\mathrm{ML}}) = \frac{1}{N}\sum_{t=1}^{N} e^2_{\boldsymbol{k}}[t, \hat{\boldsymbol{\theta}}^{\mathrm{ML}}]. \quad (28)$$

In the *homoscedastic* case we have:

$$\ln det\{\mathbf{\Lambda}(\boldsymbol{\theta})\} = \ln\bigl[\sigma_e^2(\boldsymbol{\theta})\bigr]^{M_1M_2} = M_1M_2\ln\sigma_e^2(\boldsymbol{\theta}) \quad (29)$$

and the final residual variance is given by (28).

V. CONSISTENCY ANALYSIS

The consistency of the OLS, WLS and ML estimators of the previous section is examined. For simplicity, the case of cross-sectionally uncorrelated innovations sequences with different variances (*heteroscedastic case*) is considered. The estimated model is assumed to have the exact structure of the true system, with the latter and the excitation signals satisfying the assumptions A1, A2 and A3 of section III. The proofs of the theorems are provided in [7].

For the Least Squares (LS) estimators of the previous section we have the following theorem:

Theorem 1: Least Squares estimator consistency. Let $\boldsymbol{\theta}_o$ be the true projection coefficient vector, $w_{\boldsymbol{k}}[t]$ a white zero mean process with $E\{w^2_{\boldsymbol{k}}[t]\} = \sigma_w^2(\boldsymbol{k})$ for every operating point, and $E\{\boldsymbol{\phi}_{\boldsymbol{k}}[t]\boldsymbol{\phi}_{\boldsymbol{k}}{}^T[t]\}$ a nonsingular matrix. Then:

$$\hat{\boldsymbol{\theta}}_N^{\mathrm{LS}} \xrightarrow{a.s.} \boldsymbol{\theta}_o \qquad (N \longrightarrow \infty)$$

with a.s. designating convergence in the almost sure sense [9, pp. 18-19]. □

Note that, using the Kolmogorov theorem [9, p. 32], it is easily seen that:

$$\hat{\sigma}_w^2(\boldsymbol{k}, N) \xrightarrow{a.s.} \sigma_w^2(\boldsymbol{k}) \qquad (N \longrightarrow \infty)$$

as well.

Theorem 2: Maximum Likelihood estimator consistency. Let $\bar{\boldsymbol{\theta}}_o = [\boldsymbol{\theta}_o^T \vdots \gamma_w[k_{i,j}, k_{m,n}]]$ be the true parameter vector, $w_{\boldsymbol{k}}[t]$ a normally distributed zero mean white process with $E\{w^2_{\boldsymbol{k}}[t]\} = \sigma_w^2(\boldsymbol{k})$ for every operating point, and $E\{\boldsymbol{\phi}_{\boldsymbol{k}}[t]\boldsymbol{\phi}_{\boldsymbol{k}}{}^T[t]\}$ a nonsingular matrix. Then:

$$\hat{\bar{\boldsymbol{\theta}}}_N^{\mathrm{ML}} \xrightarrow{a.s.} \bar{\boldsymbol{\theta}}_o \qquad (N \longrightarrow \infty).$$

□

VI. MONTE CARLO STUDY

The effectiveness of the OLS, WLS and ML estimators for VFP-ARX models is now examined via a Monte Carlo study. It is noted that the ML estimator is initialized by the WLS estimates, and makes use of the Gauss–Newton nonlinear optimization scheme [10] (maximum number of iterations 500; maximum number of function evaluations 5000; termination tolerance of the loss function 10^{-2}; termination tolerance of the estimated parameters 10^{-12}).

The study is based upon a VFP–ARX(4,2)$_9$ model with zero delay ($b_0 \neq 0$ in the eXogenous polynomial) and AR, X subspaces consisting of the cross-products of the first three (hence functional dimensionality $p = 9$) shifted Chebyshev polynomials of the second kind [5]. It includes 500 runs, in each one of which the first scalar operating parameter takes 16 values ($k^1_i \in [1, 16]$) and the second scalar operating parameter takes 20 values ($k^2_j \in [1, 20]$). Thus, each run includes excitation–response signals (of length equal to $N = 1024$ samples) from $M_1 \times M_2 = 320$ operating conditions.

Each response is corrupted by random noise at the 10% standard deviation level in accordance with the ARX structure expression (innovations standard deviation over the noise–free response standard deviation equal to 0.10). The innovations sequences corresponding to different operating conditions are cross-sectionally uncorrelated, but characterized by different variances (groupwise heteroscedasticity). Some of the true system coefficients of projection (out of a total of 54) are indicated in the second column of Table I.

Monte Carlo partial estimation results by the Ordinary Least Squares (OLS), the Weighted Least Squares (WLS) and the Maximum Likelihood (ML) methods are presented in Table I (mean estimates $\pm$ standard deviations). Some of these results are pictorially depicted in Fig. 2 (AR/X coefficient of projection estimates, 95% confidence intervals). As may be readily observed, the results are all very accurate. All three methods provide essentially unbiased estimates, with the WLS and ML methods expectedly providing better accuracy for the coefficients of projection (smaller standard deviations, thus narrower confidence intervals).

TABLE I

INDICATIVE MONTE CARLO ESTIMATION RESULTS FOR THE VFP-ARX$(4,2)_9$ MODEL (SELECTED PARAMETERS; 500 RUNS PER METHOD; MEAN ESTIMATE $\pm$ STANDARD DEVIATION).

COEFF.	TRUE	OLS ESTIMATE	WLS ESTIMATE	ML ESTIMATE
$a_{1,1}$	-0.0459	-0.04602 $\pm$ 0.00038	-0.04599 $\pm$ 0.00024	-0.04598 $\pm$ 0.00024
$a_{1,7}$	-0.0058	-0.00579 $\pm$ 0.00018	-0.00578 $\pm$ 0.00012	-0.00579 $\pm$ 0.00012
$a_{2,1}$	-0.3869	-0.38690 $\pm$ 0.00032	-0.38692 $\pm$ 0.00021	-0.38692 $\pm$ 0.00021
$a_{2,5}$	0.0235	0.02356 $\pm$ 0.00022	0.02349 $\pm$ 0.00015	0.02349 $\pm$ 0.00015
$a_{3,3}$	-0.0533	-0.05333 $\pm$ 0.00024	-0.05334 $\pm$ 0.00016	-0.05334 $\pm$ 0.00016
$a_{3,8}$	0.0179	0.01790 $\pm$ 0.00018	0.01788 $\pm$ 0.00012	0.01788 $\pm$ 0.00012
$a_{4,1}$	0.6046	0.60466 $\pm$ 0.00030	0.60467 $\pm$ 0.00022	0.60467 $\pm$ 0.00022
$a_{4,9}$	-0.0013	-0.00133 $\pm$ 0.00015	-0.00135 $\pm$ 0.00011	-0.00135 $\pm$ 0.00011
$b_{1,1}$	0.7453	0.74526 $\pm$ 0.00061	0.74527 $\pm$ 0.00042	0.74527 $\pm$ 0.00042
$b_{1,8}$	0.2484	0.24843 $\pm$ 0.00079	0.24844 $\pm$ 0.00035	0.24844 $\pm$ 0.00035
$b_{2,3}$	-0.1987	-0.19871 $\pm$ 0.00082	-0.19871 $\pm$ 0.00043	-0.19871 $\pm$ 0.00043
$b_{2,6}$	0.1682	0.16815 $\pm$ 0.00084	0.16826 $\pm$ 0.00045	0.16827 $\pm$ 0.00045

Fig. 3 finally depicts the true system's frequency response magnitude versus frequency and the first scalar operating parameter k^1 (for set k^2) along with their mean estimated (by the OLS, WLS and ML methods) counterparts. The agreement between each estimate and the true frequency response magnitude is excellent.

ACKNOWLEDGMENT

The authors acknowledge the funding of this work by the European Social Fund (ESF), Operational Program for Education and Vocational Training II (EPEAEK II), and particularly the program PYTHAGORAS II of the Greek Ministry of Education.

REFERENCES

[1] T. Dielman, *Pooled Cross-Sectional and Time Series Data Analysis*. Marcel Dekker, Inc., 1989.

[2] J. Sakellariou and S. Fassois, "Externally Dependent Functional ARX (FARX) Models for the Identification of Dynamical Systems Under Multiple Operating Conditions," submitted for publication.

[3] J. Sakellariou, K. Petsounis, and S. Fassois, "A Functional Model Based Method for Fault Detection and Identification in Stochastic Structural Systems," in *Proc. of the Second European Workshop on Structural Health Monitoring (SHM 2004), Munich, Germany*, 2004, pp. 679–686.

[4] J. Sakellariou and S. Fassois, "Fault Detection and Identification in a Scale Aircraft Skeleton Structure via a Functional Model Based Method," in *Proc. of the International Conference on Noise and Vibration Engineering (ISMA 2004), Leuven, Belgium*, 2004, pp. 515–529.

[5] C. Dunkl and Y. Xu, *Orthogonal Polynomials of Several Variables*. Cambridge University Press, 2001.

[6] D. Bernstein, *Matrix Mathematics*. Princeton University Press, 2005.

[7] F. Kopsaftopoulos and S. Fassois, "On the Estimation of Vector Dependent Functionally Pooled ARX (VFP-ARX) Models," under preparation for publication.

[8] W. Greene, *Econometric Analysis*. Prentice-Hall International, Inc., 1997.

[9] H. White, *Asymptotic Theory for Econometricians*, 2nd ed. Academic Press, 2001.

[10] T. Soderstrom and P. Stoica, *System Identification*. Prentice Hall, 1989.

[11] J. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications and Control*. Prentice Hall, 1995.

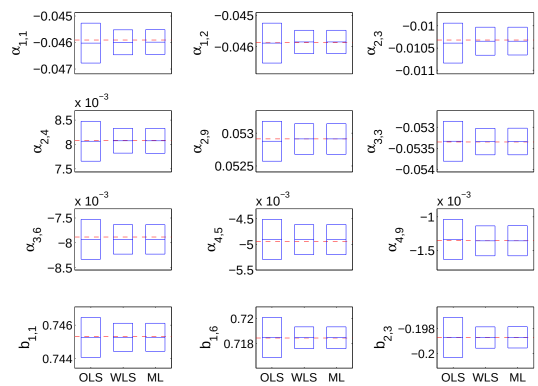

Fig. 2. VFP-ARX$(4,2)_9$ true values of the coefficients of projection (dashed lines) and Monte Carlo estimates (boxes indicate mean ± 1.96 standard deviations) based upon the OLS, WLS and ML methods (500 runs per method).

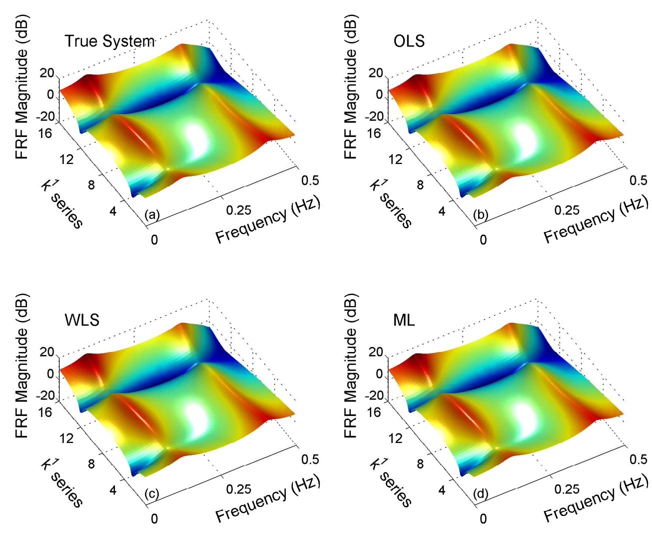

Fig. 3. VFP-ARX$(4,2)_9$ based frequency response magnitude versus frequency and k^1 (k^2 is set to k^2_9): (a) true system, (b) mean OLS estimate, (c) mean WLS estimate and (d) mean ML estimate (mean parameters over 500 runs).