**Preprints of the 16th IFAC Symposium on System Identification**
**The International Federation of Automatic Control**
**Brussels, Belgium. July 11-13, 2012**

# Vector-dependent Functionally Pooled ARX Models for the Identification of Systems Under Multiple Operating Conditions

**Fotis P. Kopsaftopoulos, Spilios D. Fassois** [1]

*Stochastic Mechanical Systems & Automation (SMSA) Laboratory*
*Dept. of Mechanical & Aeronautical Engr., University of Patras*
*26504 Patras, Greece*
*E-mail: {fkopsaf,fassois}@mech.upatras.gr, Web: www.smsa.upatras.gr*

**Abstract:** The identification of stochastic systems operating under multiple conditions is addressed based on data records obtained under a sample of these conditions. The problem is important in many practical applications and is tackled within a recently introduced Functional Pooling framework. The study focuses on the case of operating conditions characterized by several parameters. Global Vector-dependent Functionally Pooled models of the ARX type are postulated, proper estimators based on the Least Squares and Maximum Likelihood principles are formulated, and their strong consistency and asymptotic normality are established. For model structure selection a Genetic Algorithm based procedure is formulated. The performance characteristics of the methods are assessed via a Monte Carlo study.

*Keywords:* stochastic systems, identification algorithms, ARX models, asymptotic analysis, genetic algorithm optimization, Monte Carlo simulations

## 1. INTRODUCTION

Classical system identification aims at deriving a model representing a system under a specific operating condition. Yet, in many cases, a system may operate under different conditions at different occasions (time periods), with the dynamics depending in a pseudo-static fashion on certain *operating parameter(s)* – also referred to as *scheduling parameter(s)*. Typical examples include mechanical structures operating under different environmental conditions, such as temperature (Hios and Fassois [2009a]), aircraft systems under different flight conditions, such as altitude and velocity (Dimogianopoulos et al. [2009]), machinery operating under different rotating speeds, and so on.

Given a number of data records from a system operating under different conditions, it is highly desirable to establish a *single* and *global* model, that, while *compact* (parsimonious), will be capable of accurately representing the dynamics under *any* considered condition.

A "feasible" approach for solving this identification problem could be along the lines of the "multi-model" principle. This postulates identification of several conventional models (one for each operating condition), and their subsequent "interconnection" into a single global model via proper interpolation of their parameters. This is in fact the procedure followed for tackling the problem within the celebrated Linear Parameter Varying (LPV) modelling framework (see Toth [2010] and the references therein).

[1] Corresponding author. Tel./fax: +30 2610 969495 (direct), +30 2610 969492 (central).

Nevertheless, such a two-stage "multi-model" approach is statistically suboptimal and leads to decreased accuracy. The reasons for this are: First, the artificial splitting of the problem into disjoint subproblems (separate identification for each data record) leads to the estimation of an unnecessarily high number of parameters (due to the fact that each model is identified separately from any other), a fact violating the principle of statistical parsimony and leading to decreased accuracy. Second, any interrelations that may exist among the various data records are disregarded, a fact leading to further loss of information. Third, the separate treatment of the parameter interpolation stage leads to further (unnecessary) estimation errors.

In order to overcome the aforementioned drawbacks and enable the identification of compact and accurate "global" models in a *statistically optimal sense*, a novel *Functional Pooling (FP) framework* has been introduced by the authors and co-workers (Sakellariou and Fassois [2007], Hios and Fassois [2009a]). This FP framework is based on three important entities:

(i) A *stochastic Functionally Pooled (FP)* model structure that explicitly allows for system modelling under *multiple* operating conditions via a single ("global") mathematical representation. This representation is characterized by parameters that *functionally* (explicitly) depend on the operating parameter in a quasi-static fashion and allows for the proper, parsimonious, modelling of the dynamics under all possible conditions without using excessively many parameters or requiring a separate interpolation stage.

(ii) *Data pooling* techniques which simultaneously treat, as a single entity, the data records corresponding to all available operating conditions. In this way potential interrelations are also accounted for.
(iii) Properly formulated statistical inference techniques for model estimation.

The identification of univariate and multivariate (vector) FP models characterized by a *single* operating parameter has been treated in Sakellariou and Fassois [2007] and Hios and Fassois [2009b], respectively. First applications using early versions of the methods are reported in Hios and Fassois [2009a] (modelling of composite beam dynamics under different temperature conditions), Sakellariou and Fassois [2008] (modelling of different damage states in an aircraft skeleton structure), Dimogianopoulos et al. [2009] (modelling of aircraft dynamics under different flight conditions). In Hios and Fassois [2009a] the aforementioned theoretical advantages and improved achievable accuracy of the FP framework over the multi-model approach are also practically confirmed.

The *aim* of the present study is the proper formulation and extension of the FP framework to the case of stochastic systems operating under conditions characterized by *several* (more than one) operating parameters – a case of obvious practical importance. These parameters are collected into a vector, which is subsequently referred to as the *operating parameter vector*, and the models are referred to as *Vector-dependent Functionally Pooled (VFP)*. Due to space limitations, the case of AutoRegressive with eXogenous excitation (ARX) models is presently treated, although the more complete ARMAX (AutoRegressive Moving Average with eXogenous excitation models) case is also treated in Kopsaftopoulos [2012] and shall be presented elsewhere.

Following the postulation of the class of VFP models, the study proceeds with the formulation of proper estimators based on the Least Squares (LS) and Maximum Likelihood (ML) principles. For these developments, a proper vector algebra and the construction of functional subspaces comprising of polynomials of several variables (Dunkl and Xu [2001]) are employed. An additional contribution of the study is the establishment of the *strong* consistency (almost sure convergence) property of the formulated estimators (in contrast to the weaker "in probability" convergence proven in Sakellariou and Fassois [2007] and Hios and Fassois [2009b]). For this purpose a reformulation of the procedures and stronger asymptotic theorems are utilized. For the model structure (model orders and functional subspaces) estimation a Genetic Algorithm (GA) based procedure is formulated. The effectiveness and performance characteristics of the estimators are finally assessed via a Monte Carlo study.

## 2. THE VFP-ARX MODEL STRUCTURE

The identification of *Vector-dependent Functionally Pooled (VFP)* models is based on $M_1 \times M_2$ sets of excitation and/or response data records corresponding to a sample of the admissible operating conditions. Each record corresponds to a specific value of the operating parameter vector $\boldsymbol{k}$, which, without loss of generality, is presently assumed to be two-dimensional. A sample of $M_1$ values is used for the scalar operating parameter $k^1$ and a sample of $M_2$ values is used for $k^2$.

The set of all possible operating conditions for each scalar parameter belongs to the range $[k^1_{min},\ k^1_{max}] \in \mathbb{R}$ and $[k^2_{min},\ k^2_{max}] \in \mathbb{R}$, while discretized versions $\{k^1_1, k^1_2, \ldots, k^1_{M_1}\}$ and $\{k^2_1, k^2_2, \ldots, k^2_{M_2}\}$ are used for data acquisition. Hence, each experiment is characterized by a specific value of $\boldsymbol{k} = [k^1_i, k^2_j]$. 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 $k^2$). Assuming $N$ sample-long records, the complete data set is designated as [2]:

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

with $t$ designating normalized discrete time, and $x_{\boldsymbol{k}}[t], y_{\boldsymbol{k}}[t]$ the excitation and response signals, respectively, corresponding to the $\boldsymbol{k}$-th operating condition.

The VFP AutoRegressive with eXogenous excitation (VFP-ARX) representation postulated in this work is of the form:

$$y_{\boldsymbol{k}}[t] + \sum_{i=1}^{na} a_i(\boldsymbol{k}) \cdot y_{\boldsymbol{k}}[t-i] = \sum_{i=0}^{nb} b_i(\boldsymbol{k}) \cdot x_{\boldsymbol{k}}[t-i] + w_{\boldsymbol{k}}[t] \quad (2a)$$

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

$$a_i(\boldsymbol{k}) = \sum_{j=1}^{pa} a_{i,j} G_{d_a(j)}(\boldsymbol{k}),\ b_i(\boldsymbol{k}) = \sum_{j=1}^{pb} b_{i,j} G_{d_b(j)}(\boldsymbol{k}) \quad (2c)$$

$$E\big\{w_{k_{i,j}}[t] \cdot w_{k_{m,n}}[t-\tau]\big\} = \gamma_w[k_{i,j}, k_{m,n}] \cdot \delta[\tau] \quad (2d)$$

with $na$, $nb$ designating the AutoRegressive (AR) and eXogenous (X) orders, respectively, and $w_{\boldsymbol{k}}[t]$ the innovations signal, which is zero-mean, serially uncorrelated, with variance $\sigma^2_w(\boldsymbol{k})$, but potentially cross-correlated with its counterparts corresponding to different experiments. $E\{\cdot\}$ designates statistical expectation, $\delta[\tau]$ the Kronecker delta ($\delta[0] = 0, \delta[\tau] = 1$ for $\tau \neq 0$), $\mathcal{N}(\cdot,\cdot)$ Gaussian distribution, while iid stands for identically independently distributed. As (2c) indicates, the AR and X parameters $a_i(\boldsymbol{k})$, $b_i(\boldsymbol{k})$ are modelled as explicit functions of the operating parameter vector $\boldsymbol{k}$ and belong to the functional subspaces:

$$\mathcal{F}\langle a_i(\boldsymbol{k})\rangle \triangleq \{G_{d_a(1)}(\boldsymbol{k}), G_{d_a(2)}(\boldsymbol{k}), \ldots, G_{d_a(pa)}(\boldsymbol{k})\} \quad (3a)$$

$$\mathcal{F}\langle b_i(\boldsymbol{k})\rangle \triangleq \{G_{d_b(1)}(\boldsymbol{k}), G_{d_b(2)}(\boldsymbol{k}), \ldots, G_{d_b(pb)}(\boldsymbol{k})\} \quad (3b)$$

spanned by the basis functions $G_{d_a(j)}(\boldsymbol{k})$, $G_{d_b(j)}(\boldsymbol{k})$ consisting of polynomials of two variables. These may be obtained from corresponding univariate polynomials (Chebyshev, Legendre and other families – see Dunkl and Xu [2001]). $pa$, $pb$ designate the AR and X parameter subspace dimensionalities, respectively, while the indices $d_a(j)$ ($j = 1, \ldots, pa$) and $d_b(j)$ ($j = 1, \ldots, pb$) designate the specific basis functions that are included in each subspace. The constants $a_{i,j}$ and $b_{i,j}$ designate the AR and X coefficients of projection, respectively.

The VFP-ARX$(na, nb)_{[pa,pb]}$ model of (2a)–(2d) is parameterized in terms of the model's projection coefficients $a_{i,j}$, $b_{i,j}$, the innovations covariance $\gamma_w[k_{i,j}, k_{m,n}]$

[2] Bold-face upper/lower case symbols designate matrix/column-vector quantities, respectively.

$\big(\gamma_w[k_{i,j}, k_{i,j}] = \sigma_w^2[k_{i,j}]\big)$, and the model structure $\mathcal{M}$, defined by the model orders $na$, $nb$ and the functional subspaces $\mathcal{F}\langle a_i(\boldsymbol{k})\rangle$, $\mathcal{F}\langle b_i(\boldsymbol{k})\rangle$. It is worth noting that:

(1) All information in terms of interrelations among the data records in $Z^{NM_1M_2}$ is reflected in the covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{w}[t]} = E\{\boldsymbol{w}[t]\boldsymbol{w}^T[t]\}$ with $\boldsymbol{w}[t] \triangleq \big[w_{k_{1,1}}[t] \ \dots \ w_{k_{M_1,M_2}}[t]\big]^T$. This knowledge is incorporated into the parameter estimation phase to obtain statistically optimal models (see section 3).
(2) The projection of the parameters $a_i(\boldsymbol{k}), b_i(\boldsymbol{k})$ on the functional subspaces $\mathcal{F}\langle a_i(\boldsymbol{k})\rangle$, $\mathcal{F}\langle b_i(\boldsymbol{k})\rangle$ allows for models capable of representing the system dynamics everywhere within $[k_{min}^1,\ k_{max}^1] \times [k_{min}^2,\ k_{max}^2] \in \mathbb{R}^2$, and not only at the distinct values $\{k_1^1, k_2^1, \dots, k_{M_1}^1\} \times \{k_1^2, k_2^2, \dots, k_{M_2}^2\}$ involved in $Z^{NM_1M_2}$.
(3) The form of functional dependence is important. Physical insight may be used, while experience indicates that orthogonal polynomials or trigonometric functions are often sufficient (refer to Hios and Fassois [2009a] and Kopsaftopoulos and Fassois [2011] for practical applications where shifted Type II Chebyshev polynomials are employed).

Using the backshift operator $\mathcal{B}^i$ $\big(\mathcal{B} \cdot x[t] \triangleq x[t-i]\big)$ the VFP-ARX model may be expressed as follows:

$$A[\mathcal{B}, \boldsymbol{k}] \cdot y_{\boldsymbol{k}}[t] = B[\mathcal{B}, \boldsymbol{k}] \cdot x_{\boldsymbol{k}}[t] + w_{\boldsymbol{k}}[t] \tag{4a}$$

$$A[\mathcal{B}, \boldsymbol{k}] \triangleq 1 + \sum_{i=1}^{na} a_i(\boldsymbol{k})\mathcal{B}^i,\ B[\mathcal{B}, \boldsymbol{k}] \triangleq \sum_{i=0}^{nb} b_i(\boldsymbol{k})\mathcal{B}^i. \tag{4b}$$

In analogy to conventional models, this representation is assumed to satisfy the following conditions:

CD1. *Stability condition.* The poles of the AR polynomials (see Equation (4a)) lie inside the unit circle for the discretized sample values of $\boldsymbol{k}$.
CD2. *Irreducibility condition.* The polynomials $A[\mathcal{B}, \boldsymbol{k}]$ and $B[\mathcal{B}, \boldsymbol{k}]$ are left coprime for the sample values of $\boldsymbol{k}$.
CD3. The input signal $x_{\boldsymbol{k}}[t]$ is stationary, ergodic and persistently exciting with $E\{x_{k_{i,j}}[t] w_{k_{m,n}}[t]\} = 0$ $\forall\ i, j, m, n$.

## 3. VFP-ARX MODEL PARAMETER ESTIMATION

Consider the general case of VFP-ARX models with "incomplete" (that is not necessarily including all consecutive basis functions up to the specified degree) functional subspaces. The estimation of the VFP-ARX coefficient of projection vector:

$$\boldsymbol{\theta} \triangleq [a_{1,1} \dots a_{1,pa} \dots a_{na,pa} \vdots b_{0,1} \dots b_{0,pb} \dots b_{nb,pb}]^T \tag{5}$$

is considered based on available signal samples $\{x_{\boldsymbol{k}}[t]\}_{t=1}^N$, $\{y_{\boldsymbol{k}}[t]\}_{t=1}^N$ and a selected model structure $\mathcal{M}$.

The VFP-ARX model of (2a)–(2d) may be rewritten as:

$$y_{\boldsymbol{k}}[t] = \Big[\boldsymbol{\varphi}_{AR}^T[t] \otimes \boldsymbol{g}_{AR}^T(\boldsymbol{k}) \vdots \boldsymbol{\varphi}_X^T[t] \otimes \boldsymbol{g}_X^T(\boldsymbol{k})\Big] \cdot \boldsymbol{\theta} + e_{\boldsymbol{k}}[t] \Longrightarrow$$
$$y_{\boldsymbol{k}}[t] = \boldsymbol{\phi}_{\boldsymbol{k}}^T[t] \cdot \boldsymbol{\theta} + e_{\boldsymbol{k}}[t] \tag{6}$$

with $e_{\boldsymbol{k}}[t]$ designating the model's one-step-ahead prediction error (residual) corresponding to $\boldsymbol{k}$-th operating condition and:

$$\boldsymbol{\varphi}_{AR}[t] \triangleq \big[\ -y_{\boldsymbol{k}}[t-1] \ \dots \ -y_{\boldsymbol{k}}[t-na]\ \big]^T$$
$$\boldsymbol{\varphi}_X[t] \triangleq \big[\ x_{\boldsymbol{k}}[t] \ \dots \ x_{\boldsymbol{k}}[t-nb]\ \big]^T$$
$$\boldsymbol{g}_{AR}(\boldsymbol{k}) \triangleq \big[\ G_{d_a(1)}(\boldsymbol{k})\ G_{d_a(2)}(\boldsymbol{k}) \ \dots \ G_{d_a(pa)}(\boldsymbol{k})\ \big]^T$$
$$\boldsymbol{g}_X(\boldsymbol{k}) \triangleq \big[\ G_{d_b(1)}(\boldsymbol{k})\ G_{d_b(2)}(\boldsymbol{k}) \ \dots \ G_{d_b(pb)}(\boldsymbol{k})\ \big]^T.$$

Model parameter estimation requires stacking the equations of the form (6) for each distinct operating condition $\{k_{1,1}, k_{1,2}, \dots, k_{M_1,M_2}\}$. This procedure is referred to as *pooling* and gives:

$$\begin{bmatrix} y_{k_{1,1}}[t] \\ \vdots \\ y_{k_{M_1,M_2}}[t] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}_{k_{1,1}}^T[t] \\ \vdots \\ \boldsymbol{\phi}_{k_{M_1,M_2}}^T[t] \end{bmatrix} \cdot \boldsymbol{\theta} + \begin{bmatrix} e_{k_{1,1}}[t] \\ \vdots \\ e_{k_{M_1,M_2}}[t] \end{bmatrix} \Longrightarrow$$
$$\boldsymbol{y}[t] = \boldsymbol{\Phi}[t] \cdot \boldsymbol{\theta} + \boldsymbol{e}[t]. \tag{7}$$

Following substitution of the available signal samples (for $t = 1, \dots, N$) the following *dynamic* regression like expression is obtained:

$$\boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{\theta} + \boldsymbol{e}. \tag{8}$$

Notice that despite its phenomenal resemblance to standard regression, this expression includes a rich structure of interdependencies among the various variables and experiments, which need to be carefully taken into account. Furthermore, the term *functional pooling* signifies the functional dependence of each equation on the operating parameter vector $\boldsymbol{k}$. Proper estimation procedures based on the Least Squares (LS) and Maximum Likelihood (ML) principles are developed next.

### *3.1 Least Squares (LS) based estimation methods*

Minimization of the Weighted Least Squares (WLS) criterion:

$$J(\boldsymbol{\theta}) = \frac{1}{NM_1M_2} \sum_{t=1}^{N} \boldsymbol{e}^T[t] \boldsymbol{\Gamma}_{\boldsymbol{w}[t]}^{-1} \boldsymbol{e}[t] = \frac{1}{NM_1M_2} \boldsymbol{e}^T \boldsymbol{\Gamma}_{\boldsymbol{w}}^{-1} \boldsymbol{e} \tag{9}$$

with $\boldsymbol{\Gamma}_{\boldsymbol{w}} = \boldsymbol{\Gamma}_{\boldsymbol{w}[t]} \otimes \boldsymbol{I}_N$ ($\boldsymbol{I}_N$ designates the $N$ unity matrix), leads to the corresponding WLS estimator:

$$\widehat{\boldsymbol{\theta}}^{\mathrm{WLS}} = \big[\boldsymbol{\Phi}^T \boldsymbol{\Gamma}_{\boldsymbol{w}}^{-1} \boldsymbol{\Phi}\big]^{-1} \big[\boldsymbol{\Phi}^T \boldsymbol{\Gamma}_{\boldsymbol{w}}^{-1} \boldsymbol{y}\big]. \tag{10}$$

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

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{w}[t]}^{\mathrm{OLS}} = \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{e}[t, \widehat{\boldsymbol{\theta}}^{\mathrm{OLS}}] \boldsymbol{e}^T[t, \widehat{\boldsymbol{\theta}}^{\mathrm{OLS}}] \tag{11}$$

with $\boldsymbol{e}[t, \widehat{\boldsymbol{\theta}}^{\mathrm{OLS}}]$ designating the residuals $\boldsymbol{e}[t]$ for $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{\mathrm{OLS}}$.

The estimator in (10) is then expressed as:

$$\widehat{\boldsymbol{\theta}}^{\mathrm{WLS}} = \big[\boldsymbol{\Phi}^T (\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{w}}^{\mathrm{OLS}})^{-1} \boldsymbol{\Phi}\big]^{-1} \big[\boldsymbol{\Phi}^T (\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{w}}^{\mathrm{OLS}})^{-1} \boldsymbol{y}\big] \tag{12}$$

with the final residual covariance matrix obtained as:

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{w}[t]}^{\mathrm{WLS}} = \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{e}[t, \widehat{\boldsymbol{\theta}}^{\mathrm{WLS}}] \boldsymbol{e}^T[t, \widehat{\boldsymbol{\theta}}^{\mathrm{WLS}}]. \tag{13}$$

The Ordinary Least Squares (OLS) estimator is obtained as a special case by setting $\boldsymbol{\Gamma}_{\boldsymbol{w}} = \boldsymbol{I}$ in equation (10), with $\boldsymbol{I}$ designating the $NM_1M_2$ unity matrix.

### 3.2 The Maximum Likelihood (ML) estimation method

The complete parameter vector $\bar{\boldsymbol{\theta}} = \left[\boldsymbol{\theta}^T \vdots \gamma_w[k_{i,j}, k_{m,n}]\right]$ is estimated as Kopsaftopoulos [2012]:

$$\widehat{\bar{\boldsymbol{\theta}}}^{\mathrm{ML}} \triangleq \arg\max_{\bar{\boldsymbol{\theta}}} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Gamma}_{\boldsymbol{w}[t]}/\boldsymbol{e}) \tag{14}$$

with $\mathcal{L}(\cdot)$ designating the conditional likelihood function. In the general case of *normally distributed* and *contemporaneously correlated* residuals[3] $e_{\boldsymbol{k}}[t]\ \forall\ \boldsymbol{k}$ the ML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\Gamma}_{\boldsymbol{w}[t]}$ are obtained as:

$$\widehat{\boldsymbol{\theta}}^{\mathrm{ML}} \triangleq \arg\min_{\boldsymbol{\theta}} \ln\det\Big\{\frac{1}{N}\sum_{t=1}^{N} \boldsymbol{e}[t, \widehat{\boldsymbol{\theta}}]\boldsymbol{e}^T[t, \widehat{\boldsymbol{\theta}}]\Big\} \tag{15}$$

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{w}[t]} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{e}[t, \widehat{\boldsymbol{\theta}}^{\mathrm{ML}}]\boldsymbol{e}^T[t, \widehat{\boldsymbol{\theta}}^{\mathrm{ML}}] \tag{16}$$

Notice that the use of nonlinear optimization techniques is required.

It should be noted that the WLS based estimation coincides with the ML based estimation when the innovations covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{w}}$ in Equation (10) is replaced by consistent estimates.

## 4. LARGE SAMPLE PROPERTIES OF THE ESTIMATORS

The consistency and asymptotic distribution of the LS and the ML estimators are studied under the assumption of exactly known model structure. The proofs of the theorems may be found in Kopsaftopoulos [2012].

### 4.1 Consistency analysis

In contrast to previous studies referring to the scalar operating parameter case (Sakellariou and Fassois [2007], Hios and Fassois [2009b]) where a weaker "in probability" convergence was proven, *strong* consistency (almost sure convergence) of the formulated estimators is presently established. This is based on a proper reformulation of the appropriate asymptotic analysis framework and the use of proper strong convergence tools.

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

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

with a.s. designating almost sure convergence. □

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

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

[3] Contemporaneously correlated residuals: $E\{e_{k_{i,j}}[t]e_{k_{i,j}}[t]\} = \sigma^2_e[k_{i,j}]$ and $E\{e_{k_{i,j}}[t]e_{k_{m,n}}[t]\} = \gamma_e[k_{i,j}, k_{m,n}]$.

### 4.2 Asymptotic distribution

*Theorem 3.* (WLS estimator). Let $\boldsymbol{\theta}_o$ be the true projection coefficient vector, $w_{\boldsymbol{k}}[t]$ a white zero mean process with $E\{w^2_{\boldsymbol{k}}[t]\} = \sigma^2_w(\boldsymbol{k})$ for every operating condition, and $E\{\boldsymbol{\phi}_{\boldsymbol{k}}[t]\boldsymbol{\phi}_{\boldsymbol{k}}{}^T[t]\}$ a nonsingular matrix. Then:

$$\sqrt{NM_1M_2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{P}) \qquad (N \longrightarrow \infty)$$

with

$$\boldsymbol{P} = [\boldsymbol{\Phi}^T\boldsymbol{\Gamma}^{-1}_{\boldsymbol{w}}\boldsymbol{\Phi}]^{-1} \tag{17}$$

and "$\xrightarrow{d}$" designating convergence in distribution (White [2001, pp. 65–66]). □

*Theorem 4.* (ML estimator). Let $\bar{\boldsymbol{\theta}}_o = \left[\boldsymbol{\theta}_o^T \vdots \gamma_w[k_{i,j}, k_{m,n}]\right]$ be the true parameter vector, $w_{\boldsymbol{k}}[t]$ a normally distributed zero mean white process with $E\{w^2_{\boldsymbol{k}}[t]\} = \sigma^2_w(\boldsymbol{k})$ for every operating condition, and $E\{\boldsymbol{\phi}_{\boldsymbol{k}}[t]\boldsymbol{\phi}_{\boldsymbol{k}}{}^T[t]\}$ a nonsingular matrix.

Then the estimate $\widehat{\bar{\boldsymbol{\theta}}}$ follows asymptotically Gaussian distribution with mean $\bar{\boldsymbol{\theta}}_o$ and covariance matrix equal to that indicated by the Cramér-Rao lower bound:

$$\widehat{\bar{\boldsymbol{\theta}}}^{\mathrm{ML}}_N \sim \mathcal{N}(\bar{\boldsymbol{\theta}}, \boldsymbol{P}^{\mathrm{ML}}) \qquad (N \longrightarrow \infty). \qquad \square$$

The part of $\boldsymbol{P}^{\mathrm{ML}}$ that corresponds to the coefficients of projection vector $\boldsymbol{\theta}_o$ may be shown to coincide with the covariance matrix of equation (17), in analogy to the classical linear regression case with correlated residual sequences (Söderström and Stoica [1989, p. 564]). Hence, the WLS estimator of Section 3.1 achieves *efficient* estimation of $\boldsymbol{\theta}$ (reaching the Cramér-Rao lower bound) when a consistent estimate of the true covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{w}}$ is employed.

## 5. VFP-ARX MODEL STRUCTURE ESTIMATION

Model structure estimation may be based on customary statistical criteria, such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC – Ljung [1999], Söderström and Stoica [1989]), that are properly adapted to the VFP model structure. In the present study these criteria are used within the context of an introduced Genetic Algorithm (GA) based procedure which is very useful in the case of incomplete (that is not necessarily including consecutive basis functions) functional subspaces.

Given a basis function family, selection of the VFP-ARX model structure $\mathcal{M}$ refers to the estimation of the set of integers:

$$\mathcal{M} \triangleq \{na, nb, pa, pb, d_a(1), \ldots, d_a(pa), d_b(1), \ldots, d_b(pb)\}.$$

Model structure estimation may then be seen as the estimation of the integer-valued model structure vector $\boldsymbol{m}$:

$$\boldsymbol{m} \triangleq [na\ nb \vdots pa\ pb \vdots d_a(1)\ \ldots\ d_a(pa) \vdots d_b(1)\ \ldots\ d_b(pb)]^T \tag{18}$$

$$na, nb \in \{1, \ldots, \bar{n}\}, \quad pa, pb \in \{1, \ldots, \bar{p}\} \quad d_a(j), d_b(j) \in \{1, \ldots, \bar{d}\} \tag{19}$$

with $\bar{n}$, $\bar{p}$ and $\bar{d}$ designating the maximum considered orders, subspace dimensionalities and basis function indices,

respectively, which define the search space of the model structure estimation subproblem.

The estimation of $\boldsymbol{m}$ may be based on minimization of the BIC (Kopsaftopoulos [2012]):

$$\widehat{\boldsymbol{m}} = \arg\min_{\boldsymbol{m}} \text{BIC}(\boldsymbol{m}), \tag{20}$$

$$\text{BIC}(\boldsymbol{m}) = \ln det\{\boldsymbol{\Gamma}_{\boldsymbol{w}[t]}\} + \dim(\boldsymbol{\theta}) \cdot \frac{\ln(NM_1M_2)}{N}. \tag{21}$$

However, the model structure $\mathcal{M}$ is not uniquely defined in terms of the model structure vector $m$ as defined in (18). As the sub-vectors $[d_a(1) \dots d_a(pa)]$, $[d_b(1) \dots d_b(pb)]$ correspond to ordered sets of integers, any recomposition produces equivalent model structures (for example vectors $[d_a(1)\ d_a(2)\ d_a(3)]$ and $[d_a(1)\ d_a(3)\ d_a(2)]$ correspond to the same model structure). Thus, the model structure $\mathcal{M}$ is *not uniquely defined*, which implies that several global minima with respect to $\boldsymbol{m}$ exist in the BIC criterion (fitness function). Moreover, during the optimization procedure the dimension of the model structure vector $\boldsymbol{m}$ varies, as the subspace dimensionalities $pa$ and $pb$ change. Thus, the definition of the model structure vector $m$ of equation (18) is inappropriate for actual optimization.

In order for the model structure $\mathcal{M}$ to be uniquely defined, $\boldsymbol{m}$ is transformed into a binary variable vector as follows (details in Kopsaftopoulos [2012]):

$$\boldsymbol{m}_{bin} \triangleq \left[na\ nb \vdots z_{a,1} \dots z_{a,pa} \vdots z_{b,1} \dots z_{b,pb}\right]_{[3+pa+pb]} \tag{22}$$

with $z_{a,j}, z_{b,j}$ designating binary variables indicating whether the basis function $G_j(\boldsymbol{k})$ is included in the functional subspace $\mathcal{F}\langle AR\rangle$, $\mathcal{F}\langle X\rangle$, respectively. Thus:

$$z_{a,j} = 1 \iff G_j(\boldsymbol{k}) \in \mathcal{F}\langle AR\rangle$$
$$z_{a,j} = 0 \iff G_j(\boldsymbol{k}) \notin \mathcal{F}\langle AR\rangle$$

and similarly for the variables $z_{b,j}$.

By the above procedure the model structure $\mathcal{M}$ is uniquely defined by the fixed-dimensional vector $\boldsymbol{m}_{bin}$ of equation (22). Hence, the estimation of $\boldsymbol{m}_{bin}$ is achieved as:

$$\widehat{\boldsymbol{m}}_{bin} = \arg\min_{\boldsymbol{m}_{bin}} \text{BIC}(\boldsymbol{m}_{bin}). \tag{23}$$

The minimization of $\boldsymbol{m}$ constitutes a discrete variable optimization problem, which is tackled via the use of Genetic Algorithms (GAs). Alternatively, in order to reduce the dimension of the optimization problem, the AR and X model orders may be initially selected via customary model order selection techniques and the GA based procedure may be used for estimating the functional subspace dimensionalities and indices.

## 6. A MONTE CARLO STUDY

A VFP-ARX$(4,1)_{[9,9]}$ true model with zero delay is used to generate the data sets. An incomplete (that is not including all consecutively obtained functions) functional subspace of maximum polynomial degree 4 consisting of 9 shifted Type II Chebyshev polynomials (subspace dimensionality $p = pa = pb = 9$) is considered. The study includes 500 runs, in each one of which the first scalar operating parameter takes 20 values ($k_i^1 \in [1, 20]$) and the second 16 ($k_j^2 \in [1, 16]$). Thus, each run includes excitation-response signals ($N = 1024$ samples) from

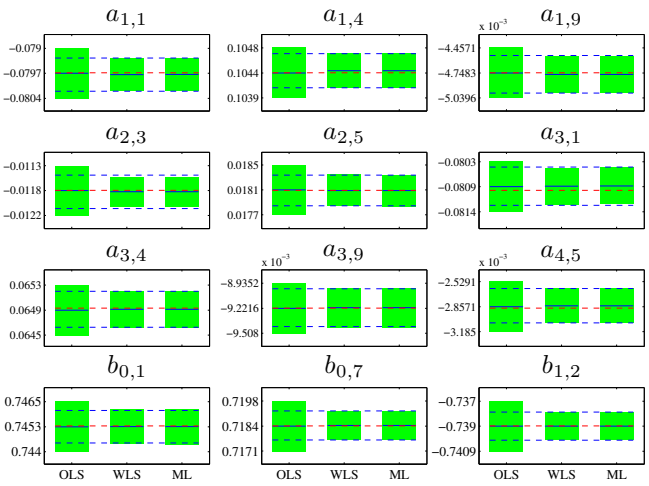


Fig. 1. Indicative VFP-ARX$(4,1)_{[9,9]}$ projection coefficient estimation results by the OLS, WLS and ML methods (500 runs per method): True values (dashed red lines) and sample mean estimates $\pm 1.96$ sample standard deviation (shaded boxes). The dashed blue lines indicate the theoretical (asymptotic) WLS/ML standard deviations.

Table 1. GA details for functional subspace estimation.

| Population | Elite | Crossover | Fitness func. tol. |
|---|---|---|---|
| 140 | 10 | 0.7 | $10^{-4}$ |

$M_1 \times M_2 = 320$ operating conditions. Each response is corrupted by random noise at the 10% standard deviation level (according to the ARX equation). The innovations sequences corresponding to different operating conditions are cross-sectionally uncorrelated, characterized by different variances $\sigma_w^2(\boldsymbol{k})$. In all cases the responses are generated by using a number of mutually independent, random sequences with zero mean and approximately flat spectra acting as excitations and innovations.

Model structure estimation consists of the determination of the functional subspace dimensionality $p$, as well as of the specific basis functions included in the subspace, and is treated via the GA procedure; details are provided in Table 1. The correct functional subspace dimensionality and specific basis functions have been thus selected in 89% of the cases (446/500 correct selections). These results may be further improved by increasing the population size, nevertheless computational burden will be also increased.

OLS and WLS estimation is based on QR implementation (Ljung [1999, pp. 318–320]). WLS estimation is non-iterative, while the ML estimation makes use of the Gauss-Newton non-linear optimization scheme (maximum number of iterations 100; maximum number of function evaluations 5000; termination tolerance of the loss function $10^{-2}$; termination tolerance of the estimated parameters $10^{-8}$). The WLS estimation is initialized by the OLS covariance matrix, while the ML estimator is initialized by the WLS estimates. Indicative parameter estimation results are presented in Fig. 1. All three methods provide accurate, effectively unbiased, estimates, with the WLS and ML methods expectedly achieving better accuracy for the coefficients of projection (smaller standard deviations).

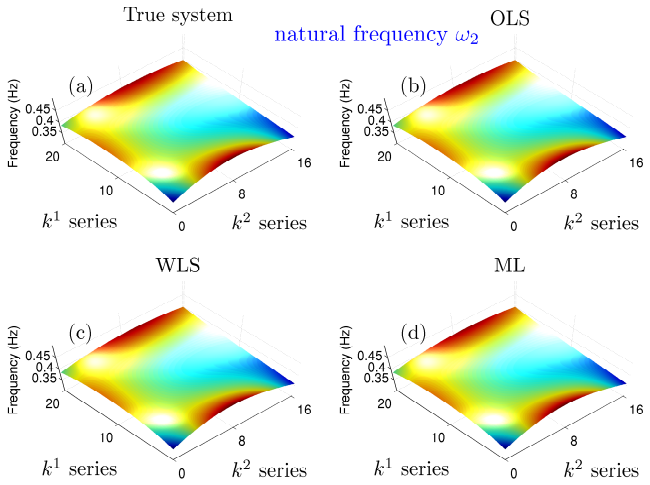


Fig. 2. VFP-ARX model second natural frequency versus $k^1$ and $k^2$: (a) true model, (b) OLS estimate, (c) WLS estimate, and (d) ML estimate (sample means over 500 runs).

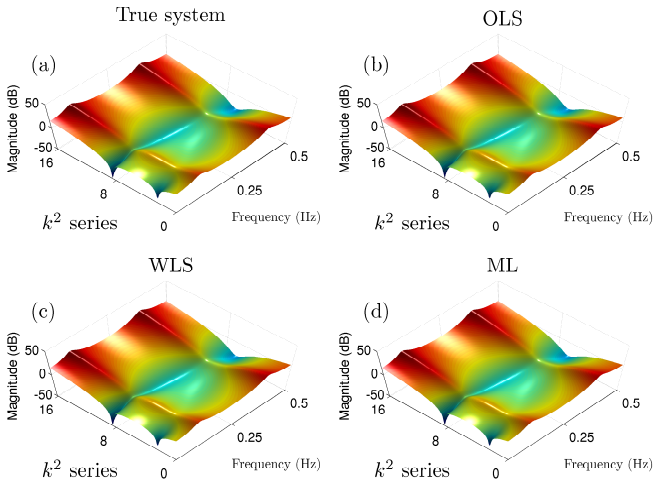


Fig. 3. VFP-ARX FRF magnitude versus frequency and $k^2$ ($k^1 = k_8^1$): (a) true model, (b) OLS estimate, (c) WLS estimate, and (d) ML estimate (sample means over 500 runs).

The validity of the derived asymptotic distribution for the projection coefficient estimators is confirmed by the excellent agreement between the sample standard deviations and their counterparts obtained through the asymptotic analysis.

The VFP-ARX$(4,1)_{[9,9]}$ second natural frequency versus the operating vector components ($k^1$, $k^2$) is depicted in Fig. 2 for the true model and the OLS, WLS, and ML estimates (sample means over 500 runs). Fig. 3 depicts the VFP-ARX$(4,1)_9$ based Frequency Response Function (FRF; $B[e^{-j\omega}, \boldsymbol{k}]/A[e^{-j\omega}, \boldsymbol{k}]$) magnitude surfaces versus frequency and $k^2$ (with $k^1 = k_8^1$), for the true system, as well as the OLS, WLS, and ML estimates (also sample means over 500 runs). The agreement between the true and all estimated curves is indeed excellent.

## 7. CONCLUDING REMARKS

The identification of global and parsimonious models based on data records obtained under a sample of different operating conditions, each one characterized by *several* parameters, was considered. The problem was tackled within a recently developed Functional Pooling (FP) framework. Vector-dependent Functionally Pooled ARX (VFP-ARX) models were postulated, and estimators based on the LS and ML principles were formulated. The estimators' *strong* consistency and asymptotic distribution were established. For model structure (including functional subspaces) estimation a Genetic Algorithm based procedure was formulated. The effectiveness of the formulated methods, along with the validity of the asymptotic results, were confirmed via a Monte Carlo study.

## REFERENCES

Dimogianopoulos, D.G., Hios, J.D., and Fassois, S.D. (2009). Fdi for aircraft systems using stochastic pooled-narmax representations: Design and assessment. *IEEE Transactions on Control Systems Technology*, 17(6), 1385–1397.

Dunkl, C. and Xu, Y. (2001). *Orthogonal Polynomials in Several Variables*. Cambridge University Press.

Hios, J.D. and Fassois, S.D. (2009a). Stochastic identification of temperature effects on the dynamics of a smart composite beam: assessment of multi–model and global model approaches. *Smart Materials and Structures*, 18(3), 035011 (15pp).

Hios, J.D. and Fassois, S.D. (2009b). Stochastic identification under multiple operating conditions: Functionally pooled varma methods. In *Proceedings of the 15th Symposium on System Identification (SYSID)*. Saint-Malo, France.

Kopsaftopoulos, F.P. (2012). *Advanced Functional and Sequential Statistical Time Series Methods for Damage Diagnosis in Mechanical Structures*. Ph.D. thesis, University of Patras, Greece.

Kopsaftopoulos, F.P. and Fassois, S.D. (2011). A stochastic functional model based method for vibration based damage detection, precise localization, and magnitude estimation. *Mechanical Systems and Signal Processing*, submitted to the special issue on interdisciplinary and integration aspects in Structural Health Monitoring.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice–Hall, 2nd edition.

Sakellariou, J.S. and Fassois, S.D. (2007). A functional pooling framework for the identification of systems under multiple operating conditions. In *Proceedings of 15th Mediterranean Conference on Control and Automation*. Athens, Greece.

Sakellariou, J.S. and Fassois, S.D. (2008). Vibration based fault detection and identification in an aircraft skeleton structure via a stochastic functional model based method. *Mechanical Systems and Signal Processing*, 22, 557–573.

Söderström, T. and Stoica, P. (1989). *System Identification*. Prentice–Hall.

Toth, R. (2010). *Modeling and Identification of Linear Parameter-Varying Systems*, volume 403 of *Lecture Notes in Control and Information Sciences*. Springer, Germany.

White, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press, 2nd edition.