# A population-based method for the identification of ARX models

Levy Batista, El-Hadi Djermoune, Thierry Bastogne, and Anne Gégout-Petit

***Abstract*—System identification is a data-driven input-output modeling approach more and more used in biology. With the advent of high-content screening studies in cell biology, new parameter estimation issues have emerged such as the identification of systems population to test the effects of numerous pharmaceutical compounds. To this end, a new data-driven modeling approach is proposed based on a hierarchical representation. An ARX (Auto-Regressive with eXogenous input) model structure is used at a low level to describe the input-output dynamics and, at a higher level, a gaussian distribution law characterizes the inter-individual dispersion between the systems of a population. Categorical covariates are introduced in the problem with the aim to finally estimate their effect on the input-output dynamics. The parameter estimation step relies on an EM (Expectation-Maximisation) algorithm. The so-formed population-based identification method is implemented, tested and compared with the classical ARX identification technique, which is repeatedly applied to each individual system. Using the Fisher information matrix, the parameter standard errors are estimated and two statistical tests are proposed to assess the effects of covariates. A Monte-Carlo simulation study is carried out and shows the practical relevance of the population-based identification method, which allows to drastically increase the statistical power of tests. Finally, an application on *in vitro* data illustrates the applicability of the innovative identification method.**

***Index Terms*—System identification, Population approach, Mixed effects models, EM algorithm, ARX structure**

## I. Introduction

FOR years, revolutions of nano- and bio-technologies have largely contributed to the advent of new modeling issues. One of them consists in identifying populations of dynamical systems. A first example in cell biology deals with the high-content screening (HCS) studies, which aim at testing several hundreds groups of pharmaceutical compounds by analyzing multidimensional data at early steps of drug design [30]. In this application context, cells are firstly incubated with the substance to be tested and the cell activity is measured in real time. A second example concerns the microfluidic systems used in nanomedicine for drug delivery [29]. Each microfluidic process is a very small unit of production that is continuously operating to provide small amounts of pharmaceuticals. The production scale-up generally requires to parallelize a few dozen or a few hundred of microfluidic systems. Another targeted biomedical application area deals with the identification from longitudinal data of prognostic biomarkers in personalized medicine [12].

In all those application fields, culture wells, manufacturing micro-units and patients are dynamic systems whose input-output variables are time signals. One common goal is to control their output trajectory in order to reach quality/efficacy specifications. Such a challenge cannot be addressed without getting empirical cause-effect models of the input-output dynamics. This is the field of system identification, a set of statistical methods that aim at building mathematical models of dynamical systems from measured data [18]. Numerous methods have been developed and successfully implemented to characterize and estimate the pharmacokinetic and pharmacodynamic effects of compounds and treatments [14], [11], [6], [7], [23], [1], [10], [9], [3], [2], [26], [25]. Every time, the proposed methods are applied to one or a few systems but does not globally address the identification issue of large sets of dynamic systems.

Indeed, another feature in common between the three emerging biomedical fields is the concept of population. Dozens, hundreds or thousands of similar systems are involved in each case. Quality considerations then require to control response reproducibility and to identify which critical risk factors affect the inter-system variability. To identify additive and synergistic effects of those risk factors, we need to integrate a new class of input variables called covariates. The latter are usually categorical variables and often correspond to qualitative information such as the cell lines, animal sex, diet type, compound category, patient history or genetic markers. Their effects on the input-output dynamics are then considered as unknown hyper-parameters to be estimated from input-output time series. This is the main issue addressed in this study.

The traditional approach to handle that problem consists in applying separately the dynamic system identification methods to each biological subject of the studied population. This strategy requires to repeat many times the identification procedure even though the input-output datasets are recorded during the same period of time and in the same experimental framework. It particularly includes the stochastic part of the dynamic model, which is systematically identified for each system when it could be assumed as identical within the studied population. Consequently, this technique unnecessary adds a large number of unknowns in the estimation problem and may finally significantly lower the statistical power of tests used to assess criticality of risk factors. Since every individual system

L. Batista is with CYBERNANO, Villers-lès-Nancy, France (e-mail: lbatista@cybernano.eu).

E.-H. Djermoune and T. Bastogne are with the Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France (e-mail: {elhadi.djermoune,thierry.bastogne}@univ-lorraine.fr).

Anne Gégout-Petit is with the Université de Lorraine, CNRS, IECL, F-54000 Nancy, France (e-mail: anne.gegout-petit@univ-lorraine.fr).

Thierry Batogne and Anne Gégout-Petit are also with INRIA, BIGS, Université de Lorraine, Vandoeuve-lès-Nancy, France.

can bring useful information for the rest of cell cultures, such a drawback could be avoided with a hierarchical modeling approach [15].

So far, the population-based identification of dynamic systems has not been addressed except in [13] in which the authors proposed a hybrid structure by adding an autoregressive part to a mixed model. However, the proposed mathematical representation is limited to a first-order dynamics and there is no probability distribution associated to the autoregressive parameters. Such prior information would allow us to regularize the estimation problem by defining the upper part of the hierarchical model. This paper aims at proposing a solution based on a class of hierarchical models composed of a given input-output model for the lower level and a probability distribution of the individual parameters for the upper level. The resulting augmented representation is also called random- or mixed-effect model. M. Lavielle and L. Aarons have recently shown that a number of pharmacokinetic models, which the parameter estimation problem is known to be ill-conditioned at an individual level, can become identifiable in practice at the population level if a number of specific assumptions on the probabilistic model hold [16]. Historically, this class of models is the result of extensive works carried out by the population approach group in the international statistical community [27], [15], but was not investigated in the system control framework despite the increase of application cases in this field.

Thereafter, we propose a population-based method for the identification of linear and time-invariant dynamic systems described by an ARX model structure whose (random) parameters are described by probability densities. The addressed issue is to estimate the hyper-parameters of this hierarchical model from input-output signals coming from a population of biological subjects. In that aim, we propose to implement an Expectation-Maximization (EM) algorithm to perform both a Bayesian estimation at the individual level and a maximum likelihood estimation at the population level. In [4], [5], we introduced a first version of solution and we assessed the proposed method on a simulation example and compared it with a classical system identification approach repetitively applied to each individual system. In this paper, we derive the complete calculations related to the estimation of the hyper-parameters and the Fisher information matrix which represents a key step to estimate the uncertainty on the parameters and the p-values used to estimate the criticality of the covariate effects. Moreover, the proposed mixed-effect ARX identification method is applied to real-world *in vitro* data to assess its usefulness in a real biological context.

The paper is organized as follows. In the next section, the hierarchical ARX model structure is presented. The EM algorithm is described in Section 3 for the parameter estimation. The uncertainty on the estimated parameters is derived from the Fisher information matrix in Section 4. In Section 5, the estimation performances are assessed on a single dataset and by performing Monte-Carlo simulations. Section 6 is devoted to the applicability assessement of the proposed method to *in vitro* data obtained from Gap-FRAP tests. Finally, we conclude by a discussion on the advantages and drawbacks of the proposed approach.

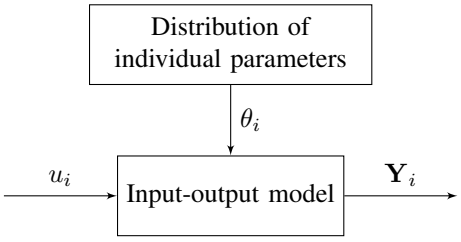



Fig. 1: Hierachical representation of population model

TABLE I: Description of the main indices used in this article

| Notation | Description |
|---|---|
| $I$ | nb of individual systems (subjects) |
| $T$ | nb of time samples per system output |
| $n_a + n_b$ | nb of parameters in the ARX model (bottom layer) |
| $n_g$ | nb of identifiable covariate effects (top layer) |
| $i \in \{1, \dots, I\}$ | subject index in a population of systems |
| $j \in \{1, \dots, n_a + n_b\}$ | parameter index in an individual system |
| $k \in \{1, \dots, n_g\}$ | covariate index |
| $t \in \{0, \dots, T-1\}$ | time-series index |

## II. Hierarchical ARX model

The hierarchical ARX model we propose to identify from input-output data is composed of two layers, as illustrated in Figure 1. The notations used in this paper are given in Table I.

### A. Bottom layer: an ARX model structure

The lower part of the hierarchy describes the input-output dynamical model based on an ARX structure, which is commonly used in the system identification literature [18]. Herein, it represents the dynamical behavior of the $i$-th subject or system (see Fig. 2) and satisfies the following equation for every $t \geq n_a$:

$$\mathcal{A}_i(q)\,\mathbf{Y}_i(t) = \mathcal{B}_i(q)\mathbf{u}_i(t) + E_i(t), \tag{1}$$

where $i \in \{1, \dots, I\}$, $\mathbf{Y}_i(t)$ is the output process until time $t$: $\mathbf{Y}_i(t) = (Y_i(s))_{0 \leq s \leq t}$ and $\mathbf{u}_i(t)$ is the input process until time $t$: $\mathbf{u}_i(t) = (u_i(s))_{s \leq t}$. Both of them are measured at discrete integer time instants $t < T$ and we suppose that $\mathbf{u}_i(t)$ is deterministic and known. $T$ is the number of time measurements of $\mathbf{Y}_i$ acquired at a constant sampling rate and $I$ is the number of subjects in a large biological sample to be analyzed. Moreover, the noises $E_i(t)$ of this equation are independent Gaussian variables

$$(E_i(t))_{t \geq 0} \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2).$$

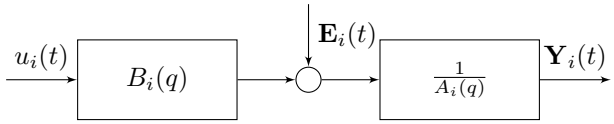



Fig. 2: ARX structure used to describe the individual input-output behaviour of the $i$-th system

The polynomials $\mathcal{A}_i(q)$, and $\mathcal{B}_i(q)$ are given by:

$$\mathcal{A}_i(q) = 1 + a_{i,1}q^{-1} + \ldots + a_{i,n_a}q^{-n_a} \quad (2)$$

$$\mathcal{B}_i(q) = b_{i,1}q^{-n_d} + \ldots + b_{i,n_b}q^{-(n_b-1+n_d)}, \quad (3)$$

where $q$ is the delay operator such that $\forall n \in \mathbb{Z}, q^{-n}g(t) = g(t-n)$ for a given time series $g(t)$. The polynomial orders $n_a$, $(n_b-1)$, and the input-output delay $n_d$ are assumed to be known; they will not be included in the estimation problem. Let us gather all the model parameters for the subject $i$ in the vector

$$\theta_i = [a_{i,1}, \ldots, a_{i,n_a}, b_{i,1}, \ldots, b_{i,n_b}]^{\mathrm{t}} \in \mathbb{R}^{(n_a+n_b)}. \quad (4)$$

At each current time instant, we assume the past values of the input-output variables are known and (1) becomes[1]:

$$Y_i(t)|\varphi_{i,t-1}, \theta_i = \varphi_{i,t-1}^{\mathrm{t}} \cdot \theta_i + E_i(t), \quad (5)$$

$$\sim \mathcal{N}(\varphi_{i,t-1}^{\mathrm{t}} \cdot \theta_i, \sigma_e^2) \qquad \forall t \geq n_a, \quad (6)$$

where $\varphi_{i,t-1}$ is a vector of size $(n_a + n_b)$:

$$\varphi_{i,t-1} = [-y_i(t-1), \ldots, -y_i(t-n_a), u_i(t-n_d), \ldots \\ \ldots, u_i(t-n_d-n_b+1)]^{\mathrm{t}} \quad (7)$$

that contains the previous (strictly before time instant $t$) observations from $\mathbf{Y}_i(t)$ and $\mathbf{u}_i(t)$ useful for the definition of $Y_i(t)$. In other words, if we consider the conditional density distribution of $Y_i$ at time $t$ given the past $\varphi_{i,t-1}$, $f_{Y|\varphi}(y|\varphi)$, it equals the Gaussian density function $\phi_{\varphi^{\mathrm{t}}\cdot\theta,\sigma_e^2}(y)$, where $\phi_{\mu,\sigma^2}(.)$ denotes the Gaussian density function with mean $\mu$ and variance $\sigma^2$. Despite the dependence between the successive values $Y_i(t)$, $t \geq 0$, one can decompose recursively the likelihood of the vector $\mathbf{Y}_i = [Y_i(n_a), \ldots, Y_i(T-1)]^{\mathrm{t}}$ leading to

$$\mathcal{L}(\mathbf{Y}_i\,|\varphi_{i,0}, \theta_i) = \prod_{t=n_a}^{T-1} f_{Y_i(t)|\varphi_{i,t-1}}(Y_i(t)) \quad (8)$$

$$= \prod_{t=n_a}^{T-1} \phi_{\varphi_{i,t-1}^{\mathrm{t}}\cdot\theta_i,\sigma_e^2}(Y_i(t)). \quad (9)$$

From this expression of the individual likelihood, each of the $I$ individual models can be estimated independently from each other using the maximum likelihood estimator that is given by

$$\hat{\theta}_i = \left(\mathbf{\Phi}_i^{\mathrm{t}}\mathbf{\Phi}_i\right)^{-1}\mathbf{\Phi}_i^{\mathrm{t}}\mathbf{Y}_i \quad (10)$$

where $\mathbf{\Phi}_i$ is a regressor matrix of dimension $(T-n_a)\times(n_a+n_b)$:

$$\mathbf{\Phi}_i = [\varphi_{i,n_a-1}, \ldots, \varphi_{i,T-2}]^t \in \mathbb{R}^{(T-n_a)\times(n_a+n_b)}. \quad (11)$$

The vector $\hat{\theta}_i$ is an unbiased and minimum-variance estimate of $\theta_i$ and its variance is given by [18]

$$\mathrm{var}(\hat{\theta}_i) = \left(\frac{\mathbf{\Phi}_i^{\mathrm{t}}\mathbf{\Phi}_i}{\sigma_e^2}\right)^{-1} \quad (12)$$

It is worth mentioning that even though the variance is minimum among the linear estimators, it is not necessary small. For that to be possible, the following conditions should be satisfied:

- $\sigma_e^2$ needs to be small;
- $T$ and consequently $T-n_a$, the number of observations used for the estimation and number of rows in $\mathbf{\Phi}_i$, should be as large as possible;
- $\mathbf{\Phi}_i$ has to be full column rank and well-conditioned, that is, $u_i$ should be a persistent excitation of order $n_a+n_b$ to collect relevant information about the system through the output measurements [18].

These conditions are often in opposition with most of the biological experimental set-ups. In practice, the noise level is usually high, $T$ is generally small and the choice of the input signal is very limited, if not impossible.

### *B. Top layer: a mixed-effect model*

The idea is to add prior information owing to the upper part of the hierarchical model. Despite the variability of individuals in each group, all their responses remain similar and individual parameters should be all in the same vicinity. To encode such a proximity between subjects' parameters, we define the upper level of the hierarchy with the following distribution:

$$\theta_i \sim \mathcal{N}(\theta_0 + Bc_i, \Omega), \quad i = 1, \ldots, I \quad (13)$$

where $\theta_0 = [\theta_{0,1}, \ldots, \theta_{0,(n_a+n_b)}]^{\mathrm{t}}$ is the vector of mean parameters associated with the reference group, $c_i = [c_{i,1}, \ldots c_{i,n_g}]^{\mathrm{t}}$ contains the $n_g$ values or levels of the covariates for subject $i$ and the $(n_a+n_b)\times n_g$ matrix $B = (\beta_{j,k})_{1\leq j\leq(n_a+n_b),1\leq k\leq n_g}$ with $\beta_{j,k}$ is the fixed effect of the $k$-th covariable $c_{i,k}$ on $\theta_{i,j}$. The positive semi-definite covariance matrix $\Omega \in \mathbb{R}^{(n_a+n_b)\times(n_a+n_b)}$ describes the random variability of parameters between subjects in the same group. Equation (13) can be detailed for each single parameter. The mixed-effect model of the $j$-th parameter $\theta_{i,j}$ is given by:

$$\theta_{i,j} = \theta_{0,j} + \sum_{k=1}^{n_g}\beta_{j,k}c_{i,k} + W_{i,j}, \qquad j \in \{1, \ldots, n_a+n_b\}, \quad (14)$$

in which $\theta_{0,j}$ denotes the intercept term describing the mean effect of the reference group while $\beta_{j,k}$ is the fixed effect of the $k$-th covariable on $\theta_{i,j}$. $\mathbf{W}_{i,\cdot} \sim \mathcal{N}(\mathbf{0}_{n_a+n_b}, \Omega)$ describes the random effects, *i.e.* the intra-group variability and the correlation between the $\theta_{i,j}$'s. Covariates $c_{i,k}$ can be either quantitative variables or binary variables. In the latter case, modalities 0 and 1 stand for the reference level and the level to be tested, respectively. Similarly, if one covariable has more than two levels ($l > 2$) then it can be encoded by $l-1$ binary variables.

The unknown parameters $\theta_0$ and $B$ in (13) may be gathered in a single vector. Let us denote by $\beta$ the vectorization of matrix $B$: $\beta = \mathrm{vec}(B)$, and the vector of parameters by $\Upsilon = [\theta_0^{\mathrm{t}}, \beta^{\mathrm{t}}]^{\mathrm{t}}$. It is straightforward to show that

$$\theta_0 + Bc_i = C_i\Upsilon \quad (15)$$

where $C_i = [1,\, c_i^{\mathrm{t}}] \otimes \mathrm{I}_{n_a+n_b} \in \mathbb{R}^{(n_a+n_b)\times(n_a+n_b)(n_g+1)}$, $\otimes$ and $\mathrm{I}_n$ stand for the Kronecker product and the identity matrix

[1]Note that the input signal has to be known for $t < 0$ if $n_b+n_d-1 > n_a$.

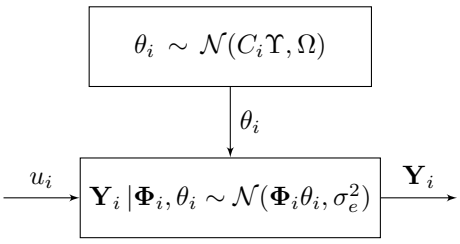



Fig. 3: Hierachical representation of population model

of size $n$, respectively. The parameters in $\Upsilon$ are clearly divided into two parts:

- the upper part $\theta_0 \in \mathbb{R}^{(n_a+n_b)}$ is a vector containing the mean parameters of the reference group,
- the lower part $\beta \in \mathbb{R}^{n_g(n_a+n_b)}$ contains the hyperparameters estimating the relative effects of the $n_g$ covariable levels on the $(n_a + n_b)$ model parameters.

### C. Estimation Problem

Finally, the hyper-parameters to be estimated are gathered in:

$$\Theta = (\Upsilon, \Omega, \sigma_e^2). \tag{16}$$

Using (6), one can express for each $t$ the conditional density of $Y_i(t)$ according to $\Theta$, the past and covariates $c_i$ by

$$Y_i(t)|\varphi_{i,t-1}, \Theta \sim \mathcal{N}(\varphi_{i,t-1}^{\mathrm{t}} \cdot C_i\Upsilon, \varphi_{i,t-1}^{\mathrm{t}}\Omega\varphi_{i,t-1} + \sigma_e^2), \tag{17}$$

$\forall t \geq n_a$. $\Theta$ can be determined by maximizing the likelihood $\mathcal{L}(\mathbf{Y}_i; \Theta|\varphi_{i,0})$. However in the presence of latent variables (the $\theta_i$'s here), it is usual to use a EM algorithm. It is the object of the next section, where we propose to implement an EM algorithm to maximize the likelihood of the upper level using the posterior expectation of $\theta_i|y_i$ from the lower level of the hierarchical model (see Figure 3).

## III. Expectation-maximization algorithm for the maximum likelihood estimation

The expectation-maximisation (EM) algorithm was first introduced by Dempster [8] who proved that each iteration of the algorithm increases the likelihood of the hyper-parameters. This algorithm relies on a Bayesian approach in which a prior information is used on the missing parameters $\theta_i$ and maximizes the likelihood of the hyper-parameters $\Theta$. It is well known for its ability to deal with missing data [8]. Indeed, the EM algorithm estimates the hidden variables, in our case $\theta_i$, by its expectation conditionally to $y_i$, a realization of $\mathbf{Y}_i$, and $\hat{\Theta}^{(\ell-1)}$, the estimation of $\Theta$ after $\ell - 1$ iterations of the algorithm.

### A. Maximization step

Let us define $\mathcal{L}(\mathbf{Y}_i^{(c)}|\Theta)$ the likelihood of the complete data vector for the $i$-th system, $\mathbf{Y}_i^{(c)} = [\mathbf{Y}_i^{\mathrm{t}}, \theta_i^{\mathrm{t}}]^{\mathrm{t}}$, corresponding to the concatenation of the output signal $\mathbf{Y}_i$ and the vector $\theta_i$ considered as the hidden variable,

$$\begin{aligned}\mathcal{L}(\mathbf{Y}_i^{(c)}; \Theta) &= \mathcal{L}(\mathbf{Y}_i, \theta_i; \Theta) \\ &= \mathcal{L}(\mathbf{Y}_i\,|\theta_i)\mathcal{L}(\theta_i; \Theta).\end{aligned} \tag{18}$$

At each iteration $\ell$, the principle of the EM algorithm is to maximize the conditional expectation of the complete log-likelihood given the observations $(\mathbf{Y}_i)_{1\leq i\leq I}$ and the current estimate of the parameter $\hat{\Theta}^{(\ell-1)}$ that is to maximize:

$$\begin{aligned}&\mathbb{E}_{\hat{\Theta}^{(\ell-1)}}\left[\log\Big(\prod_{i=1}^{I}\mathcal{L}(\mathbf{Y}_i^{(c)}; \Theta)\Big)\Big|(\mathbf{Y}_i)_{1\leq i\leq I}\right] \\ &= \sum_{i=1}^{I}\mathbb{E}_{\hat{\Theta}^{(\ell-1)}}\left[\log\big(\mathcal{L}(\mathbf{Y}_i\,|\theta_i)\big)|(\mathbf{Y}_i)_{1\leq i\leq I}\right] \\ &+ \sum_{i=1}^{I}\mathbb{E}_{\hat{\Theta}^{(\ell-1)}}\left[\log\big(\mathcal{L}(\theta_i; \Theta)\big)|(\mathbf{Y}_i)_{1\leq i\leq I}\right].\end{aligned} \tag{19}$$

The first term in the RHS of (19) concerns $\Theta_{\mathbf{Y}} = \sigma_e^2$, the part of $\Theta$ that parametrizes the law of $Y_i|\theta_i$ and the second one $\Theta_\theta = (\Upsilon, \Omega)$ parametrizes the law of the $\theta_i$'s. The two parts can be maximized independently leading to equations (20)-(22) given at the top of the next page, where $\mathrm{tr}(\cdot)$ stands for the trace of a matrix.

The fact to maximize the conditional expectation of the complete log-likelihood is like replacing the missing variable $\theta_i$ by its posterior expectation. This is the technical solution of the EM algorithm to iteratively compute estimations of $\theta_i$ and $\Theta$ that maximize the complete log-likelihood. The computation of the posterior expectation is the object of the Expectation step that is explained in the next section.

### B. Expectation step

The E-step aims at computing the posterior expectation of the individual parameter $\theta_i$ with (13) being the prior distribution. We use again the distribution of the complete data vector $\mathbf{Y}_i^{(c)}$ for the $i$-th system. We use the different conditional factorizations of the likelihood of $\mathbf{Y}_i^{(c)}$ to show that

$$\mathcal{L}(\theta_i|\,\mathbf{Y}_i, \Theta) = \frac{\mathcal{L}(\mathbf{Y}_i\,|\theta_i, \Theta)\mathcal{L}(\theta_i|\Theta)}{\mathcal{L}(\mathbf{Y}_i\,|\Theta)} \tag{23}$$

We combine Equations (6), (13), (15) and (17) and use classical computations on normal distributions to show that the posterior distribution of $\theta_i|y_i$ is Gaussian with mean and variance given by:

$$\mathbb{E}_\Theta(\theta_i|\,\mathbf{Y}_i) = C_i\Upsilon + \Omega\mathbf{\Phi}_i^{\mathrm{t}}V_i^{-1}(\mathbf{Y}_i - \mathbf{\Phi}_i C_i\Upsilon) \tag{24}$$

$$\mathrm{var}_\Theta(\theta_i|\,\mathbf{Y}_i) = \Omega - \Omega\mathbf{\Phi}_i^{\mathrm{t}}V_i^{-1}\mathbf{\Phi}_i\Omega, \tag{25}$$

with $V_i = \mathbf{\Phi}_i\Omega\mathbf{\Phi}_i^{\mathrm{t}} + \sigma_e^2\mathrm{I}_{T-n_a}$. Using the Woodbury identity, we can rewrite :

$$\mathrm{var}_\Theta(\theta_i|\,\mathbf{Y}_i) = \left(\frac{\mathbf{\Phi}_i^{\mathrm{t}}\mathbf{\Phi}_i}{\sigma_e^2} + \Omega^{-1}\right)^{-1}. \tag{26}$$

It is interesting to compare (12) with (26). We observe that a new term is added. The latter is the inverse of the *a priori* variance on $\theta_i$. For instance, if the prior is noninformative ($\Omega \to +\infty$), it will not affect the estimation of $\theta_i$ compared to the individual approach.

$$\hat{\Upsilon}^{(\ell)} = \left( \sum_{i=1}^{I} C_i^{\mathrm{t}} (\hat{\Omega}^{(\ell-1)})^{-1} C_i \right)^{-1} \sum_{i=1}^{I} C_i^{\mathrm{t}} (\hat{\Omega}^{(\ell-1)})^{-1} \mathbb{E}_{\Theta^{(\ell-1)}} \left[ \theta_i | \, \mathbf{Y}_i \right] \tag{20}$$

$$\hat{\Omega}^{(\ell)} = \frac{1}{I} \sum_{i=1}^{I} \left( C_i \Upsilon^{(\ell)} \Upsilon^{(\ell)\mathrm{t}} C_i^{\mathrm{t}} - 2\mathbb{E}_{\Theta^{(\ell-1)}} \left[ \theta_i | \, \mathbf{Y}_i \right] \Upsilon^{(\ell)\mathrm{t}} C_i^{\mathrm{t}} + \mathbb{E}_{\Theta^{(\ell-1)}} \left[ \theta_i \theta_i^{\mathrm{t}} | \, \mathbf{Y}_i \right] \right) \tag{21}$$

$$\hat{\sigma}_e^{2(\ell)} = \frac{1}{(T - n_a) I} \sum_{i=1}^{I} \left( \mathbf{Y}_i^{\mathrm{t}} \mathbf{Y}_i - 2\mathbb{E}_{\Theta^{(\ell-1)}} \left[ \theta_i | \, \mathbf{Y}_i \right] \mathbf{\Phi}_i^{\mathrm{t}} \mathbf{Y}_i + \mathrm{tr}(\mathbf{\Phi}_i^{\mathrm{t}} \mathbb{E}_{\Theta^{(\ell-1)}} \left[ \theta_i \theta_i^{\mathrm{t}} | \, \mathbf{Y}_i \right] \mathbf{\Phi}_i) \right). \tag{22}$$

Let us define the sufficient statistics needed to maximize the likelihood of the hyper-parameters at iteration $\ell$ by:

$$\begin{aligned} S_{i,1}^{(\ell)} &\triangleq \mathbb{E}_{\Theta^{(\ell-1)}} \left[ \theta_i | \, \mathbf{Y}_i \right] \\ &= C_i \Upsilon^{(\ell-1)} + \Omega^{(\ell-1)} \mathbf{\Phi}_i^{\mathrm{t}} (V_i^{(\ell-1)})^{-1} (\mathbf{Y}_i - \mathbf{\Phi}_i C_i \Upsilon^{(\ell-1)}), \end{aligned} \tag{27}$$

the same for the sum of squares $\theta_i \theta_i^{\mathrm{t}}$:

$$\begin{aligned} S_{i,2}^{(\ell)} &\triangleq \mathbb{E}_{\Theta^{(\ell-1)}} \left[ \theta_i \theta_i^{\mathrm{t}} | \, \mathbf{Y}_i \right] \\ &= S_{i,1}^{(\ell)} S_{i,1}^{(\ell)\mathrm{t}} + \Omega^{(\ell-1)} - \Omega^{(\ell-1)} \mathbf{\Phi}_i^{\mathrm{t}} (V_i^{(\ell-1)})^{-1} \mathbf{\Phi}_i \Omega^{(\ell-1)}. \end{aligned} \tag{28}$$

In the maximization step we use them to replace the hidden variables by their statistics $S_{i,1}^{(\ell)}$, and $S_{i,2}^{(\ell)}$. We can initialize the upper part of $\hat{\Upsilon}^{(0)}$ (*i.e.* $\hat{\theta}_0^{(0)}$) and $\hat{\sigma}_e^{2(0)}$ by an ARX identification applied to one individual, and $\hat{\Omega}^{(0)}$ with reasonable arbitrary values. In the case where covariables are used, the lower part of $\hat{\Upsilon}^{(0)}$ (*i.e.* $\hat{\beta}^{(0)}$) is set to 0. It can be shown that the procedure increases the conditional likelihood of the complete data (18) at each iteration.

## IV. Uncertainty of estimates

### A. Fisher information matrix

In the Gaussian framework, we know that the maximum likelihood estimator is unbiased and reaches the Cramér-Rao bound, which is the lower limit of the variance. The Cramér-Rao bound is the inverse of the Fisher information matrix. If we maximize the likelihood $\mathcal{L}(\mathbf{Y}; \Theta) = \prod_{i=1}^{I} \mathcal{L}(\mathbf{Y}_i; \Theta)$ of the observed data $(\mathbf{Y}_i)_{1 \leq i \leq I}$ then the Fisher information matrix is expressed as:

$$\mathcal{I}_{\mathbf{Y}}(\Theta) = -\mathbb{E}_{\Theta} \left[ \frac{\partial^2 \log \mathcal{L}(\mathbf{Y}; \Theta)}{\partial \Theta \partial \Theta^{\mathrm{t}}} \right]. \tag{29}$$

However, in order to keep the hierarchical meaning of the model we maximized $\mathcal{L}(\mathbf{Y}^{(c)}; \Theta) = \prod_{i=1}^{I} \mathcal{L}(\mathbf{Y}_i^{(c)}; \Theta)$ with the EM algorithm. In this case, T. A. Louis proposed in [19] a solution to compute the Fisher information matrix of the measured data from the vector of the complete data (completed by the $\theta_i$'s). He showed that the information on the data is equal to the information of the complete data minus the missing information:

$$\mathcal{I}_{\mathbf{Y}}(\Theta) = \mathcal{I}_{\mathbf{Y}^{(c)}}(\Theta) - \mathcal{I}_{\theta | \, \mathbf{Y}}(\Theta), \tag{30}$$

where $\theta = (\theta_i)_{1 \leq i \leq I}$. Thereby using the missing information principle, the observed information may be expressed as:

$$\mathcal{I}_{\mathbf{Y}}(\Theta) = -\mathbb{E}_{\Theta} \left[ \frac{\partial^2 \log \mathcal{L}(\mathbf{Y}^{(c)}; \Theta)}{\partial \Theta \partial \Theta^{\mathrm{t}}} \right] - \mathrm{var}_{\Theta} \left[ \frac{\partial \log \mathcal{L}(\theta; \Theta)}{\partial \Theta} \right]. \tag{31}$$

Note that these statistics can be computed in the last iteration of the EM algorithm using the gradient and the Hessian matrix for the complete-data problem. The derivatives in (31) have analytical forms but computing the expectation and the variance-covariance matrix require to calculate high order moments of $\theta_i$. To assess these quantities we used the moment-generating function of a multivariate normal distribution. Once the Fisher information matrix is estimated, we can compute the standard errors by inverting this matrix and taking the square root of the diagonal elements. Details of the calculation of the Fisher information matrix are provided in the supplementary material associated with the paper.

### B. Statistical tests

Differences between groups can be tested at the two different levels, either on the lower level (*i.e.* testing directly the individual parameters $\theta_i$ with a t-test) or on the upper level (testing the hyper-parameters $\beta$ of being different from 0). We have used the t-test to compare the results between the individual approach and the population approach.

*1) Two-sample t-test:* The two-sample t-test is a parametric test that compares the mean parameter of two independent samples. Assuming the two data samples, $x$ and $y$, are drawn from populations with equal variances, the associated test statistics is defined as:

$$t = \frac{\overline{x} - \overline{y}}{s}, \tag{32}$$

with

$$s = \sqrt{\frac{(I_1 - 1) s_x^2 + (I_2 - 1) s_y^2}{I_1 + I_2 - 2}}, \tag{33}$$

where $\overline{x}$ and $\overline{y}$ are the sample means. In our case, $x$ corresponds to the $\theta_i$ associated with the first group and $y$ the one of the second group. $s_x$ and $s_y$ are the sample standard deviations, $I_1$ and $I_2$ are the sample sizes. The test statistic under the null hypothesis follows a Student's t-distribution with $I_1 + I_2 - 2$ degrees of freedom.

Note that this test is usually too optimistic since it does not take into account that the parameters $\theta_i$ are estimated and

therefore uncertain. On the contrary, the Wald test is computed with the parameter uncertainty, but this test is much more complicated for the individual approach because it requires to determine the uncertainty of the mean group from the uncertainty of the individual estimations. In this paper, only the t-test is used to compare the two approaches, and the Wald test is only computed for the population approach.

*2) Wald Test:* As in [21], once the standard errors (SE) are estimated, we propose to use the Wald test to assess the effects of covariables, coded in $\Upsilon$ (see (15) and (14), on the input-output model parameters. As defined in Section 2, $\beta_{j,k}$ is the fixed effect of a given covariable $k$ on the model parameter $j$. For any $\beta_{j,k}$ we compute its estimate $\hat{\beta}_{j,k}$. The null hypothesis to test is $H_0 : \{\beta_{j,k} = 0\}$.

The Wald statistics is defined as:

$$S_W = \frac{\hat{\beta}^2_{j,k}}{SE(\hat{\beta}_{j,k})^2}, \tag{34}$$

and follows a $\chi^2_1$ distribution with one degree of freedom under $H_0$. In the results presented later on, we display the p-value of this test.

## V. Performance assessment

### A. Experimental design

The objective is to compare the estimation performances of the classical identification strategy applied independently to all the tested subjects (systems) and the population-based identification approach developed in the previous sections. The simulated input-output data of the $i$-th system are generated by the following ARX model:

$$Y_i(t) = -a_{i,1}Y_i(t-1) - a_{i,2}Y_i(t-2) + b_{i,1}u_i(t-1) + E_i(t) \tag{35}$$

with $n_a = 2$, $n_b = 1$, $n_d = 1$, $E_i(t) \sim \mathcal{N}(0, \sigma_e^2 = 0.01)$ (SNR $= 7.33\,\text{dB}$) and $t \in [0, ..., T-1]$ with $T = 30$ (small dataset). The noise level was chosen high to be conformed with most of situations observed in *in vivo* and clinical trials. The model parameters are gathered in $\theta_i = [a_{i,1}\ a_{i,2}\ b_{i,1}]^{\text{t}}$. These parameters are distributed according to a gaussian distribution with a mean vector $C_i\Upsilon$ and a diagonal covariance matrix $\Omega$ describing the inter-individual variability matrix in each group of subjects. Let $\omega^2 = (\omega^2_{a_1}, \omega^2_{a_2}, \omega^2_{b_1})$ denote the vector composed of the diagonal entries of $\Omega$, corresponding with the random effects on the model parameters, *i.e.* the unexplained intra-group variability. We set $\omega^2_{a_1} = \omega^2_{a_2} = 0.001$ and $\omega^2_{b_1} = 0.01$. We have considered two groups of $I_1 = I_2 = 20$ subjects. The first group is used as reference and the number of identifiable covariable effects is equal to $n_g = 1$. Parameters of the reference group are centered on $a_{ref,1} = 0.7$; $a_{ref,2} = 0.01$; $b_{ref,1} = 1$ and the ones of the second group are shifted by quantities $\delta a_{g,1} = -0.2$; $\delta a_{g,2} = 0.1$; $\delta b_{g,1} = 0.2$. The covariate matrix $C_i$ is given by:

$$C_i = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right], \tag{36}$$

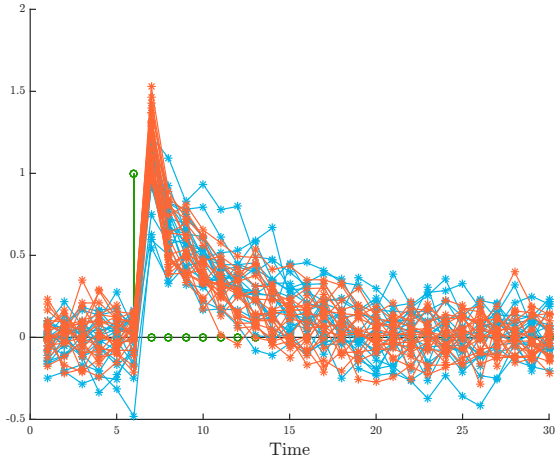



Fig. 4: Example of a simulation dataset. Input signal (green); output signals from group 1 (blue); output signals from group 2 (orange).

if the subject $i$ belongs to the reference group; otherwise (group 2), the expression of $C_i$ becomes:

$$C_i = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]. \tag{37}$$

The two elements of the vector of unknown parameters $\Upsilon$ are:

$$\theta_0 = \left[a_{ref,1}, a_{ref,2}, b_{ref,1}\right]^{\text{t}} \tag{38}$$

$$\beta = \left[\delta a_{g,1}, \delta a_{g,2}, \delta b_{g,1}\right]^{\text{t}}. \tag{39}$$

In (37), the content of the second block assumes that the change of group can affect all the ARX model parameters. A Kronecker pulse signal at $t = 5$ is chosen as input for all subjects. Data are simulated with a constant sampling rate. The same ARX model structure is used in the two identification strategies (individually repeated *vs* population-based) to be compared.

### B. Results and discussion

*1) Test on one dataset:* A simulation dataset is shown in Figure 4 and the corresponding estimates obtained with the population-based identification method only are presented in Table II. We note that the elements of $\Upsilon$ are well estimated with low standard-errors, the same goes for $\sigma^2_e$. About the components of the covariance matrix $\Omega$, we observe a slight over-estimation of $\omega^2$ compared with the simulation values. This behavior was expected because the resulting variance is a combinaison of the true variance and the variance of the estimates $\theta_i$. The p-values clearly indicate that two parameters, $a_1$ and $b_1$, are more likely to be affected by the group covariable. The p-value for $\beta_2$ is larger but is still significant, emphasizing that the group factor also affects $a_2$. These results are coherent with the simulation set up since the offset for $a_2$ was twice lower. The estimates for $\theta$ and $\beta$ are also acceptable compared with the simulation values. Moreover, the proposed approach succeeded in detecting the real effects of the group covariable.

TABLE II: Estimation results of the population-based identification method applied to one input-output dataset shown in Figure 4. SE is the standard error and RSE (%) is the relative standard error. Wald test p-values are also indicated to assess which model parameters are affected by the change of group (covariable).

| Parameters | | True Values | Estimates | SE | RSE | p-value |
|---|---|---|---|---|---|---|
| $\hat{\Upsilon}$ | $\hat{\theta}_{0,1} = \hat{a}_{ref,1}$ | 0.6 | 0.6821 | 0.0414 | 6 | |
| | $\hat{\theta}_{0,2} = \hat{a}_{ref,2}$ | 0.1 | 0.0963 | 0.0415 | 43 | |
| | $\hat{\theta}_{0,3} = \hat{b}_{ref,1}$ | 1 | 1.0012 | 0.0358 | 4 | |
| | $\hat{\beta}_{1,1} = \delta\hat{a}_{g,1}$ | -0.2 | -0.2039 | 0.0488 | 24 | 2.89e-05 |
| | $\hat{\beta}_{2,1} = \delta\hat{a}_{g,2}$ | 0.1 | 0.1133 | 0.0489 | 43 | 0.0205 |
| | $\hat{\beta}_{3,1} = \delta\hat{b}_{g,1}$ | 0.2 | 0.2169 | 0.0506 | 23 | 1.85e-05 |
| $\hat{\omega}^2$ | $\hat{\omega}^2_{a_1}$ | 0.001 | 0.0026 | 0.00218 | 82 | |
| | $\hat{\omega}^2_{a_2}$ | 0.001 | 0.0027 | 0.00228 | 84 | |
| | $\hat{\omega}^2_{b_1}$ | 0.01 | 0.0151 | 0.00575 | 38 | |
| $\hat{\sigma}^2_e$ | | 0.01 | 0.0103 | 0.00045 | 4 | |

TABLE III: Relative bias (%) computed on 1000 datasets with $I = 10$ and $I = 100$ using the individual and the population ARX methods.

| $\hat{\Theta}$ | Bias (%) | | | |
|---|---|---|---|---|
| | $I = 10$ | | $I = 100$ | |
| | Individual | Population | Individual | Population |
| $\hat{\theta}_0$ | -1.104 | -1.452 | -1.164 | -1.476 |
| | -14.106 | -7.323 | -12.483 | -5.551 |
| | 0.155 | 0.097 | 0.213 | 0.143 |
| $\hat{\beta}$ | -0.647 | -1.733 | -1.015 | -1.893 |
| | 2.857 | -0.304 | 0.158 | -2.429 |
| | -2.931 | -2.920 | -0.575 | -0.525 |
| $\hat{\omega}^2$ | 867.638 | 288.912 | 685.482 | 190.298 |
| | 861.193 | 291.376 | 676.921 | 189.704 |
| | 132.949 | 44.193 | 98.986 | 22.380 |
| $\hat{\sigma}^2_e$ | -10.748 | -5.080 | -10.760 | -4.089 |

TABLE IV: Relative RMSE (%) computed on 1000 datasets with $I = 10$ and $I = 100$ using basic ARX method and the EM algorithm with hierarchical model

| $\hat{\Theta}$ | RMSE (%) | | | |
|---|---|---|---|---|
| | $I = 10$ | | $I = 100$ | |
| | Individual | Population | Individual | Population |
| $\hat{\theta}_0$ | 5.663 | 5.611 | 2.156 | 2.312 |
| | 42.406 | 39.811 | 17.796 | 13.568 |
| | 4.586 | 4.563 | 1.384 | 1.367 |
| $\hat{\beta}$ | 27.129 | 26.604 | 8.307 | 8.285 |
| | 54.789 | 53.804 | 15.975 | 15.876 |
| | 80.159 | 80.002 | 9.743 | 9.696 |
| $\hat{\omega}^2$ | 1149.289 | 408.437 | 702.883 | 199.441 |
| | 1167.671 | 423.010 | 693.312 | 199.053 |
| | 200.200 | 106.922 | 102.780 | 30.596 |
| $\hat{\sigma}^2_e$ | 11.190 | 5.773 | 10.808 | 4.190 |

*2) Individual-ARX vs population-ARX:* Figure 5 shows estimates of the two competitive approaches applied to the simulation dataset shown in Figure 4. We clearly observe a larger dispersion of the estimates provided by the individual ARX technique (in red). In contrast, the distributions of the estimated parameters obtained with the population-based method (yellow) are close to that of the simulated ones (blue). Moreover, the population approach yields significantly lower t-test p-values, emphasizing an increased statistical power of the population-based identification method in comparison with the individual one. As a result, when the datasets are small, the population approach appears to be more effective in detecting small effects caused by covariates.

*3) Monte-Carlo simulations:* In order to compare our population estimator with the reference individual estimator, we computed the relative bias and the relative root mean square error (RMSE) of the estimates from estimation on 1000 independent simulations of the model. Two cases were tested, one with $I = 10$, and the other with $I = 100$ in order to assess the impact of the sample size.

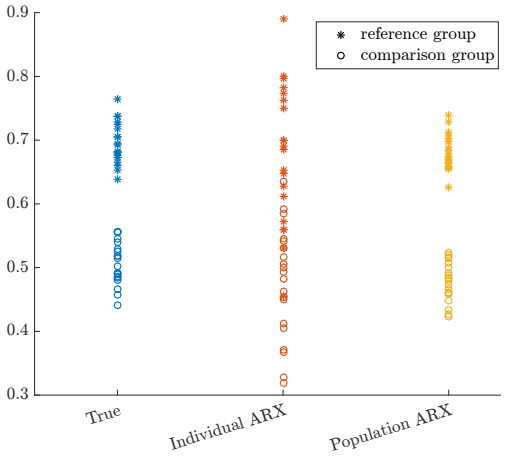
Table III and Table IV show respectively the relative bias and the relative RMSE in percentage over 1000 input-output datasets obtained with $I = 10$ and $I = 100$ subjects. We observe that both methods exhibit a low bias on all the mean parameters but this bias is lower for the population approach. We mainly note that the population-based technique divides by three the bias on $\omega^2$ and by two the bias on $\sigma^2_e$ compared with the individual procedure. The performance improvement is of the same order of magnitude for the RMSE.

Concerning the effect of the sample size ($I$), increasing the number of subjects seems not to affect the estimation bias on $\theta_0$ and $\beta$ but has a significant effect on their RMSE. However the inflation of $I$ allows to reduce the bias and RMSE on the variance parameters.

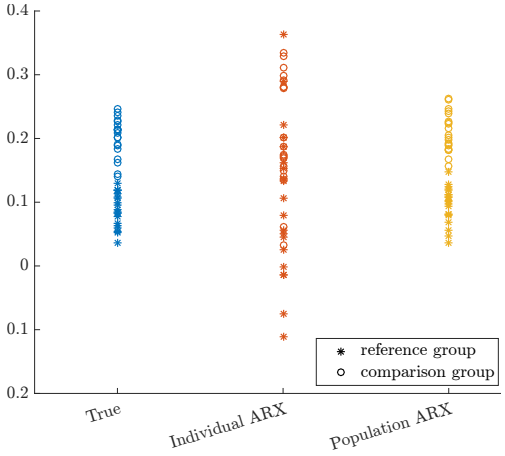
Finally, the simulation variability is used over the 1000 simulation replicates to assess the *true* SE. Figure 6 shows the histograms of the 1000 SE estimates on each parameter compared to the *true* SE. values. For all parameters, the estimates are close to the true values but they are always overestimated. This result can be explained by the fact that $\theta_i$ is not measured but estimated by the expectation of $\theta_i|y_i$ in the EM algorithm. Note that in the case of the individual identification method these SE values are not available, rendering implementation of the Wald test impossible.

## VI. Application to the analysis of Gap-FRAP data

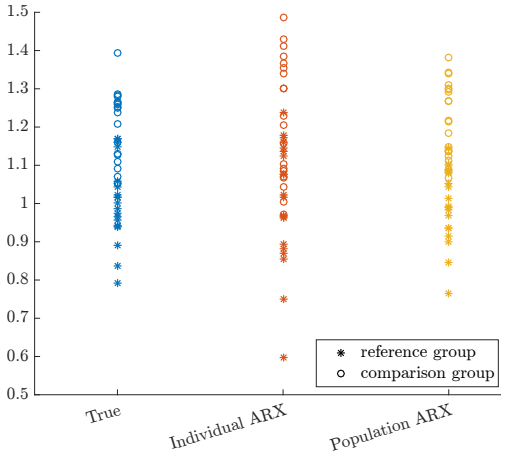
Developed in the 70s, the Fluorescence Recovery After Photobleaching (FRAP) technique is based on the progressive increase of fluorescence intensity in a photobleaching area obtained after illumination with a LASER beam. This widely used method is principally dedicated to study fluorescent constituents (as fluorescent chimeric protein) mobility in cellular membranes and cytoplasm at microscopic scale [17]. Gap-FRAP technique enlarges the FRAP technique to study the




(a) $a_{i,1}$, t-test (*ind*): $1.27 \cdot 10^{-7}$, t-test (*pop*): $2.13 \cdot 10^{-23}$




(b) $a_{i,2}$, t-test (*ind*): $2.30 \cdot 10^{-3}$, t-test (*pop*): $1.24 \cdot 10^{-14}$




(c) $b_{i,1}$, t-test (*ind*): $1.40 \cdot 10^{-4}$, t-test (*pop*): $1.77 \cdot 10^{-7}$

Fig. 5: Dispersions of the estimates for two identification approaches (individual ARX in red *vs* population ARX in yellow), compared to *true* parameters (blue). Two groups of parameters: $(*)$ reference group; $(\circ)$ comparison group.

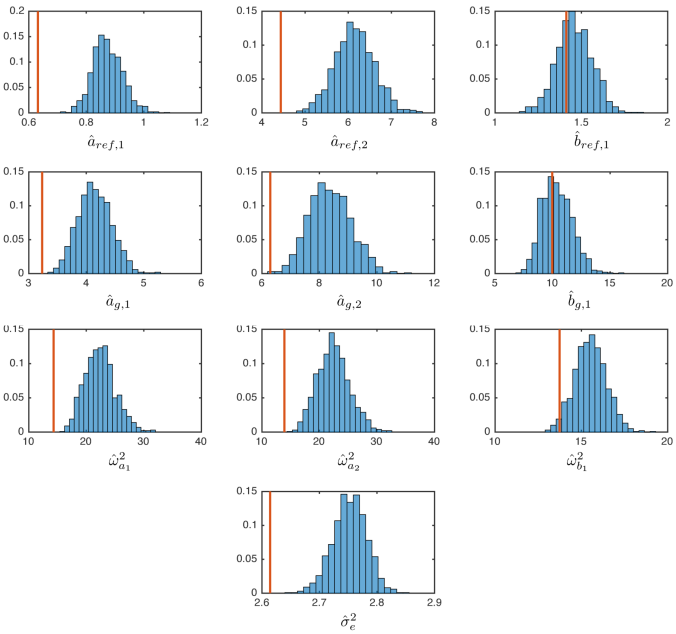



Fig. 6: Histograms of the SE estimates (%) on all parameters over the 1000 simulation with $I = 100$. The red line corresponds to the *true* SE.

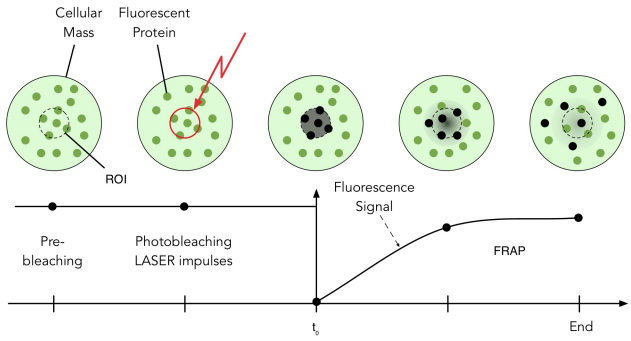



Fig. 7: Illustration of the fluorescent molecules transfers through GJIC. In grey the photobleached area, in light green the intermediate area and in dark green the intact area.

functionality of Gap Junctional Intercellular Communication (GJIC) channels *in vitro* [28] according to an experimental protocol illustrated in Figure 7. GLIC is the mainly form for direct contact between cytoplasms of adjacent cells. Each gap junction is made of twelve protein sub-units called connexins (Cx). The Cx family is composed of 21 types of proteins classified according to their molecular weight [22]. It is assumed that GJIC is of great importance in regulation of development, differentiation and growth. Actually, several physiological roles characterize them, including electrical coupling, answer of tissues to hormones, regulation of embryonic development, the homeostasis and the balance of the cellular proliferation. GJIC is also partially implicated in resistance in drugs and ionizing radiation by contact effect between tumor cells in multicellular 3-dimensional culture [20]. As a consequence the characterization of GJIC functionality is a challenging application for the determination of new biomarkers in medicine.

In [24], authors demonstrated the efficiency of system identification methods to describe GJIC dynamics from experimental data but the proposed procedure required to be applied to each cell culture individually. This study case is thus a good example to test our population-based system identification method. We used an ARX model form with the following structural indices: $n_a = n_b = n_d = 1$:

$$(1 - a_i q^{-1})\ \mathbf{Y}_i(t)|\theta_i = b_i\, u(t-1) + \mathbf{E}_i(t). \qquad (40)$$

The input variable $u(t)$ is a step signal corresponding to the LASER switch-off period. $b_i$ is a gain between the output and input variables while $a_i$ is the pole of the first-order transfer function. The vector of unknown ARX parameters is: $\theta_i = [a_i, b_i]^{\mathrm{t}}$. We wish to test the effects of two independent covariables ($n_g = 2$) on the two previous model parameters:

- $c_{i,1}$ is the cell type. Two head and neck cancer cell lines: KB cells and FaDu cells were used. KB cells (positive line) are known to express Cx43 proteins responsible of GJIC unlike FaDu cells (negative line). This knowledge is used to test the ability of our data-driven modeling approach to detect this property;
- $c_{i,2}$ is the type of *in vitro* culture (monolayer *vs* spheroid) with the objective to test whether or not this factor influences the *in vitro* response.

Each of the two cofactors takes two modalities. As a consequence, the total number of identifiable parameters in the top layer model is equal to $n_g = 2$. The structure of the mixed-effect model is defined as follows:

$$\theta_{i,j} = \theta_{0,j} + \beta_{j,1} c_{i,1} + \beta_{j,2} c_{i,2} + W_{i,j}, \qquad (41)$$

for the $j$-th parameter of $\theta_i$, with $j \in \{1, 2\}$. To estimate the hyperparameters of the top layer model, a $2^2$ full factorial design of experiments was carried out and each of the four experimental conditions was repeated six times, *i.e.* $I = 24$ individual responses to be analyzed. Among the four main groups of experiments, one is used as reference and corresponds to the FaDu cell line in monolayer (2D) culture. Each Gap-FRAP assay takes 900s and is recorded with a sampling period of 15s, *i.e.* $T = 60$ time samples for each response. For more details about the experimental set up, the reader is invited to refer to [24]. The distribution of individual parameters are supposed to be Gaussian with a mean value given by: $C_i\Upsilon$. The upper part of $\Upsilon$ is composed of parameters associated with the reference group: $a_{ref}$ and $b_{ref}$. The other elements of $\Upsilon$ are the effects (deviations from the reference group) induced by the two covariables $(c_{i,1}, c_{i,2})$ on the two model parameters $(b_i, a_i)$:

$$\Upsilon = \begin{bmatrix} \theta_{0,1} \\ \theta_{0,2} \\ \beta_{1,1} \\ \beta_{2,1} \\ \beta_{1,2} \\ \beta_{2,2} \end{bmatrix} = \begin{bmatrix} a_{ref} \\ b_{ref} \\ \delta a_{\mathrm{KB}} \\ \delta b_{\mathrm{KB}} \\ \delta a_{\mathrm{SPHERO}} \\ \delta b_{\mathrm{SPHERO}} \end{bmatrix}. \qquad (42)$$

The matrix $C_i$ is organized in three blocks of two columns:

$$C_i = \left[\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right]. \qquad (43)$$

TABLE V: Results of the population-based system identification method applied to Gap-FRAP data

| Parameters | | Estimates | SE | RSE (%) | p-value |
|---|---|---|---|---|---|
| $\Upsilon$ | $\theta_{0,2}$ | 0.9299 | 0.0075 | 1 | |
| | $\theta_{0,1}$ | 0.0083 | 0.0007 | 9 | |
| | $\beta_{1,1}$ | 0.0615 | 0.0103 | 17 | 2.05e-09 |
| | $\beta_{2,1}$ | -0.0079 | 0.0010 | 12 | 2.25e-16 |
| | $\beta_{1,2}$ | 0.0014 | 0.0095 | 682 | 0.883 |
| | $\beta_{2,2}$ | -0.0007 | 0.0009 | 123 | 0.417 |
| $\omega^2$ | | 0.00039 | 0.000159 | 41 | |
| | | 3.706e-06 | 1.212e-06 | 33 | |
| $\sigma_e^2$ | | 6.12e-07 | 2.64e-08 | 4 | |

The first block corresponds to the reference group (FaDu-monolayer). The other two blocks indicate the presence of effects (coded by 1) between the two culture covariables ($c_{i,1}$ = KB, $c_{i,2}$ = SPHERO) on the model parameters. As a reminder, the number of rows in $C_i$ coincides to the number of ARX model parameters that could be affected by the studied covariables. The estimation results are presented in Table V. For the reference group, we observe a low gain and a pole close to one. Values of $\beta_{1,1} = \delta a_{\mathrm{KB}}$ and $\beta_{2,1} = \delta b_{\mathrm{KB}}$ are clearly different from zero but this is not the case for $\beta_{1,2} = \delta a_{\mathrm{SPHERO}}$ and $\beta_{2,2} = \delta b_{\mathrm{SPHERO}}$. Those observations are confirmed by the respective p-values. It confirms the two cell lines have two distinct dynamics and influence both parameters of the ARX model as illustrated in Figures 10 and 9. By contrast, the type of culture (2D *vs* 3D) does not significantly affect the *in vitro* responses.

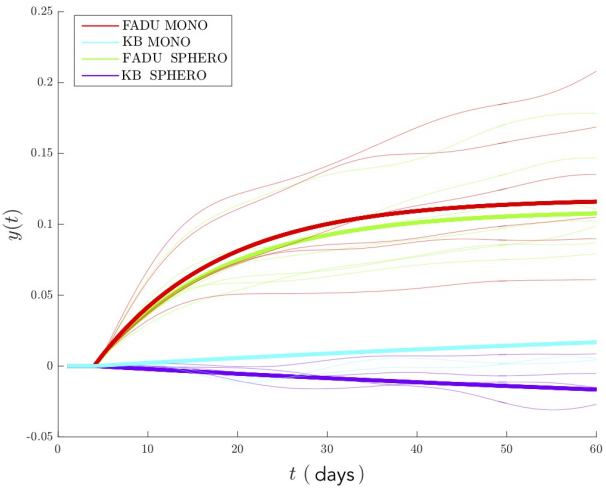



Fig. 8: Individual measured responses (thin lines) and population responses (bold lines) of the ARX model for each group.

## VII. Discussion

The main advantage of the ARX structure is that it is expressed linearly with respect to the model parameters through the matrix $\Phi_i$. This allows to write the model such as :

$$\begin{cases} y_i|\theta_i \sim \mathcal{N}(\Phi_i\theta_i, \sigma_e^2) \\ \theta_i \sim \mathcal{N}(C_i\Upsilon, \Omega), \end{cases} \qquad (44)$$

where $\Phi_i$ is composed of known input-output signals, even if $y_i$ is a stochastic variable, its realizations are known. The ARX

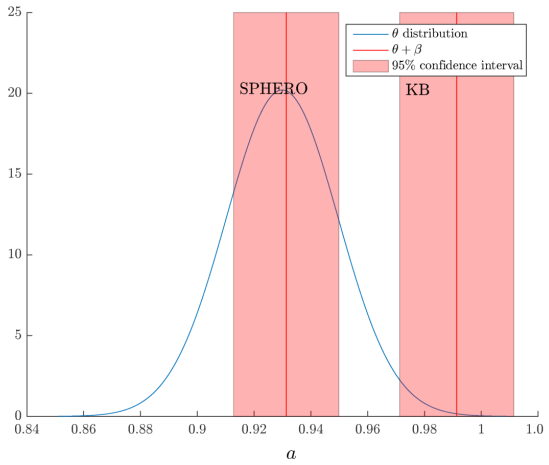



Fig. 9: Distribution of $a$. Blue: estimated distribution for the reference group: FaDu-monolayer. Red: mean values and 95% uncertainty intervals for the spheroid effect ($\beta_{1,2} = \delta a_{\text{SPHERO}}$) and KB effect ($\beta_{1,1} = \delta a_{\text{KB}}$).

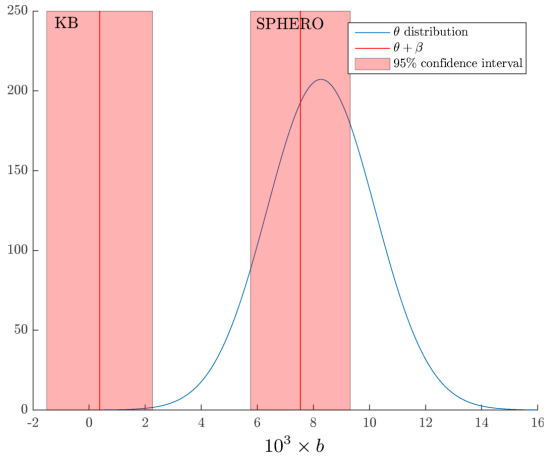



Fig. 10: Distribution of $b$. Blue: estimated distribution for the reference group: FaDu-monolayer. Red: mean values and 95% uncertainty intervals for the spheroid effect ($\beta_{2,2} = \delta b_{\text{SPHERO}}$) and KB effect ($\beta_{2,1} = \delta b_{\text{KB}}$).

structure is the only model with this property. Indeed, only the polynomials $A(q)$ and $B(q)$ are composed of regressors completely known. $\Phi_{A,i}$ is composed of past values of $y_i$ while $\Phi_{B,i}$ contains current and past values of $u_i$. All other structures involved in the Box-Jenkins model sketched in Figure 11 are composed by at least one of polynomials $C(q)$, $D(q)$ or $F(q)$ and their respective matrices $\Phi_{C,i}(\theta_i)$, $\Phi_{D,i}(\theta_i)$ and $\Phi_{D,i}(\theta_i)$. In this case, it becomes impossible to express $\theta_i|y_i$ as a linear function of $\theta_i|y_i$ with respect to $y_i$, which is necessary in the expectation step of the expectation-maximization algorithm. Accordingly, the challenge is to develop statistical tools able to sample stochastic variables distributed according to non-trivial probability functions such as the MCMC algorithm. One promising solution is the adaptation of the EM algorithm with the stochastic approximation of the E-step, called SAEM [15].

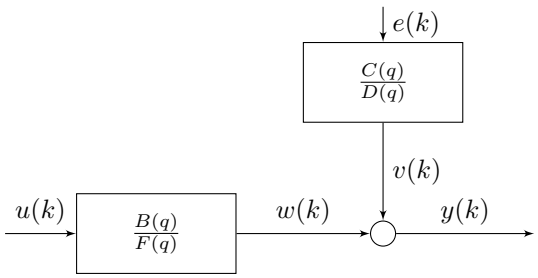



Fig. 11: Block diagram of the Box-Jenkins structure

## VIII. Conclusion

A population-based identification method was developed, implemented and evaluated in this paper. The proposed approach relies on a hierarchical representation composed of an ARX model structure at a low level to describe the input-output dynamics and a normal distribution law at a higher level to characterize the inter-individual dispersion between the systems. An EM (Expectation-Maximisation) algorithm is used to estimate the model parameters and their uncertainty is computed from the Fisher information matrix. For the analysis of input-output data provided by screening studies in biology, categorial covariables have been introduced in the hierarchical representation and two statistical tests were proposed to assess the effects of covariables on the input-output dynamics.

Several simulations were carried out and have emphasized the practical relevance of the population-based identification method. We clearly show that the proposed estimation approach significantly reduces the estimation errors in terms of bias and variance. Moreover, compared with a classical system identification procedure repeatedly applied to each individual system, the new approach drastically increases the statistical power of tests and is therefore able to detect smaller effects of covariables. This is a significant competitive advantage to analyse data from *in vitro* screening studies.

The proposed identification method was finally applied to *in vitro* data obtained from Gap-FRAP tests. Results not only shows the applicability of the proposed approach but also the easy interpretation of results. As short-term perspectives, we are currently working on the extension of the population approach to other model structures, such as output-error and Box-Jenkins models, that are suited to describe more complex dynamics.

## References

[1] S. Audoly, G. Bellu, L. D'Angio, M.P. Saccomani, and C. Cobelli. Global identifiability of nonlinear models of biological systems. *IEEE Trans. Biomed. Eng.*, 48:55–65, 2001.

[2] T. Bastogne, S. Mézières-Wantz, N. Ramdani, P. Vallois, L. Tirand, D. Bechet, and M. Barberi-Heyob. Parameter estimation of pharmacokinetics models in the presence of random timing errors. *European Journal of Control*, 14(2), 2008.

[3] T. Bastogne, L. Tirand, D. Bechet, M. Barberi-Heyob, and A. Richard. System identification of photosensitiser uptake kinetics in photodynamic therapy. *Biomedical Signal Processing and Control*, 2:217–225, 2007.

[4] L. Batista, T. Bastogne, and E.-H. Djermoune. Mixed-effects ARX model identification of dynamical systems. In *25th Meeting of the Population Approach Group in Europe, PAGE 2016*, page PAGE 25 (2016) Abstr 5807, Lisboa, Portugal, June 2016. Poster presentation.

[5] L. Batista, E.-H. Djermoune, and T. Bastogne. Identification of dynamical systems population described by a mixed effect ARX model structure. In *Proc of the 20th IFAC World Congress*, Toulouse, France, July 2017.

[6] C. Cobelli, G. Toffolo, and D. Foster. *Tracer Kinetics in Biomedical Research: From Data to Model*. Kluwer Academic Publishers, 2000.

[7] J. Delforge, A. Sirota, and M.B. Mazoyer. Identifiability analysis and parameter identificaion of an *in vivo* ligand-receptor model from PET data. *IEEE Trans. Biomed. Eng.*, 37(7):653–661, 2000.

[8] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data *via* the EM algorithm. *Journal of the Royal Statistical Society, Serie B*, 39(1):1–38, 1977.

[9] N.D. Evans, R.J. Errington, M.J. Chapman, P.J. Smith, M.J. Chappell, and K.R. Godfrey. Compartmental modelling of the uptake kinetics of the anti-cancer agent topotecan in human breast cancer cells. *International Journal of Adaptive Control and Signal Processing*, 19:395–417, 2005.

[10] N.D. Evans, R.J. Errington, M. Shelley, G.P. Feeney, M.J. Chapman, K.R. Godfrey, P.J. Smith, and M.J. Chappell. A mathematical model for the *in vitro* kinetics of the anti-cancer agent topotecan. *Mathematical Biosciences*, 189:185–217, 2004.

[11] D. Feng, S.C. Huang, Z.Z. Wang, and D. Ho. An unbiased parametric imaging algorithm for nonuniformly sampled biomedical system parameter estimation. *IEEE Transactions on Medical Imaging*, 15(4):512–518, Aug 1996.

[12] Holger Fröhlich, Rudi Balling, Niko Beerenwinkel, Oliver Kohlbacher, Santosh Kumar, Thomas Lengauer, Marloes H Maathuis, Yves Moreau, Susan A Murphy, Teresa M Przytycka, et al. From hype to reality: data science enabling personalized medicine. *BMC medicine*, 16(1):150, 2018.

[13] T. Funatogawa, I. Funatogawa, and M. Takeuchi. An autoregressive linear mixed effects model for the analysis of longitudinal data which include dropouts and show profiles approaching asymptotes. *Statistics in Medicine*, 27(30):6351–6366, 2008.

[14] R. Gomeni, H. Piet-Lahanier, and E. Walter. Study of the pharmacokinetics of betaxolol using membership set estimation. *Biomed. Meas. Infor. Contr.*, 2(4):207–211, 1988.

[15] M. Lavielle. *Mixed Effects Models for the Population Approach. Models, Tasks, Methods & Tools.* Chapman & Hall/CRC Biostatistics Series, 2014.

[16] M. Lavielle and L. Aarons. What do we mean by identifiability in mixed effects models? *Journal of Pharmacokinetics and Pharmacodynamics*, 43(1):111–122, 2016.

[17] J. Lippincott-Schwartz and G.H. Patterson. Development and use of fluorescent protein markers in living cells. *Science*, 300(5616):87–91, 2003.

[18] L. Ljung. *System Identification, Theory for the User*. PreWiley-Hall, 1987.

[19] T.A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2):pp. 226–233, 1982.

[20] P.L. Olive and R.E. Durand. Drug and radiation resistance in spheroids: cell contact and kinetics. *Cancer Metastasis Rev*, 13:121–38, 1994.

[21] A. Samson, M. Lavielle, and F. Mentré. The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Statistics in Medicine*, 26(27):4860–75, Nov 2007.

[22] G. Söhl and K. Willecke. Gap junctions and the connexin protein family. *Cardiovasc. Res.*, 62:228–232, 2004.

[23] G. Sparacino, C. Tombolato, and C. Cobelli. Maximum-likelihood versus maximum a posteriori parameter estimation of physiological system models: the C-peptide impulse response case study. *IEEE Trans. Biomed. Eng.*, 47(6):801–811, 2000.

[24] J.-B. Tylcz, M. Abbaci, T. Bastogne, W. Blondel, D. Dumas, and M. Barberi-Heyob. System identification of the fluorescence recovery after photobleaching in gap junctional intracellular communications. In *Proc of the 52nd IEEE Conference on Decision and Control*, Florence, Italy, December 2013.

[25] J.-B. Tylcz, T. Bastogne, H. Benachour, D. Bechet, E. Bullinger, H. Garnier, and M. Barberi-Heyob. A model-based pharmacokinetics characterization method of engineered nanoparticles for pilot studies. *IEEE Transactions on NanoBioscience*, 14(4):368–377, April 2015.

[26] J.-B. Tylcz, K. El Alaoui-Lasmaili, E.-H. Djermoune, N. Thomas, B. Faivre, and T. Bastogne. Data-driven modeling and characterization of anti-angiogenic molecule effects on tumoral vascular density. *Biomedical Signal Processing and Control*, 20:52–60, July 2015.

[27] G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. Springer, 2009.

[28] M.H. Wade, J.E. Trosko, and M. Schindler. A fluorescence photobleaching assay of gap junction-mediated communication between human cells. *Science*, 232:525–8, 1986.

[29] J. Xu, S. Zhang, A. Machado, S. Lecommandoux, O. Sandre, and A. Gu, F.and Colin. Controllable microfluidic production of drug-loaded plga nanoparticles using partially water-miscible mixed solvent microdroplets as a precursor. *Scientific Reports*, 7(1):4794, 2017.

[30] F. Zanella, J.B. Lorens, and W. Link. High content screening: seeing is believing. *Trends in Biotechnology*, 28(5):237–245, 2010.