Cellular and Molecular Biology ™ **53**, N°2, 11-21
DOI 10.1170/T783

ISSN 1165-158X
2007 Cell. Mol. Biol.™

PARAMETER ESTIMATION AND ORDER SELECTION FOR AN EMPIRICAL MODEL OF VO_2 ON-KINETICS.

O. ALATA[1] AND O. BERNARD[2]

[1]Laboratoire Signal Image Communication (SIC), EA 4103, BP 30179, 86962 Futuroscope Chasseneuil Cedex, France
Université de Poitiers, UFR Sciences Fondamentales et Appliquées
Fax : 00 (33) 5 49 49 65 70, Email : alata@sic.univ-poitiers.fr
[2]Laboratoire des Adaptations Physiologiques aux Activités Physiques (LAPHAP), EA 3813, 4, allée Jean Monnet, 86000 Poitiers,
France Université de Poitiers, Faculté des Sciences du Sport

Received November 22 rd, 2005; Accepted March 23rd, 2006; Published April 27th, 2007

Abstract – In humans, $\dot{V}O_2$ on-kinetics are noisy numerical signals that reflect the pulmonary oxygen exchange kinetics at the onset of exercise. They are empirically modelled as a sum of an offset and delayed exponentials. The number of delayed exponentials, i.e. the order of the model, is commonly supposed to be one for low-intensity exercises and two for high-intensity exercises. As no ground truth has ever been provided to validate these postulates, physiologists still need statistical methods to verify their hypothesis about the number of exponentials of the $\dot{V}O_2$ on-kinetics especially in the case of high-intensity exercises. Our objectives are first to develop accurate methods for estimating the parameters of the model at a fixed order, and then, to propose statistical tests for selecting the appropriate order. In this paper, we provide, on simulated Data, performances of Simulated Annealing for estimating model parameters and performances of Information Criteria for selecting the order. These simulated Data are generated with both single-exponential and double-exponential models, and noised by white and Gaussian noise. The performances are given at various Signal to Noise Ratio (*SNR*). Considering parameter estimation, results show that the confidences of estimated parameters are improved by increasing the SNR of the response to be fitted. Considering model selection, results show that Information Criteria are adapted statistical criteria to select the number of exponentials.

Key words: $\dot{V}O_2$ on-kinetics, parametric modelling, parameter estimation, model selection, optimisation, stochastic methods, general linear model, simulated annealing, information criteria.

INTRODUCTION

In the context of biomedical signal processing based on Bayesian methods, this paper presents theoretical results about modelling the oxygen uptake ($\dot{V}O_2$) on-kinetics with empirical models. These results concern both the estimation of parameters for a given empirical model and the selection of its actual form in a finite set of hypothesised models.

In humans, $\dot{V}O_2$ on-kinetics are signals that reflect the pulmonary oxygen exchange kinetics at the onset of exercise. Their study may provide insight in the metabolic behaviour of muscular cells and may be used for the evaluation of the physical fitness of a subject [19]. In order to provide help for the interpretation of these signals, $\dot{V}O_2$ on-kinetics are currently characterised with empirical models that are a weighted sum of an offset and delayed exponentials [3, 10]:

$$\dot{V}O_2(t) = A_0 + \sum_{m=1}^{o} A_m \left(1 - e^{-\left(\frac{t-td_m}{\tau_m}\right)}\right) U(t-td_m) \qquad \text{Eq. 1}$$

with A_0, the offset, o, the order of the model, A_m, td_m, τ_m, respectively the weight, the time delay, and the time constant of the exponential $m = 1, \ldots, o$ and U, the unit step function $U(t) = \begin{vmatrix} 0 \text{ if } t < 0 \\ 1 \text{ if } t \geq 0 \end{vmatrix}$. The parameter set of the model of order o is:

$$\theta_o = \left\{A_0, \{A_m, td_m, \tau_m\}_{m=1,\cdots,o}\right\} \qquad \text{Eq. 2}$$

In practice, the obtained sequences of numerical data are noised by breathing irregularities and these noisy signals have a Signal to Noise Ratio (*SNR*) relatively low. Parameter estimation, i.e. the estimation of θ_o, is currently based on the optimisation of statistical criteria computed from these noisy numerical Data called "observations". In the general context, the parameter estimation of this kind of function presents several difficulties:

1. The parametric function (see Eq. 1) is not continuously differentiable.
2. Fitting a sum of exponentials to numerical Data is known to be an ill-conditioned problem [14] i.e. slight fluctuations in the observed data can result in very large fluctuations in the estimated parameters.

3. Constraints are imposed on the value of parameters due to physiological considerations (for example, weighted terms must be positive).

Considering $\dot{V}O_2$ on-kinetics, the use of non-linear regression methods based on Gradient-Descent (*GD*) principle can be found in some papers [4, 5 & 10]. Nevertheless, *GD* based methods are not perfectly adapted to the mathematical properties of these above-mentioned exponential functions, and may impair the estimation of the kinetic parameters. To estimate θ_o alternative approaches may be proposed. In the following, we will present the stochastic optimisation method we have implemented in order to optimise the Likelihood function of the observations: the Simulated Annealing (*SA*) [9, 14]. In a recent paper [6], a comparison of the performances of a *GD* method and *SA* was presented on simulated Data with ground truth and real Data. The model used was double exponential, i.e. with $o = 2$ (see Eq. 1). The main conclusions of this study were:

1. Compared to *GD*, *SA* improves the estimation of the parameters in the case of simulated Data.
2. In the case of real Data, *SA* provides lower Residual Sum of Squares (*RSS*) of the difference between the raw data and the model than *GD*.
3. The parameters of the second exponential are estimated with low accuracy.

In this last paper [6], only one level of *SNR* was studied (it was around 20 dB). In the present paper, we provide performances of *SA* at various *SNR*, in the case of $o = 1$ and $o = 2$, by using a new set of simulated Data. This study is done in order to complete the previous paper and to verify that the estimation accuracy of the parameters of the second exponential is improved at high *SNR* (around 30 dB).

In the literature, step-increases in power from light to moderate work rates (< 60% $\dot{V}O_{2\max}$) are considered to be well described by the exponential model of first order (Eq. 1 with $o = 1$) while second-order model is used to describe transitions from light to higher work rates (> 60% $\dot{V}O_{2\max}$) [3] where $\dot{V}O_{2\max}$ represents the maximum oxygen uptake of a subject. The choice of the order of the model is sometimes argued by statistical approaches based on an F test [3, 11]. Nevertheless, no ground truth has ever been provided to check the pertinence of these approaches. We recall that, when using empirical models, the simplest model that describes the data satisfactorily is sought. The second purpose of the present paper is therefore to re-examine the question of model order selection by testing statistical criteria with appropriate properties: the information criteria (*IC*).

METHODS

Estimation problem and numerical Data

The breath-by-breath values of $\dot{V}O_2$ are sampled with a variable sampling period as the respiratory frequency changes due to breath-to-breath irregularities, and also, due to the need for increasing the ventilatory flow rate. $\dot{V}O_2$ values can be viewed as a vector $\mathbf{vo_2} = [vo_{2,1}, \cdots, vo_{2,N}]^T$, sampled over a time vector $\mathbf{t} = [t_0, \cdots, t_N]^T$ (see Fig. 1). Each $t_k - t_{k-1}, k = 1, \ldots, N$, corresponds to the laps of time of a breath and the first value of $\mathbf{vo_2}$ is obtained at t_1. $\mathbf{vo_2}$ is supposed to incorporate an underlying physiological response (i.e. the oxygen exchanges occurring at the alveolar site) plus an additive noise $\mathbf{e}_o = [e_{o,1}, \cdots, e_{o,N}]^T$.

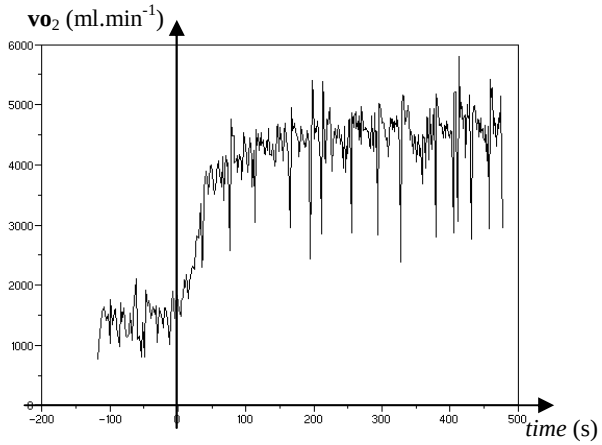

Figure 1. Example of $\dot{V}O_2$ on-kinetics of high intensity exercises.

In the present context, Eq. 1 is supposed to describe this underlying $\dot{V}O_2$ response. Then, it is possible to write $\mathbf{vo_2}$ following the General Linear Model (*GLM*) of order $o, o \geq 1$ [14]:

$$\mathbf{vo_2} = \sum_{m=0}^{o} A_m \mathbf{g}_m + \mathbf{e}_o = \mathbf{G}_o \mathbf{a}_o + \mathbf{e}_o \qquad \text{Eq. 3}$$

with $\mathbf{G}_o = [\mathbf{g}_0, \cdots, \mathbf{g}_o]$ and $\mathbf{a}_o = [A_0, \cdots, A_o]^T$. The vectors $\mathbf{g}_m, m = 0, 1, \ldots, o$, have a length of N and can be written $\mathbf{g}_0 = [1, \cdots, 1]^T$ and $\mathbf{g}_m = [g_{m,1}, \cdots, g_{m,N}]^T$, $m = 1, \ldots, o$, and

$$g_{m,k} = \begin{cases} 0 & \text{if } t_k < td_m \\ 1 - \exp\left(-\dfrac{t_k - td_m}{\tau_m}\right) & \text{if } t_k \geq td_m \end{cases}, k = 1, \ldots, N.$$

The estimation of θ_o (see Eq. 2) at fixed order o from $\mathbf{vo_2}$ is done classically by procedures that minimise the *RSS* between the experimental data points and the model [3, 4, 5, 6 & 11] i.e. the procedures are searching for the set

of parameters that minimises the Euclidean norm of $\mathbf{e}_o$ (see Eq. 3). This minimisation must be done with the following physiological constraints on parameters:

1. $\mathbf{a}_o > 0$ [1],
2. $td_1 \le \cdots \le td_o$,
3. $0 < \tau_1 \le \cdots \le \tau_o$.

In the following, the noise $\mathbf{e}_o = [e_{o,1}, \cdots, e_{o,N}]^T$ is supposed to be a realisation of a Gaussian white noise with variance ${\sigma_{e,o}}^2$ for the model of order *o*.

We provide now with the mathematical background for understanding our methods for estimating the model parameters and the order of the model.

Parameter estimation based on Simulated Annealing (SA)

Bayesian methods are based on the fundamental theorem proposed by Bayes. In the context of parametric estimation, Bayes' theorem can be written [14]:

$$P(\theta_o|\mathbf{vo}_2, M_o) = \frac{P(\mathbf{vo}_2|\theta_o, M_o)P(\theta_o|M_o)}{P(\mathbf{vo}_2|M_o)} \quad \text{Eq. 4}$$

M_o represents the Data model of order *o*. $P(\theta_o|\mathbf{vo}_2, M_o)$ and $P(\theta_o|M_o)$ are the posterior probability and the prior probability of θ_o, respectively, conditionally to the observation vector, $\mathbf{vo}_2$, and the Data model of order *o*. $P(\mathbf{vo}_2|\theta_o, M_o)$ is the well-known likelihood function $L(\theta_o; \mathbf{vo}_2)$ of the parameter sets at a fixed order *o*. $P(\mathbf{vo}_2|M_o)$ is called the Bayesian evidence of the observation vector.

Maximising the likelihood function at fixed order is a well-known statistical criterion for estimating a parameter set called Maximum Likelihood criterion (*ML*):

$$\hat{\theta}_o = \arg\max_{\theta_o}(L(\theta_o; \mathbf{vo}_2)) \quad \text{Eq. 5}$$

This criterion is known to be consistent, i.e. it is asymptotically unbiased and the estimation variance tends to zero asymptotically (maximal precision). At fixed order, $P(\mathbf{vo}_2|M_o)$ is a constant value and if $P(\theta_o|M_o)$ is supposed to be approximately uniform, we obtained:

$$L(\theta_o; \mathbf{vo}_2) \approx P(\theta_o|\mathbf{vo}_2, M_o) \quad \text{Eq. 6}$$

This result is of importance as the posterior probability is often easier to express literally than the other terms of Eq. 4. We present now the basis of *SA*, the algorithm we use for optimising *ML* criterion.

SA belongs to the set of stochastic global methods of optimisation. *SA* is stochastic as it is based on the simulation of a Markov Chain (*MC*). Indeed a new state of a *MC* is sampled each iteration using a probabilistic transition kernel that models the transition between two states. When sufficient iterations have been done (and so sufficient states of the Chain have been sampled) the probability law of the states converges if the transition kernel has some properties [9, 12 & 17].

In our case, a state is defined by a given θ_o set: from a random initial set $\theta_{o,0}$, a new set of parameters $\theta_{o,i}$ is sampled each iteration *i* of the algorithm. To estimate θ_o with *SA*, the convergent probability law of the states above mentioned must be chosen uniform on the sets of parameters which provide global maxima of Eq. 5 and zero elsewhere. To this end, *SA* doesn't optimise directly the Likelihood function but an energy function of the states, i.e. the model parameter values, which is proportional to the Likelihood function. This energy function can be defined as (see Eq. 6):

$$En(\theta_o) = -\log(p(\theta_o|\mathbf{vo}_2, M_o)) \quad \text{Eq. 7}$$

and has a Gibbs distribution [14]:

$$p(En(\theta_o)) = \frac{1}{Z}\exp\left(-\frac{En(\theta_o)}{BT}\right) \quad \text{Eq. 8}$$

where *Z* is a constant of normalisation, *B*, the Boltsmann's constant and *T*, the "temperature". *SA* consists in sampling θ_o according to Eq. 8, with decreasing values of *T*.

In the case of $\dot{V}O_2$ kinetics, $p(\theta_o|\mathbf{vo}_2, M_o)$ can be literally expressed when $\mathbf{e}_o$ (see Eq. 3) is supposed to be a realisation of a white and Gaussian (and then independently and identically distributed (i.i.d.)) noise [14]:

$$p(\theta_o|\mathbf{vo}_2, M_o) \propto \frac{\left[\mathbf{vo}_2{}^T\mathbf{vo}_2 - \mathbf{vo}_2{}^T\mathbf{G}_o\left(\mathbf{G}_o{}^T\mathbf{G}_o\right)^{-1}\mathbf{G}_o{}^T\mathbf{vo}_2\right]^{-\frac{(N-(o+1))}{2}}}{\sqrt{\det\left(\mathbf{G}_o{}^T\mathbf{G}_o\right)}} \quad \text{Eq. 9}$$

According to Eq. 6, 7 & 8, the states that maximise $L(\theta_o; \mathbf{vo}_2)$ are the states of lowest energies. Then, minimising $En(\theta_o)$ provides *ML* estimation of θ_o. For the special case $T = 0$, the only possible states are the ones that produce the global minima of the energy (see Eq. 8). Therefore, the Gibbs distribution at $T = 0$ is the desired convergent probability law above mentioned.

To iterate *SA*, two main approaches exist in order to sample a Markov Chain with transition kernel which has appropriate properties: the Metropolis-Hastings algorithm and the Gibbs sampler. To estimate θ_o, we use Metropolis-Hastings algorithm (for implementation details see Appendix).

Now, let see how selecting an order.

[1] $\mathbf{a}_o > 0$ means that all the elements of vector $\mathbf{a}_o$ are strictly positive.

Order selection based on Information Criteria (IC)

In the present case, the general form of *IC* is [2]:

$$IC(o) = -2\log L\left(\hat{\theta}_o; \mathbf{vo}_2\right) + C(N)\left(|\theta_o| + 1\right) \quad \text{Eq. 10}$$

where $\hat{\theta}_o$ is the *ML* estimation of θ_o at order *o*, $|\theta_o|$, the cardinal of θ_o, and *C*(*N*) is a term which depends on the size of the observations. The value of *IC* at a given order corresponds to the opposite of the optimal value of logLikelihood function with an added penalisation $C(N)\left(|\theta_o| + 1\right)$. This penalisation in the minimisation procedure against *o* should prevent over parameterisation as it becomes higher when the order of the model increases. In our context, *IC* seem appropriate as they provide a model which represents a compromise between a model with low complexity and the fully description of the Data. This is the basis for fitting an empirical model to Data as it has been previously recalled in introduction.

From the definition given in Eq. 10, the order selection between *o* = 1 and *o* = 2 using *IC* (see Eq. 10) is done by the following procedure:

1. Compute $\hat{\theta}_1$ with *SA*.
2. Compute $\hat{\theta}_2$ with *SA*.
3. Compute $IC(1)$ and $IC(2)$.
4. $\hat{o} = \arg\min_{o=1,2} IC(o)$.

We now provide the literal form of *IC* in the case of *GLM* and white and Gaussian assumption for noise. The likelihood function can be expressed as follows:

$$L\left(\hat{\theta}_o; \mathbf{vo}_2\right) = P(\hat{\mathbf{e}}_o) = \prod_{k=1}^{N} p\left(\hat{e}_{o,k}\right) \quad \text{Eq. 11}$$

Using the log-likelihood function, the previous equation becomes:

$$LL\left(\hat{\theta}_o; \mathbf{vo}_2\right) = \log\left(P(\hat{\mathbf{e}}_o)\right) \approx -\left(\log\left(\hat{\sigma}_{e,o}{}^2\right) + 1 + \log(2\pi)\right)\frac{N}{2} \quad \text{Eq. 12}$$

as $p\left(\hat{e}_{o,k}\right) = \frac{1}{\sqrt{2\pi\hat{\sigma}_{e,o}{}^2}} \exp\left(-\frac{\hat{e}_{o,k}{}^2}{2\hat{\sigma}_{e,o}{}^2}\right), k = 1, \ldots, N.$

Order selection using *IC* can finally be written as follows (see Eq. 10 & 12):

$$\hat{o} = \arg\min_{o=1,2}\left(N\log\hat{\sigma}_{e,o}{}^2 + C(N)\left(|\theta_o| + 1\right)\right) \quad \text{Eq. 13}$$

as we removed all constant values of Eq. 12 in Eq. 13. To estimate *o* in the case of $\dot{V}O_2$ kinetics, we use a particular *IC* called $\phi_\beta IC$ with $C(N) = C_{\phi_\beta}(N) = N^\beta \log(\log(N))$, $0 < \beta < 1$ [8]. In the part *Results and Discussion*, we will discuss the choice of an appropriate value of β using simulated Data.

MATERIALS

Data

Simulated $\dot{V}O_2$ responses were generated in order to mimic first-order and second-order exponential kinetics over a time period of 8 minutes. For *o* = 1, the set of parameters was comparable to the set used by Lamarra & al. [10]: $A_0 = 1500$, $A_1 = 2000$, $td_1 = 15$, and $\tau_1 = 35$. For *o* = 2, it is recognized that the $\dot{V}O_2$ time course described by the second component of Eq. 1 either delays the attainment of a steady state or makes $\dot{V}O_2$ increase inexorably until the end of the exercise (or until $\dot{V}O_{2\max}$ is reached) [3, 4 & 5]. Therefore, both kinds of time course were simulated. The first set of parameters leads to a steady state over the duration of the exercise: $A_0 = 1500$, $A_1 = 2000$, $td_1 = 15$, $\tau_1 = 25$, $A_2 = 600$, $td_2 = 180$, and $\tau_2 = 80$; the second set uses a high τ_2 value in order to simulate a drift in $\dot{V}O_2$ over the entire duration of exercise: $A_0 = 1500$, $A_1 = 2000$, $td_1 = 15$, $\tau_1 = 25$, $A_2 = 600$, $td_2 = 180$, and $\tau_2 = 320$.

In order to test the sensitivity of parameter estimation and order selection methods against noise, the 3 reference kinetics were noised with different levels of white and Gaussian noise. In previous articles [10 & 15], this noise has been frequently quantified by the coefficient of variation (*CV*%) of the breath-by-breath data i.e. the ratio between the standard deviation of the $\dot{V}O_2$ fluctuations and the magnitude of the response:

$$CV\% = 100 \times \frac{\sigma_{e,o}}{A_1} \quad \text{Eq. 14}$$

Four values of *CV*% were used: 20, 15, 10 and 5. These different levels of signal-to-noise ratio (*SNR*) may be expressed in a more conventional unity (i.e. in decibel, dB) by the following formula:

$$SNR_{\text{dB}} = -10\log_{10}\frac{\sum_{k=1}^{N}\left(vo_{2,k} - e_{o,k}\right)^2}{\sum_{k=1}^{N} e_{o,k}{}^2} \quad \text{Eq. 15}$$

The numerator and the denominator of Eq. 15 represent the energy of the signal without noise and the energy of the noise (see Eq. 3), respectively. For each value of *CV*%, 40 stochastic vectors of noise were generated per set of parameters in order to mimic the condition of repetitive exercise testing. So, we had 160 simulated responses with *o* = 1 and 320 simulated responses with *o* = 2 (80 per *CV*%). Table 1 provides the correspondence values between *CV*% and SNR_{dB} obtained on the whole simulated Data.

Table 1. Values of mean SNR_{dB} against *CV*% for simulated Data with *o* = 1 (model with 1 exponential) and *o* = 2 (models with 2 exponentials).

	CV%	SNR_{dB}
o = 1	20	18,05
	15	20,52
	10	24,04
	5	30,06
o = 2	20	18,93
	15	21,41
	10	24,96
	5	30,96

Statistical Analysis

In order to evaluate the performances of the parameter estimation at various *SNR*, each simulated response was modelled with the model order corresponding to the ground truth. Then, for each parameter, the sensitivity of estimation was evaluated by two classical indices:

1. The bias (*b*) as an index of accuracy of the estimation. It is the mean of the estimation errors, i.e. the mean difference between the reference value and the estimates.

2. The standard deviation of the estimation errors ($\hat{\sigma}_{e,o}$) as an index of the precision of the estimation.

In order to evaluate order selection performances at various *SNR*, each signal was fitted with both models (o = 1 and o = 2). Then, the sensitivity of order selection was evaluated by expressing:

1. The percentage of correct order selection for the simulated responses generated with the single exponential model.
2. The percentage of correct order selection for the simulated responses generated with the double exponential model.

These percentages were expressed depending on the level of noise and the level of β.

RESULTS AND DISCUSSION

Performances of SA

Fig. 2 and Fig. 3 provide results of estimation performances for model with o = 1, and for models with o = 2, respectively. These figures show biases and 95% confidence intervals of the data for each parameter. The lower *CV*% (higher is SNR_{dB} – see Tab. 1), the higher the confidence for each parameter. Equivalent performances are obtained for A_0, td_1 and τ_1 whatever the model order and *CV*%. At *CV*% = 5 compared to *CV*% = 20, the confidence intervals are ±6 versus ±25 $mlO_2.min^{-1}$ for A_0 (i.e. about ±0.35% versus ±1.0% of the reference magnitude), and ±1.5 s versus ±7.5 s for td_1 and τ_1 (i.e. about ±7.5% versus ±35% of the respective reference times), respectively. In accordance with our previous paper [6], these results reinforce the need for improving the *SNR* of the raw data before modelling $\dot{V}O_2$ kinetics. This could be done by improving the algorithms used to compute breath-by-breath $\dot{V}O_2$ data [7].

Concerning the performances of the parameter estimation for the single exponential model (o = 1), a direct comparison is possible with the results published by Lamarra & al. [10] as their methodology for generating simulated data is identical to ours. The difference between both studies concerns the estimation algorithm: *SA* in the present one; a non-linear algorithm based on Gradient descent method in the study by Lamarra et al. [10] (and their study only tested the estimation of td_1 and τ_1). Although we originally developed *SA* for parameter estimation of double exponential model [6], this comparison shows, in the case of a single exponential model, that *SA*, compared to a *GD* method, tends to improve the parameter confidences: at *CV*% = 20, the 95% confidence interval of td_1 and τ_1 is about ±10 s with *GD* method while it is reduced to about ±6 s with *SA*. At *CV*% = 5, the 95% confidence interval of td_1 and τ_1 is about ±4 s with *GD* method while it is reduced to about ±1.5 s with *SA*. We hypothesise that this improvement in parameter estimation is due to the fact that *SA* is more adapted to the mathematical properties of the functions to be optimised (i.e. the exponential functions – see introduction).

Concerning the performances of the parameter estimation for the double exponential model (o = 2), a direct comparison of the present results at *CV* = 20% is possible with the results of our previous study [6]. A trend towards a smaller confidence interval for each parameter is observed in the present study. As an example, the confidence intervals are ±450 versus ±1800 $mlO_2.min^{-1}$ for A_2, ±110 versus ±183 s for td_2 and ±400 versus ±1470 s for τ_2 in the present study compared to the previous one, respectively. This may be due to the fact that the present simulated data contain about twice more samples than the signals of the previous study.

As we hypothesised in our previous study, the present results demonstrate that an improvement in *SNR* largely reduces the confidence intervals of the estimated parameters. Unfortunately, the confidence intervals for the second-exponential parameters (A_2, td_2 and τ_2) remain excessively large even at the highest *SNR* (i.e. *CV*%= 5). As an example, the confidence intervals remains at ±57 $mlO_2.min^{-1}$ for A_2 (i.e. ±10% of the reference magnitude), ±19 s for td_2 (i.e. ±10% of the reference magnitude) and ±61 s for τ_2 (i.e. ±47% of the reference magnitude, in average). Therefore, this limits the interpretation of these parameters. We suppose that the different shapes of $\dot{V}O_2$ response influence the confidence intervals of the estimated parameters. More precisely, one of both sets of reference parameters we employed to simulate data leads $\dot{V}O_2$ to increase linearly rather than asymptotically over the exercise duration corresponding to the second-exponential component. Then, the A_2/τ_2 ratio may be seen as a good approximation of the slope of a linear function describing this increase in $\dot{V}O_2$. As infinity of values for both parameters may produce this ratio, this kind of $\dot{V}O_2$ time course tends to increase the confidence intervals of the second-exponential parameters.

Model order selection

Fig. 4 and 5 show the evolution, against both β and *SNR*, of percentages of correct model order selection for models with o = 1 and o = 2, respectively. The first information brought by these figures is that the higher β, the lower the obtained order of the model. Consequently, when using the single-exponential model as the reference response, 100% of correct model order selection is obtained for $\beta > 0.15$; when using the double-exponential model as the reference response, 100% of correct model order selection is obtained for $\beta < 0.3$. In a general context, a range of β values equalled to [0.15, 0.3] allows us to select the correct model order whatever the *SNR* ($5 \leq CV\% \leq 20$).

In the literature, different penalisation terms $C(N)$ (see Eq. 10 & 13) than the formulation we used ($\phi_\beta IC$ using $C_{\phi_\beta}(N) = N^\beta \log(\log(N))$, $0 < \beta < 1$) have been proposed. The main penalisations found in literature are:

1. $C_{\mathrm{AIC}}(N) = 2$, Akaike *IC* (*AIC*) [1]
2. $C_{\mathrm{BIC}}(N) = \log(N)$, Bayesian *IC* (*BIC*) [18] or Minimum Description Length (MDL) [16].

About $\phi_\beta IC$, β can be chosen between $\beta_{\min}$ and $\beta_{\max}$ [13]:

$$\beta_{\min} = \frac{\log\log(N)}{\log(N)} \le \beta \le 1 - \beta_{\min} = \beta_{\max} \qquad \text{Eq. 16}$$

Let's notice that it is also possible to compute *AIC* and *BIC* from $\phi_\beta IC$. Indeed, two values of β allow the computation of the penalisations of *AIC* (β_{AIC}) and *BIC* (β_{BIC}) from $C_{\phi_\beta}(N)$:

$$\beta_{\mathrm{AIC}} = \frac{\log(2) - \log\log\log(N)}{\log(N)} \qquad \text{Eq. 17a}$$

$$\beta_{\mathrm{BIC}} = \frac{\log\log(N) - \log\log\log(N)}{\log(N)} \qquad \text{Eq. 17b}$$

When $N > 15$, $\beta_{\mathrm{AIC}} < \beta_{\mathrm{BIC}} < \beta_{\min} < \beta_{\max}$. Fig. 4 and 5 show the effect of these penalisation values on the model order selection. It appears that:

$$\hat{o}_{\mathrm{AIC}} \ge \hat{o}_{\mathrm{BIC}} \ge \hat{o}_{\phi_\beta}(\beta = \beta_{\min}) \ge \hat{o}_{\phi_\beta}(\beta = \beta_{\max}) \qquad \text{Eq. 18}$$

with $\hat{o}_{\mathrm{IC}}$ the estimated order using a given *IC*. This hierarchy was predictable because the higher the penalisation, the lower the complexity of the selected model. From these different *IC*, only two are appropriate to select the correct model order in the present case: *BIC* and $\phi_\beta IC$ with $\beta = \beta_{\min}$ because both are included in the range [0.15, 0.3].

These results clearly show that the selection of the model order (and so the choice of the adapted penalisation term in order to describe $\dot{V}O_2$ kinetics appropriately) is a complex mathematical problem. Therefore, this questions the validity of another statistical approach used in previous physiological studies for the same goal [3, 11]. This method retains the second-order model if the decrease in *RSS* is sufficient to offset the loss in degrees of freedom associated with the increased number of model parameters as determined by an F test. Table 2 shows the percentages of correct model order selection against both *CV*% and the order of the reference model, using this method. It is well apparent that it reacts like Information Criteria incorporating small β values and penalisation terms: comparable results are obtained with $0 < \beta < 0.1$. Then, this approach tends to overestimate the model order whatever *SNR* and may have led investigators to erroneous results in previous studies.

Table 2. Model order selection results using F test.

	CV%	*o* = 1	*o* = 2	average
P=0,05	20	87,5	100	93,75
	15	82,5	100	91,25
	10	95	100	97,5
	5	92,5	100	96,25
P=0,025	20	92,5	100	96,25
	15	90	100	95
	10	97,5	100	98,75
	5	95	100	97,5
P=0,01	20	97,5	100	98,75
	15	97,5	100	98,75
	10	100	100	100
	5	97,5	100	98,75

CONCLUSION

The first conclusion of the present paper is that, when using *SA*, the confidence of the estimated kinetic parameters, for exponential models describing $\dot{V}O_2$ kinetics (*o* = 1, 2), is proportional to the *SNR* of the fitted $\dot{V}O_2$ response. Nevertheless, concerning the second-order model, the confidence intervals of the second-exponential parameters remain excessively large to allow their interpretation. This may be due to the different tested shapes of $\dot{V}O_2$ responses, and then further studies are needed in order to precise which shapes of $\dot{V}O_2$ response lead to confidence intervals small enough to be interpreted.

The second conclusion of the present paper is that the selection of the model order (*o* = 1, 2) using *IC* needs to use an appropriate penalisation term i.e. a β value $C_{\phi_\beta}(N)$ in the range [0.15, 0.3] in order to obtain 100% of correct model order selection. Comparatively, the approach based on an F test and previously used in physiological studies leads to an overestimation of the model order. This questions the validity of some previous published results.

APPENDIX

Implementation of SA in the case of o = 2

- Initialisation:
  - $i = 0$, $T = T_0$.
  - Sample an initial state $\mathbf{s}_0 = [td_{1,0}, \tau_{1,0}, td_{2,0}, \tau_{2,0}]^T$ that defines a matrix $\mathbf{G}_0$ (see Eq. 3). $\theta_{o,0}$ is then computed from $\mathbf{s}_0$ using a least-squares estimation of $\mathbf{a}_0$, $\mathbf{a}_0 = (\mathbf{G}_0{}^T \mathbf{G}_0)^{-1} \mathbf{G}_0{}^T \mathbf{vo}_2$.

- Resample $\mathbf{s}_0$ until all the elements of $\mathbf{a}_0$ are positive, according to physiological constraints (see Methods-*Estimation problem and numerical Data*).
- $\theta_{\mathrm{vo}_2,0} = \left\{A_{0,0}, \left\{A_{m,0}, td_{m,0}, \tau_{m,0}\right\}_{m=1,2}\right\}$.

- Step 1:
  - Produce a new state using a random perturbation ς : $\mathbf{s}_p = \mathbf{s}_i + \varsigma$.
  - Compute $\mathbf{G}_p$ and $\mathbf{a}_P = \left(\mathbf{G}_P{}^T \mathbf{G}_P\right)^{-1} \mathbf{G}_P{}^T \mathbf{vo_2}$.
  - Resample $\mathbf{s}_p$ from $\mathbf{s}_i$ until all the elements of $\mathbf{a}_p$ are positive.
  - Then, make the state proposal $\theta_{o,p} = \left\{A_{0,p}, \left\{A_{m,p}, td_{m,p}, \tau_{m,p}\right\}_{m=1,2}\right\}$ for $\theta_{o,i+1}$.
- Step 2:

  Compute the acceptance ratio $\alpha = \min\{1, r\}$ with $r = \dfrac{p\left(En\left(\theta_{\mathrm{vo}_2,p}\right)\right) T\left(\theta_{\mathrm{vo}_2,p} \middle| \theta_{\mathrm{vo}_2,i}\right)}{p\left(En\left(\theta_{\mathrm{vo}_2,i}\right)\right) T\left(\theta_{\mathrm{vo}_2,i} \middle| \theta_{\mathrm{vo}_2,p}\right)}$.

  T is the transition kernel:

$$\frac{T\left(\theta_{\mathrm{vo}_2,p} \middle| \theta_{\mathrm{vo}_2,i}\right)}{T\left(\theta_{\mathrm{vo}_2,i} \middle| \theta_{\mathrm{vo}_2,p}\right)} = \prod_{m=1}^{2} \frac{p\left(td_{m,p}\right) p\left(\tau_{m,p}\right)}{p\left(td_{m,i}\right) p\left(\tau_{m,i}\right)} = \prod_{m=1}^{2} \frac{p\left(\tau_{m,p}\right)}{p\left(\tau_{m,i}\right)}$$

  since we assume uniform probability laws for time delays.

  State $\theta_{o,i+1}$ has a probability α to be equal to $\theta_{o,p}$ and 1-α to be equal to $\theta_{o,i}$. So,

  If $r \geq 1$,

  $\alpha = 1$ and $\theta_{o,i+1} = \theta_{o,p}$.

  Else

  sample a realisation x of a random variable $X \sim U_{[0,1]}$ where $X \sim U_{[a,b]}$ signifies that X follows a continuous uniform law on [a, b].

  If $x \leq \alpha$,

  $\theta_{o,i+1} = \theta_{o,p}$,

  Else $\theta_{o,i+1} = \theta_{o,i}$.
- Step 3:

  If $T > T_{\min}$

  Decrease T. $i = i + 1$. Return to step 1.

  Else ($T = T_{\min}$)

  $\hat{\theta}_o = \theta_{o,i}$ and stop the iterations.

  Hence $\hat{\theta}_o$ obtained, $\mathbf{e}_o$ is estimated as follows:

$$\hat{\mathbf{e}}_o = \mathbf{vo_2} - \hat{\mathbf{G}}_o \hat{\mathbf{a}}_o$$

Estimated RSS can be directly computed as the sum of the elements of $\hat{\mathbf{e}}_o$.

The decreasing scheme of T, i.e. the annealing schedule, is of importance. If it is sufficiently low, the convergence to the global minimum is insured. Of course, it is impossible to decrease T until 0 and annealing schedules that guarantee convergence are intractable because they take a huge amount of time. This is the reason we stop the algorithm to a minimum temperature $T_{\min}$ reached in a finite number of iterations. To estimate θ_o with *SA* algorithm, the configuration of B, T_0, $T_{\min}$, the temperature decreasing scheme, the sample laws of initial state and state proposal $\theta_{o,p}$ are described in the following:

- $B = \dfrac{2}{N-3}$ equal to the inverse of the power of Eq. 8 in the case $o = 2$.
- $T_0 = 0{,}1$ and $T_{\min} = 0{,}000001$ in order to have coherent values for $\dfrac{p\left(En\left(\theta_{\mathrm{vo}_2,p}\right)\right)}{p\left(En\left(\theta_{\mathrm{vo}_2,i}\right)\right)} = \exp\left(-\dfrac{En\left(\theta_{\mathrm{vo}_2,p}\right) - En\left(\theta_{\mathrm{vo}_2,i}\right)}{BT}\right)$ and so, for α (see *step 2* of the algorithm).
- T is decreased with a geometric annealing schedule: $T_i = \lambda^i T_0$. The number of iterations $I_{\max}$ has been chosen equal to 1500000 and $\lambda = \exp\left(\dfrac{\log\left(T_{\min}/T_0\right)}{I_{\max}}\right)$.
- sample laws: for the time values, td_1 and td_2, we use uniform laws of the form $td_l \sim U_{[val_{\min}, val_{\max}]}$, $l = 1, 2$, and for scale parameters, τ_1 and τ_2, uniform laws for logarithm [14]:

$$p(\tau_l) = \begin{cases} 0 & \text{if } \tau_l < val_{\min} \\ 1/\left(\tau_l \log\left(val_{\max}/val_{\min}\right)\right) & \text{if } val_{\min} \leq \tau_l \leq val_{\max} \\ 0 & \text{if } \tau_l > val_{\max} \end{cases}$$

, $l = 1, 2$. Then, for the initial state $\mathbf{s}_0$, $val_{\min}$ and $val_{\max}$ values for each parameter are:
  - $td_{1,0}$: $val_{\min}$ = -50 and $val_{\max}$ = 50.
  - $\tau_{1,0}$: $val_{\min} = \tau_{\min}$ and $val_{\max} = \tau_{\max}$.
  - $td_{2,0}$: $val_{\min} = td_{1,0}$ and $val_{\max} = t_{\max}$.
  - $\tau_{2,0}$: $val_{\min} = \tau_{1,0}$ and $val_{\max} = \tau_{\max}$.

  and for $\mathbf{s}_p$:
  - $td_{1,p}$: if $td_{1,i}$ - 2,5 < -50, $val_{\min}$ = -50,

    else $val_{\min} = td_{1,i} - 2{,}5$,

    if $td_{1,i}$ + 2,5 > 50, $val_{\max}$ = 50,

else $val_{\max} = td_{1,i} + 2{,}5$.

- $\tau_{1,p}$: if $\tau_{1,i} - 20 < \tau_{\min}$, $val_{\min} = \tau_{\min}$,
  else $val_{\min} = \tau_{1,i} - 20$,
  if $\tau_{1,i} + 20 > \tau_{\max}$, $val_{\max} = \tau_{\max}$,
  else $val_{\max} = \tau_{1,i} + 20$.
- $td_{2,p}$: if $td_{2,i} - 5 < td_{1,p}$, $val_{\min} = td_{1,p}$,
  else $val_{\min} = td_{2,i} - 5$,
  if $td_{2,i} + 5 > t_{\max}$, $val_{\max} = t_{\max}$,
  else $val_{\max} = td_{2,i} + 5$.
- $\tau_{2,p}$: if $\tau_{2,i} - 20 < \tau_{1,p}$, $val_{\min} = \tau_{1,p}$,
  else $val_{\min} = \tau_{2,i} - 20$,
  if $\tau_{2,p} + 20 > \tau_{\max}$, $val_{\max} = \tau_{\max}$,
  else $val_{\max} = \tau_{2,i} + 20$.

These values have been chosen to obtain coherent values of acceptance ratio, to visit many states at the beginning of iterations and to converge at the end of iterations.

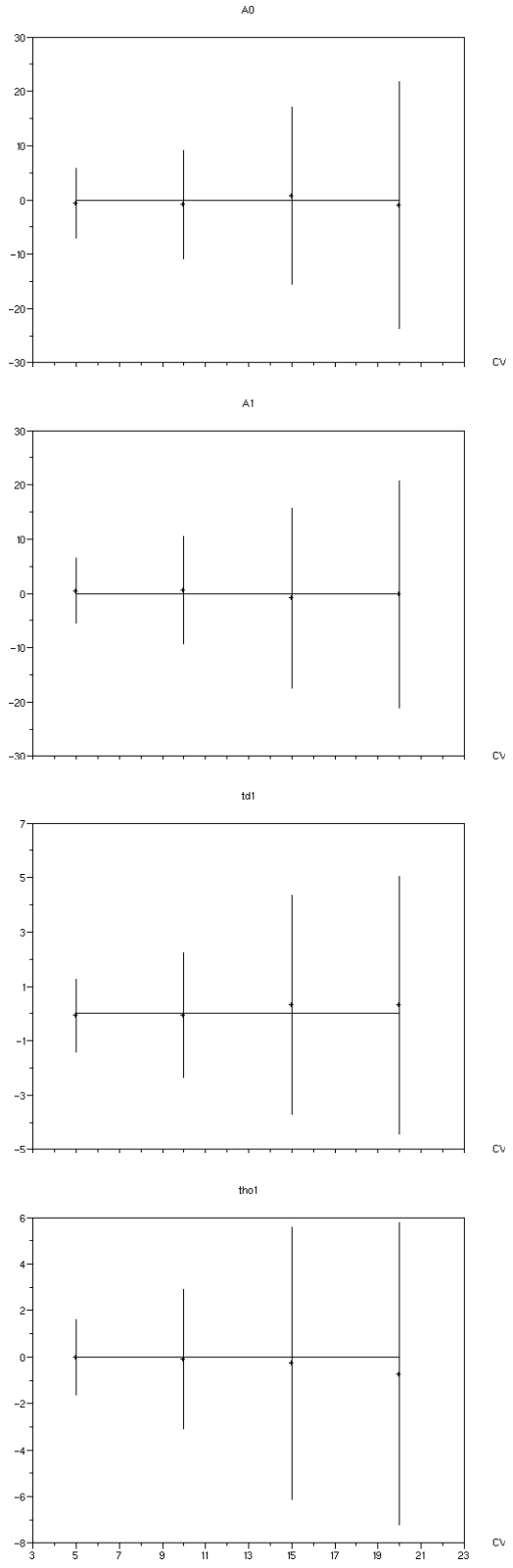

Figure 2. Biases and 95% confidence intervals of the estimated parameters for a single exponential model (o = 1) expressed as a function of *CV*% (or SNR_{dB} see Tab. 1). Top to bottom: offset A_0 ($mlO_2.min^{-1}$), parameters of the exponential: amplitude A_1 ($mlO_2.min^{-1}$), time delay td_1 (sec), time constant τ_1 (sec).

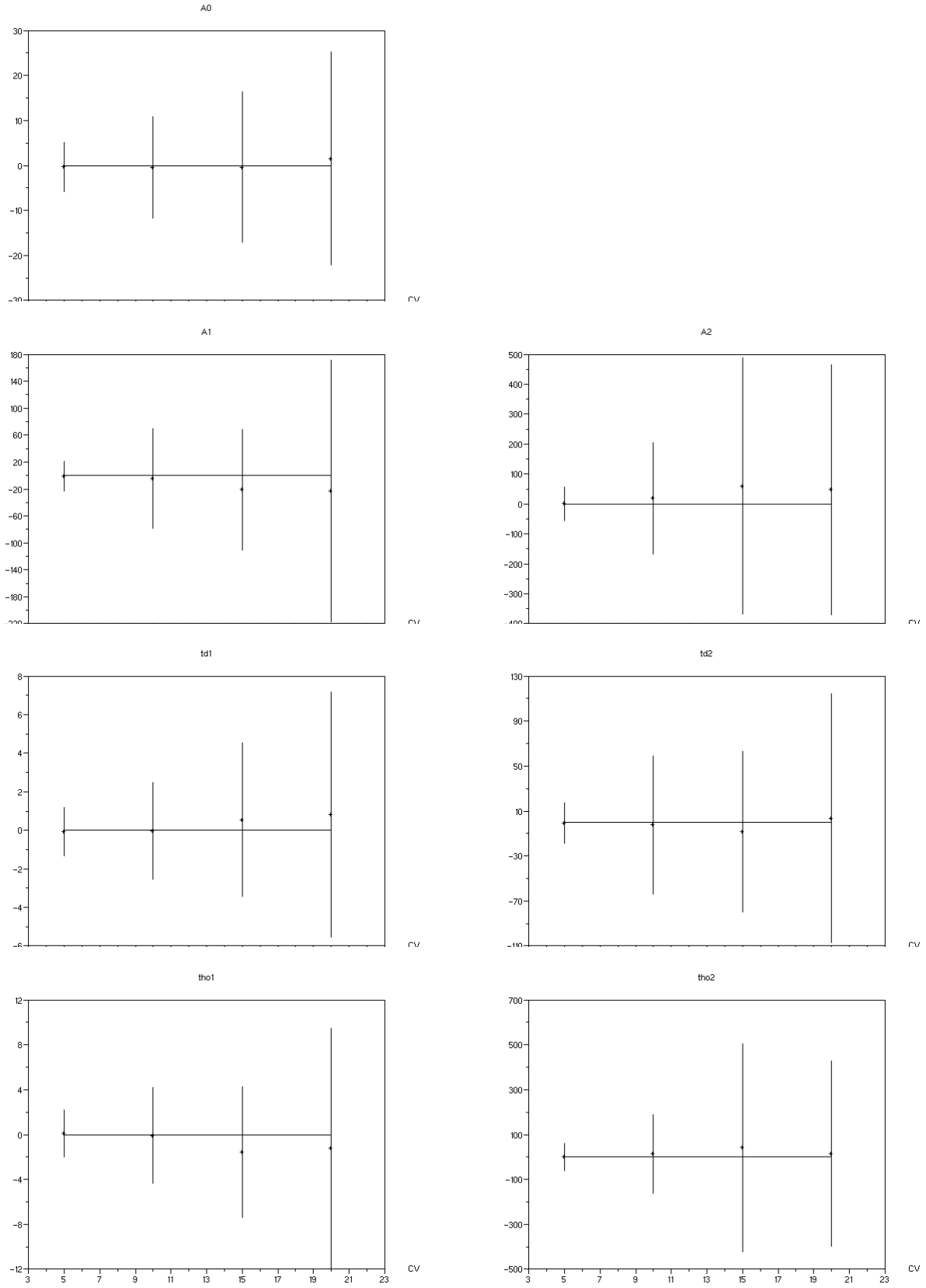

Figure 3. Biases and 95% confidence intervals of the estimated parameters for double exponential models ($o = 2$) expressed as a function of CV% (or SNR_{dB} see Tab. 1). Top to bottom, left to right: offset A_0 ($mlO_2.min^{-1}$), parameters of the 1st exponential: amplitude A_1 ($mlO_2.min^{-1}$), time delay td_1 (sec), time constant τ_1 (sec), parameters of the 2nd exponential: amplitude A_2($mlO_2.min^{-1}$), time delay td_2 (sec), time constant τ_2 (sec).

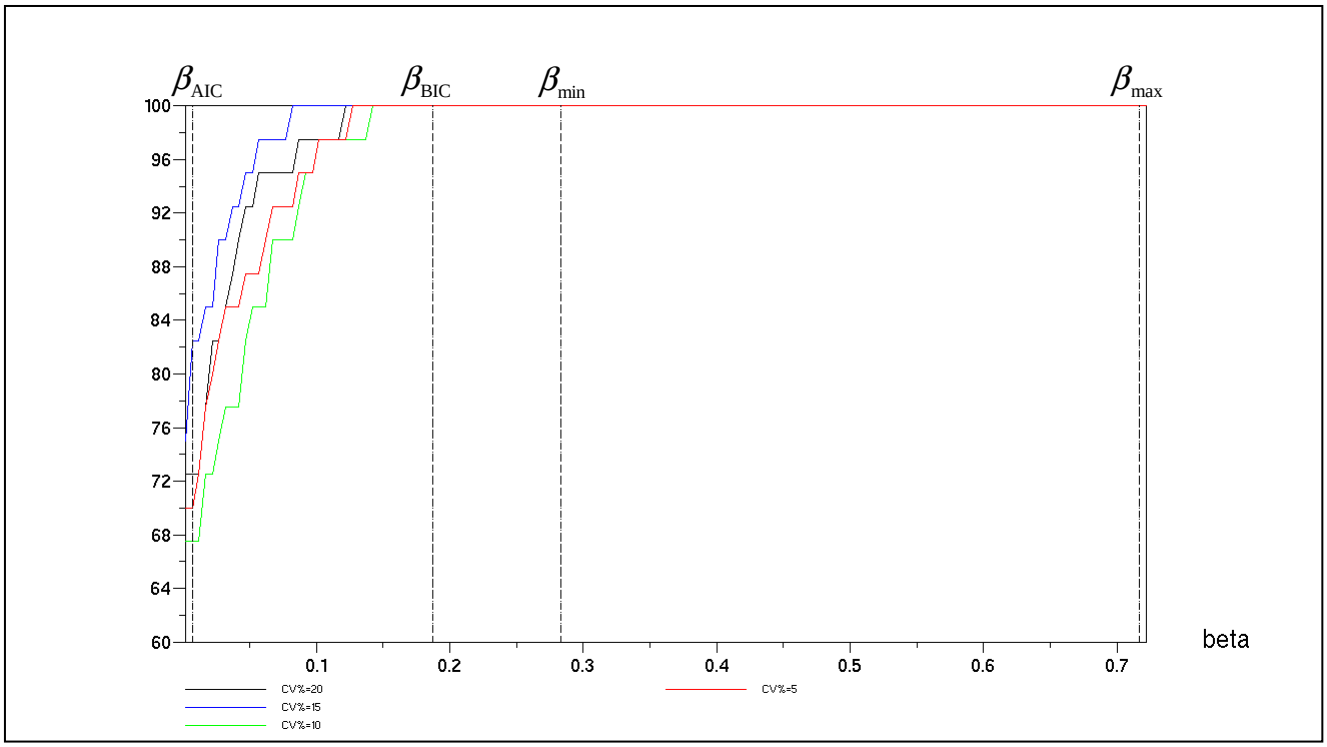

Figure 4. Percentages of correct model order selection for the simulated responses with $o = 1$. The percentages are expressed as a function of $CV\%$ (or SNR_{dB} see Tab. 1) and for β values varying from 0 to about β_{max}.

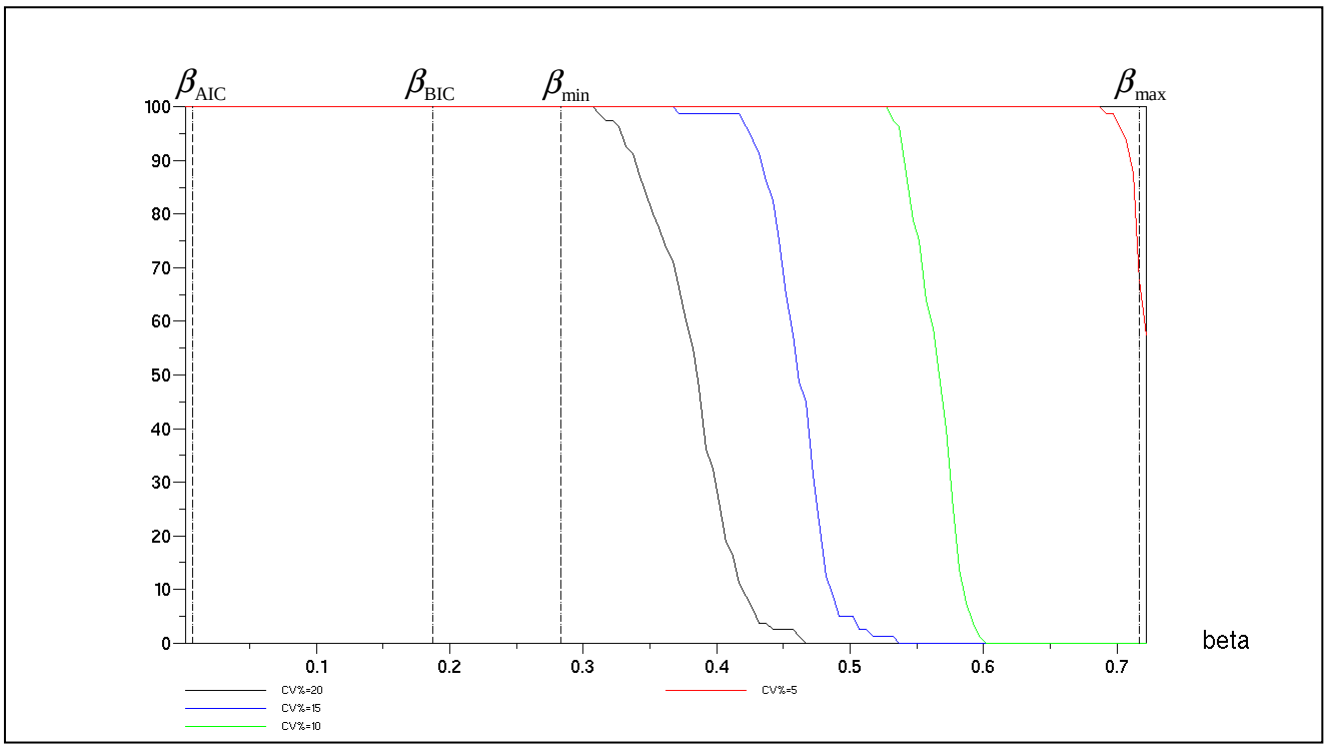

Figure 5. Percentages of correct model order selection for the simulated responses with $o = 2$. The percentages are expressed s a function of $CV\%$ (or SNR_{dB} see Tab. 1) and for β values varying from 0 to about β_{max}.

REFERENCES

1. Akaike, H., A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 1974, **19** (6): 716-723.
2. Alata, O. and Olivier, C., Choice of a 2-D causal autoregressive texture model using information criteria. *Pattern Recognition Letters*, 2003, **24**: 1191-1201.
3. Barstow, T.J. and Molé, P.A., Linear and non-linear characteristics of oxygen uptake kinetics during heavy exercise. *Journal of Applied Physiology*, 1991, **71**: 2099-2106.
4. Bernard, O., Maddio, F., Ouattara , S., Jimenez, C., Charpenet, A., Melin, B. and Bittel, J., Influence of the oxygen uptake slow component on the aerobic energy cost of high-intensity sub-maximal treadmill running in humans. *European Journal of Applied Physiology*, 1998, 78: 578-585.
5. Bernard, O., Ouattara, S., Maddio, F., Jimenez, C., Charpenet, A., Melin, B. and Bittel, J., Determination of the velocity associated with $\dot{V}O_{2max}$. *Medicine and Science in Sports and Exercise*, 2000, 33: 464-470.
6. Bernard, O., Alata, O. and Francaux, M., On the modelling of breath-by-breath oxygen uptake kinetics at the onset of high-intensity exercises: Simulated Annealing versus GRG_2 method. *Journal of Applied Physiology*, Volume 100, pages 1049-1058 - March 2006.
7. Cautero, M., Beltrami, A.P., di Prampero, P.E. and Capelli, C., Breath-by-breath alveolar oxygen transfer at the onset of step exercise in humans: methodological implications. *European Journal of Applied Physiology*, 2002, 88: 203–213.
8. El-Matouat, A. and Hallin, M., Order selection, stochastic complexity and Kullback-Leibler information. In: *Time Series Analysis*, Hannan, E.J. (ed), Springer Verlag, New York, 1996, **2**: 291-299.
9. Fitzgerald, W.J., Markov chain Monte Carlo methods with applications to signal processing. *Signal Processing*, 2001, **81**: 3-18.
10. Lamarra, N., Whipp, B.J., Ward, S.A. and Wasserman, K., Effect of interbreath fluctuations on characterizing exercise gas exchange kinetics. *Journal of Applied Physiology*, 1987, **62**: 2003-2012.
11. Morton, R.H., Two-dimensional short-term model of oxygen uptake kinetics. *Journal of Applied Physiology*, 1985, 58: 1736-1740.
12. Neal, R.M., *Probabilistic Inference using Markov Chain Monte Carlo Methods*, Technical Report CRG-TR-91-1 University of Toronto, Ontario, Canada, 1993.
13. Olivier, C., Jouzel, F. and El-Matouat, A., Choice of the number of component clusters in mixture models by information criteria. In: *Proc. Vision Interface*, 1999, pp. 74-81.
14. O'Ruanaid, J.K. and Fitzgerald, W.J., In: *Numerical Bayesian Methods Applied to Signal Processing*, Springer Verlag, New York, 1996.
15. Puente-Maestu, L., Sánz, M.L., Sánz, P., Nuñez, A., González, F. and Whipp, B.J., Reproducibility of the parameters of the on-transient cardiopulmonary responses during moderate exercise in patients with chronic obstructive pulmonary disease. *European Journal of Applied Physiology*, 2001, 85: 434-441.
16. Rissanen, J., Stochastic complexity and modeling. *Annals Statist.*, 1986, **14** (3): 1080-1100.
17. Robert, C.P., and Casella, G., In: *Monte Carlo Statistical Methods*, Springer Verlag, New-York, 1999.
18. Schwarz, G., Estimating the dimension of a model. *Annals Statist.*, 1978, **6**: 461-464.
19. Tschakovsky, M.E. and Hughson, R.L., Interaction of factors determining oxygen uptake at the onset of exercise. *Journal of Applied Physiology*, 1999, 86: 1101-1113.