On the modeling of breath-by-breath oxygen uptake kinetics at the onset of high-intensity exercises: simulated annealing vs. GRG$_2$ method

Olivier Bernard, Olivier Alata, and Marc Francaux

1Laboratoire des Adaptations Physiologiques aux Activités Physiques, Université de Poitiers, Faculté des Sciences du Sport, Poitiers; 2Laboratoire Signal Image Communications, Université de Poitiers, Unité de Formation et de Recherche Sciences Fondamentales et Appliquées, Futuroscope Chasseneuil, France; and 3Institut d’Éducation Physique et de Réadaptation, Faculté de Médecine, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Submitted 16 June 2005; accepted in final form 19 October 2005

Bernard, Olivier, Olivier Alata, and Marc Francaux. On the modeling of breath-by-breath oxygen uptake kinetics at the onset of high-intensity exercises: simulated annealing vs. GRG$_2$ method. J Appl Physiol 100: 1049–1058, 2006. First published October 27, 2005; doi:10.1152/japplphysiol.00712.2005.—Modeling in the time domain, the non-steady-state O$_2$ uptake on-kinetics of high-intensity exercises with empirical models is commonly performed with gradient-descent-based methods. However, these procedures may impair the confidence of the parameter estimation when the modeling functions are not continuously differentiable and when the estimation corresponds to an ill-posed problem. To cope with these problems, an implementation of simulated annealing (SA) methods was compared with the GRG$_2$ algorithm (a gradient-descent method known for its robustness). Forty simulated V$_{O2}$ on-responses were generated to mimic the real time course for transitions from light- to high-intensity exercises, with a signal-to-noise ratio equal to 20 dB. They were modeled twice with a discontinuous double-exponential function using both estimation methods. GRG$_2$ significantly biased two estimated kinetic parameters of the first exponential (the time delay $d_1$ and the time constant $\tau_1$) and impaired the precision (i.e., standard deviation) of the baseline $A_0$, $d_1$, and $\tau_1$ compared with SA. SA significantly improved the precision of the three parameters of the second exponential (the asymptotic increment $A_2$, the time delay $d_2$, and the time constant $\tau_2$). Nevertheless, $d_2$ was significantly biased by both procedures, and the large confidence intervals of the whole second component parameters limit their interpretation. To compare both algorithms on experimental data, 26 subjects each performed two transitions from 80 W to 80% maximal O$_2$ uptake on a cycle ergometer and O$_2$ uptake was measured breath by breath. More than 88% of the kinetic parameter estimations done with the SA algorithm produced the lowest residual sum of squares between the experimental data points and the model. Repeatability coefficients were better with GRG$_2$ for $A_1$, although better with SA for $A_2$ and $\tau_2$. Our results demonstrate that the implementation of SA improves significantly the estimation of most of these kinetic parameters, but a large inaccuracy remains in estimating the parameter values of the second exponential. O$_2$ uptake kinetics; parametric modeling; parameter estimation

FOR SEVERAL DECADES, the study of O$_2$ uptake (V$_{O2}$) kinetics at the onset of step exercises has attracted the attention of physiologists. With the introduction of automatic systems for computing V$_{O2}$ on a breath-by-breath basis (1, 4), there has been an increasing interest in characterizing the kinetics of V$_{O2}$ in the time domain. As the breath-by-breath analysis allows a high time resolution, the non-steady-state V$_{O2}$ kinetics have been currently modeled with different mathematical approaches (2). Two different types of mathematical models can be derived to characterize V$_{O2}$ kinetics: structural and empirical models. “A structural model consists of a series of mathematical relationships that have identifiable physiological correlates and describe the intermediate components of the physiological response to an input stress. Integration of these intermediate responses leads to a prediction of the output response of interest, which then can be compared with observed responses.” Inversely, in empirical models, “the best description of the output response is chosen by some (usually statistical) criteria from a set of competing mathematical models. Insight into the nature of the underlying intermediate physiological processes is sought from analysis of the model components and parameters” (2). Then, the strategy for fitting empirical models to experimental data differs from fitting structural models in that the simplest model that describes the data satisfactorily is sought. This implies that no physiological significance can a priori be assigned to the parameters of these empirical models, albeit physiological correspondence may be a posteriori hypothesized.

In the context of empirical models applied to V$_{O2}$ kinetics to step increases in power, the estimation of the kinetic parameters has been commonly performed by nonlinear regression algorithms. These methods typically refer to gradient-descent-based (GD) algorithms (6, 7, 11, 17, 35), validated on problems with continuously differentiable functions and in the absence of numerical difficulties such as degeneracy or ill conditioning (i.e., slight fluctuations in the observed data can result in very large fluctuations in the estimated parameters) (14, 20, 21). For exercise transitions from light to moderate work rates [$<60\%$ maximal V$_{O2}$ ($V_{O2\max}$)], the V$_{O2}$ time course is assumed to be well described by a first-order exponential function (2, 19, 26, 29, 35). Then, GD methods seem to be appropriate, because the model is well conditioned and continuously differentiable over the period of work. For exercise transitions from light to higher work rates ($>60\%$ V$_{O2\max}$), the estimation of the kinetic parameters refers to a more complex mathematical problem because the V$_{O2}$ kinetics are generally best described with a second-order exponential model incorporating unit step functions (2, 3, 6, 7, 25). The second exponential is introduced in the model to describe a drift (rather than a steady state) in V$_{O2}$ that is observed beyond the third minute of exercise. This model is not continuously differentiable owing to the introduc-

Address for reprint requests and other correspondence: O. Bernard, Laboratoire des Adaptations Physiologiques aux Activités Physiques (LAPHAP EA 3813), Université de Poitiers, Faculté des Sciences du Sport, 4 allée Jean Monnet, 86000 Poitiers, France (e-mail: olivier.bernard@univ-poitiers.fr).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Innovative Methodology

1050 MODELING THE VO2 ON-KINETICS OF HIGH-INTENSITY EXERCISES

tion of step functions. And the problem of fitting a sum of exponentials to numerical data is known to be an ill-conditioned optimization problem (24). In this context, GD methods do not seem perfectly adapted to the mathematical properties of this double exponential function and may impair the estimation of the kinetic parameters. An additional drawback is that GD methods need to be initialized by the investigator (6, 7, 11, 17): the complex properties of the double exponential model may compromise or influence the convergence of the estimation when the investigator introduces starting values far from the optimal solution. An alternative approach to solve this fitting problem consists in using stochastic methods of optimization such like simulated annealing (SA), which belongs to Monte Carlo Markov chain algorithms (15, 24): they are designed for locating the global optimum of the function to be optimized regardless of its mathematical properties. They are also unsupervised in the sense that there is no need for intervention of the investigator because the starting values are randomly chosen in a parameter space.

The purpose of the present paper is to propose an implementation of SA specifically dedicated to fit second-order exponential models with step functions to breath-by-breath VO2 data, in exercise transitions from light- to high-intensity work rates. Applications on simulated and experimental data are performed and compared with the results obtained with a supervised GD algorithm [the generalized reduced gradient 2 algorithm (GRG2)] already used in previous studies (6, 7). The choice for this last algorithm has been dictated by the fact that it has a reputation of robustness on problems involving numerical difficulties, compared with other nonlinear optimization methods (14).

METHODS

Empirical Model to Describe the VO2 Kinetics

To compare SA and GRG2, 40 simulated VO2 responses and 52 experimental VO2 responses are fitted by both algorithms, with a second-order exponential function of the form

\[
VO2(t) = A_0 + A_1 \times \left[1 - \exp\left(-\frac{t - td_1}{\tau_1}\right)\right]U(t - td_1) + A_2 \times \left[1 - \exp\left(-\frac{t - td_2}{\tau_2}\right)\right]U(t - td_2)
\]

where

\[
U(t) = \begin{cases} 
1 & \text{if } t \geq 0 \\
0 & \text{if } t < 0
\end{cases}
\]

is the unit step function. VO2 is the rate of O2 uptake at time t. A1 and A2 are the asymptotic increments above the baseline A0, \(\tau_1\) and \(\tau_2\) are the time constants, and \(td_1\) and \(td_2\) are the time delays for each exponential term after the start of the exercise. This function is commonly used to describe VO2 on-responses from light- to high-intensity (>60% VO2max) work rates (2, 3, 6, 7, 25).

When fitting the experimental data with this model, previous studies usually excluded the first 30 s after the start of exercise (19, 35). The abrupt increase in VO2 occurring during this first period reflects cardiodynamic gas exchange. The rationale for its exclusion is that it is functionally distinct from the subsequent exponential response (33, 35). Its magnitude is also recognized to decrease dramatically when the step transition is from a light-to-moderate exercise intensity compared with a rest-to-exercise transition (35). And the higher the baseline of the transition, the less distinguishable the cardiodynamic component will be from the subsequent exponential VO2 increase (18). In the present study, the VO2 baseline is fixed at \(=1.5\) L/min (i.e., a light-intensity exercise). Then, the cardiodynamic component is considered to be negligible, and all VO2 data measured during the exercise transition are included in the curve fitting.

Estimation Algorithms

The mathematical concepts of both estimation methods (SA and GRG2) and their implementations are extensively documented in the APPENDIX. In the present study, they are used to fit VO2 on-responses acquired breath by breath. Owing to breathing irregularities, these VO2 data fluctuate from breath to breath and so can be viewed as the sum of an underlying physiological information and “noise.” Because this noise is mainly white and Gaussian (19, 27, 30), the criterion to be optimized by SA and GRG2 is the residual sum of squares (RSS) between the raw data and the model (RSS has to be minimized). In a physiological sense, the configuration of both algorithms imposes some constraints on the estimated parameters: 1) \(A_0, A_1, \text{ and } A_2\) have real positive values, 2) \(td_1 \leq td_2\) and 3) \(0 < \tau_1 \leq \tau_2\).

Furthermore, the initialization of GRG2 is made with the same starting values of kinetic parameters, whatever the response to fit. However, to minimize the influence of initialization on the “optimum” solution, each VO2 response is fitted three times with three different sets of starting values (Table 1). Over the three “optimum” solutions obtained for a given response, the solution retained is the one producing the lowest RSS.

Comparison of Both Methods on Simulated Data

Simulated data. Simulated VO2 responses are generated on the basis of real time course and temporal responses. First, experimental VO2 data have been measured during a period of 2 min of a light exercise followed by 8 min of constant-load exercise (80% VO2max) on two different subjects (see Experimental data). Then, the second-order exponential model (Eq. 1) has been used to fit the VO2 responses of both subjects to generate two different sets of kinetic parameters for the transition from light to high intensity (Table 2). At high intensity (>60% VO2max), it is recognized that the VO2 time course described by the second component of Eq. 1 (and called the “slow VO2max component”) either delays the attainment of a steady state or leads VO2 to increase inexorably until the end of the exercise (or until VO2 is reached) (2, 3, 6, 7). Therefore, both subjects have been carefully chosen to take into account both shapes of time course. Set 1 of kinetic parameters produces a second exponential component during which VO2 reaches the asymptotic value of the component before the end of the exercise. Set 2 produces a second exponential component during which VO2 increases almost linearly until the exercise stops. Therefore, both reference VO2 kinetics have been noised by generating 20 Gaussian white noise vectors with a variance equal to 52,000 for set 1 and 20 Gaussian white noise vectors with a variance equal to 125,000 for set 2. As a whole, this produced 40 different simulated VO2 on-responses with an approximately constant SNR (SNR in decibels) equal to 20 dB. Figure 1 shows examples of simulated VO2.

Table 1. Sets of starting values used for initialization of GRG2 algorithm

<table>
<thead>
<tr>
<th>Parameter Sets</th>
<th>A0</th>
<th>A1</th>
<th>td1</th>
<th>td2</th>
<th>(\tau_1)</th>
<th>(\tau_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>1500</td>
<td>2500</td>
<td>0</td>
<td>30</td>
<td>500</td>
<td>120</td>
</tr>
<tr>
<td>Set 2</td>
<td>1300</td>
<td>2300</td>
<td>-10</td>
<td>25</td>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td>Set 3</td>
<td>1700</td>
<td>2700</td>
<td>10</td>
<td>35</td>
<td>700</td>
<td>140</td>
</tr>
</tbody>
</table>

GRG2, generalized reduced gradient 2; \(A_0, A_1, A_2\), baseline and asymptotic increments for each exponential term of the empirical model, expressed in ml O2 min\(^{-1}\); \(\tau_1\) and \(\tau_2\), time constants; \(td_1\) and \(td_2\), time delays for the first and second exponential terms of the empirical model, respectively, expressed in seconds.
Comparison of Both Methods on Experimental Data

Overview. The purpose of this section is to verify that the results obtained on simulated data are in accordance with those obtained on experimental data. As presented above, SA and the GRG2 algorithms determine the kinetic parameters by minimizing RSS. The method that provides the lowest RSS could be considered as the better one (see the Appendix). Therefore, the first criterion used to compare both algorithms on experimental data is RSS obtained on the same signals.

In parallel, it is interesting to compare the repeatability of each estimation procedure. Several repetitions of the same exercise by humans offer an intrapersonal variability of the response due, for one part, to the variability of the underlying time course of \( \dot{V}_{O_2} \) and, for another part, to the variability of the noise. We make the assumption that, in healthy subjects, the intrapersonal variability of the underlying physiological response is low. Therefore, the most pertinent algorithm should be the one that is the less sensible to the intrapersonal variability of the noise; this is the estimation method that produces the lowest difference of estimation between test 1 and test 2. The second criterion of comparison is therefore the coefficient of repeatability adopted by the British Standards Institution (8).

Subjects. Twenty-six healthy young men, students in physical education, volunteered to participate in this study after being informed of the purpose, the requirements, and their rights as subjects. They were (means \( \pm \) SD) 22 \( \pm \) 1 yr of age, their height was 182 \( \pm \) 8 cm, and their body mass was 73.0 \( \pm \) 7.9 kg. The experimental protocol was approved by the Ethics Committee of the Faculty of Medicine of the Université Catholique de Louvain, and all subjects gave informed, written consent before participating in the study. The subjects were fully familiar with laboratory exercise.

Experimental protocol. The subjects were instructed to arrive at the laboratory in a rested and normally hydrated state, at least 2 h postprandial, and to avoid strenuous exercise in the 24 h preceding a test session. For each subject, tests took place at the same time of day. All exercises were performed on a friction-loaded cycle ergometer (Monark type 818E, Stockholm, Sweden). For all tests, seat height was held constant within a subject and the feet were strapped to the pedals. In a preliminary test session, the subjects performed a ramp exercise test (20 W/min; pedal rate 80 rpm) to volitional exhaustion, preceded by a 3-min unloaded cycling work. The breath-by-breath data were collected and displayed at 30-s intervals to determine peak \( \dot{V}_{O_2} \). During each subsequent session, subjects performed a square-wave exercise transition from a 4-min baseline at 80 W to an 8-min heavy exercise at 80% of peak \( \dot{V}_{O_2} \) (pedal rate 80 rpm). Pulmonary gas exchange was measured breath by breath during the entire duration of the square-wave exercise. This test was realized twice in separate sessions to test the repeatability of the kinetic parameters.

Pulmonary gas-exchange analysis. The subjects breathed through a face mask during the entire duration of the square-wave exercise, and pulmonary gas exchanges were monitored breath by breath by use of an integrated computerized system (Quark Cosmed, Rome, Italy). Precision-analyzed gas mixtures and a 3-liter Rudolph syringe were
Statistical analysis. Each of the 52 experimental responses (26 for the first test and 26 for the second test) is fitted with the second-order exponential function (Eq. 1) by using GRG2 and the SA algorithms with the configurations presented above. For RSS analysis, a linear regression analysis is computed to compare the RSS values produced by both algorithms on the 52 experimental responses. For the analysis of the repeatability of each kinetic parameter, each algorithm produces the mean of the 26 intraday differences of repeated measurements and the corresponding coefficient of repeatability. This coefficient of repeatability is defined as twice the standard deviation of the 26 individual differences of repeated measurements. The prerequisite condition for repeatability analysis is that the mean of the intraday differences be not significantly different from zero. For each estimation procedure, the mean difference for each kinetic parameter is compared with zero by a Student’s t-test. Comparison of the repeatability of both estimation procedures is done with an F-test used on the variances of the intraday differences. The critical level accepted for statistical significance is set at $P = 0.05$.

RESULTS

Simulated Data

For each parameter of the second-order exponential function, Fig. 2 and Table 3 present the bias, the limits for its 95% confidence interval, and the standard deviation of the estimation error produced either by GRG2 or by the SA algorithm on the whole 40 simulated responses.

Concerning the baseline and the first exponential component of the model (called phase II), the results demonstrate that the SA algorithm produces better estimates than GRG2. For three of four kinetic parameters ($A_0$, $t_1$, $\tau_1$), the precision of the estimations is significantly higher with the SA than with the GRG2 algorithm ($P < 0.05$), and two of four parameters are significantly biased by the GRG2 algorithm ($t_1$ and $\tau_1$; $P < 0.05$). Moreover, the bias for $t_1$ and $\tau_1$ is significantly lower with the SA than with the GRG2 algorithm ($P < 0.05$). The bias for $A_1$ is significantly higher with the SA than with the GRG2 algorithm ($P < 0.05$), but its value remains small compared with the mean magnitude of phase II (<4%).

Fig. 2. Biases and their 95% confidence intervals for the estimation errors of the kinetic parameters of the simulated $V\dot{O}_2$ responses. The bias is the mean difference between the value estimated with either the simulated annealing (SA) algorithm or the generalized reduced gradient 2 (GRG2) algorithm, less the reference value. The 95% confidence intervals are expressed as the inferior and superior limits of confidence. $A_0$, $A_1$, $A_2$, baseline and asymptotic increments for each exponential term of the empirical model (in ml O2/min); $t_1$ and $\tau_1$, time constants; $t_d1$ and $t_d2$, time delays for the first and second exponential terms of the empirical model, respectively (in s).
Concerning the second exponential component of the model (called the slow component), the precision is significantly higher with the SA algorithm for all three parameters ($A_2$, $td_2$, and $\tau_2$; $P < 0.05$). Nevertheless, the high values for the precision intervals of the biases and for the standard deviations of the estimation errors of the three kinetic parameters indicate that all estimations are impaired, even when produced by the SA algorithm. Finally, $td_2$ is significantly biased by both algorithms ($P < 0.05$).

**Experimental Data**

Figure 3 presents the linear relationship between RSS values ($r^2 = 0.9995; P < 0.05$), with the values produced by the SA algorithm being the independent variables and the values produced by GRG2 being the dependent variables. The $y$-intercept ($b$) is significantly higher than 0 ($b = 278.244; P < 0.05$), whereas the slope of the relationship ($a$) is not significantly different from the unity ($a = 0.9999; P > 0.05$). In details, 46 of the 52 exercise tests show a lower RSS value when produced by the SA algorithm. On these 46 tests, the SA values of RSS are on average $4\%$ lower than with GRG2. Conversely, in the six other situations, the average value produced by the GRG2 algorithm is only $0.2\%$ lower than the value produced by SA.

For each kinetic parameter, Table 4 presents the mean values (SD) obtained on the first and second repetition of the exercise, the mean intrindividual difference, and the repeatability coefficient for each estimation procedure. It must be emphasized that the estimated average SNR is $\approx 21\text{ dB}$ for each test (test 1 and test 2), which is comparable to the SNR used to generate simulated data. Except for $A_0$, the repeated measurements (test 1 vs. test 2) are not significantly different from each other, whatever the estimation procedure ($P > 0.05$). The repeatability coefficients are not significantly different between estimation procedures for $td_1$, $\tau_1$, and $td_2$ ($P > 0.05$). GRG2 gives a significantly better coefficient of repeatability for $A_1$ ($P < 0.05$). The repeatability coefficients are significantly better with SA for $A_2$ and $\tau_2$ ($P < 0.05$). As for simulated data, the large repeatability coefficients for the kinetic parameters of the second-exponential component show that both estimation procedures may produce large differences of parameter values from the same (or comparable) $V\dot{O}_2$ on-response.

**DISCUSSION**

The main result of the present study is that the SA algorithm, compared with the GRG2 algorithm, improves the estimation of the kinetic parameters of the second-order exponential function describing the $V\dot{O}_2$ time course in a transition from light- to high-intensity exercises. This is shown by results on simulated data and also by the global better estimation of parameters on experimental data (SA returns lower RSS than GRG2). To our knowledge, this study is the first to focus specifically on the improvement of the estimation procedure for these second-order kinetic parameters.

Comparison of both algorithms (SA vs. GRG2) on simulated data is linked to the postulate that the intersample fluctuations of pulmonary gas exchanges are white and Gaussian. Previous studies have already characterized these fluctuations as being white and Gaussian in healthy and chronic obstructive pulmonary disease adults (19, 27, 30) but focused only on exercises of moderate intensity (for which the pulmonary gas exchanges are appropriately described by a first-order exponential function at the onset of a step exercise). In these studies, the restriction to this intensity domain facilitated the statistical analysis of the breath-to-breath fluctuations because it could be performed at $V\dot{O}_2$ steady state. In the present study (with a more complex time course of pulmonary gas exchanges than at moderate intensity), the same statistical analysis was difficult to perform because the attainment of a $V\dot{O}_2$ steady state was a priori not known. Therefore, we made the assumption that the statistical characteristics of the noise of both intensity domains (moderate and high) were comparable, and we generated simulated data using additive white Gaussian noise. This postulate

![Fig. 3. Linear relationship between residual sum of squares (RSS) values produced by the SA and GRG2 algorithms, on experimental $V\dot{O}_2$ responses.](image-url)
Innovative Methodology

MODELING THE \( \dot{V}O_2 \) ON-KINETICS OF HIGH-INTENSITY EXERCISES

Table 4. Kinetic parameters estimated on experimental \( \dot{V}O_2 \) responses

<table>
<thead>
<tr>
<th>Experimental Data</th>
<th>( A_0 ) (mL/min(^{-1}))</th>
<th>( A_1 ) (mL/min(^{-1}))</th>
<th>( t_d ) (s)</th>
<th>( \tau_1 ) (s)</th>
<th>( A_2 ) (mL/min(^{-1}))</th>
<th>( t_d ) (s)</th>
<th>( \tau_2 ) (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRG2 Test 1</td>
<td>1,422 (181)</td>
<td>1,603 (398)</td>
<td>6 (9)</td>
<td>42 (22)</td>
<td>3,812 (10,919)</td>
<td>244 (148)</td>
<td>1,543 (3,435)</td>
</tr>
<tr>
<td>Test 2</td>
<td>1,501 (140)</td>
<td>1,536 (367)</td>
<td>11 (7)</td>
<td>36 (16)</td>
<td>86,205 (427,653)</td>
<td>252 (116)</td>
<td>925 (2,631)</td>
</tr>
<tr>
<td>Mean difference</td>
<td>-79*</td>
<td>67</td>
<td>-4</td>
<td>7</td>
<td>-82,393</td>
<td>-8</td>
<td>618</td>
</tr>
<tr>
<td>Repeatability coefficient</td>
<td>307</td>
<td>370</td>
<td>22</td>
<td>39</td>
<td>856,914</td>
<td>350</td>
<td>7,294</td>
</tr>
<tr>
<td>SA Test 1</td>
<td>1,423 (181)</td>
<td>1,352 (486)</td>
<td>10 (13)</td>
<td>24 (11)</td>
<td>878 (1,038)</td>
<td>165 (92)</td>
<td>622 (1,332)</td>
</tr>
<tr>
<td>Test 2</td>
<td>1,506 (140)</td>
<td>1,346 (361)</td>
<td>14 (10)</td>
<td>25 (13)</td>
<td>1,691 (5,470)</td>
<td>178 (116)</td>
<td>369 (704)</td>
</tr>
<tr>
<td>Mean difference</td>
<td>-83*</td>
<td>7</td>
<td>-4</td>
<td>-1</td>
<td>-813</td>
<td>-13</td>
<td>253</td>
</tr>
<tr>
<td>Repeatability coefficient</td>
<td>321</td>
<td>812†</td>
<td>32</td>
<td>34</td>
<td>11,062†</td>
<td>274</td>
<td>3,026†</td>
</tr>
</tbody>
</table>

*Mean (SD) value of kinetic parameters estimated on experimental \( \dot{V}O_2 \) responses, either with GRG2 or with SA algorithms given for test 1 and test 2. The mean difference and the repeatability coefficient of these repeated measurements are also presented for each variable and each estimation method. †Significantly different from 0; ‡significantly different from the corresponding value obtained with the GRG2 algorithm. Significance at 0.05 level

could have then produced different results between simulated and experimental data if the statistical characteristics of both kinds of data had been dramatically different. However, the best estimations produced by the SA algorithm (compared with GRG2) on simulated \( \dot{V}O_2 \) kinetics are confirmed on experimental data.

The superiority of SA may be explained by two main factors. The pertinence of the SA algorithm depends on the configuration of intrinsic parameters, such as the transition kernel (see the appendix). Our results suggest that the configuration chosen for these parameters is appropriate enough to improve the estimations made by the GRG2 algorithm. The second explaining factor is that the SA algorithm is a stochastic unsupervised method, whereas GRG2 needs to be initialized. Therefore, the estimated parameters found by GRG2 are supposed to be influenced by the initialization. It must be kept in mind that, in the present study, an attempt to minimize this drawback was made by testing several initializations. Despite this precaution, the SA algorithm produces better estimates than GRG2.

Our results on simulated data also demonstrate that the confidence levels observed for the kinetic parameters of the first exponential component (\( t_{d1} \) and \( \tau_1 \)) are in accordance with the results published elsewhere for the corresponding kinetic parameters of moderate-intensity step exercises, i.e., for kinetic parameters of phase II (19). With an average SNR equal to 20.3 dB, the SA algorithm produces a standard deviation of the estimation error equal to 11.4 and 9.0 s for \( t_{d1} \) and \( \tau_1 \), respectively. Lamarra et al. (19) have previously proposed a prediction equation of this standard deviation incorporating two factors: the SNR value of the fitted signal and a constant \( L \) that depends on the value of the reference time constant and on the amount of data available for fitting. This equation, validated for a first-order exponential model when fitted with a GD method, predicts standard deviations around 10.5 and 8.8 s for \( t_{d1} \) and \( \tau_1 \), respectively, when taking into account our average SNR of 20.3 dB (Lamarra et al. did not directly express the SNR in dB but used an index of the noise-to-signal ratio defined as the ratio between the standard deviation of the fluctuations and the magnitude of the response; in the present case, this gives an average value of \( \approx 21\% \)). This is an interesting result because it is well known that the parameter confidences of a model degrade with increasing the model order. Because we focused on a second-order exponential function describing \( \dot{V}O_2 \) time course, this quality of estimation was not expected compared with the estimations made by Lamarra et al. for a first-order function. This is in evidence an improvement due to the SA algorithm because the quality of estimation for \( t_{d1} \) and \( \tau_1 \) is comparatively less with GRG2 (the standard deviations of the estimation error are 30.8 and 24.9 s, respectively). Concerning the experimental data, the repeatability coefficients of phase II parameters seem a priori larger than those previously published. Puente-Maestu et al. (26) have tested the repeatability of the time constant (\( \tau \)) and of the magnitude of the asymptotic response (\( A \)) for the phase II of the \( \dot{V}O_2 \) time course of moderate-intensity step exercises. The noise-to-signal ratio [expressed as defined by Lamarra et al. (19)] was 10%. Using the equation of Lamarra et al., a prediction of their repeatability coefficient for the same SNR of 20.3 dB gives values equal to 21.2 s and 220 ml/min for \( \tau \) and \( A \), respectively. Our values of repeatability coefficient for the corresponding kinetic parameters of phase II are found to be larger (at least equal to 34 s and 370 ml/min for \( \tau_1 \) and \( A_1 \), respectively). A possible explanation may be linked to the time constant value. Lamarra et al. have showed that the higher the time constant value, the higher the confidence level of a kinetic parameter for a given SNR. Because the population studied by Puente-Maestu et al. exhibits its time constant values close to 80 s (more than twice the values of the present population), this may partly explain the present worst result of repeatability.

Concerning the second exponential component of the model, the SA algorithm improves the confidence intervals of the biases and the standard deviations of the estimation errors of the three parameters (\( A_2 \), \( t_{d2} \), and \( \tau_2 \)) and also the repeatability of \( A_2 \) and \( \tau_2 \) (\( P < 0.05 \) compared with the GRG2 algorithm). Unfortunately, these values remain high whatever the algorithm used. Therefore, their interpretation is experimentally limited. One possible explanation for this poor confidence level is linked to the shape of the underlying \( \dot{V}O_2 \) kinetics. On simulated data, set 2 of reference kinetic parameters induces a linear increase of \( \dot{V}O_2 \) (rather than an exponential one), over the duration corresponding to the second exponential. This is due to both high values for the magnitude and the time constant of this exponential component (Table 2). In this context, the poor confidence level may be explained by the fact that the \( A_2 \)-to-\( \tau_2 \) ratio is a good approximation of the slope of a linear function describing this increase in \( \dot{V}O_2 \); thus infinity of values

\[ J \text{ Appl Physiol} \cdot \text{VOL} \ 100 \cdot \text{MARCH} \ 2006 \cdot \text{www.jap.org} \]
for both parameters may produce this ratio (2, 3). A second possible explanation is linked to the low SNR_{dB} value of the data. We simulated breath-by-breath \( \dot{V}O_2 \) time courses with a large breath-by-breath variability of the data (corresponding to a low SNR_{dB}) to mimic the typical breath-by-breath variability of \( \dot{V}O_2 \) data obtained with commercial gas-exchange analysis systems. Although this does not prevent doing a comparison of both algorithms (which is the purpose of the present study), it is well recognized that the lower SNR_{dB}, the lower the confidence level of the estimation of a model parameter. To improve further this confidence level, reducing the magnitude of the noise of \( \dot{V}O_2 \) data has therefore to be envisaged before fitting the kinetics. Concerning the repeatability of the estimated parameters, a previous study has already tested it using a GD method (25). But it neither expressed the repeatability in the same way than in the present study nor provided SNR_{dB} values; thus only qualitative comparisons are possible. Their results are in accordance with the present ones: these authors stated that the repeatability of \( \dot{A}_2 \) and \( \tau_1 \) is well higher than for \( \dot{A}_2 \) and \( \tau_2 \) and that the interpretation of these two last indexes is limited. An interesting finding of the present study is that the SA algorithm improves the repeatability of \( \dot{A}_2 \) and \( \tau_2 \) compared with GRG2. Concomitantly, the better precision for SA in estimating \( \dot{A}_2 \) and \( \tau_2 \) on simulated data (compared with GRG2) means that SA is less sensible than GRG2 to the intersignal differences of noise. This may partly explain the better repeatability coefficients produced by SA if we make the assumption that the within-subject variability of the \( \dot{V}O_2 \) response is due rather to the variability of the noise than to the variability of the underlying physiological response (19). Thus the less sensibility of SA to the intersignal differences of noise on simulated data enables a lesser sensibility to the within-subject variability between repeated experimental measurements.

Another result of the present study is the significant overestimation of \( \tau_2 \) on simulated data, whatever the estimation algorithm (see Table 3). This may be due to the characteristics of the simulated noise rather than an intrinsic incapacity of both algorithms to estimate \( \tau_2 \) accurately. Readers must be aware of the fact that a realization of a Gaussian white noise over a long time period may contain short series of values that look like “local perturbations” (i.e., the mean residual substantially deviates from zero on very short local periods). Therefore, the beginning of the second exponential (\( \tau_2 \)) may be attributed to the time of appearance of a local perturbation when it looks like a “step” in the \( \dot{V}O_2 \) time course. In the present study, the probability that these local perturbations appear beyond the reference \( \tau_2 \) (\( \tau_{2_{ref}} \)) seems high because the exercise duration beyond \( \tau_2 \) is long compared with the time spent from the start of exercise to \( \tau_2 \); this may explain the overestimation of this time delay by both algorithms. An example of this phenomenon is observed in Fig. 4. Figure 4A shows the reference \( \dot{V}O_2 \) kinetics produced by set 1 of parameters (dark line) and the associated simulated data. Figure 4B shows the residual differences between these kinetics and data. Note a local perturbation well apparent between 200 and 220 s (i.e., beyond the time of \( \tau_{2_{ref}} \)). Consequently, a value of 215 s is attributed to \( \tau_2 \) for the fitted \( \dot{V}O_2 \) kinetics (dotted line in A).

In real contexts, these local perturbations may happen from perturbations of the breathing pattern such as coughing, sneezing, or swallowing, or from a temporary incapacity for some algorithms to compute accurate \( \dot{V}O_2 \) values (\( \dot{V}O_2 \) is locally overestimated or underestimated) (5). Thus the probability for \( \tau_2 \) to be biased in real contexts may be high.

The aim of the present paper was to demonstrate the superiority of SA (compared with GRG2) for fitting a given \( \dot{V}O_2 \) signal, but not to discuss the validity of this raw signal for insight into the nature of the underlying physiological processes. However, the above discussion points out that not only the estimation procedure may influence the confidence level of the estimated parameters; the SNR level and the appearance of local perturbations in the raw signal are also of importance. Therefore, special attention must be paid to the quality of this raw signal (i.e., as little noise as possible) before fitting it. A frequently used procedure to reduce the magnitude of the noise and local perturbations is the superposition of several repeated measurements (2, 3, 29). Unfortunately, one of the disadvantages of this filtering procedure is that a great number of repeated measurements is sometimes needed to obtain a desired confidence level in an estimated kinetic parameter. For example, Lamarra et al. (19) proposed a prediction equation of the number of square-wave repetitions required to achieve a desired confidence level when the modeling function is a first-order exponential function. Applied to our data, it could be predicted that \( \sim 40 \) repeated measurements would have been needed to reduce the standard deviation of \( \tau_1 \) from 9.0 to 1.0 s. Rather than filtering the data a posteriori, attempts may be tried to improve the methodology for computing breath-by-breath \( \dot{V}O_2 \) data. At this time, two main approaches for calculating...


Innovative Methodology

$\dot{V}O_2$ have been developed and used in publications. The first one calculates breath-by-breath $O_2$ taken up at the mouth by considering the difference between inspired and expired $O_2$ volumes (4, 22). The second one estimates the breath-by-breath alveolar-capillary transfer of $O_2$ by accounting for the breath-to-breath changes in alveolar $O_2$ stores (1, 5, 9, 16, 31, 34). Some of these algorithms (1, 5, 10, 12, 34) have been shown to largely reduce the magnitude of the breath-by-breath variability of the data (especially those that compute the alveolar gas exchanges). Unfortunately, their use for insight into the nature of the underlying physiological processes is still unsure because none of them induces the same $\dot{V}O_2$ time course during transient states (1, 5, 11, 12), and so none of them can be considered as a gold standard. For example, when comparing the different algorithms that compute $\dot{V}O_2$ at the alveolar level, the time constant of the phase II ($\tau$) at the onset of a square wave exercise may vary from 34.3 to 46.8 s (11). These considerations lead us to suggest that future research should concomitantly deal with the improvement of both the procedures for assessing $\dot{V}O_2$ data and the procedures for estimating the kinetic parameters, when the final goal is to provide insight into the nature of the physiological processes underlying the $\dot{V}O_2$ kinetics.

In conclusion, the present investigation demonstrates that the proposed implementation of a global stochastic optimization method (SA) improves the estimations of the $\dot{V}O_2$ kinetic parameters of square-wave high-intensity exercises, compared with a GD method (GRG2). However, despite tight limits of confidence for the baseline and the parameters of the first exponential component of the model, a large inaccuracy remains in estimating the parameter values of the second exponential component. This may be due to the shape of the underlying $\dot{V}O_2$ response showing, in some cases, linear rather than exponential increase in $\dot{V}O_2$ and also due to the large breath-to-breath irregularities of the $\dot{V}O_2$ response. Therefore, one of the perspectives of this work concerns an attempt to improve the SNR, by reenvisioning the algorithms used to compute $\dot{V}O_2$ data.

APPENDIX

Estimation Problem

The breath-by-breath values of $\dot{V}O_2$ are sampled with a variable sampling period as the respiratory frequency changes due to the breath-to-breath irregularities and also due to the need for increasing the ventilatory flow rate. $\dot{V}O_2$ values can then be viewed as a vector $\mathbf{v}_0 = [\dot{V}O_{2,1}, \ldots, \dot{V}O_{2,N}]$, sampled over a time vector $t = [t_0, \ldots, t_N]^T$. Each $t_{k} - t_{k-1}$, $k = 1, \ldots, N$, corresponds to the laps of time of a breath, and the first value of $\mathbf{v}_0$ is obtained at $t_1$. $\mathbf{v}_0$ is supposed to incorporate an underlying physiological response (i.e., the $O_2$ exchanges occurring at the alveolar site) plus an additive noise $\mathbf{e} = [e_1, \ldots, e_N]^T$. In the present context, Eq. 1 is supposed to describe this underlying $\dot{V}O_2$ response. Thus it is possible to write $\mathbf{v}_0$ following the general linear model (GLM) (24):

$$\mathbf{v}_0 = \sum_{m=0}^{2} A_m \mathbf{g}_m + \mathbf{e} = \mathbf{G} \mathbf{a} + \mathbf{e}$$

(2)

The vectors $\mathbf{g}_m$, $m = 0, 1, 2$, have a length of $N$ and can be written as follows:

$$\mathbf{g}_0 = [1, \ldots, 1]^T,$$

(3)

$$\mathbf{g}_m = [g_{0m}, \ldots, g_{Nm}]^T, m = 1, 2,$$

(4)

and

$$\mathbf{G} = [\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}_2]$$

(6)

The estimation of $\theta_{\dot{V}O_2}$ is done classically by procedures that minimize the RSS between the experimental data points and the model, i.e., the procedures are searching for the set of parameters that minimizes the Euclidean norm of $\mathbf{e}$ (see Eq. 2). In the present case, this is a constrained optimization problem (some constraints exist on the parameter values; see Estimation algorithm in METHODS) on a function that is not continuously differentiable. Moreover, the problem of fitting a sum of exponential to numerical data is known to be ill conditioned. In the following, we present the two algorithms we compared on this estimation problem.

Simulated Annealing

SA is based on theory of Markov chains (MC) and belongs to the set of stochastic global methods of optimization. SA is an estimation procedure in the maximum likelihood (ML) sense that consists in optimizing the likelihood function $L$. In the case of $\dot{V}O_2$ data, $L$ can be expressed as follows:

$$L(\theta_{\dot{V}O_2}, \dot{V}O_2) = p(\dot{V}O_2 | \theta_{\dot{V}O_2})$$

(9)

where $L(\theta_{\dot{V}O_2}; \dot{V}O_2)$ is the likelihood function and $p(\dot{V}O_2 | \theta_{\dot{V}O_2})$ is the probability density function of $\dot{V}O_2$ conditionally to $\theta_{\dot{V}O_2}$. Optimizing $L$ is known to be adapted to the most general contexts in optimization and provides consistent parameter estimation, whatever the probabilistic characteristics of the noise $\mathbf{e}$. In the present case, $\mathbf{e}$ can be considered as a realization of zero mean independently and identically distributed temporal series of Gaussian random variables, i.e., $\mathbf{e}$ is a realization of a Gaussian white noise with variance $\sigma^2$ (19, 27, 30).

Then, each $e_k$ has the same probability density function, which is of the form:

$$p(e_k) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{e_k^2}{2\sigma^2}\right), k = 1, \ldots, N$$

(10)

Because $\mathbf{e}$ is additive and independently and identically distributed, $L$ can be expressed as:

$$L(\theta_{\dot{V}O_2}, \dot{V}O_2) = p(e) = \prod_{k=1}^{N} p(e_k)$$

(11)

Because $\mathbf{e}$ is Gaussian, using the log-likelihood function can do simplification:

$$LL(\theta_{\dot{V}O_2}, \dot{V}O_2) = \log[p(e)] = -[\log(\sigma^2) + 1 + \ln(2\pi)]^N \frac{N}{2}$$

(12)

Therefore, the ML optimization criterion becomes:

$$\hat{\theta}_{\dot{V}O_2} = \arg \min_{\theta_{\dot{V}O_2}} (\hat{\sigma}^2)$$

(13)

where $\hat{\theta}_{\dot{V}O_2}$ is the estimation of $\theta_{\dot{V}O_2}$ and $\hat{\sigma}^2$ is the estimation of the variance of the noise.

J Appl Physiol • VOL 100 • MARCH 2006 • www.jap.org
Let us remark that RSS can be directly obtained from $\hat{\sigma}_v^2$:

$$RSS = N \times \hat{\sigma}_v^2$$

(14)

Used in that context, SA is a stochastic iterative method of RSS minimization. SA is stochastic as it is based on the simulation of a MC. Indeed a new state of a MC is sampled each iteration by 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 (13, 23, 28).

In our case, a state is defined by a given set of $\theta_{v_0}$: from a random initial set $\theta_{v_0}$, a new set of parameters $\theta_{v_0}$ is sampled each iteration $i$ of the algorithm. To estimate $\theta_{v_0}$ with SA, the convergent probability law of the states above mentioned must be chosen uniform on the sets of parameters that provide global maxima of Eq. 9 (or global minima of Eq. 13) and zero elsewhere. To this end, SA does not optimize directly the likelihood function but an energy function of the minimization. SA is stochastic as it is based on the simulation of a special case of the Metropolis-Hastings algorithm (for more details, see Ref. 24).

The energy function $E_n$ can now be defined as:

$$E_n(\theta_{v_0}) = -\log[p(\theta_{v_0} | v_0)]$$

(18)

According to Eq. 18, the states that maximize $L(\theta_{v_0}; v_0)$ are the states of lowest energies. Then, minimizing $E_n(\theta_{v_0})$ provides ML estimation of $\theta_{v_0}$, and so RSS minimization (see Eqs. 13 and 14). For the special case $T = 0$, the only possible states according to Eq. 15 are the ones that produce the global minima of the energy. Therefore, the Gibbs distribution at $T = 0$ is the desired probability law above mentioned.

To iterate SA, two main approaches exist to sample a MC with a transition kernel that has appropriate properties: the Metropolis-Hastings algorithm and the Gibbs sampler. In this paper, we used the Metropolis-Hastings algorithm (for more details, see Ref. 24).

Gradient-Descent-Based Method (the GRG2 Algorithm)

The principle of GD methods like the GRG2 algorithm is to find iteratively an optimum for $\theta_{v_0}$ by the use of gradient information. The GRG method (20, 21) is an evolution of the Wolfe’s reduced gradient method. Although the validity of the reduced gradient method is limited to unconstrained nonlinear problem, with first-order information only, GRG2 is adapted to constrained nonlinear optimizations and incorporates second-order information based on reduced Hessian. Although it is guaranteed to find an optimum only on problems with continuously differentiable functions, and then only in the absence of numerical difficulties (such as degeneracy or ill conditioning), GRG2 has a reputation for robustness, compared with other nonlinear optimization methods, on difficult problems where these conditions are not fully satisfied. This is the case of estimating $\theta_{v_0}$ (see Estimation Problem).

GRG2 is based on Newton-Raphson iterations to find a root of the gradient of the function $f(\theta_{v_0})$, $\nabla f(\theta_{v_0})$, where $f$ is the cost function to optimize. Let suppose $\theta_{v_{0,j}} + \delta_i$ is the location of the minimum along the direction of $\theta_{v_{0,j}}$, then

$$\delta_i = (\nabla \nabla^T f(\theta_{v_{0,j}}))^{-1} \nabla f(\theta_{v_{0,j}})$$

(20)

where $\{\nabla \nabla^T f(\theta_{v_{0,j}})\}^{-1}$ is the inverse of the Hessian matrix. By definition, a variable metric algorithm estimates iteratively this matrix. In GRG2, the variable metric algorithm in use is the Broyden-Fletcher-Goldfarb-Shanno algorithm. The algorithm converges when $\|\delta_i\|$ undergoes a threshold fixed by the user. We fixed with the threshold at $10^{-12}$. The GRG2 method is implemented in the EXCEL solver (14).

ACKNOWLEDGMENTS

We are grateful to Hervé Billette for technical assistance, as well as the subjects for patient participation.

GRANTS

This work was supported by the Association de Recherche en Physiologie des Activités Physiques et Sportives.

REFERENCES


