South African Statist. J. (1984) **18**, *161-176*

THE CHOICE OF AN APPROPRIATE NONLINEAR MODEL FOR THE RELATIONSHIP BETWEEN CERTAIN VARIABLES IN RESPIRATORY PHYSIOLOGY

S.H.C. du Toit: Institute for Statistical Research, Human Sciences Research Council, Pretoria.

and

R. Gonin: Institute for Biostatistics, Medical Research Council, Tygerberg.

Key words: General transformation; growth curves; initial parameters; nonlinear least squares; oxyhaemoglobin.

Summary: The Richards family of growth curves has proved to be of great use in practical situations. In this paper an extension of this family is obtained by considering a general transformation of the abscissa. This transformation is equivalent to fitting a Richards curve with varying parameters. The behavioural pattern of these parameters provides additional insight into oxyhaemoglobin equilibrium problems.

1. OXYGEN SATURATION IN RESPIRATORY PHYSIOLOGY

Pulmonary disorders such as asthma, chronic bronchitis and emphysema produce specific patterns when lung function tests are performed. One such lung function test involves the relationship between oxygen saturation in arterial blood (sO_2) and the partial pressure of oxygen in the pulmonary capillary (pO_2). Oxygen saturation is expressed as a percentage (%) whilst pO_2 is measured in torr (mm Hg). pO_2 can only be

measured directly in arterial blood. This requires catheterization, an invasive procedure which carries the risk of infection as well as causing discomfort to the severely ill patient.

Over the last decade a new device, the fibre-optic ear oximeter has been developed. The oximeter which is attached to the lobe of the patient's ear is able to measure the absorption spectrum of oxyhaemoglobin in the arterial blood of the ear. From this spectrum the oxygen saturation can be calculated. Once the saturation is known the pO_2 can be calculated by means of the oxygen dissociation curve. This obviates the use of invasive techniques to measure pO_2.

It is our purpose to find a mathematical relationship between oxygen saturation sO_2 (dependent variable) and oxygen tension (partial pressure) pO_2 (independent variable). An important point which has to be borne in mind, is the invertibility of the relationship, i.e. pO_2 (dependent variable) and sO_2 (independent variable) in order to use non-invasive techniques.

In the next section we shall discuss the Richards family of growth curves. In section 3 we shall discuss the effect of general transformations on the abscissa and derive the extended Richards family of growth curves. In the last section an application of this theory to oxyhaemoglobin equilibrium data will be given.

2. THE RICHARDS FAMILY OF GROWTH CURVES

A family of growth curves which has proved to be of great use in practical situations is the so-called Richards (1959) family of growth curves. Du Toit (1979) has written this family as a four parameter function:

$$(1) \quad f(t) = \alpha(1 + s\beta\rho^{t})^{\lambda}, \quad t > 0$$

where the parameter λ determines the point of inflection of the curve and the constant, s, denotes the sign of the term $\beta\rho^t$.

If the function increases monotonically in t, then $s = -1$ for $\lambda \geq 0$ and $s = +1$ for $\lambda < 0$. Opposite signs are allocated when the function increases monotonically in t. The following linear inequality constraints are imposed upon the parameters:

$$
\begin{aligned}
(2)\quad & \alpha \geq 0 \\
& 0 \leq \beta \qquad \text{if } s = 1 \\
& 0 \leq \beta \leq 1 \qquad \text{if } s = -1 \\
& 0 \leq \rho \leq 1 ;
\end{aligned}
$$

Under the constraints (2), $f(t)$ is a monotonic function in t and possesses a uniquely determined inverse.

Three well-known growth curves are special cases of (1), viz.:

$$
\begin{aligned}
(3)\quad & f(t) = \alpha(1 - \beta\rho^t) && \text{(the modified exponential, } \lambda = 1) \\
(4)\quad & f(t) = \alpha/(1 + \beta\rho^t) && \text{(logistic curve, } \lambda = -1) \\
(5)\quad & f(t) = \alpha \exp(-\beta^*\rho^t) && \text{(Gompertz, } \beta^* = |\beta\lambda|,\ |\lambda| \to \infty).
\end{aligned}
$$

These curves have the advantage that each parameter has a definite physical meaning.

For example, if we take

t = the value of the independent variable (e.g. time)
and $f(t)$ = characteristic being measured
then α = asymptote or limiting value of f, i.e. as $t \to \infty$ $f \to \alpha$
β = potential increase or decrease (i.e. change) in f as t varies
ρ = slope parameter which characterises the rate at which f varies with a change in t.

In general a decrease in ρ coincides with an increase in growth rate and vice versa.

3. THE EFFECT OF GENERAL TRANSFORMATIONS ON THE ABSCISSA (x)

Suppose we make the linear transformation

$$t = mx + c \tag{6}$$

and investigate its effect on (1). Substitution of (6) into (1) yields:

$$f(mx + c) = \alpha(1 + s\beta\rho^{mx+c})^{\lambda} = \alpha(1 + s\beta^*(\rho^*)^x)^{\lambda}$$

where $\beta^* = \beta\rho^c$, $\rho^* = \rho^m$.

We therefore conclude that the class of models (1) is closed with respect to linear transformations of the argument. For example a Gompertz curve will remain a Gompertz curve, but instead of defining the function in terms of the parameters β and ρ it is defined in terms of the new parameters β^* and ρ^*.

Consider a general monotonic transformation $t = g(x)$ where $g(x)$ is a continuous function in x. The previous invariance property will enable us to vary the values of β and ρ in a continuous fashion. For this purpose consider the graphical representation in figure 1.

This figure shows that $g(x)$ may be approximated by a straight line or tangent

$$t = c_i + m_i x$$

in the neighbourhood of each point x_i. From (6) it follows that

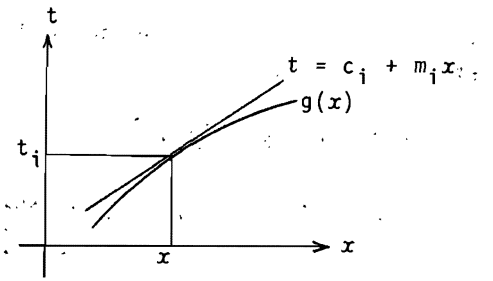

FIGURE 1. Effect of a continuous transformation $t = g(x)$ on the abscissa.

$$(7) \quad f(x_i) = (1 + s\beta_i \rho_i^{x_i})^{\lambda}$$

at point $x = x_i$ where

$$(8) \quad \beta_i = \beta \rho^{c_i} \text{ and}$$

$$(9) \quad \rho_i = \rho^{m_i}.$$

Conversely, given a set of k values (β_i, ρ_i, x_i), $i=1,2,\ldots,k$, then k points $g(x_i)$ on the curve $g(x)$ can be determined provided that the location and scale of $t = g(x)$ are fixed by the constraint $g(x) = x_1$ for all $x \leq x_1$.

This restriction implies that

$$(10) \quad c_1 = 0 \text{ and } m_1 = 1.$$

Substitution of (10) into (8) and (9) yields

$$\beta_1 = \beta \text{ and } \rho_1 = \rho.$$

We observe that the usual Richards curve coincides with the transformation $g(x) = x$, i.e. β and ρ remain constant with increasing x. The *extended* Richards curve is now defined as follows:

$$(11) \quad y = f(t) = f(g(x)) = \alpha(1 + s\beta\rho^{g(x)})^{\lambda}$$

with $g(x)$ a continuous function in x. With the aid of (8) and (9) we can determine c_i and m_i, $i=1,\ldots,k$ using the recursive equations:

$$(12) \quad c_i = (\ell n \beta_i - \ell n \beta_1)/\ell n \rho_1$$

$$m_i = \ell n \rho_i / \ell n \rho_1 .$$

In section 4 we consider a method whereby the initial parameters in (11) may be determined.

4. THE DETERMINATION OF THE INITIAL PARAMETERS FOR THE EXTENDED RICHARDS CURVE

A general procedure to determine the functional form of $g(x)$ is as follows:

Given n pairs of observations (x_i, y_i), $i=1,2,\ldots,n$, divide the data set into $k = n - m + 1$ overlapping subsets of size m:

Subset 1 $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$

Subset 2 $(x_2, y_2), (x_3, y_3), \ldots, (x_{m+1}, y_{m+1})$

$\vdots$

Subset k $(x_{n-m+1}, y_{n-m+1}), \ldots, (x_n, y_n)$

A Richards curve (with λ and α constant) is then fitted to each subset i. The values of $\hat{c}_i$ and $\hat{m}_i$ are calculated

by means of (12). If the mean value of the x's of each subset is denoted by $\bar{x}_i$, the following will hold approximately:

(13) $g(\bar{x}_i) \approx \hat{c}_i + \hat{m}_i \bar{x}_i$, $i = 1,\ldots,k$.

By inspection of the plotted points $(\bar{x}_i, g(\bar{x}_i))$, one can then choose a mathematical function $g(x)$. Note that $\bar{x}_1$ corresponds to the x_1 in the restriction of section 3.

General procedure

(a) Fit a Richards curve to the observed data to see whether it describes the data adequately.

(b) If the fit is not good enough, fit Richards curves to the subsets as described previously (α and λ as determined in (a)).

(c) Calculate the $k=n-m+1$ points $(\bar{x}_i, g(\bar{x}_i))$ using (12) and (13) and fit a curve $g(x)$ to these points.

(d) Substitute $g(x)$ into (11) and fit this curve to the whole data set as before.

5. THE APPLICATION OF THE EXTENDED RICHARDS FAMILY GROWTH CURVES TO OXYHAEMOGLOBIN SATURATION DATA

The measurement of sO_2 (y) and pO_2 (x) values in blood samples is standard practice in respiratory units all over the world. One such source of data ($n = 46$ cf. table 1) was reported by Severinghaus (1979) and is presently being used as a standard by the respiratory unit of Tygerberg Hospital.

The authors were requested to derive a mathematical model which would describe the data adequately. The requirements of such a model are:

(a) that it be invertible with respect to the variables sO_2 (y) and pO_2 (x);

(b) that it be implementable on a portable calculator for clinical use; and

(c) that the model parameters be physiologically meaningful.

In figure 2 the relationship between the sO_2 and pO_2 values given in table 1 is displayed graphically.

Initially we fitted the Richards curve to the data. The estimated value of the shape parameter λ was very large. Therefore a Gompertz curve was refitted to the data in table 1 (cf. (5), section 2). Using nonlinear least squares we found:

(14) $\hat{y} = 98{,}001 \exp(-4{,}60586\ (0{,}931615)^{x})$

Residual mean square = 0,5571

R^2 = 0,9995 (coefficient of determination).

This equation, however, yielded residuals ($\pm$ 1%) which render (14) unsuitable for clinical purposes despite the low residual mean square error. The data set was therefore divided into 41 subsets of size 6 each (see section 4):

Subset 1 Observations 1 to 6
Subset 2 Observations 2 to 7
⋮
Subset 41 Observations 41 to 46.

A Gompertz curve was fitted to each subset. A subset size of 6 was chosen because 6 observations are sufficient for the individual curve fitting while also ensuring that an adequate number of subsets are obtained. In table 2 the values of

β_i, ρ_i (see (8) and (9)), $\bar{x}_i$, the mean value of the x-values for subset i and corresponding $g(\bar{x}_i)$ (see (13)) and residual mean square are given.

From table 2 we conclude that:

(a) The residual mean square values indicate that a Gompertz curve adequately describes the variation in each subset.

(b) The $\hat{\beta}$ and $\hat{\rho}$ values vary systematically with increasing $\bar{x}$.

(c) At x-values near 20, the maximal rate of growth occurs, i.e. where ρ has its minimum value 0,927.

Conclusions (a) and (b) therefore suggest that a Gompertz curve with varying values of β and ρ will yield a very good fit over the region of x-values.

A Richards curve was fitted by means of nonlinear least squares to the data given in the last two columns of table 2 (n = 41). This yielded:

(15) $\hat{g}(\bar{x}) = 90,13\ (1-0,9389\ (0,98006)^{\bar{x}})^{1,51}$

Residual mean square = 0,1038

$R^2 = 0,9998$.

Instead of fitting a Gompertz curve we shall fit a modified curve as described by (14) and (15). Direct substitution of (14) into (15) leads to an expression containing, amongst others, two parameters in the form ρ^{α}. To avoid the ensuing indeterminacy in estimating these parameters we set $\rho^{\alpha} = \exp(-K_1)$. This leads to the modified curve (16) in which the parameters found in (14) and (15) were used as initial estimates in the least squares fitting:

(16) $\hat{y} = A \exp(-B \exp(-K(x)))^{K_3}$

where $K(x) = K_1(1-\exp(-K_2 x))$.

The following transformations were made to alleviate the computational burden:

$$C_1 = \exp(-B)$$
$$C_2 = \exp(-K_1)$$
$$C_3 = \exp(-K_2)$$
$$C_4 = K_3$$
$$C_5 = A.$$

The following curve was therefore fitted

$$(17) \quad y = C_5 C_1^{C_2 (1-C_3^x)^{C_4}}$$

Residual mean square = 0,0321

$R^2 = 1,0000$

where $C_1 = 0,0124904951$
$C_2 = 0,000740817624$
$C_3 = 0,984303206$
$C_4 = 1,280779869$
$C_5 = 99,607959.$

Note that a high number of significant digits have to be retained to ensure the desired degree of accuracy required for the present application.

The inverse of (16) is:

$$x = \ell n(1-K'(y)^{1/K_3})/K_2$$

where

$$(19) \quad K'(y) = \ell n\{-\ell n(y/A)/B\}K_1.$$

In figure 3 a plot of the residuals $y-\hat{y}$ vs. x is given. Severinghaus *op. cit.* states that interpolation was used between points. This could possibly account for the observed pattern in the residuals. The inverse relationship yields residuals within clinically acceptable limits.

6. CONCLUDING REMARK

The technique of fitting curves described in this paper was developed to solve a specific problem. However, the authors have found other applications for this technique and it is conceivable that researchers may find further applications for the methodology described here.

ACKNOWLEDGEMENT

This paper represents a portion of the research which the authors carried out in conjunction with drs R.I. Stewart and C.M. Lewis of the Departments of Medical Biochemistry and Internal Medicine respectively, Medical School, University of Stellenbosch, Tygerberg, RSA.

The authors are also grateful to prof. J.S. Maritz of La Trobe University, Australia and prof. M.W. Browne of the University of South Africa for their valuable comments and suggestions for improving the manuscript.

TABLE 1. RELATIONSHIP BETWEEN OXYGEN SATURATION (sO_2) AND OXYGEN TENSION (pO_2)

pO_2 (x)	2	4	6	8	10	12	14	16	18	20
sO_2 (y)	1,19	2,56	4,37	6,68	9,58	12,96	16,89	21,40	26,50	32,12
pO_2 (x)	22	24	26	28	30	32	34	36	38	40
sO_2 (y)	37,60	43,14	48,27	53,16	57,54	61,69	65,16	68,63	71,94	74,69
pO_2 (x)	42	44	46	48	50	52	54	56	58	60
sO_2 (y)	77,29	79,55	81,71	83,52	85,08	86,59	87,70	88,93	89,95	90,85
pO_2 (x)	65	70	75	80	85	90	95	100	110	120
sO_2 (y)	92,73	94,06	95,10	95,84	96,42	96,88	97,25	97,49	97,91	98,21
pO_2 (x)	130	140	150	175	200	225	250			
sO_2 (y)	98,44	98,62	98,77	99,03	99,20	99,32	99,41			

TABLE 2. ESTIMATED VALUES FOR THE DETERMINATION OF A SET OF TANGENT EQUATIONS

β_i	ρ_i	Residual mean square	$\bar{x}_i$	$g(\bar{x}_i)$
0,00480	0,93148	0,011	9,00	9,00
0,00476	0,93200	0,003	11,00	11,01
0,00480	0,93133	0,013	13,00	13,02
0,00492	0,92971	0,053	15,00	15,05
0,00504	0,92836	0,040	17,00	17,11
0,00514	0,92738	0,014	19,00	19,21
0,00514	0,92744	0,016	21,00	21,31
0,00503	0,92837	0,035	23,00	23,40
0,00481	0,93017	0,049	25,00	25,46
0,00460	0,93181	0,053	27,00	27,47
0,00427	0,93426	0,066	29,00	29,41
0,00403	0,93610	0,049	31,00	31,30
0,00387	0,93726	0,016	33,00	33,16
0,00376	0,93807	0,012	35,00	34,97
0,00367	0,93867	0,009	37,00	36,77
0,00365	0,93887	0,014	39,00	38,52
0,00350	0,93982	0,009	41,00	40,29
0,00330	0,94111	0,002	43,00	42,02
0,00317	0,94206	0,009	45,00	43,69
0,00299	0,94327	0,007	47,00	45,35
0,00271	0,94525	0,017	49,00	46,94
0,00245	0,94715	0,007	51,00	48,50
0,00230	0,94827	0,004	53,00	50,02
0,00218	0,94919	0,004	55,00	51,49
0,00198	0,95087	0,004	57,50	53,28
0,00178	0,95266	0,015	60,50	55,30
0,00150	0,95538	0,014	64,00	57,57
0,00127	0,95781	0,023	68,00	59,97
0,00106	0,96043	0,031	72,50	62,49
0,00077	0,96463	0,017	77,50	65,08
0,00058	0,96815	0,011	82,50	67,40
0,00042	0,97196	0,007	87,50	69,46
0,00029	0,97585	0,009	93,33	71,50
0,00021	0,97952	0,007	100,00	73,51
0,00016	0,98224	0,004	107,50	75,48
0,00013	0,98409	0,002	115,83	77,41
0,00011	0,98530	0,002	125,00	79,33
0,00008	0,98758	0,002	137,50	81,59
0,00006	0,98946	0,002	152,50	83,87
0,00005	0,99084	0,002	170,00	86,16
0,00004	0,99186	0,001	190,00	88,46

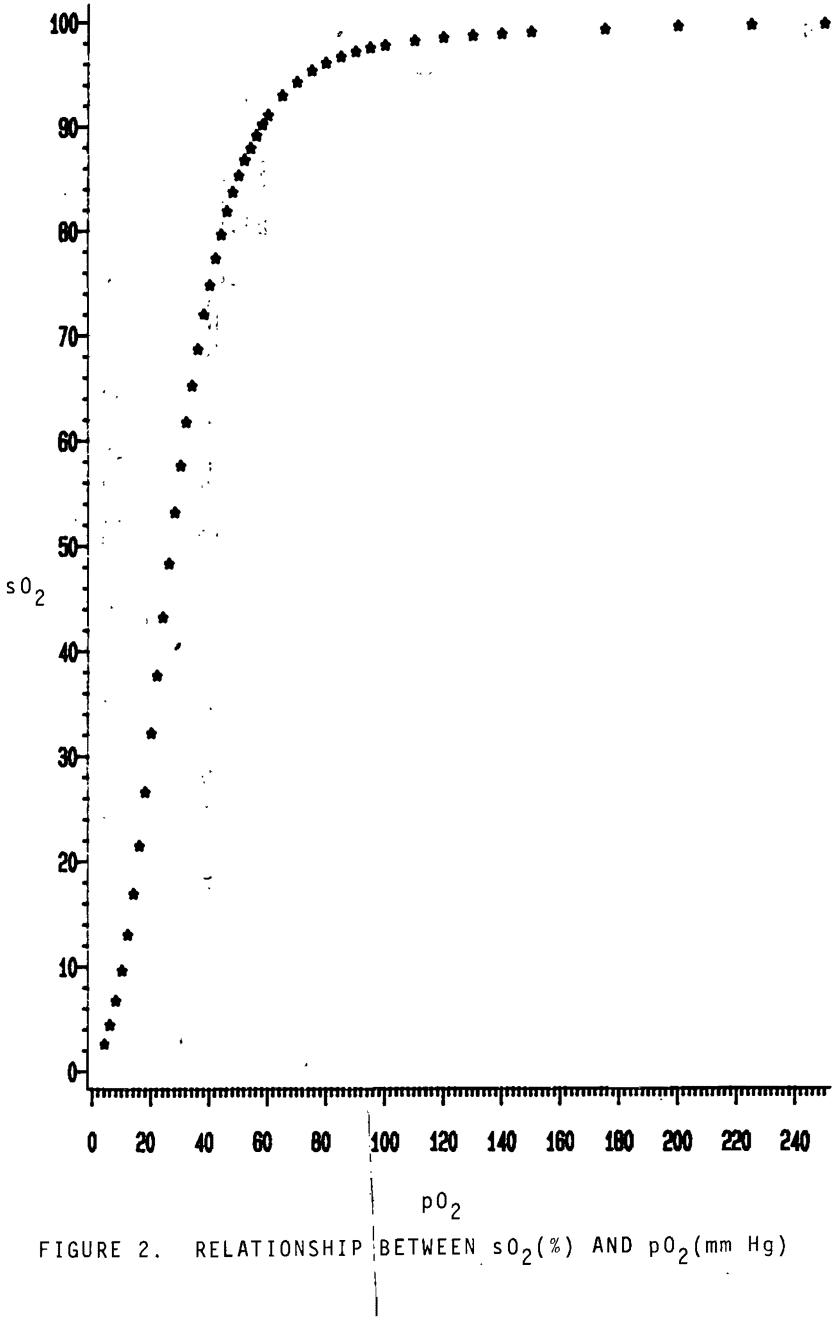

FIGURE 2. RELATIONSHIP BETWEEN sO_2(%) AND pO_2(mm Hg)

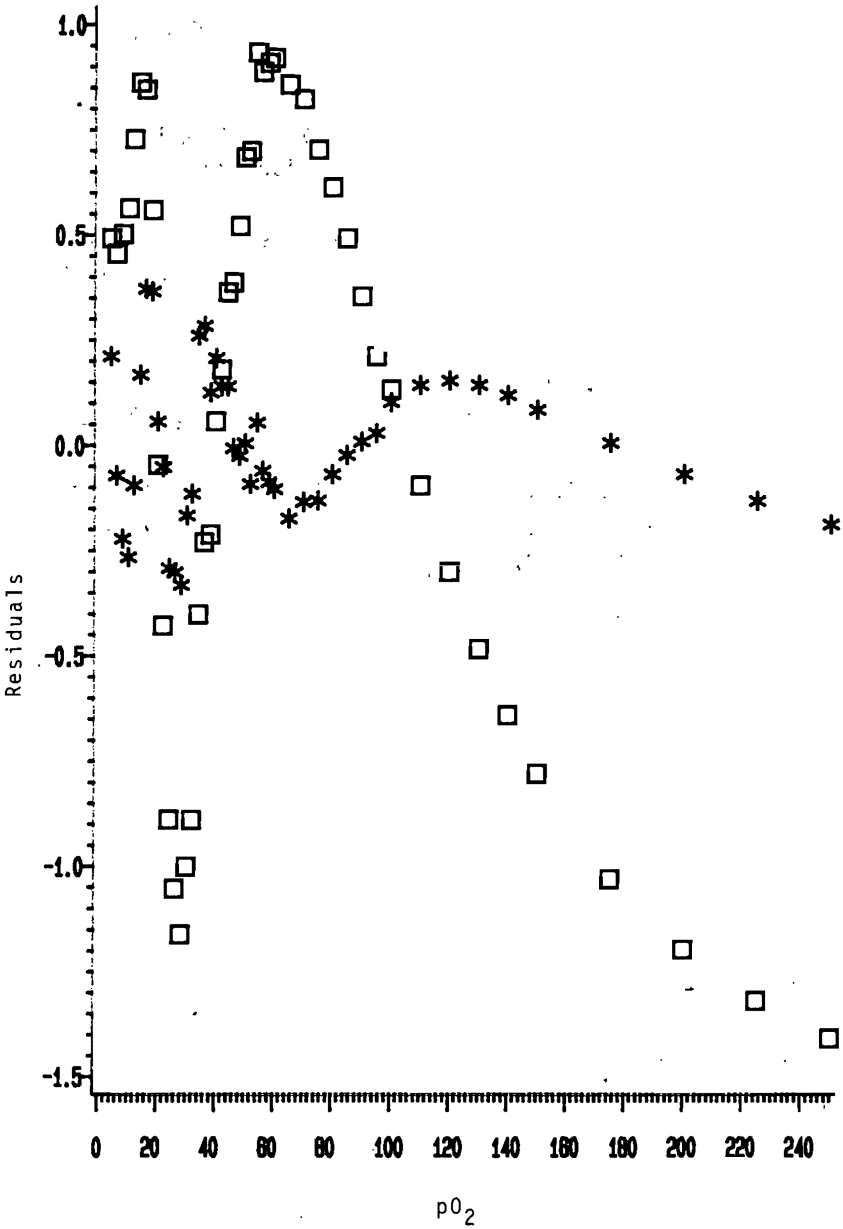

FIGURE 3. y (sO_2) residuals vs x (pO_2)

REFERENCES

DU TOIT, S.H.C. (1979). *The analysis of growth curves.* Ph.D. Thesis, University of South Africa.

GILL, P.E. and MURRAY, W. (1978). Algorithms for the solution of the nonlinear least-squares problem. *SIAM J. Numer. Anal.*, **15**, 977-992.

RICHARDS, F.J. (1959). A flexible growth function for empirical use. *Journ. of Experimental Botany*, **10**, 290-300.

SEVERINGHAUS, J.W. (1979). Simple, accurate equations for human blood O_2 dissociation computations. *J. Appl. Physiol.*, **46**, 599-602.

Manuscript received 1982.09.29, revised 1983.03.04, 1984.09.19.