REGRANS: A "BASIC" PROGRAM FOR AN EXTENSIVE ANALYSIS OF RELATIVE GROWTH

PAULO RICARDO PEZZUTO
Departamento de Oceanografia - Universidade do Rio Grande,
Caixa Postal 474 - CEP 96201-900 - Rio Grande - RS - Brasil

ABSTRACT

REGRANS - Regression Analysis - was developed as a tool to deal with bivariate relative growth studies. The program calculates linear predictive and functional regressions up to 1000 pairs of data and works with 8 non-linear models. It also comprises 11 statistical tests and an iterative routine that detects transition points on the data, corresponding to changes on the relative growth rates. REGRANS provides three statistics that allow direct comparisons amongst the fits obtained by each model and method of regression. The use of REGRANS results in a significant reduction on the subjectivity of the results and on the effort spent on the analysis. The Portuguese or English compiled version of REGRANS is provided on a floppy disk with a README file containing the "user's guide", and a SAMPLE file to test the program. The author would appreciate if requests of copies are sent along with the respective 5 1/4" or 3 1/2" disk. Requests may also be done to the editor of Atlântica.

KEY WORDS: Relative growth; allometry; software.

RESUMO

Recentemente, e sobretudo entre os carcinologistas, tem-se questionado a utilização do modelo potencial ($Y=aX^b$) no estudo do crescimento relativo, visto não haver qualquer justificativa biológica para sua adoção indiscriminada. Estatisticamente, as múltiplas formas de abordar a questão e ainda a variedade de testes aplicáveis às regressões demandariam do pesquisador tempo em demasia e utilização seguida de vários programas para se efetuar uma análise mais acurada dos dados. Visando a sanar tais problemas, elaborou-se o presente programa, que incorpora diversas técnicas de análise bivariada do crescimento relativo. A partir da leitura de arquivos ou de dados fornecidos via teclado, são calculadas regressões lineares preditivas e funcionais para até 1000 observações. Como normalmente se verificam alterações nos níveis de alometria ao longo da ontogenia representadas por pontos de inflexão ou descontinuidades nas retas, o programa faz a determinação estatística dos mesmos através de um processo iterativo, e apresenta um teste para avaliar sua significância, juntamente com mais 10 testes relacionados às regressões e correlações. Estão disponíveis rotinas de transformações para mais sete modelos não-lineares além do potencial, possibilitando o cálculo posterior das respectivas regressões. A aplicação do programa, além de minimizar o esforço despendido no trabalho, vem reduzir ou mesmo eliminar o elevado grau de subjetividade dos resultados que normalmente é observado na literatura. A versão compilada em português ou inglês do REGRANS é fornecida em disquete, juntamente com um arquivo LEIA-ME contendo o "guia do usuário" e um arquivo de dados para teste do programa. Solicita-se que pedidos de cópias do mesmo sejam acompanhadas do respectivo disquete de 5 1/4" ou 3 1/2".

PALAVRAS-CHAVE: crescimento relativo; alometria.

1 - INTRODUCTION

The potential equation $Y=aX^b$ has been used to describe the relationship between measurements of organs or body parts in most of animal and vegetal taxa, and especially in crustacea. The **b** parameter is the "equilibrium constant" (Teissier, 1960) or allometric growth constant (Hartnoll, 1982), and **a** is the "initial growth index" (Teissier, 1960), X and Y corresponding respectively to the reference and the studied dimensions. Through logarithmization of X and Y values, the expression assumes the simple linear form logY= log a + b logX, allowing the **a** and **b** parameters to be estimated by a minimum squares regression analysis.

This practice has been used as a standard method for relative growth studies since the paper of Huxley in 1932 (Jolicoeur, 1963). However, the following methodological problems associated to this technique are not solved yet:

1 - The indiscriminated choice of the potential model itself is a matter of doubt, because there is no biological or theoretical justification to apply it *a priori*, in detriment to any other function (Lovett & Felder, 1989). Hartnoll (1982) pointed out that its application "is only empirical and is not based on any understanding of the mechanism which controls the relative growth".

2 - According to Zar (1968), the logarithmic transformation of the data, though mathematically correct to linearize the equation, could lead only to an approximation of the real values of **a** and **b**, which

sometimes could be very poor.

3 - The minimum squares method (or predictive regression) assumes that the independent variable is measured without errors, a condition somewhat difficult to fulfill in morphometric studies (Finney & Abelle, 1981). When it is not possible to choose between two variables which one should be considered independent (Clarke, 1980), it is recommended the use of the Model II of regression (Sokal & Rohlf, 1969) as the Functional Regression or the Reduced Major Axis (Ricker, 1973, 1975) and the Major Axis (Jolicoeur, 1963, 1975).

4 - The allometric growth coefficient for one specific variable often does not remain constant during the ontogeny, especially in crustacea. The changes in growth can be observed as discontinuities or inflection points in determined sites of the graphs, which may reflect "critical molts", corresponding to the pre-puberal, puberal and mature stages of the species' life cycle (Teissier, 1960; Hartnoll, 1978). However, these points may be so tenuous that their determination is prevented, resulting in a great degree of subjectivity in the location of the respective curves and in the estimation of their parameters (see Haley, 1973; Du Preez & McLachlan, 1984; Haefner Jr., 1985; Rodrigues, 1985), since "the inclusion of a few inappropriate data points can significantly alter the slope and level of allometry exhibited by the measured variable" (Clayton & Snowden, 1991).

Lovett & Felder (1989) discussed these questions extensively. Through the comparative application of various regression techniques to the study of the relative growth of chelae in *Callianassa louisianensis* (Decapoda, Thalassinidea), these authors concluded that it is necessary some statistical austerity to determine accurately the model that provides the best description of the relationship between the selected variables. However, a more complete analysis (as suggested by them) demands extensive periods of time and effort to be performed, especially due to the absence of a statistical package that encloses all the specific functions that could make a thorough analysis possible.

The REGRANS system was developed as a tool to minimize the subjectivity in the studies of relative growth, attaining reliable results in a small period of time.

2 - THE PROGRAM

REGRANS was originally written in BASIC language and consists of 5 submenus (Fig. 1), interconnected by a main menu with capacity to process up to 1000 pairs of data, which can be loaded from a file, or from the keyboard.

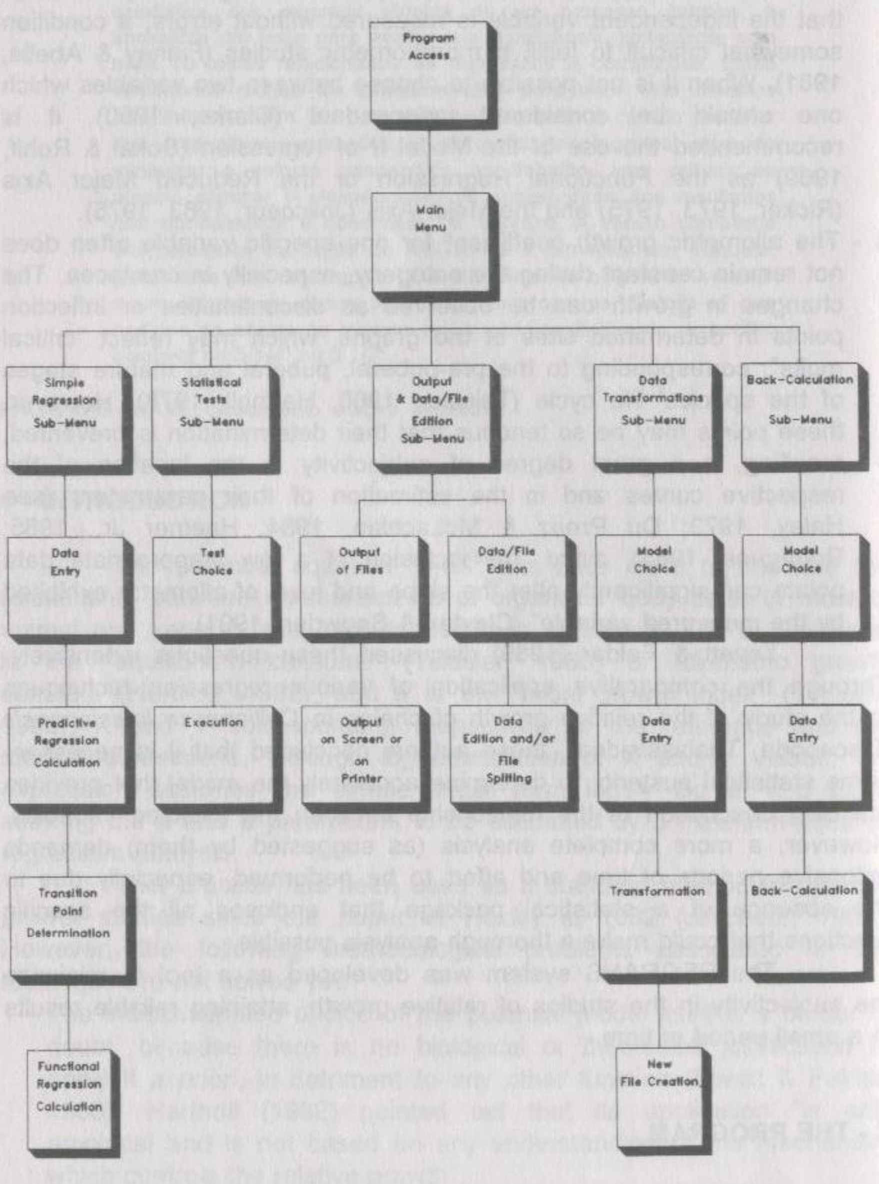

FIGURE 1 – Simplified flowchart of the REGRANS system, showing its structure based on 5 submenus and their main options and steps.

On the first submenu (Simple Regression), the program calculates predictive regressions. Functional Regressions are also available. In addition to the standard statistics provided by predictive regressions (intercept, slope and its standard error and significance, and the correlation and determination coefficients), the program calculates three sums for the predictive and functional methods, which consist in one of the main ways to evaluate the fit of them to the data, besides the normality of the residuals. They are the sum of squared residuals on Y ($SSRy = \Sigma (y-\hat{y})^2$), the sum of squared residuals on X ($SSRx = \Sigma (x-\hat{x})^2$), and the sum of the products of the residuals on X and Y ($SPRxy = \Sigma (y-\hat{y}) (x-\hat{x})$).

The minimum squares method provides the line that minimizes the SSRy (Sokal & Rohlf, 1969), as it assumes the inexistence of error on the X measurements. Nevertheless, functional regressions tend to minimize the SPRxy, considering the errors on the two variables.

The coefficient of determination (R^2) and the t-statistic have been conventionally used as measures of fitness for predictive regressions but, as pointed out by Chatterjee & Price (1977), their high values do not necessarily imply a good fit between variables (see Fig. 1.1 and Tab. 1.1 of Chatterjee & Price, 1977). Moreover, the R^2 and t parameters cannot be used as indicative to choose between predictive and functional regressions. On the other hand, the difficulty of choosing the best expression does not occur with the three sums, because they are calculated for the two methods, allowing direct comparisons between them.

All the regressions residuals calculated by REGRANS can be stored into files and the normality of these residuals can be checked through a specific test enclosed on the second submenu (see below). The normality consists in an alternative measure of fit, because it is a basic assumption of the regression models (Sokal & Rohlf, 1969; Chatterjee & Price, 1977; Lovett & Felder, 1989).

One of the most important advantages presented in REGRANS is the adaptation of a routine idealized by Yeager & Ultsch (1989), called Iterative Process for the Determination of the Transition Point in this study. This routine helps to find objectively the point where the data set can be divided in two subsets, looking for the best fit in both. The program sorts the data in ascending order of X values, and calculates a predictive regression with the first three pairs of data and another one with the remaining, storing the resulting sum of the two SSRy. Then, REGRANS takes the first four pairs of data and proceeds in the same way. This process is repeated step by step until it recognizes the last but two pairs of data. Afterwards, REGRANS looks for the lowest calculated sum of SSRy among all, and assumes the corresponding X value as the transition point. It provides all the statistics for both lines, and gives to the user the option of saving the respective residuals. Lovett & Felder (1989) used a similar process to determine the transition points. However, this process was not executed point to point as above, but at fixed intervals of X values. As pointed out by Nickerson et al. (1989), this method cannot provide the best fit, because the intervals in which the data are splited are arbitrarily determined, although it probably also results on a valid description of data.

Assuming that the data display more than one transition point, the program allows the division of the file to remove one of the two determined segments, followed by another run of the Iterative Process. This function, as well as options to edit the data and to print the files, can be done by the third submenu.

All regressions are calculated in the linear form. However, REGRANS can perform the corresponding data transformations for seven non-linear models besides the potential model, through the fourth submenu, enhancing its applicability as a general package for regressions analysis. The seven models are: Exponential I ($Y=ab^X$), Exponential II ($Y=ae^{bX}$), Exponential III ($Y=e^{a+bX}$), Exponential IV ($Y=aXe^{-bX}$), Exponential V ($Y=1-e^{-aX^{\wedge}b}$), Hyperbolic (reciprocal) ($1/Y=a+bX$) and Parabolic ($Y=aX-bX^2$).

When a choice between two models (e.g. linear and potential) has to be made, the type of parameter to be used as measure of fit constitutes a hard task, because the statistics frequently used do not often reflect the real fit of the model in many cases, since they can be affected by the transformations or by the number of observations, amongst other factors. The sums calculated by the program also present this problem, at least for the transformations, because the residuals correspond to the distances from each point to their respective estimated values, thus reflecting the measurement scale of the studied variables.

Hence, the SSRy, SSRx and SPRxy calculated for the non-linear models are in a corresponding scale and magnitude to the transformed data, which prevents comparisons either between them or with the linear sums and residuals. Lovett & Felder (1989) have proposed an efficient and simple method to solve this problem, which was incorporated to REGRANS as its fifth submenu.

Taking into account that the data transformations only search for the estimation of **a** and **b** parameters of the original models through the linear regressions, REGRANS requests the user to inform which non-linear model is adopted, the non-transformed values of X and Y, and the **a** and **b** figures estimated by the regression calculated with the

linearized data. Then, the two SSR, SPRxy and the residuals in the original data measurement scale will be back-calculated, allowing a direct comparison between the fit obtained by each model, through these standardized sums and the normality of the deviations. This routine is available to the eight non-linear models performed by REGRANS.

Finally, the statistical tests listed below were included on the second submenu, which intends to cover all the functions performed by the program: a) comparison between the slopes of predictive (Sokal & Rohlf,1969) and functional regressions (Clarke, 1980); b) significance test for predictive regression slope (Sokal & Rohlf, 1969); c) confidence intervals for predictive (Sokal & Rohlf, 1969) and functional slopes (Ricker, 1975); d) significance test for the coefficient of correlation (Sokal & Rohlf, 1969); e) comparison of the coefficient of correlation against another value (Sokal & Rohlf, 1969); f) comparison between two coefficients of correlation (Sokal & Rohlf, 1969); g) K-S test for normality of residuals (Siegel, 1975) - it applies a Kolmogorov-Smirnov test to the residuals, comparing their distribution to a theoretical normal distribution; and h) covariance analysis (Snedecor & Cochran, 1971) - it tests up to six lines with up to 1000 pairs of data each one for equality on slopes and intercepts.

The "test for the total and combined SSRy" is also available in REGRANS, which compares the SSRy obtained for the data as a whole, against the sum of SSRy values for the lines at the left side and at the right side of the transition point (Somerton, 1980). As emphasized by Yeager & Ultsch (1989), there is a very frequent possibility of splitting a set of data with a resultant sum of SSRy lower than the SSRy for the whole range of them. However, because this difference may not be statistically significant, it is recommended the application of this test before concluding about the existence of the transition point indicated by the Iterative Process.

3 - FINAL REMARKS

In relation to the commonly available regression programs that may be used to perform relative growth or similar studies, REGRANS offers the following advantages:

1 - Through the analysis of the three sums (SSRy, SSRx and SPRxy), the user can validate his choice between the predictive and functional methods by the comparative analysis of the respective fit to the data, and not only by their theoretical assumptions.
2 - Reduction or elimination of the subjectivity in the determination of the

transition point through the Iterative Process, the test for differences between the total and combined SSRy, and the Covariance Analysis.

3 - Possibility of multiple comparisons amongst the linear and non-linear models, by the direct analysis of the three sums.

4 - Variety of non-linear models.

5 - Disponibility of statistical tests, including those for functional regressions, which often are not included in popular statistical programs.

6 - Facility of use and data exchange with worksheets and graphics packages, through ASCII files.

REGRANS' best characteristic is that it has been developed with the main purpose of including in a single package most of the statistical techniques potentially useful to bivariate studies of relative growth, preventing the application of several softwares to perform an accurate analysis.

REGRANS can be loaded on any IBM-PC or compatible microcomputer. The Portuguese or English compiled version of the program is provided on a floppy disk, with a README file that contains a "user's guide" explaining all steps that the user must perform to obtain a complete analysis of the data. The author would appreciate if requests of REGRANS copies are sent along with the respective 5 1/4" or 3 1/2" disk. Requests may also be done to the editor of Atlântica.

The routines and results provided by REGRANS were exhaustively checked against other statistical and worksheet packages. However, this does not imply any warranty by the author and Atlântica. Please, in case you find any problem in the program, write to the author specifying the details and steps of the analysis.

Keeping the authorship, REGRANS may be freely copied and distributed for scientific purposes, and is provided without any charges. By no means, anyone is authorized to charge for REGRANS.

AKNOWLEDGMENTS

I would like to thank to Dr. E. A. Santos Filho, A. G. Soares, P. R. Pagliosa Alves, P. J. P. Santos and T. L. Almeida by the valuable suggestions, criticism and bibliographic support, and to E.G.Vidal and Dr. C. E. Bemvenuti by the reading of the manuscript.

I am also indebted to E. S. Alves for her continuous assistance and encouragement, to A. B. Greig, A. Leite and especially to Dr. E. G. Reis for the help on the English versions of the program and manuscript.

This work was supported by a graduate research scholarship of the Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq).

REFERENCES

CHATTERJEE, S. & PRICE, B. 1977. Regression analysis by example. New York: John Wiley & Sons. 228 p.

CLARKE, M. R. B. 1980. The reduced major axis of a bivariate sample. *Biometrika*, 67(2):441-466.

CLAYTON, D. A. & SNOWDEN, R. J. 1991. Allometric growth in *Ilyoplax stevensi* (Decapoda, Brachyura, Ocypodidae). *Crustaceana*, 61(1):1-10.

DU PREEZ, H. H. & McLACHLAN, A. 1984. Biology of the three-spot swimming crab, *Ovalipes punctatus* (De Haan). I. Morphometrics and relative growth (Decapoda, Portunidae). *Crustaceana*, 47(1):72-82.

FINNEY, W. C. & ABELE, L. G. 1981. Allometric variation and sexual maturity in the obligate coral commensal *Trapezia ferruginea* Latreille (Decapoda, Xanthidae). *Crustaceana*, 41(2):113-130.

HAEFNER JR., P. A. 1985. Morphometry, reproduction, diet and epizoites of *Ovalipes stephensoni* Williams, 1976 (Decapoda, Brachyura). *J. Crustacean Biol.*, 5(4):658-672.

HALEY, S.R. 1973. On the use of morphometric data as a guide to reproductive maturity in the ghost crab, *Ocypode ceratophthalmus* (Pallas) (Brachyura, Ocypodidae). *Pacif. Sci.*, 27(4):350-362.

HARTNOLL, R. G. 1978. The determination of relative growth in crustacea. *Crustaceana*, 34(3):281-293.

________. 1982. Growth. In: ABELE, L. G. (ed.). The biology of Crustacea. Embryology, morphology, and genetics. New York: Academic Press. v. 2. p. 111-146.

JOLICOEUR, P. 1963. The multivariate generalization of the allometry equation. *Biometrics*, 19:497-499.

________. 1975. Linear regressions in fishery research: some comments. *J. Fish. Res. Bd. Can.*, 32:1491-1494.

LOVETT, D. L. & FELDER, D. L. 1989. Application of regression techniques to studies of relative growth in crustaceans. *J. Crustacean Biol.*, 9(4):529-539.

NICKERSON, D. M.; FACEY, D. E. & GROSSMAN, G. D. 1989. Estimating physiological thresholds with continuous two-phase regression. *Physiol. Zool.*, 62(4):866-887.

RICKER, W. E. 1973. Linear regressions in fishery research. *J. Fish. Res. Bd. Can.*, 30:409-434.

________. 1975. A note concerning Professor Jolicoeur's comments. *J. Fish. Res. Bd. Can.*, 32:1494-1498.

RODRIGUES, S. A. 1985. Sobre o crescimento relativo de *Callichirus major* (Say, 1818) (Crustacea, Decapoda, Thalassinidea). *Bolm. Zool.*, Univ. S. Paulo, 9:195-211.

SIEGEL, S. 1975. Estatística não-paramétrica (para as ciências do comportamento). McGraw-Hill. 350p.

SNEDECOR, G. W. & COCHRAN, W. G. 1971. Métodos Estadísticos. Mexico: Compañía Editorial Continental. 703 p.

SOKAL, R. R. & ROHLF, F. J. 1969. Biometry. The principles and practice of statistics in biological research. San Francisco: W. H. Freeman. 776 p.

SOMERTON, D. A. 1980. A computer technique for estimating the size of sexual maturity in crabs. *Can. J. Fish. Aquat. Sci.*, 37:1488-1494.

TEISSIER, G. 1960. Relative Growth. In: WATERMAN, T. H. (ed.). The physiology of Crustacea. New York: Academic Press. p. 537-560.

YEAGER, D. P. & ULTSCH, G. R. 1989. Physiological regulation and conformation: a BASIC program for the determination of critical points. *Physiol. Zool.*, 62(4):888-907.

ZAR, J. H. 1968. Calculation and miscalculation of the allometric equation as a model in biological data. *Bioscience*, 18:1118-1120.

APPENDIX

A simple application of the REGRANS, built on an hypothetic row of 198 measurements, is presented below. These data are included on REGRANS disk as a SAMPLE file allowing the user to test the program and to become familiar with its routines.

The following example of analysis should not be understood as a rule or a new proposal of methodology, but it is only a simplification of the method used by the author, when working on his relative growth data.

Through the steps of this example, the user will look for a transition point on the variables (corresponding to a change of the growth pattern), testing its statistical significance, and will also determine the best model (e.g. linear or potential) and method (predictive or functional) of regression, according to their fitting to the data.

For simplicity, the linear and potential models only are compared, and the SPRxy is chosen as fit measure, ignoring the normality of the residuals.

The example is organized in steps of analysis, which should be observed along with Figure 1.

Step 1. On the 1.st submenu, the original data are read and the predictive regression calculation for the whole set is performed.

Step 2. The user runs the Iterative Process and REGRANS searches the transition point, splitting the original line into two segments.

Step 3. Through the input of the slopes and correlation coefficients obtained for the three lines above, the program calculates the corresponding functional regressions.

Step 4. The first submenu is closed, and the data are log-transformed on the 4.th submenu, with the creation of a new linearized sample file.

Step 5. The steps 1 to 3 are run again, but with the transformed data.

Although the results for all steps performed above are summarized on the Table 1, it is worth mentioning that the corresponding statistics for the regressions are provided by the program after each run.

The acceptance of the transition point must follow three conditions, as explained below:

Step 6. The significance of the transition points should be checked on the 2.nd submenu, with the "Test for the Total and Combined SSRy" (see text for explanation). For the two models of the example, the splitting of the data in two segments improved the fit by reducing significantly the SSRy around the lines. (For both models, F=73.20;

TABLE 1 - Summary of regression statistics calculated by REGRANS for the sample data. P = Predictive Regression. F = Functional Regression. The intercepts of the potential lines are showed as LOG a. Note the difference among the SSRs and SPRxy for the potential lines in respect of the linear ones, due to the lower scale of measurement of the former.

MODEL	SEGMENT	METHOD	INTERCEPT	SLOPE(+-SE)	r	R2	t	DF	SSRy	SSRx	SPRxy
	TOTAL	P	0.07	0.49(0.01)	0.94	0.89	39.09	196	239.85	1008.26	-491.76
		F	-0.64	0.52					247.08	1420.20	-491.76
LINEAR	LEFT	P	-2.26	0.61(0.02)	0.98	0.95	26.88	35	8.27	22.16	-13.54
	(x≤18.3)	F	-2.45	0.63					8.37	25.60	-13.54
	RIGHT	P	4.58	0.32(0.02)	0.81	0.66	17.50	159	128.41	1269.08	-377.38
	(x≥18.3)	F	2.69	0.39					141.79	6971.33	-403.69
	TOTAL	P	-0.54	1.17(0.02)	0.96	0.92	48.88	196	0.48	0.35	-0.41
		F	-0.60	1.21					0.49	0.93	-0.41
POTENTIAL	LEFTH	P	-0.84	1.42(0.05)	0.98	0.96	27.94	35	0.05	0.02	-0.04
	(x≤18.3)	F	-0.87	1.45					0.05	0.05	-0.04
	RIGHT	P	0.21	0.64(0.03)	0.82	0.67	18.03	159	0.15	0.38	-0.23
	(x≥18.3)	F	0.01	0.78					0.17	15.76	-0.24

P<0.001; 2,194 DF).

Step 7. Analysis of the predictive slopes significance through the t-values (Table 1). This process is very important because it is possible, for example, that one of the two lines has a slope not different from zero, invalidating the transition point.

Step 8. Through the 3.rd submenu, the original and the linearized files must be divided on the transition point with the creation of four new files, each one corresponding to the left or right lines of the linear and potential models, allowing the comparison between their slopes and intercepts in the step 9.

Step 9. For the predictive regressions, the 2.nd submenu performs an ANCOVA and also a comparison between the functional slopes, which intends to test if the two segments are significantly different for each method and model, and to definitively confirm the existence of a transition point, as can be seen on the results for the sample data, showed on Table 2.

TABLE 2 — Comparison between intercepts and slopes of the left and right lines determined by the Iteractive Process for the sample data.

			MODEL	
METHOD	PARAMETER		LINEAR	POTENTIAL
		F	80.50	35.93
	INTERCEPTS	P	<0.001	<0.001
		DF	1;195	1;195
PREDICTIVE		F	46.95	190.43
	SLOPES	P	<0.001	<0.001
		DF	1;194	1;194
		t	7.99	10.96
FUNCTIONAL	SLOPES	P	<0.001	<0.001
		DF	76	77

Step 10. Although other tests can be performed such as calculation of confidence intervals for the slopes through the 2.nd submenu, for the sake of simplicity they are not showed here.

Step 11. Through the 5.th submenu, the user informs the untransformed data file and the **a** and **b** values of all lines of the potential model (predictive and functional). The program back-calculates the SSRs

and SPRxy in the original scale of measurement for this model (Table 3).

Step 12. As the functional method minimizes the SPRxy, the comparison between the SPRxy calculated by the functional and predictive regressions for each model is made; e.g. on Table 3, it can be seen that the functional regression fits better the data for both lines (left and right) in the potential model than the minimum squares ones. Otherwise, the predictive regressions in the linear model fit better than the functional ones (Table 1).

Step 13. The most suitable fits of each model are compared and the user can finally decide by the adoption of the best of them. In this case, the potential model adjusted by functional regression provided the best fit by accurately describing the two segments of the sample data, as can be seen comparing Tables 1 and 3.

TABLE 3 – Back-calculated SSRs and SPRxy for the potential regressions of the sample data. P = Predictive Regression, F = Functional Regression.

SEGMENT	METHOD	SSRy	SSRx	SPRxy
TOTAL	P	301.86	924.52	- 527.97
	F	335.75	936.64	- 560.38
LEFT	P	7.82	20.14	- 12.46
	F	7.76	19.11	- 12.08
RIGHT	P	127.68	1273.76	- 402.53
	F	140.79	934.89	- 362.65

It is necessary to stress that this process is unique among several types of analysis liable to be performed by the program and that can be done according to the necessity and methodology of the user. Moreover, REGRANS can be used for many applications other than relative growth studies, becoming this example only a simple illustrative exposition of some routines of the REGRANS system.