Journal of
APPLIED
BIOMEDICINE

J. Appl. Biomed.
4: 87–94, 2006
ISSN 1214-0287

ORIGINAL ARTICLE

GOSA, a simulated annealing-based program for global optimization of nonlinear problems, also reveals transyears

Jerzy Czaplicki[1], Germaine Cornélissen[2] and Franz Halberg[2]

[1]Institute of Pharmacology and Structural Biology, CNRS UMR 5089, Toulouse, France
[2]Halberg Chronobiology Center, University of Minnesota, Minneapolis, MN, USA

Received 12th January 2006.
Revised 14th March 2006.
Published online 10th April 2006.

Summary
Transyears in biology have been documented thus far by the extended cosinor approach, including linear-nonlinear rhythmometry. We here confirm the existence of transyears by simulated annealing, a method originally developed for a much broader use, but described and introduced herein for validating its application to time series. The method is illustrated both on an artificial test case with known components and on biological data. We provide a table comparing results by the two methods and trust that the procedure will serve the budding sciences of chronobiology (the study of mechanisms underlying biological time structure), chronomics (the mapping of time structures in and around us), and chronobioethics, using the foregoing disciplines to add to concern for illnesses of individuals also a budding focus on diseases of societies, like crime, and of nations and civilizations, like war.

Key words: simulated annealing – linear-nonlinear rhythmometry – transyears – cosinor

INTRODUCTION

Fitting multidimensional experimental data to complex multi-parameter nonlinear models presents a host of serious problems, the most important of which is finding a set of starting values for the searched parameters. This requirement is necessary in the case of deterministic minimization techniques, based on calculation of gradients of the target function with respect to the adjusted parameters. In this context, the target function is a sum of squared deviations between the measured points and those calculated on the basis of a selected theoretical model. The objective is to find the global minimum of the error function. The use of deterministic search techniques, such as the steepest descent, conjugate gradients or the Newton-Raphson method (Press et al., 1992) is widespread mainly because of their efficiency, i.e. rapid convergence to the minimum. The necessity to provide the starting point at the beginning of the calculation, i.e. concrete values for all model parameters, however, is a serious obstacle.

In the multi-dimensional parameter space, the landscape of the minimized function may be very complex. Since the deterministic algorithms use the sign and the value of the target function's gradients at any given point to determine the direction of the search and the size of the step in that direction, they are bound to follow the path, which descends from the starting point to the nearest minimum. There is

✉ Jerzy Czaplicki, Université Paul Sabatier, Institut de Pharmacologie et Biologie Structurale CNRS, 205, route de Narbonne, 31077 Toulouse, France
☎ 🖷 +33 (0)5 61 17 54 04
💻 cgeorge@ipbs.fr

no guarantee that the minimum reached corresponds to the global minimum of the minimized function. These algorithms are not capable of jumping over a barrier, resulting from the function's profile, to find a deeper minimum.

One way to overcome this obstacle is to run the minimization procedure several times, each time with a different starting point, and to retain the best result, corresponding to the smallest value of the target function. By increasing the number of runs with different starting points, we increase our confidence that the best fit will approach the global minimum (within the limits of a predetermined accuracy). We can never be certain, however, that the data set cannot be described more satisfactorily by a different combination of parameters. Moreover, this procedure increases significantly the overall time necessary to accomplish the task, which greatly reduces the appeal of this approach in the multidimensional case.

Unfortunately, when working with high-dimensional experimental data, it is nearly impossible to find initial values by a trial-and-error method. The correct solution to the problem can be found only if the starting point is close enough to the minimum, which means that one must have some *a priori* knowledge of at least the order of magnitude of the parameters involved. This prerequisite, however, is contrary to the reason for which we do the fitting; we prefer to obtain the results without having to guess anything. Therefore, we decided to develop an alternative, or rather a complement to the currently existing software packages: an easy-to-use program, based on simulated annealing, which is a stochastic minimization algorithm devoid of the drawbacks of the deterministic algorithms. It should be noted that currently many efforts are underway to achieve global optimization by deterministic algorithms. At present, this approach requires the use of auxiliary functions, whose form depends on the nature of the investigated problem. This lack of generality prevented us from opting for this solution: our software was designed to deal with a broad class of problems.

MATERIALS AND METHODS

Most of the commercially available programs are useful for analyses of one-dimensional cases. In practice, we often deal with multi-dimensional studies, where several independent variables determine the outcome. In such cases a number of curves are obtained, a circumstance which increases the number of unknowns and the time required to find the best fit to the experimental data. Taking into account all these factors, we chose to work with stochastic algorithms, of which the simulated annealing (SA) technique (Vanderbilt and Louie 1984, Bohachevsky et al. 1986, Corana et al. 1987) is the most powerful one. The SA algorithm is based on the concept of attaining the lowest energy states through slow cooling (e.g. annealing of metals) and is currently used in molecular modeling (Nilges et al. 1988). Much of its success is due to random sampling of the parameter space, based on the probabilistic Monte Carlo method (Metropolis et al. 1953). A comprehensive overview of the SA algorithm has been given by Goffe et al. (1994).

The stochastic character of the SA algorithm provides one of its main advantages: it is no longer necessary to make choices concerning the starting point. In fact, the initial set of parameters is generated randomly in order to avoid any bias in the choice of the subsequent search trajectory. On the other hand, since the sampling of the parameter space must be adequate, the time necessary to complete a single SA run is longer than that required by a single run of a deterministic algorithm. Since deterministic programs have to be run many times with different initial data, the total working time of the two methods, however, becomes comparable, with the advantage in favor of the stochastic methods, because the procedure is totally automated.

Herein, we illustrate a practical implementation of the SA algorithm, using the GOSA (**G**lobal **O**ptimization by **S**imulated **A**nnealing) software, applied to the nearly 16-year record of systolic blood pressure analyzed chronobiologically in Figures 3a-f in Halberg et al. (2006a) and to a 15-year record of the urinary excretion of 17-ketosteroids.

RESULTS AND DISCUSSION

Both the absence of a calendar year and the presence instead of transyears are corroborated in the blood pressure series, Tables 1 and 2. The extent of agreement between the two approaches can be seen from the overlapping 95% confidence intervals of the periods of the different components resolved by the two methods. Components resolved by only one of the two approaches are indicated by a '+' on the right. Table 3 further compares the performance of the GOSA software with linear-nonlinear rhythmometry (Halberg 1980, Cornélissen and Halberg 2005, Halberg et al. 2006a, b), originally tested on a simulated data series consisting of two cosine curves with close periods (Rummel et al. 1974), Fig. 1.

A stepwise analysis by GOSA of a 15-year series of daily (with gaps) steroid metabolite excretions by a healthy man also confirmed the

Table 1. **Results from the GOSA software applied to 16-year record of blood pressure and heart rate of a man** (70 years of age at start of monitoring); see also Figs 3a–f in Halberg et al. (2006a).

MESOR (95% CI)	Slope (95% CI)	Period (95% CI)	Amplitude (95% CI)	Acrophase (95% CI)
		Systolic Blood Pressure (mm Hg)		
132.4 (131.9, 132.9)		13.024 (11.970, 14.078)	4.2 (3.4, 5.0)	97.4 (80.2, 114.6)
		5.656 (5.333, 5.979)	3.6 (2.9, 4.3)	217.7 (189.1, 246.4)
		2.940 (2.877, 3.003)	3.5 (2.7, 4.3)	80.2 (57.3, 103.1)
		2.103 (2.037, 2.168)	2.0 (1.3, 2.7)	126.1 (85.9, 166.2)
		1.662 (1.637, 1.687)	2.5 (1.8, 3.2)	97.4 (68.8, 126.1)
		1.281 (1.265, 1.298)	2.4 (1.7, 3.1)	5.7 (-28.6, 40.1)
		0.528 (0.526, 0.531)	1.5 (0.9, 2.1)	263.6 (223.5, 303.7) +
		Diastolic Blood Pressure (mm Hg)		
73.1 (72.8, 73.4)		12.408 (11.567, 13.248)	3.3 (2.8, 3.8)	80.2 (63.0, 97.4)
		5.593 (5.290, 5.897)	2.1 (1.7, 2.5)	223.5 (194.8, 252.1)
		3.023 (2.951, 3.094)	2.2 (1.7, 2.7)	63.0 (40.1, 85.9)
		2.130 (2.078, 2.182)	1.6 (1.1, 2.1)	108.9 (74.5, 143.2)
		1.733 (1.700, 1.766)	1.6 (1.1, 2.1)	45.8 (17.2, 74.5)
		Heart Rate (beats/min)		
73.5 (72.9, 74.1)	-0.004 (-0.006, -0.002)	9.081 (8.654, 9.509)	2.0 (1.6, 2.4)	131.8 (114.6,149.0)
		4.331 (4.230, 4.433)	3.2 (2.8, 3.6)	338.0 (320.9,355.2)
		3.017 (2.935, 3.099)	1.7 (1.3, 2.1)	269.3 (246.4,292.2)
		2.379 (2.330, 2.428)	1.5 (1.1, 1.9)	91.7 (68.8,114.6)
		1.741 (1.717, 1.766)	1.7 (1.4, 2.0)	212.0 (189.1,234.9)

+ component resolved by the GOSA software but not by linear-nonlinear rhythmometry; decimals do not imply corresponding precision

CI, confidence interval

Model consisted of a sum of cosine curves. Their number, as well as their amplitudes, periods and phases constituted fitting parameters. Acrophase, expressed in degrees, with 360°=period length, is referenced with respect to time t=0, set to Dec. 21, 1989

Transyears are shown in bold

Table 2. **Results from linear-nonlinear rhythmometry (extended cosinor) applied to 16-year record of blood pressure and heart rate of a man** (70 years of age at start of monitoring); see also Figs 3a–f in Halberg et al. (2006a).

MESOR (95% CI)	Slope	Period (95% CI)	Amplitude (95% CI)	Acrophase (95% CI)
		Systolic Blood Pressure (mm Hg)		
132.5 (131.1, 133.9)		12.640 (9.103, 16.180)	3.57 (1.21, 5.93)	-137 (-108, -166)
		5.743 (4.743, 6.744)	3.67 (1.73, 5.61)	-271 (-239, -303)
		2.910 (2.725, 3.095)	3.48 (1.14, 5.81)	-206 (-175, -238)
		2.103 (1.935, 2.271)	2.02 (0.08, 3.96)	-286 (-231, -342)
		1.707 (1.618, 1.796)	2.41 (0.44, 4.38)	-277 (-225, -329)
		1.275 (1.225, 1.324)	2.26 (0.44, 4.08)	-285 (-232, -339)
		Diastolic Blood Pressure (mm Hg)		
73.3 (72.4, 74.3)		12.460 (10.070, 14.850)	3.29 (1.81, 4.78)	-110 (-88, -132)
		5.684 (4.908, 6.459)	2.38 (1.11, 3.65)	-282 (-249, -315)
		3.017 (2.844, 3.190)	2.65 (1.09, 4.20)	-185 (-158, -211)
		2.148 (1.974, 2.321)	1.46 (0.16, 2.76)	-257 (-205, -308)
		1.717 (1.640, 1.795)	1.99 (0.64, 3.35)	-261 (-224, -298)
		1.343 (1.278, 1.408)	1.43 (0.19, 2.68)	-192 (-145, -240) +
		Heart Rate (beats/min)		
62.8 (61.9, 63.6)	-0.004	8.894 (7.667, 10.120)	1.98 (0.76, 3.20)	-173 (-142, -204)
		4.371 (4.108, 4.634)	3.33 (2.15, 4.52)	-51 (-32, -71)
		2.994 (2.769, 3.219)	1.81 (0.63, 2.98)	-30 (-353, -67)
		2.365 (2.197, 2.533)	1.52 (0.48, 2.56)	-247 (-206, -288)
		1.723 (1.644, 1.803)	1.48 (0.41, 2.55)	-72 (-29, -116)
		1.404 (1.334, 1.475)	1.06 (0.00, 2.12)	-66 (-13, -118) +

+ components resolved by linear-nonlinear rhythmometry but not by the GOSA software; decimals do not imply corresponding precision
other symbols as in Table 1
Fit of model consisting of cosine curves with anticipated trial periods of 10.5, 5.25, 3.0, 2.0, 1.7, and 1.3 years. Acrophase (phase of maximum), expressed in (negative) degrees, with 360°≡period length and phase reference set to January 1st, 1989
Transyears are shown in bold

Table 3. **Comparison of GOSA software and linear – nonlinear rhytmometry applied to test series** in Fig. 1. Data simulated according to model: $Y_i = 100 + 10 \cos(2\pi t_i/24 - \pi) + 2 \cos(2\pi t_i/24.8 - \pi) + 5\,R$, where i = 1, …, 336 ($\Delta t$=1 hour; T=14 days) and R is uniformly distributed with zero mean and range =1 (± 0.5).

Method	Period (hours)	MESOR	Amplitude	Acrophase	P-value
1a Linear (L) step	24.00	100.18 (99.98, 100.37)	10.47 (10.20, 10.74)	-98 (-97, -100)	<0.01
1b Nonlinear	23.97 (23.83, 24.10)	100.13 (99.83, 100.43)	9.75 (7.67, 11.84)	-91 (-78, -105)	<0.05[@]
(NL) step	24.63 (24.04, 25.21)		2.49 (0.34, 4.63)	-98 (-47, -149)	<0.05[@]
2 Simulated	23.96 (23.89, 24.03)	100.4 (100.1, 100.7)	9.6 (8.4 , 10.8)	-94 (-83, -105)	<0.05[@]
Annealing	24.6 (24.3 , 24.9)		2.5 (1.3 , 3.7)	-94 (-60, -128)	<0.05[@]
(GOSA)					

[@] P-value inferred from non-overlap of zero-amplitude by 95% CI; decimals do not imply corresponding precision

1a, least squares spectrum detects peak at anticipated period of 24 hours and small sidelobes with periods of 26.7 and 21.8 hours, resolved nonlinearly as presence of second component with period slightly longer than 24 hours

95% confidence intervals listed in parentheses

acrophase expressed in (negative) degrees, with 360°≡period length and 0°=00:00 at start of series

These close components could not be resolved by other approaches such as Fourier, or the Enright and Lomb-Scargle periodograms (Czaplicki et al. 2006)

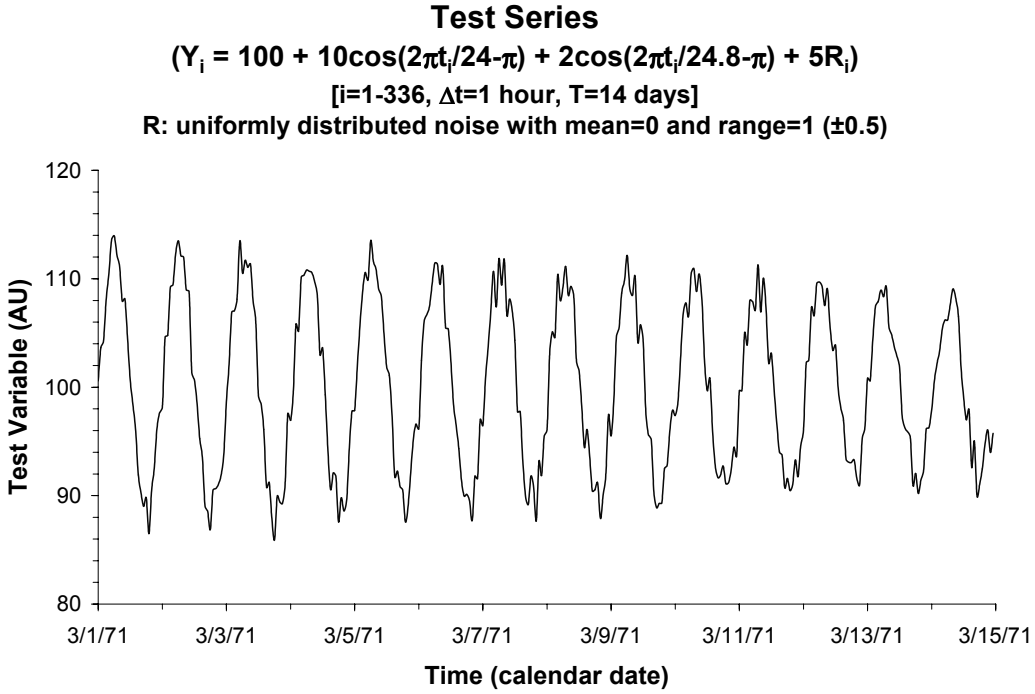

Fig. 1. **Test series** originally discussed in Rummel et al. (1974). See also Halberg (1980).

Table 4: **Stepwise analysis by GOSA* of a series of daily (with gaps) steroid metabolite excretions by a healthy man****

Components***	1	2	3	4	5	6	7	8
Amplitude	90.1 ± 1.7	26.6 ± 1.4	8.5 ± 1.4	7.4 ± 1.3	6.8 ± 1.3	5.6 ± 1.3	6.0 ± 1.3	7.0 ± 1.3
Period (years)	10.52 ± 0.09	5.62 ± 0.13	3.61 ± 1.00	1.555 ± 0.016	1.295 ± 0.014	1.015 ± 0.008	0.925 ± 0.008	0.544 ± 0.003
Phase	5.80 ± 0.04	3.8 ± 0.2	1.6 ± 0.4	4.7 ± 0.4	0.07 ± 0.42	4.5 ± 0.5	4.4 ± 0.4	4.3 ± 0.4

* Global Optimization by Simulated Annealing (www.bio-log.biz)
** see Halberg et al. (2004)
*** Around mean values of 92±1 (mg/24h). Estimates listed with their 95% confidence intervals. Phase expressed in radians

presence of transyears, among others, Table 4. These and an about 10.5-year component similar to the solar activity cycle had been reported earlier (Halberg et al., 2004). Of interest is the about 200-day variation detected with the GOSA software (Czaplicki et al. 2006), very close to the 205-day change in spontaneous abortions in Padua reported by Valandro et al. (2004), which the authors associate with the extreme lunar perigeal positions, a periodicity of the moon in reaching its greatest distance from the earth, later extended to spontaneous abortions in Nové Zámky (Matuška and Mikulecky, 2006).

The aim of this work was to provide biologists and biochemists with an easy-to-use, reliable program capable of finding global minima of an arbitrary function. The results of test cases reported above indicate that this goal has been achieved. The program, based on the simulated annealing algorithm, finds the global minimum of a specified function in a given range of variability of unknown parameters. The values of the resulting parameters are reported along with their uncertainties, estimated on the basis of the corresponding covariance matrix. Absolute values of summed squared deviations and the number of degrees of freedom for the studied model are reported, which render comparisons between different models possible.

The GOSA program has been developed and tested on several different platforms: Unix-based SGI workstations, PC computers running Microsoft Windows OS and Intel-based Linux clusters. GOSA is capable of solving a wide array of problems from different domains of research. The software is available from Bio-Log (www.bio-log.biz). The accompanying documentation contains numerous examples of its use. A combination of random search methods, such as GOSA, with the linear-nonlinear cosinor and other Minnesotan methods (Halberg 1980, Cornélissen and Halberg 2005) remains a challenge.

ACKNOWLEDGEMENTS

Support: GM-13981 (FH), Dr. h.c. mult. Earl Bakken Fund (GC, FH)

JC wishes to thank Didier Fournier for his significant contribution to the development of the software through many stimulating discussions and extensive testing of the program.

REFERENCES

Bohachevsky I.O., Johnson M.I., Stein M.L.: Generalized simulated annealing for function optimization. Technometrics 28:209–217, 1986.

Corana A., Marchesi M., Martini C., Ridella S.: Minimizing multimodal functions of continuous variables with the 'simulated annealing algorithm'. ACM Trans. Math. Softw. 13:262–280, 1987.

Cornélissen G., Halberg F.: Chronomedicine. In Armitage P., Colton T. (eds.): Encyclopedia of Biostatistics, 2nd ed., John Wiley & Sons Ltd., Chichester 2005, pp. 796–812.

Czaplicki J., Cornélissen G., Bakken E.E., Halberg F.: Simulated annealing software for resolving spectral components close in frequency validates endocrine transyears. Abstract, Meeting on Current Problems of Adaptation under Polar Conditions, Khanty-Manoijsk, Siberia, Russia, 2006.

Goffe W.L., Ferrier G.D., Rogers J.: Global optimization of statistical functions with simulated annealing. J. Econom. 60:65–99, 1994.

Halberg F.: Chronobiology: methodological problems. Acta Med. Rom. 18:399–440, 1980.

Halberg F., Cornélissen G., Bakken E.E. et al.: Transyears: new endpoints for gerontology and geriatrics or confusing sources of variability? J. Gerontol. A Biol. Sci. Med. Sci. 59:1344–1347, 2004.

Halberg F., Cornélissen G., Katinas G. et al.: Chronobiology's progress. Part I, season's appreciations 2004–2005: time-, frequency-, phase-, variable-, individual-, age- and site-specific chronomics. J. Appl. Biomed. 4:1–38, 2006a.

Halberg F., Cornélissen G., Katinas G. et al.: Chronobiology's progress. Part II, chronomics for an immediately applicable biomedicine. J. Appl. Biomed. 4:73–86, 2006b.

Matuška T., Mikulecky M.: Chronobiology of spontaneous abortion: Halberg's paraseasonality is dominating again. Abstract, 27th Seminar on Man in his Terrestrial and Cosmic Environment, Úpice, Czech Republic, May 16–18, 2006.

Metropolis N., Rosenbluth A.W., Rosenbluth M.N. et al.: Equations of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092, 1953.

Nilges M., Clore G.M., Gronenborn A.M.: Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. FEBS Lett. 239:129–136, 1988.

Press W.H., Flannery B., Teukolsky S., Vetterling W.T.: Numerical Recipes in Fortran. Cambridge University Press, New York 1992.

Rummel J.A., Lee J.K., Halberg F.: Combined linear-nonlinear chronobiologic windows by

least-squares resolve neighboring components in a physiologic rhythm spectrum. In Ferin M., Halberg F., Richart R.M., Vande Wiele R. (eds.): Biorhythms and Human Reproduction, Int. Inst. for the Study of Human Reproduction Conf. Proc., John Wiley & Sons, New York 1974, pp. 53–82.

Valandro L., Zordan M., Polanska M. et al.: Relevance of lunar periodicity in human spontaneous abortions. Gynecol. Obstet. Invest. 58: 179-182, 2004.

Vanderbilt D., Louie S.G.: A Monte Carlo simulated annealing approach to optimization over continuous variables. J. Comp. Phys. 56: 259–271, 1984.