Application of genetic algorithm optimization technique for fitting non-linear Richards growth model*

M A IQUEBAL[1], PRAJNESHU[2] and HIMADRI GHOSH[3]

Indian Agricultural Statistics Research Institute, New Delhi 110 012

Non-linear growth models generally describe the development of the variable of interest over the time, and are applicable in almost all disciplines related to plants, animals, fisheries etc. (Seber and Wild 2003). The type of model needed in a specific situation depends on the type of growth that occurs. A heartening aspect of these models is that these are 'mechanistic' in nature as the parameters have meaningful biological interpretation. This type of model usually arises as a result of making assumptions about the type of growth, expressing it in terms of differential or difference equations that represent these assumptions, and then solving these equations to obtain a growth model. Prajneshu and Das (1998) used the logistic and Gompertz non-linear models for studying the comparative performance of various states in respect of wheat (*Triticum aestivum* L. emand. Flori & Paol.) production in the post-green revolution era.

Here, a generalization of above 2 models, viz the Richards growth model is considered. If X(t) denotes the variable of interest, say production, or productivity at time t, the model is given by the differential equation

$$dX/dt = rX\left(K^m - X^m\right)/\left(mK^m\right) \qquad ...(1)$$

where, r is the intrinsic growth rate, K is the carrying capacity, and m is a parameter. To solve equation (1), let $y = X^{-m}$, then

$$dy/dt + ry = rK^{-m} \qquad ...(2)$$

which is a linear differential equation. Solving it, we get

$$X(t) = \frac{KX_0}{\left[X_0^m + \left(K^m - X_0^m\right)e^{-rt}\right]^{1/m}} \qquad ...(3)$$

*Short note

where, X_0 indicates the value of X(t) at the initial time, i e t = 0. There are 4-parameters in this model, *viz* r, K, X_0, and m. The range of the first 3-parameters is positive while that of m is from $-\infty$ to $+\infty$. Equation (3) reduces to the monomolecular, Gompertz and logistic models when m = -1, 0, 1 respectively. For Richards model, the graph of X(t) versus t is sigmoid, i e elongated S-shaped. The point of inflexion, i e the point at which the tangent changes its direction is at $X = K/(m+1)^{1/m}$, m $\neq -1$, and occurs at

$$t = \frac{1}{r}\log\left(\frac{K^m - X_0^m}{mX_0^m}\right), m \neq -1 \qquad ...(4)$$

Thus Richards model is very flexible as the point of inflexion is not fixed but can occur at any fraction of the carrying capacity K depending on the value of m.

To apply equation (3) to data, the usual procedure is to assume an additive error term on the right hand side, thereby yielding a non-linear statistical model. The parameters of the model are then estimated by using the procedures of non-linear estimation. To this end, Seber and Wild (2003) provide a good description of various iterative procedures, like Levenberg-Marquardt procedure and does not use derivative procedure. Fortunately, most of the standard statistical packages contain computer programmes for fitting the models by these methods. For example, SPSS has NLR option and SAS has NLIN option to accomplish the task. However, global convergence is not always guaranteed and the success rate is quite low if the functional form of the model is somewhat complicated.

India's annual foodgrain production data for the post-green revolution era, viz 1966 to 2005 is utilized for present study. It increased from a level of 74.23 million tonnes in 1966–67 to the level of 208.6 million tonnes in 2005–06, i e an increase of almost 3 times in nearly 3 decades. A perusal of the data also showed that, in the initial phase, the rate of growth was quite high but subsequently the foodgrain production data had a retarded pace. Accordingly, the Richards model seemed

to be a logical choice for modelling the data. In the first instance, attempts were made to fit equation (3) to this data by using the SAS (2006) software package. Several initial values for each parameter were tried. However, global convergence did not take place. Thus, the Richards growth model could not be fitted to the given data by employing non-linear estimation procedures.

Fortunately a very powerful and versatile optimization technique of genetic algorithm (GA), motivated by the principles of genetics and natural selection, has recently been developed (Goldberg 1989). In this methodology, some fundamental ideas of genetics are borrowed and used artificially to construct search algorithms that are robust and require minimal problem information. The 3 operators, viz selection, crossover, and mutation make GA an important tool for optimization. When a string (parameter solution) is created by GA, it is evaluated in terms of its fitness, which is taken to be the residual sum of squares (RSS).

Selection operator of GA is performed to identify good solutions, to make its multiple copies and to eliminate bad solutions from the population. The most common method of selection that has better convergence and computational time efficiency is the tournament selection (Deb 2002). Since selection operator cannot create a new solution, the crossover and mutation operators are used to create a new population so that global minima may be achieved in succeeding generations. There exists a number of crossover operators due to which different string pairs are expected to have each good bit representation. Such an operator should not be allowed to use all strings in a population to preserve some good strings selected during the selection operator. If a crossover probability P_c is used, $100 p_c \%$ strings in the population are used in the crossover and rest of the population is simply copied to the new population. If B denotes string length, only B–1 pairs of different string representations, which have already survived selection pressure, are possible to be created during crossover.

In real-coded GA, a pair of real-parameter decision variable vector is used to create a new pair of offspring vectors by applying crossover operator. Here, an important operator of this type, called Simulated binary crossover operator, was developed by Deb and Agrawal (1995). Two off-spring $x_i^{(1, t+1)}$ and $x_i^{(2, t+1)}$ are produced from two parent solutions $x_i^{(1, t)}$ and $x_i^{(2, t)}$, where, $x_i^{(j, t)}$ is the value of ith variable of jth parent in t^{th} generation, i = 1, 2,..., p; j = 1, 2; and t ≥ 1. To this end, after drawing the spread factor β_i from the probability distribution with mode unity (Deb 2002), the offspring are calculated as follows.

$$x_i^{(1, t+1)} = 0.5\left\{ \left(1+\beta_{qi}\right)x_i^{(1, t)} + \left(1-\beta_{qi}\right)x_i^{(2, t)} \right\}$$
$$x_i^{(2, t+1)} = 0.5\left\{ \left(1-\beta_{qi}\right)x_i^{(1, t)} + \left(1+\beta_{qi}\right)x_i^{(2, t)} \right\} \quad ...(5)$$

These off-spring are symmetrically distributed and the points of symmetry are equispaced from the mid-point of parent solutions. Thus, biasedness towards any particular parent solution is avoided.

The mutation operator is used to ensure diversity in the population. It alters a string locally to create a better string (parameter solution). If certain genes (digits) of all chromosomes (strings) in the population are identical, their values will never change after selection and crossover. This will reduce the chance for some new chromosomes to enter this population, thereby lending the GA-procedure into the trap of local optima. To avoid this situation, mutation with very small probability, say 0.01 is required. In real-coded GA, mutation operator does same task as performed by real parameter crossover operator. The polynomial mutation operator has the advantage that the probability distribution does not change with generations, thereby avoiding local optima and so this mutation operator is used in present study. It mutates the ith variable to $y_i^{(1, t+1)}$ by the transformation:

$$y_i^{(1, t+1)} = x_i^{(1, t+1)} + \left\{ x_i^{(U)} - x_i^{(L)} \right\} \overline{\delta}_i \quad ...(6)$$

where $\overline{\delta}_i$ follows polynomial probability distribution.

Attempts are then made to fit non-linear Richards growth model given by equation (3) to India's total foodgrain production time-series data through GA-methodology. To this end, the objective function to be minimized can be written as:

$$\sum_{t=1}^{T} \left[X(t) - \frac{K}{\left[1+\exp\left(b-rt\right)\right]^{1/m}} \right]^2 \quad ...(7)$$

where b = $\log_e (K^m/X_0^m - 1)$. The GA parameters, viz. population size, crossover probability, and mutation probability for minimization of equation (7) are respectively computed as 40, 0.9, 0.01 with number of generations as 100. Using above parameter set up, GA has terminated with accuracy level (ε=10^{-3}) in 99 out of 100 runs in general. This confirms that Richards growth model is successfully fitted by GA-optimization technique. Relevant computer programmes are developed using C-language in Microsoft visual C++ compiler and could be obtained from first author on request. The run test statistic is computed as 0.32, whose absolute value being less than 1.96, indicates that the assumption of independence of error terms is not violated at 5% level. The goodness of fit statistics for fitted model, viz R^2 and Root mean square error are respectively computed as 0.93 and 11.36, which reflect that the fit is quite good. Finally, the fitted non-linear Richards growth model is obtained as

$$X(t) = \frac{290.26}{\left[1 + \exp\left(4.76 - 0.09\ t\right)\right]^{1/4.17}} \quad ...(8)$$
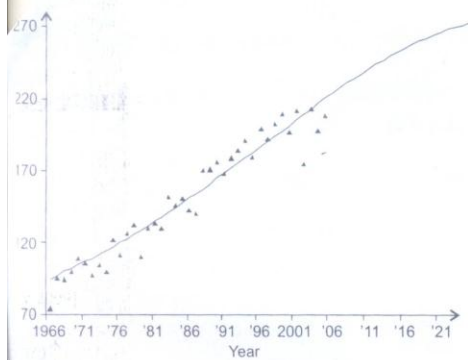
1 Fitted non-linear Richards growth model along with data points

get a visual idea, the fitted model along with data points hibited in Fig 1.

SUMMARY

lity of genetic algorithm optimization methodology for

fitting of non-linear Richards growth model is highlighted. The proposed procedure is successfully applied for modelling and forecasting of India's foodgrain production time-series data. The importance of this work is that this methodology is applicable even in those cases in which non-linear estimation procedures fail to converge. The possibility of modifying the methodology in the presence of outliers is currently being explored and the results shall be reported separately in due course of time.

REFERENCES

Deb K. 2002. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Singapore.

Deb K and Agrawal R B. 1995. Simulated binary crossover for continuous search space. *Complex Systems* **9**: 115–48.

Goldberg D E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Pub. Co., U.S.A.

Prajneshu and Das P K. 1998. Modelling wheat (*Triticum aestivum*) production at state level in post green revolution era. *Indian Journal of Agricultural Sciences* **68**: 23–26.

SAS 2006. *SAS/IML User's Guide*. SAS Institute, U.S.A.

Seber G A F and Wild C J. 2003. *Non-linear Regression*. 2nd edn., John Wiley & Sons, USA.