

World and regional populations

Henry C. Tuckwell*^a, James A. Koziol^b

^aStatistics Research Section, Institute of Advanced Studies, Australian National University, Canberra, ACT 2601, Australia

^bScripps Clinic and Research Foundation, 10666 N. Torrey Pines, La Jolla, CA 92037, USA

(Received 21 July 1992; revision received 8 March 1993)

Abstract

Non-linear regression indicates that the logistic model provides excellent fits to world and continental human population data of the period 1950–1985. If the present logistic regime persists, the world population is likely to double in 47 years. Moreover, the eventual world population would be close to its final value of 23.8 billion by the year 2200, with about 11 billion in Asia and 10.5 billion in Africa. An ~95% confidence interval for the eventual world population is found to be between 11.9 and 35.7 billion. The logistic estimates of future populations are much higher than those considered most likely by the United Nations.

Key words: World population; Population growth; Logistic model

1. Introduction

The human population predictions made by the United Nations (1988, 1991a, 1992) for the world and the various continents are obtained by the component method. This method (Leslie, 1945; Keyfitz, 1968; Pollard, 1973) requires a knowledge of birth and death rates for various age groups and, for the continental projections, migration rates. The United Nations (1988, 1991a) lists the countries within specified regions. A large number of parameters are required for these calculations. The UN predictions, which are published from time to time, consist of three estimated population trajectories, called low, medium and high variants.

These reflect different assumptions concerning fertility, mortality and migration rates, the medium variant being considered to be the most likely. For example, for the year 2025 these estimates for the world population are 7.59, 8.47 and 9.42 billion according to United Nations (1988), with slight changes of the last two figures to 8.50 and 9.44 billion in United Nations, 1991a.

However, an alternative method, which usually involves only a few parameters, is to postulate an underlying mathematical model for population growth, based for example on difference or differential equations. The simplest such model, which assumes that the rate of change of population is proportional to existing numbers, results in a Malthusian or exponential curve. This has been employed for human populations by Westing (1981a, 1981b) to estimate the total number of in-

* Corresponding author.

dividuals who have currently or previously existed. A more realistic alternative to the Malthusian model is that of Verhulst (1838), which employs the logistic differential equation. The rate of change of population is then assumed to be proportional to the product of the existing population and the difference between the existing and maximum possible populations. As a consequence, the population grows towards a finite upper limit, thereby excluding the possibility of exponential growth to infinity as in the Malthusian model.

This logistic, which is both simple and natural, gained popularity after Pearl and Reed (1920) employed it to predict likely population trends in the United States. They used census data to 1910 and obtained the parameters of the logistic curve by insisting that it pass through three data points. A subsequent statistical analysis was performed by Schultz (1930), who obtained estimates of the standard deviations of the parameters. Several similar models were later examined by Feller (1939).

More recently the logistic equation has been derived from first principles as being satisfied by the expectation in a stochastic model of population growth in a habitat with a finite number of viable sites (Tuckwell and Koziol, 1987). It is commonly employed as an interesting and meaningful example of a non-linear differential equation in modern courses of applied mathematics (Haberman, 1977; Hoppensteadt and Peskin, 1992) and is employed by biologists in their quantitative analyses (Cauteaudier and Steinberg, 1990).

Interest in the logistic curve for human populations was revived by Leach (1981). He found that the populations of England, Scotland and the United States could be well fitted with logistic curves. In the case of the United States, however, a jump in the parameters r , which describes the natural rate of increase, and K , which represents the upper limit, occurred just after the Second World War. The same phenomenon occurred in the populations of Australia and New Zealand (Koziol and Tuckwell, 1986) at the same time, as many fled from Europe after the extended period of unrest, and birth rates increased after hostilities ceased. We may thus speak of logistic regimes that persist until the parameters are changed by, for example, territorial expansion or loss or by the inci-

dence of major epidemics, or by major events of a climatological, economic, military, political or sociological nature.

2. Results and discussion

We decided to examine the population data of the world and its various regions to see if present regimes were logistic, and if so, to estimate the corresponding parameters. In a logistic regime the number $N(t)$ of individuals at time t is given by the formula

$$N(t) = \frac{K}{1 + \left(\frac{K}{N_0} - 1\right) \exp(-rt)}$$

where N_0 is the population at a reference time $t = 0$. In this formula r is positive, so that as $t \rightarrow \infty$, the population asymptotically approaches K .

Population data at 5-year intervals from 1950 to 1985 (United Nations, 1988) was employed. From these data, parameters of the best-fitting curves were obtained using a non-linear least squares algorithm (Levenberg, 1944; Marquardt, 1963; Gallant, 1987). This iterative procedure was used to find the parameters r , K and N_0 so as to minimize the simultaneous sums of squares for the individual logistic functions with respect to the associated data for the world and the various continental populations. Although eight functions were fitted simultaneously, the use of weights was unnecessary, because the data sets, and hence the various residuals sums of squares, were all of the same scale. The resulting overall fits might be informally judged to be quite good from inspection of predicted and residual plots and more formally so judged from the determination coefficient $R^2 = 0.995$. (Recall that R^2 is the fraction of the total unweighted sums-of-squares, corrected for the mean, that is accounted for by fitting the model to the data).

An example of the fits obtained is provided by the world population trajectory shown in Fig. 1. Here the data points are shown for 1950–1985 and

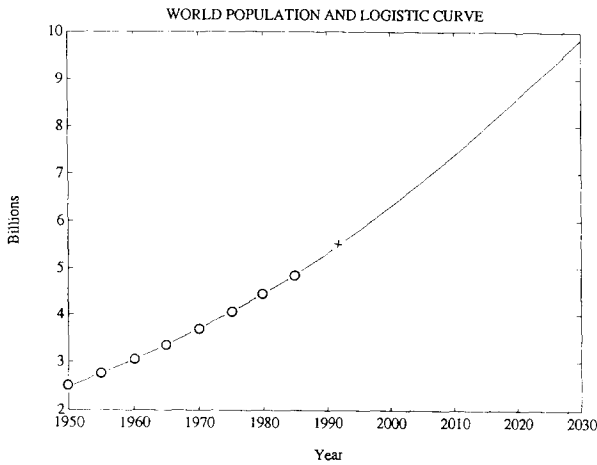


Fig. 1. World population data (circles) from United Nations (1988) for 1950–1985 and best-fitting logistic curve until 2030. The cross marks the UN figure for April 1992.

the best-fitting logistic curve to 2030. It can be seen that all the data points (circles) lie very close to the logistic. When the best-fitting logistic was used to calculate the current world population (using data prior to and up to 1985), the value obtained was 5.48 billion, which is precisely the figure released on April 29th, 1992, by the United Nations. The estimate was made without prior knowledge of this current UN figure. The data points for the continental regions (not shown) were also very close to the calculated logistics, and were obtained under the constraint that their total should equal the world value at each time point. Fig. 2A shows the logistic curve for the continental regions with K values less than 1 billion, and Fig. 2B shows the curves for Africa, Asia and the world and, for comparison, Latin America, the latter having the largest K value of the regions of Fig. 2A.

Table I shows the parameters r , K and N_0 for the best-fitting logistic curves for the world population and the various continental regions. The values of N_0 correspond to populations in 1950. By making certain simplifying but reasonable assumptions, it is also possible, based on the assumption of an underlying logistic, to estimate confidence limits for the parameters and hence for the population at any given time point (Gallant,

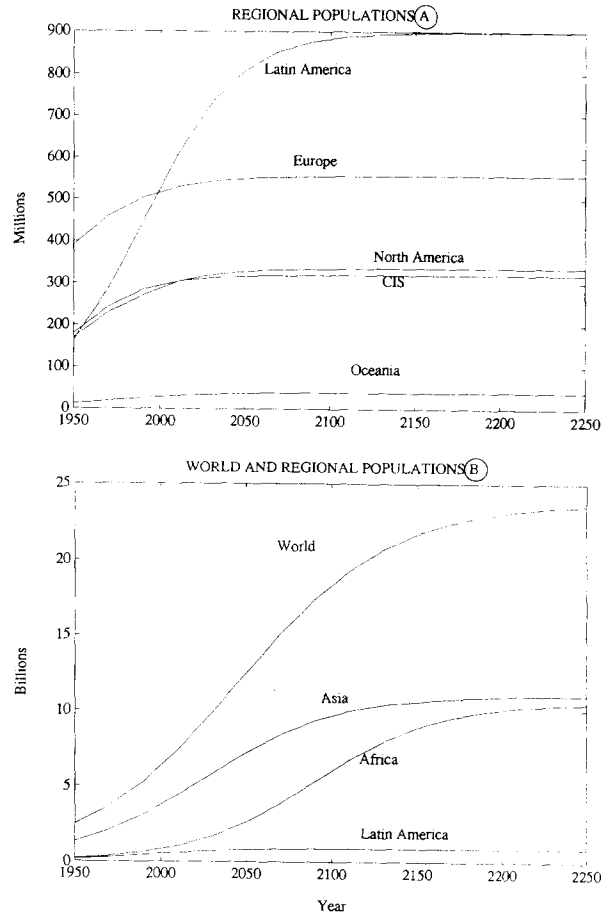


Fig. 2. (A) Population trajectories for various continental regions given by best-fitting logistic curves. (B) Population trajectories for the world and major population regions. (Note that here population numbers are in billions.)

Table 1

Values of the parameters that give the best-fitting logistic curves (Levenberg-Marquardt algorithm) for the populations of the world and the various continental regions. Based on United Nations (1988) data for 1950–1985. K and N_0 in millions, r in year⁻¹.

Region	K	N_0	r
Africa	10 571	215	0.028
Asia	11 077	1349	0.026
Europe	556	390	0.035
Latin America	899	163	0.037
North America	336	167	0.037
Oceania	40	13	0.036
CIS (USSR)	320	180	0.045
World	23 799	2495	0.022

1987, Chapter 1). Fig. 3 shows one and two standard deviations from the estimated logistic curve for the world population. The standard deviation of the K value for the world is 6.07 billion; thus there is about a 95% chance that the eventual population will lie between 11.9 and 35.7 billion. The standard deviation for the world r value is 0.00107 and that of the value of N_0 is 0.00864 billion.

The United Nations previously predicted (United Nations, 1983) that the world population would become stationary at 10.2 billion in 2095. More recently (United Nations, 1992) the projection was modified to 10 billion by 2050, with a stationary population of 11.6 billion in 2150. (The UN predictions cited are those for their medium variant.) The present logistic regime gives the population as 17.91 billion in 2095, with a stationary population of 23.8 billion achieved approximately by 2200. The UN high variant, however, puts the population at 12.5 billion in 2050 and 20.7 billion in 2150; the corresponding logistic figures are 12.5 and 21.7 billion, respectively. However, if we use a confidence interval of plus or minus two standard deviations (outer curves of Fig. 3), the logistic predictions are between 11.3 and 14.8 billion in 2050 and between 12.5 and 31.7 billion by 2150. The last figure is 11 billion more than the UN high variant.

A prediction from the logistic is that the present world population is likely to double in the alarm-

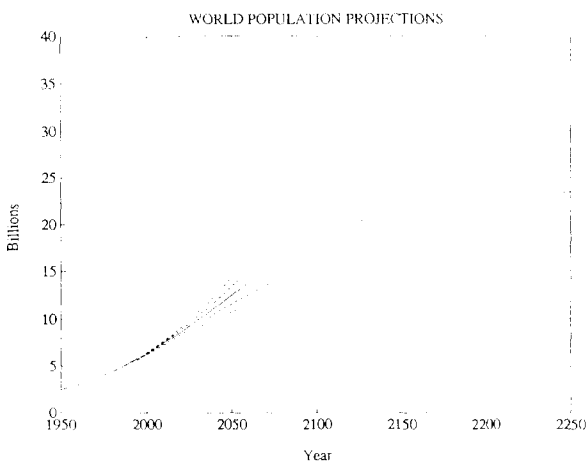


Fig. 3. Logistic curve for the world population (solid line) with confidence limits marked at plus and minus one and two standard deviations.

ingly short period of 47 years. The most highly populated regions will be Asia, with eventually 11.08 billion or 46.6% of total population, and Africa, with 10.57 billion (44.4%), giving these two regions jointly 91%. It is noteworthy that the CIS (formerly USSR) and North America are likely to grow by relatively small amounts, with the CIS reaching within 5% of its K value of 320 million within 20 years and North America reaching within 5% of its maximum of 336 million within 40 years. Similarly, Europe is predicted to reach within 5% of its final population of 556 million within 18 years. The predictions for Africa are probably the least reliable, as there is doubt about the accuracy of the data. There is support for the view that Africa is unlikely to have the capacity to support a population of a magnitude similar to that of Asia.

The UN world population projections for the year 2000 are 6.09, 6.25 and 6.41 billion for the low, medium and high variants, respectively, whereas our estimate from the logistic gives this population as 6.3 billion. The UN predictions for 2025 have been given above and can be compared with the logistic figure of 9.21 billion.

It can be seen that the UN projections and the logistic estimates increasingly differ as time increases. In fact the logistic prediction is roughly equal to the calculated high variant of the United Nations. Clearly neither the logistic nor the UN projection has guaranteed accuracy. Despite the fact that the UN prediction takes account of aging and concomitant changes in death rates, there is no reliable method of predicting what changes might occur in birth rates. The extremely accurate fit of the logistic to the present data enables one to have reasonable confidence in its predictions for the immediate future. As pointed out before (Leach, 1981; Koziol and Tuckwell, 1986), the logistic model has the very advantageous feature of pooling several population-dynamical and environmental effects and subsuming them into just a few parameters.

Our approach is open to the criticism that many other curves can fit the data just as well. Davidson (1985), for example, showed that some countries' populations were well fitted with cubic functions of time (Feller, 1939). However, of all such curves

only the logistic has been derived from first principles as a model for population growth in a bounded environment (Tuckwell and Koziol, 1987). A brief summary of the present findings has appeared (Tuckwell and Koziol, 1992).

It is probable that the actual human population trajectory for the next 100 years or so will lie between the UN and logistic predictions. However, according to the present logistic regime, the predicted number of humans will eventually be about five times as many as at present, and this could have a disastrous effect on food and energy supplies, pollution, and global weather patterns, and lead to increased risks of major epidemics and military conflicts, especially over territory and resources. It has been conjectured that rapid population growth in developing countries must lead to a poor economic outlook (United Nations, 1991b). Indeed the present work points to over-population as a major problem whose consideration should not be postponed (Lieth, 1991). The concept of optimum population levels is in need of exploration and execution (Woodwell, 1981).

Finally, it is noteworthy that the population trajectory has become approximately logistic during the latter part of the twentieth century after about 500 000 years (Diamond, 1991) in which growth was comparatively slow and erratic. We conjecture that before this period a relatively small population was kept in check by environmental hazards, military conflicts, epidemics, and poor food-gathering methods, but that now, with the relatively low frequency of major epidemics, modern methods of food production and relatively small effects of military conflicts, the world population has gained a large momentum and is likely to nearly achieve saturation value in the relatively short period of the next 200 years. Thus, although some authors have claimed that stabilization of the world's population is in progress (Merrick, 1986), there is little evidence that supports that claim.

3. References

- Cauteaudier, Y. and Steinberg, C., 1990, Biological and mathematical description of the growth pattern of *Fusarium oxysporum* in a sterilized soil. *Microbiol. Ecol.* 74, 253–260.
- Davidson, S.M., 1985, Population growth. *Nature* 314, 398.
- Diamond, J., 1991, *The Rise and Fall of the Third Chimpanzee: Evolution and Human Life* (Hutchinson, London).
- Feller, W., 1939, On the logistic growth law and its empirical verifications in biology. *Acta. Biotheor.* 5, 51–66.
- Haberman, R., 1977, *Mathematical Models* (Prentice-Hall, Englewood Cliffs).
- Gallant, A.R., 1987, *Nonlinear Statistical Models* (Wiley, New York).
- Hoppensteadt, F.C. and Peskin, C.S., 1992, *Mathematics in Medicine and the Life Sciences* (Springer-Verlag, New York).
- Keyfitz, N., 1968, *Introduction to the Mathematics of Populations* (Addison-Wesley, Reading).
- Koziol, J.A. and Tuckwell, H.C., 1986, Population projections for Australia and New Zealand by the logistic method. *NZ Stat.* 21, 35–40.
- Leach, D., 1981, Re-evaluation of the logistic curve for human populations. *J.R. Stat. Soc. A.* 144, 94–103.
- Leslie, P.H., 1945, On the use of matrices in certain population mathematics. *Biometrika* 33, 183–212.
- Levenberg, K., 1944, A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.* 2, 164–168.
- Lieth, H.H.F., 1991, in: News update. *BioScience* 41, 11–12.
- Marquardt, D.W., 1963, An algorithm for least-squares estimation of non-linear parameters. *J. Soc. Indust. Appl. Math.* 11, 431–444.
- Merrick, T.W., 1986, World population in transition. *Pop. Bull.* 41, 1–51.
- Pearl, R. and Reed, L.J., 1920, On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proc. Natl. Acad. Sci. USA* 6, 275–288.
- Pollard, J.H., 1973, *Mathematical Models for the Growth of Human Populations* (Cambridge University Press, Cambridge).
- Schultz, H., 1930, The standard error of a forecast from a curve. *J. Amer. Statist. Assoc.* 25, 139–185.
- Tuckwell, H.C. and Koziol, J.A., 1987, Logistic population growth under random dispersal. *Bull. Math. Biol.* 49, 495–506.
- Tuckwell, H.C. and Koziol, J.A., 1992, World population. *Nature* 359, 200.
- United Nations, 1983, *Pop. Bull.* 14, 17–30.
- United Nations, 1988, *World Population Prospects, 1988* (United Nations, New York).
- United Nations, 1991a, *World Population Prospects, 1990* (United Nations, New York).
- United Nations, 1991b, *Consequences of Rapid Population Growth in Developing Countries* (Taylor and Francis, New York).
- United Nations, 1992, *State of the World Population* (United Nations, New York).
- Verhulst, P.F., 1838, Notice sur la loi que la population suit dans son accroissement. *Corr. Math. Phys.* 10, 113–121.
- Westing, A.H., 1981a, A note on how many humans that have ever lived. *BioScience* 31, 523–524.
- Westing, A.H., 1981b, A note on how many humans that have ever lived. *BioScience* 32, 6.
- Woodwell, G.M., 1981, A postscript for the old boys of the subversive science. *BioScience* 31, 518–522.