

The Dynamics of the Lynx–Hare System: an Application of the Lotka–Volterra Model

L. V. Nedorezov

Research Center for Interdisciplinary Environmental Cooperation, Russian Academy of Sciences,
nab. Kutuzova 14, St. Petersburg, 191187 Russia

e-mail: l.v.nedorezov@gmail.com

Received September 6, 2015

Abstract—The Lotka–Volterra model of predator–prey dynamics was used for approximation of the well-known empirical time series on the lynx–hare system in Canada that was collected by the Hudson Bay Company in 1845–1935. The model was assumed to demonstrate satisfactory data approximation if the sets of deviations of the model and empirical data for both time series satisfied a number of statistical criteria (for the selected significance level). The frequency distributions of deviations between the theoretical (model) trajectories and empirical datasets were tested for symmetry (with respect to the Y-axis; the Kolmogorov–Smirnov and Lehmann–Rosenblatt tests) and the presence or absence of serial correlation (the Swed–Eisenhart and “jumps up–jumps down” tests). The numerical calculations show that the set of points of the space of model parameters, when the deviations satisfy the statistical criteria, is not empty and, consequently, the model is suitable for describing empirical data.

Keywords: the Lotka–Volterra model, model parameter estimation, deviation analysis

DOI: 10.1134/S000635091601019X

One of the most interesting problems (which is important for the development of mathematical ecology) is approximation of the well-known data that was collected by the Hudson Bay Company [1–5] on the fluctuations in the numbers of hare (*Lepus americanus*) and lynx (*Lynx lynx*) using the trajectories of the Lotka–Volterra model [6–9]:

$$\frac{dx}{dt} = ax - bxy, \quad \frac{dy}{dt} = -cy + dxy, \quad (1)$$

where $x(t)$ is the number of prey (hare) and $y(t)$ is the number of predators (lynx) at the time t ; a is the Malthusian parameter of the prey population (hereinafter, it is supposed that $a > 0$; if $a \leq 0$, then the system of two interacting species degenerates asymptotically at any initial sizes of the populations); c is the Malthusian parameter of predator population (mortality rate); and b , d are the coefficients of interaction between the populations. In model (1), all of the coefficients are positive, $a, b, c, d = \text{const} \geq 0$. Figure 1 presents the data on the fluctuations in the numbers of hare and lynx in Canada (in thousands of skins delivered yearly).

The approximation of the presented data using model (1) has been attempted more than once but failed each time. The major causes of failure were specified as follows: the noncoarseness of model (1), a priori inconsistency between the data and the theoretical results (in some years, the arrangement of maxi-

num numbers is “incorrect”: according to model (1), the maximum number of predators must always be observed after the maximum number of prey), “incorrect” movement of the empirical trajectory over the “prey number–predator number” plane (in some cases, the empirical trajectory is “twisted” in the opposite direction), etc. In addition, (possible) peculiarities of data collection have been indicated [3–5].

In particular, M. Gilpin [3, 4], when estimating the parameters of a more general model than (1) (including (1) as a special case), followed the procedure that was proposed by F. J. Ayala [10, 11]. According to this procedure, the left parts of equations (1) are replaced by the annual increments of the respective numbers; the parameters are then estimated with minimization of the sum of squared deviations (the squared differences of the left and right parts of the model equations).

Let us note that this approach to parameter estimation involves not the initial model per se but its discrete analog as obtained using the Euler scheme. The Adams formula could be used to obtain a countable set of discrete analogs of the initial model, which could be potentially used to obtain more reasonable results. In addition, with this approach, the empirical trajectory is not at all approximated by the model trajectory, which is “closer” to the initial data than other trajectories. Finally, the least-squares method (LSM) very often gives incorrect results: we must admit the model

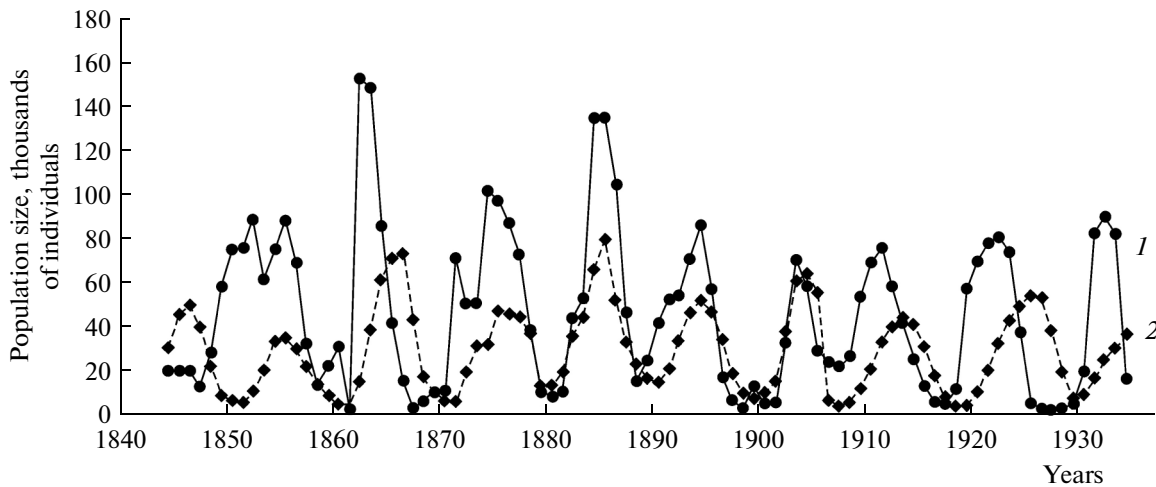


Fig. 1. The fluctuations of lynx and hare populations in Canada from 1845 to 1935: 1, the curve of hare-number variations; 2, the curve of lynx-number variations.

to be unfit for data approximation when it gives a satisfactory description with parameter estimates close to the LSM [12–16].

In this work, the initial data [1–4] are approximated by the trajectories of model (1): the system is integrated by the Runge–Kutta fourth-order method in increments of 10^{-3} ; it is considered that a time unit of the model (1000 steps) is equal to 1 year. The parameters of model (1) and the initial population sizes were selected at random (uniform distribution) in the Δ region: $x_0, y_0 \in [0.220]$, $a, b, c, d \in [0.1]$, $c \in [0.20]$. After using a random point and finding the respective trajectory of model (1), the two sequences of deviations of empirical data from the model trajectory were calculated. The series of statistical tests (for a confidence level of 5%) was checked for each sequence of deviations. The point of the space of parameters was considered to belong to the *admissible set* only if all of the tests that were used for each sequence of deviations showed the required result.

The model should be regarded as unfit for data approximation only if the admissible set for the selected confidence level proves to be empty. If the admissible set is not empty, then there are no grounds for the statement about the unfitness of the model. This work was aimed at demonstrating the unsoundness of the widespread view of the unfitness of model (1) for approximation of any data at all. A hare cannot use a lynx for food: fairly good correspondence between the model and empirical trajectories is observed also at the nonnegative values of the parameters.

STATISTICAL CRITERIA

Let there be given a sample $\{(x_k, y_k)\}$, $k = 0, 1, \dots, N$, where x_k is the prey (hare) population size at a time

k , y_k is the predator (lynx) population size at the same time, $N + 1$ is the sample size ($N = 90$). Let it also be given that $x = x(t, x_0, y_0, \vec{\alpha})$, $y = y(t, x_0, y_0, \vec{\alpha})$, i.e., the solution of system (1) at the given initial population sizes x_0, y_0 and the given values of model parameters $\vec{\alpha}$.

Let e_k^1 and e_k^2 to be the following values:

$$e_k^1 = x_k - x(k, x_0, y_0, \vec{\alpha}), \quad e_k^2 = y_k - y(k, x_0, y_0, \vec{\alpha}).$$

As well, let $\{e_k^{j+}\}$ be a set of positive deviations and $\{e_k^{j-}\}$ be a set of negative deviations, $j = 1, 2$.

Let us consider that the model trajectory gives a positive data approximation if:

1. The frequency distributions of the deviations $\{e_k^1\}$ and $\{e_k^2\}$ are symmetrical relative to the Y-axis. Symmetry relative to the Y-axis implies that the samples $\{e_k^{j+}\}$ and $\{e_k^{j-}\}$, $j = 1, 2$, have the same distribution function (these functions are different for different j values). Therefore, symmetry can be checked using the Kolmogorov–Smirnov, Wald–Wolfowitz, Mann–Whitney, and Lehmann–Rosenblatt homogeneity tests [17–19]. However, the use of 10^7 random points in Δ prevented us from detecting the values of parameters when the two sequences of deviations simultaneously satisfied all of these criteria. Hence, for further calculations we used a “weakened” variant with only the Kolmogorov–Smirnov and Lehmann–Rosenblatt tests.

2. The branches of the density function must be monotonous: the density value must increase at negative values of the argument and decrease in the domain of its positive values (greater deviations must be

observed with lesser probabilities). The monotonous behavior of the branches of the density function was checked using the Spearman's and Kendall's rank correlation coefficients [17].

Let $\{e_k^{j*+}\}$, i.e., the sample arranged in ascending order $\{e_k^{j*+}\}$: $\{e_1^{j*+}\} < \{e_2^{j*+}\} < \dots$. During the monotonous decrease in the density function, *in the ideal case*, the lengths of intervals $[0, e_1^{j*+}]$, $[e_1^{j*+}, e_2^{j*+}]$, ... must also be arranged in ascending order (and can be ranked as 1, 2, ...). Let ρ be the Spearman's correlation coefficient. When comparing the ranks of the lengths of intervals for the sample $\{e_k^{j*+}\}$ with the ideal case, the null hypothesis $H_0: \rho = 0$ (at the alternative hypothesis $H_1: \rho > 0$ and the selected significance level) should be rejected. Further calculations were made using the "weakened" variant: the null hypothesis was verified two times for each point of the space: for the samples $\{e_1^{j*+}\} \cup \{e_2^{j*+}\}$, $j = 1, 2$.

3. The sequence of deviations must have no serial correlation. For this purpose, the Swed–Eisenhart Runs Test was used [20, 21]. In some cases, the "jumps up–jumps down" runs test was also used [18].

If any of the used statistical criteria gave a negative result, then the respective point of the space of model parameters was considered as not belonging to the admissible set. If all of the criteria gave positive results (i.e., the hypothesis of symmetry relative to the Y-axis of the distribution of deviations was not rejected, then the hypothesis of the absence of serial correlation and the hypothesis of monotonous behavior of the branches of density functions were not rejected) this implied that the model can be used quite well for initial data approximation and explanation of the peculiar features of population dynamics at the tested values of parameters and that the respective point belongs to the admissible set.

The next stage includes the selection of points with extreme characteristics in the admissible set (the extreme point method [14–16]), i.e., the points for which the required result can be obtained with the maximum (or, in some cases, minimum) level of significance. If, for example, the hypothesis of distribution symmetry cannot be rejected at the significance level of 5%, then it may be a basis for accepting the null hypothesis; but a much more significant result is observed in the case where the same hypothesis cannot be rejected with a 95% significance level. The increase (in some cases, decrease) in the significance level characterizes the better theoretical and empirical data fit.

COMPUTATIONAL RESULTS

Some authors believe [1–5] that the numbers of interacting species fluctuate with an approximately 10-year cycle. However, such a regularity is not observed in the sequence of maximum numbers: in particular, the maximums can repeat every 2 years (Fig. 1). Of course, if we take some maximums into account but do not take others, it is quite possible to find a 10-year cycle.

The analysis of the elements in the admissible set shows the presence of 10-year cycles among the possible dynamic regimes (Fig. 2). The regime presented in Figure 2 is implemented at $x_0 = 115.0546$, $y_0 = 45.7497$, $a = 0.631315$, $b = 0.036912$, $c = 2.49187$, and $d = 0.104151$ and has the following characteristics: the hypotheses of distribution symmetry cannot be rejected by the Kolmogorov–Smirnov test at the significance levels of 43.57 and 10.19% (for $\{e_k^1\}$, and $\{e_k^2\}$, respectively) and by the Lehmann–Rosenblatt test with the significance levels of 49.54 and 6.4%.

The Spearman's rank correlation coefficients are 0.4352 and 0.3113 with the respective p -levels of $1.6 \cdot 10^{-5}$ and $2.67 \cdot 10^{-3}$. The Kendall rank correlation coefficients are 0.3153 and 0.2181 with the respective p -levels of $1.0 \cdot 10^{-5}$ and $2.2 \cdot 10^{-3}$. The null hypotheses (that the coefficient is equal to 0) are rejected with very low significance levels. With the significance levels of 14.7 and 35.4%, the hypothesis of the absence of serial correlation is not rejected by the Swed–Eisenhart criterion. It may also be noted that one can visually observe a rather good fit between the theoretical curve and empirical data (Figs. 2a, 2b): the maximums of the model trajectory are very close to the maximum values of the initial sample.

The admissible set has elements for which the respective deviations satisfy the greater number of statistical tests. In particular, the point with the coordinates $x_0 = 53.3244$, $y_0 = 107.0217$, $a = 0.80522$, $b = 0.025355$, $c = 6.476933$, and $d = 0.128812$ has the following characteristics: the hypotheses of distribution symmetry cannot be rejected by the Kolmogorov–Smirnov test at the significance levels of 99.81 and 6.46%, by the Lehmann–Rosenblatt test at the significance levels of 97.43 and 9.83%, by the Wald–Wolfowitz test at the significance levels of 44.23 and 26.72%, and by the Mann–Whitney test at the significance levels of 77.1 and 14.08%.

The Spearman's correlation coefficients are 0.27605 and 0.4459, with the respective p -levels of $8.08 \cdot 10^{-3}$ and $9.0 \cdot 10^{-6}$. The Kendall rank correlation coefficients are 0.1912 and 0.3118, with the respective p -levels of $7.27 \cdot 10^{-3}$ and $1.2 \cdot 10^{-5}$. The null hypotheses (of the equality of the coefficient to zero) are rejected. At the significance levels of 16.22 and 35.18%, the hypothesis of the absence of serial correlation is not rejected by the Swed–Eisenhart test. The critical values at a significance level of 20% for the

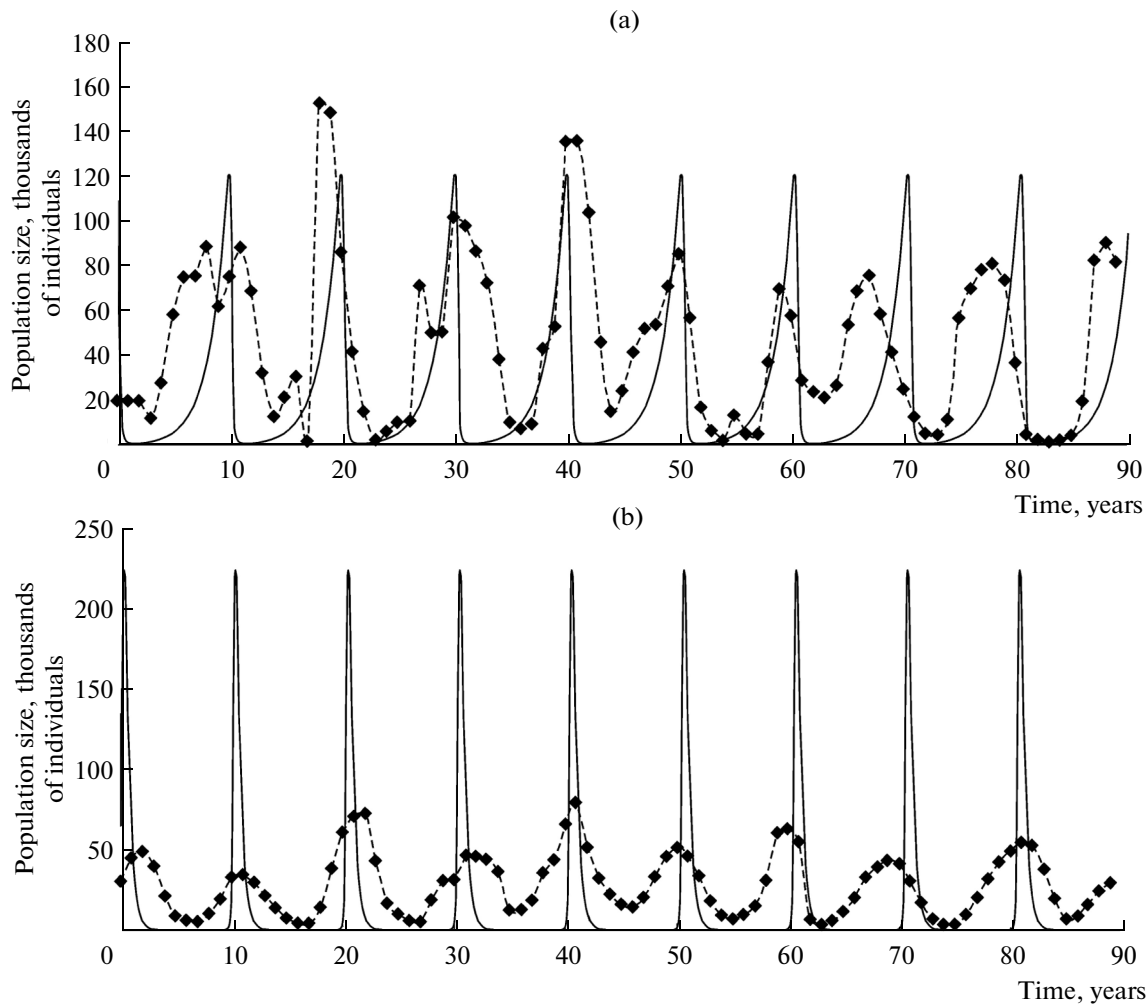


Fig. 2. The results of data approximation (10-year cycle): (a) for hare numbers; (b) for lynx numbers. The solid line is the trajectory of model (1). The dashed line is empirical data.

“jumps up—jumps down” test [18] are 54 and 66. The deviations have 56 serial numbers for the hare time series and 59 serial numbers for the other time series. Thus, the hypotheses about the presence of a serial correlation in the sequences of deviations must be rejected.

Figure 3 shows the empirical data and the trajectory of model (1) for the variant under consideration. The cycle length is 3.06 (years). Visually, the results that are presented in Fig. 3 are slightly worse than those presented in Fig. 2: the amplitude of fluctuations obviously is not in agreement with the empirical data (for the trajectory of changes in the hare numbers, the amplitude is much less than in the time series; for the trajectory of change in the lynx numbers, the opposite occurs) and the amplitude of fluctuations is much less than the empirical one.

The search for an element of the admissible set that would correspond to the minimum sum of squared

deviations for both time series has shown that it is a cycle of 2.06 years, which also does not agree with a priori assumptions. The pattern of the time series for the hare numbers is similar to that presented in Figure 3a (the amplitude of the fluctuations has a similar value but the frequency is obviously higher). The amplitude of the fluctuations of the lynx numbers is much less than the empirical one.

The analysis of 30 000 points that were selected at random from the admissible set shows the occurrence of cycles with a length from the half open interval [1, 2) with the maximum probability (0.37107); from [2, 3) with the probability of 0.2972, and from [3, 4) with the probability of 0.1595. The other probabilities are much lower. In particular, the cycles with a length from the interval [10, 11) occur with a probability of 0.001439. Cycles with a length of more than 12 were not revealed.

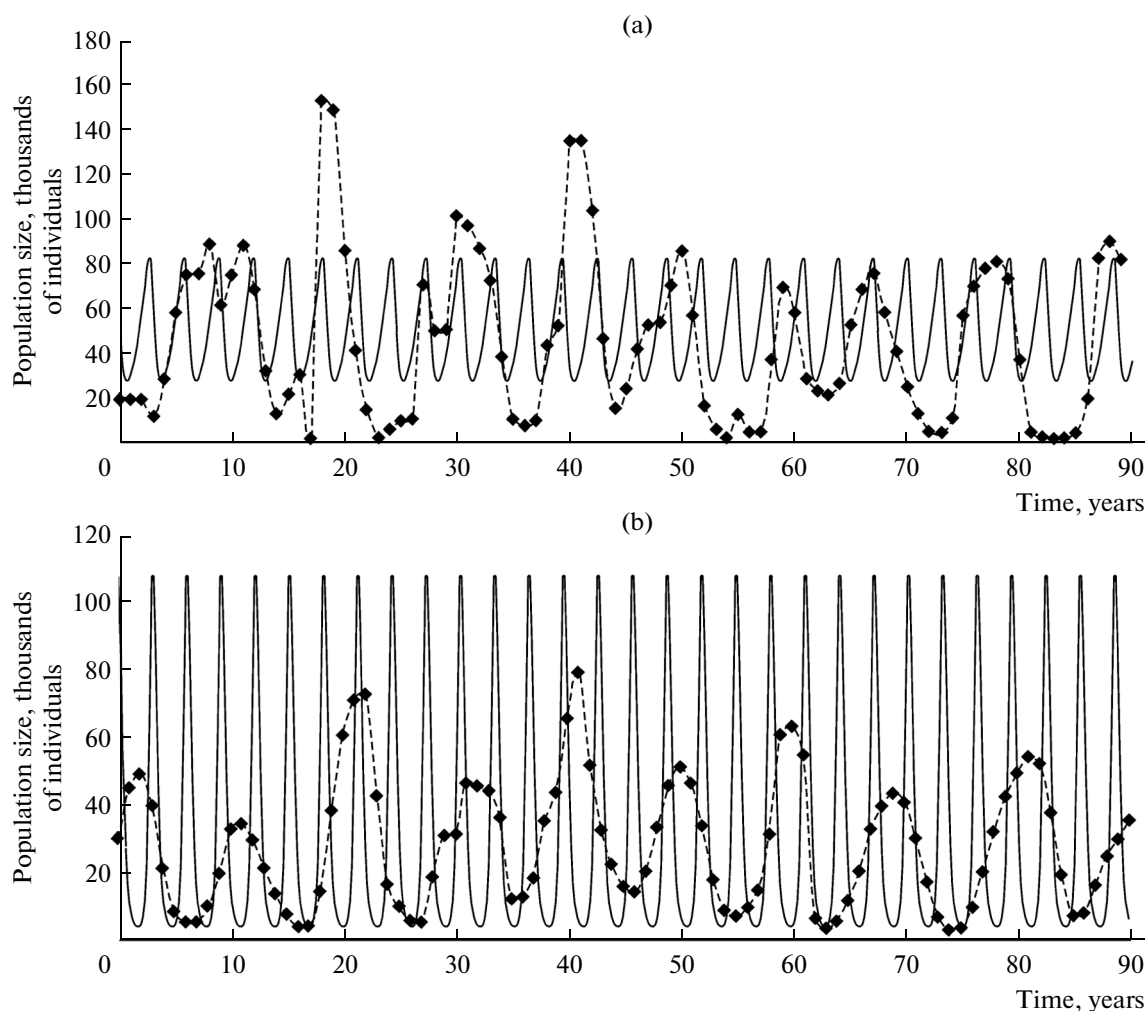


Fig. 3. The results of data approximation (3-year cycle): (a) for hare numbers; (b) for lynx numbers. The solid line is the trajectory of model (1). The dashed line is empirical data.

CONCLUSIONS

The least-squares method assumes that first some estimates of the model parameters are found and only then are the properties of the sequences of deviations from theoretical and empirical data verified. The approach that was used in this work follows exactly the opposite pattern. First, the points of the space of parameters are found where the sequences of deviations from theoretical and empirical data have the required properties (the admissible set); only then are points with extreme properties selected among the points that were found.

The least-squares method is rejected due to a number of factors. These are the absence of criteria for selecting the type of minimized functional (which, as a rule, has nothing to do with the biological problem, or the available data, or the model) and the final result, i.e., the decision on the applicability or inapplicability of the model to data approximation is rendered on the basis of properties of the single point of the space of

parameters (the global minimum of the functional). Consequently, there is an abundance of incorrect results from nonlinear models. One such parameter has just been considered in this work.

The analysis of correspondence between the Lotka–Volterra model (without autoregulation in the populations) and the empirical data on fluctuations of the hare and lynx numbers in Canada shows the existence of numerous points in the space of model parameters that yield rather good approximations of the data. At the same time, not only are a number of statistical criteria satisfied but, in some cases, correspondence between theoretical and empirical data can be visually observed.

When using the least-squares method to obtain a certain data fit, it is necessary to change the signs of the coefficients that describe the interaction between the species. In other words, a fit is possible if hares become bloodthirsty and eat the lynx. However, this

analysis shows that actually nothing of the kind is required (either from the hares or from the model).

REFERENCES

1. D. A. MacLulich, *Fluctuations in the Numbers of Varying Hare (Lepus americanus)* (Univ. of Toronto Press, Toronto, 1937).
2. C. S. Elton and M. Nicholson, *J. Anim. Ecol.* **11**, 215 (1942).
3. M. E. Gilpin, *Am. Nat.* **107** (957), 727 (1973).
4. M. Gilpin, Samarskaya Luka: *Probl. Glob. Region. Ecol.* **19** (3), 177 (2010).
5. G. S. Rozenberg, Samarskaya Luka: *Probl. Glob. Region. Ecol.* **19** (3), 180 (2010).
6. V. Volterra, *Mathematical Theory of Struggle for Life* (Nauka, Moscow, 1976) [in Russian].
7. V. Volterra, *Lecons sur la theorie mathematique de la lutte pour la vie* (Gauthiers-Villars, Paris, 1931).
8. A. I. Lotka, *J. Am. Chem. Soc.* **42** (8), 1595 (1920).
9. A. I. Lotka, *Elements of Physical Biology* (Williams & Wilkins, Baltimore, 1925).
10. F. J. Ayala, M. E. Gilpin, and J. G. Ehrenfeld, *Theor. Pop. Biol.* **4**, 331 (1973).
11. M. E. Gilpin and F. J. Ayala, *Proc. Nat. Acad. Sci. USA.* **70** (12), 3590 (1973).
12. L. V. Nedorezov, *Zh. Obshch. Biol.* **73** (2), 114 (2012).
13. L. V. Nedorezov, *Population Dynamics: Analysis, Modelling, Forecast* **1** (1), 47 (2012).
14. L. V. Nedorezov, *Chaos and Order in Population Dynamics: Modeling, Analysis, and Forecast* (Lambert Acad. Publ., Saarbrücken, 2012).
15. L. V. Nedorezov, *Biophysics (Moscow)* **60** (3), 457 (2015).
16. L. V. Nedorezov, *Biophysics (Moscow)* **60** (5), 862 (2015).
17. L. N. Bol'shev and N. V. Smirnov, *Mathematical Statistical Tables* (Nauka, Moscow, 1983) [in Russian].
18. I. Likeš and J. Laga, *Basic Statistical Tables* (SNTL, Prague, 1978; Finansy i Statistika, Moscow 1985).
19. M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods* (New York, Wiley, 1973; Finansy i Statistika, Moscow, 1983).
20. N. R. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed. (Wiley, New York, 1981; Finansy i Statistika, Moscow, 1986), Vol. 1.
21. N. R. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed. (Wiley, New York, 1981; Finansy i Statistika, Moscow, 1987), Vol. 2.

Translated by E. Makeeva