Article

Application of generalized discrete logistic model for fitting of pine looper moth time series: Feasible sets and estimations of model parameters

L.V. Nedorezov
Research Center for Interdisciplinary Environmental Cooperation RAS, Saint-Petersburg, Russia
E-mail: lev.nedorezov@gmail.com

Received 23 October 2014; Accepted 30 November 2014; Published online 1 March 2015

Abstract
New approach to estimation of ecological model parameters is considered and applied to analysis of well-known pine looper moth time series (Klomp, 1966). Within the framework of approach it is assumed that before constructing and minimizing of loss-function basic requirements to model and to deviations between empirical and theoretical (model) datasets must be formulated. After that respective statistical criterions must be determined, and with the help of these criterions structure of feasible set in space of model parameters (where these criterions are satisfied) must be obtained. Structures of feasible sets were determined for generalized discrete logistic model with known datasets of pine looper moth population dynamics. Results were compared with estimations obtained with Least Square Method.

Key words: pine looper moth; time series; discrete logistic model; statistical criterions.

Proceedings of the International Academy of Ecology and Environmental Sciences
ISSN 2220-8860
URL: http://www.iaees.org/publications/journals/piaees/online-version.asp
RSS: http://www.iaees.org/publications/journals/piaees/rss.xml
E-mail: piaees@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Let's consider the following discrete model of population dynamics:

$$x_{k+1} = F(x_k, \vec{\alpha}) . \qquad (1)$$

In (1) x_k is population size (or population density) at time moment k, $k = 0,1,2...$; $\vec{\alpha}$ is vector of model parameters. Additionally we'll assume that at initial time moment $k = 0$ population size x_0 is unknown model parameter too. Let $\{x_k^*\}$, $k = 0,1,..., N$, be a sample, empirical time series of changing of size of (certain) population; $N + 1$ is sample size. Using this sample $\{x_k^*\}$ we have to find estimations of model (1) parameters $\vec{\alpha}$ and x_0.

For this reason we can use various approaches. In particular, we can use Least Square Method (LSM) (Bard, 1974; Bolshev and Smirnov, 1983; Draper and Smith, 1998). In this occasion we have to choose loss-function, and, for example, we can choose it in the following form:

$$Q(\vec{\alpha}, x_0) = \sum_{k=0}^{N} \left(x_k(\vec{\alpha}, x_0) - x_k^* \right)^2 . \qquad (2)$$

In (2) $x_k(\vec{\alpha}, x_0)$ is solution of equation (1) which is determined for parameters $\vec{\alpha}$ and initial value x_0. *It is assumed* that *best* estimations of parameters can be obtained minimizing functional (2). After determination of values of parameters we have to check properties of set of deviations

$e_k = x_k(\vec{\alpha}, x_0) - x_k^*$.

Basic (traditional) ideas about deviations are following (Draper, Smith, 1998): e_k must be values of independent normally distributed stochastic variables with zero averages. Correspondence of deviations to Normal distribution can be checked with Kolmogorov – Smirnov, Lilliefors, Shapiro – Wilk and other tests (Shapiro et al., 1968; Lilliefors, 1967; Bolshev and Smirnov, 1983; Bard, 1974). Checking of absence/existence of serial correlation in sequence of residuals can be provided with Durbin – Watson test and non-parametric Swed – Eisenhart test (Draper and Smith, 1998).

If serial correlation is observed in a sequence of residuals we have a background for conclusion that model isn't suitable for fitting of considering time series. The same conclusion we can get in a situation when distribution of residuals isn't Normal (for fixed significance level). It means that final conclusion about suitability of model for fitting of time series we make using *one point from the space of model parameters*.

Best parameters give minimum for minimizing functional (2). But what is a background for assumption that estimations of model parameters must give minimum for (any) functional form? The answer is rather obvious: this assumption has no background. Moreover, it has no relation to biological object and biological problem we have to solve (to determine of law of population size changing). It is possible to point out the only explanation: we want to find one point and we don't want to operate with sets.

One more assumption about Normality of deviations has no background too. Moreover, in real situations this assumption doesn't correspond to reality. For example, if we estimate weights of insects, obtained dataset cannot correspond to Normal distribution: we'll never have insect with negative weight, we cannot have error in several tons of kilograms. But if we postulate that errors of measurements correspond to Normal distribution it means that a'priori we assume that we may have *insects with negative weight with positive probability*. With positive probability we may also have insects with weight of several tons.

Counter-evidences on these remarks are following: probabilities of these events (to obtain negative weights and very big weights) are very small and we can ignore such events… Of course, distribution of residuals isn't Normal but it is very close to Normal distribution etc… But we have to note that two expressions "to have Normal distribution" and "to be close to Normal distribution" are qualitatively different. And it isn't obligatory that properties which were proved for Normal distribution must be observed for distribution which is close to Normal.

Finally, summarizing presented above about LSM we can conclude that there is a lot of problems in application of this method to solution of real problems. It is a problem of selection of loss-function which has no relation to biological problem (but it has strong influence on final results). This is a problem with assumption about normality of deviations, and this is a problem with creation of final conclusion about suitability of model for fitting of time series which is *based on one point from a space of model parameters*. These problems appeared in a result of logic mistake in order of providing analysis. Before all calculations, before estimation of model parameters we have to formulate basic requirements to model and to deviations between theoretical and empirical values.

Below we compare results which were obtained for pine looper moth time series with traditional approach (LSM) and non-traditional approach. In all considering situations we got different estimations (in quantitative and qualitative sense).

2 Basic Requirements to Model

All requirements to model can be conditionally divided onto three groups:

1. Deviations between theoretical (model) and empirical time series must have symmetric distribution (symmetric density function) with respect to origin; branches of density function must have monotonic behavior (in positive part of straight line density function must decrease monotonously, and it must increase in negative part). This requirement is not strong as requirement for deviations to have Normal distribution. It is possible to point out a lot of distributions which satisfy to considering conditions (see, for example, Korn and Korn, 1973).

Let $\{e_k^+\}$ be a set of positive deviations, and $\{-e_k^-\}$ be a set of negative deviations with sign "minus" (i.e. $\{-e_k^-\}$ is a set of positive values). Symmetry of density function with respect to origin means that for selected significance level there are no reasons for rejecting hypothesis about equivalence of distributions of two samples $\{e_k^+\}$ and $\{-e_k^-\}$. For checking of symmetry of distributions Kolmogorov – Smirnov test, Lehmann – Rosenblatt test, and Mann – Whitney U-test were used (Bolshev and Smirnov, 1983; Hollander and Wolfe, 1973; Likesh and Laga, 1985).

Monotonic behavior of branches of density function was tested using Spearmen rank correlation coefficient (Bolshev, Smirnov, 1983). Let $\{e_k^{*+}\}$ be ordered sample $\{e_k^+\}$: $e_1^{*+} < e_2^{*+} < \ldots$. In the case of monotonic decreasing of right branch of density function in *ideal situation* lengths of intervals $[0, e_1^{*+}]$, $[e_1^{*+}, e_2^{*+}]$, … must be ordered too and in the same manner. In ideal case intervals can be ranked 1, 2,… (from smallest to biggest interval). This ideal variant must be compared with sequence of ranks determined by the sample. Let ρ be a Spearmen rank correlation coefficient. It is obvious that Null hypothesis $H_0: \ \rho = 0$ (with alternative hypothesis $H_1: \ \rho > 0$ and fixed significance level) must be rejected. If we have to reject Null hypothesis we have a guarantee that branch of density function has a monotonic behavior. If we cannot reject this hypothesis we haven't a guarantee.

2. In a sequence of residuals serial correlation cannot be observed. If we have to reject hypothesis about absence of serial correlation it means that some important mechanisms were not taken into account in model. Thus, we have to conclude that model cannot be used for fitting of time series. For testing of absence/existence of serial correlation Swed – Eisenhart test (Draper, Smith, 1998) and "jumps up – jumps down" (Likesh and Laga, 1985) were used.

3. Application for analysis of deviations all pointed out above tests cannot give us cogent argument for conclusion about suitability of model for fitting of empirical time series. We have no reasons to say that considering model is good if for every increasing intervals in time series model demonstrates decreasing and vice versa. Thus, we have to check hypothesis about quota q of successive variants "increasing in time series – increasing in model" and "decreasing in time series – decreasing in model" among all observed situations. It is obvious, if model demonstrates good correspondence with empirical dataset, we have to reject Null hypothesis $H_0: \ q = 0.5$ with alternative hypothesis $H_1: \ q > 0.5$.

Checking of properties of points of space of parameters (with finite steps of changing of values of parameters) will allow obtaining of *feasible set* for model where we can and/or have to find minimum of any minimizing functional.

Finally, the following statistical criterions and characteristics were used for testing of properties of points of model parameter space (in pointed out order):

1. Kolmogorov – Smirnov test (used for testing of symmetry of distribution),
2. Lehmann – Rosenblatt test (used for testing of symmetry of distribution),
3. Mann – Whitney U-test (used for testing of symmetry of distribution),
4. Spearmen rank correlation coefficient (used for testing of monotonic behavior of branches of density function),
5. Swed – Eisenhart test (used for testing of absence/existence of serial correlation),
6. “jumps up – jumps down” test (used for testing of absence/existence of serial correlation),
7. test for correspondence of changing in time series and model.

If one of pointed out statistical tests showed negative result respective point of space of model parameters was marked with green color; if all statistical tests demonstrated required results respective point of space of model parameters was marked with red color. Results of calculations are presented as projections of feasible sets on the plane (a,b). If for fixed values a and b it is possible to find any initial value x_0 when point (x_0, a, b) was marked in red color, respective point on plane (a,b) was marked with red color too. Partly this plan of estimation of model parameters was realized in our publication (Sadykova and Nedorezov, 2013).

3 Generalized Discrete Logistic Model

For fitting of datasets on pine looper moth dynamics generalized discrete logistic model (Nedorezov, 2012) was used:

$$x_{k+1} = \begin{cases} ax_k(b - x_k),\ 0 \le x_k \le b, \\ 0,\ b < x_k. \end{cases},\quad a,b = const \ge 0. \qquad (3)$$

In (3) ab is maximal birth rate; b is carrying capacity. This model has rich set of dynamic regimes, and its application for fitting of various time series allowed obtaining good results (Nedorezov, 2011; Nedorezov, Sadykova, 2010; Nedorezov, Utyupin, 2011). Model (3) has following basic properties: if $ab < 1$ population eliminates for all initial values of population size; if $1 < ab < 2$ regime of monotonic stabilization of population size is observed in model (3); if $2 < ab < 3$ there is a regime of fading fluctuations near non-zero stationary state; if $3 < ab < 4$ cyclic regimes of various lengths and chaotic regimes can be observed in phase space; if $ab > 4$ trajectories of model (3) can intersect level b, and after that trajectory becomes equal to zero identically. In last case model (3) cannot be used for forecasting of population size dynamics. These bifurcation curves $ab = 1$, $ab = 2$, $ab = 3$, and $ab = 4$ are presented on pictures.

4 Datasets

In current publication time series on pine looper moth were used (Klomp, 1966). In numerical format datasets can be found and free downloaded in Internet (NERC Centre for Population Biology, Imperial College (1999) The Global Population Dynamics Database, № 2727, 2728 and 2729). In first time series (№ 2727) values are presented in units «average number of eggs per square meter»; in second time series (№ 2728) values are presented in units «average number of larvae per square meter»; in third case (time series № 2729) values are presented in units «average density of pupae per square meter». In first case sample size is equal to 15 (initial value was obtained in 1950); in the second case sample size is equal to 14 (initial value was also obtained in 1950 but information about density of larvae in 1962 is absent); in the third case we have 14 values (initial value was obtained in 1951).

Data were collected in Netherlands in national park De Hoge Veluwe. It was demonstrated in our previous publications (Nedorezov, 2011, 2012) that discrete logistic model is unique which allowed obtaining sufficient approximation for all time series (with the framework of traditional approach to estimation of model parameters). In current publication we compare results which can be obtained for pine looper moth dynamics time series within the framework of traditional approach (LSM) and within the framework of non-traditional approach described above.

5 Time Series № 2727

For time series № 2727 following LSM-estimations for model (3) parameters were obtained: $x_0 = 38.23$, $a = 0.0437$, $b = 92.158$, $Q_{\min} = 7290.8$ where $Q_{\min}$ is minimal value of functional (2).

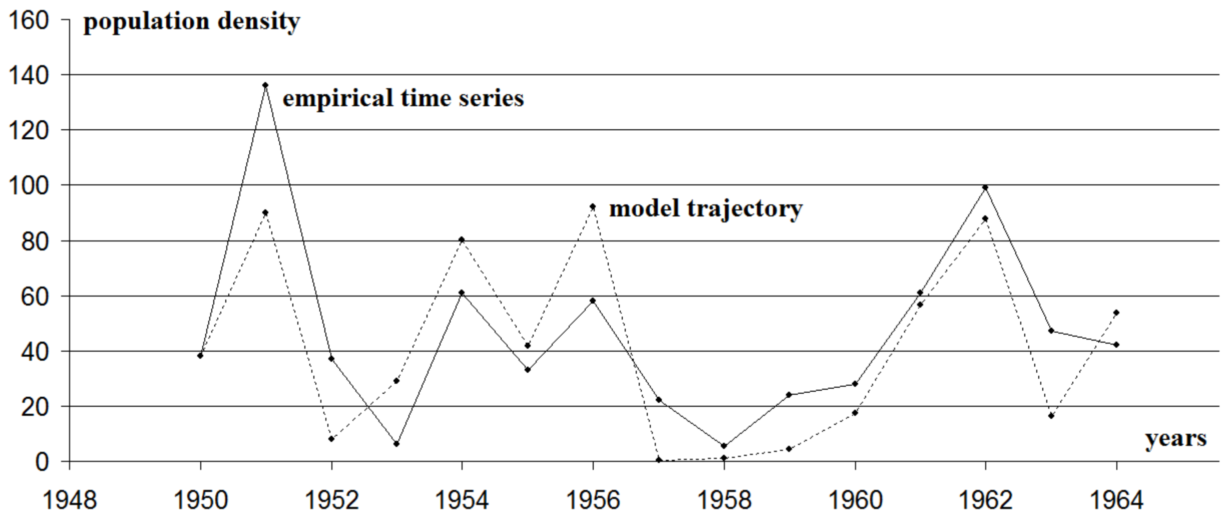

Fig. 1 Empirical dataset (2727, solid line) and model trajectory (broken line) calculated with LSM-estimations of model parameters.

For pointed out parameters behavior of empirical time series and model (3) trajectory are presented on figure 1. For these estimations of model parameters probability of event that distribution of deviations is Normal is greater than 0.2, $p > 0.2$ (Kolmogorov – Smirnov test and Lilliefors' test); Shapiro – Wilk test showed that this probability is rather big, $p = 0.99816$. Thus, hypothesis about Normality of distribution cannot be rejected even with 10% significance level (Bolshev and Smirnov, 1983; Shapiro et al., 1968; Lilliefors, 1967; Hollander and Wolfe, 1973). For deviations average plus/minus standard error is equal to -5.345 ± 5.716. Hypothesis about equivalence of average to zero cannot be rejected.

Durbin – Watson test showed that $d = 1.628$; critical value for 5% significance level $d_U = 1.36$, $d_U < d$. Swed – Eisenhart test showed that $p = 0.063$. Consequently, both tests showed that serial correlation is absent in sequence of residuals (there are no reasons for rejecting of the respective hypotheses with fixed significance level) (Draper, Smith, 1998). It means that within the framework of traditional approach to model parameter estimations (LSM) we have to conclude that generalized discrete logistic model (3) is suitable for fitting of considering time series 2727.

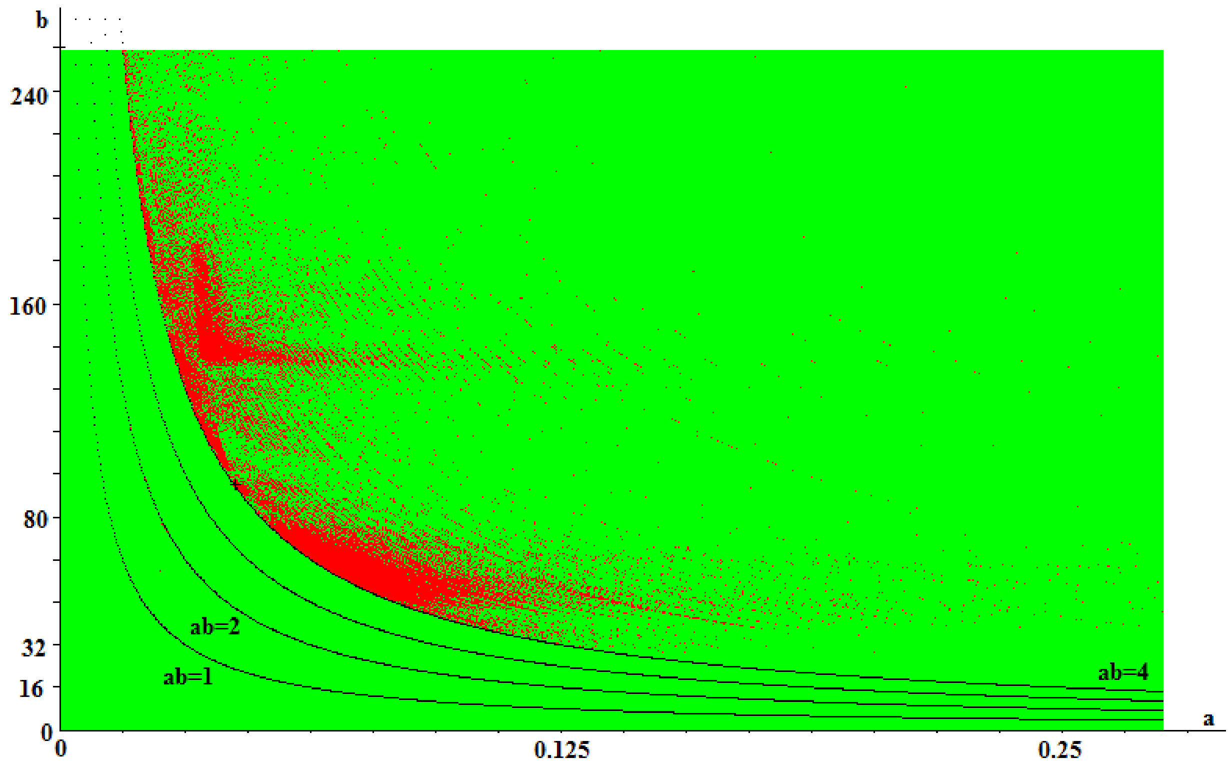

Fig. 2 Projection of feasible set (for 2727) onto the plane (a,b). Crest corresponds to point of minimum of functional (2). Functions $ab=1$, $ab=2$, $ab=3$, and $ab=4$ are basic bifurcation curves for model (3).

On Fig. 2 projection of feasible set on plane (a,b) is presented. As we can see on this picture, minimum of functional (2) doesn't belong to any set of maximum concentration of red points. It belongs to domain $\{(a,b): ab > 4\}$, and as it was pointed out above with these estimations of model parameters equation (3) cannot be used for forecast.

It is interesting to note that rather small number of red points belongs to domain $\{(a,b): 1 < ab \le 3\}$ (Fig. 2). It means that with 5% significance level we can conclude that observed time series corresponds to dynamic regime of asymptotic stabilization of population density on non-zero level. If it is true, model (3) can be used for forecast of population density changing. We have also to note that there are no red points in the domain $\{(a,b): ab < 1\}$ (observed dynamics doesn't correspond to the regime of asymptotic elimination of population). And no points were found in the domain $\{(a,b): 3 < ab < 4\}$: there are no reasons for conclusion that observed regime is cyclic or chaotic.

6 Time Series № 2728

For time series № 2727 following LSM-estimations for model (3) parameters were obtained: $x_0 = 6.5$, $a = 0.227$, $b = 19.33$, $Q_{\min} = 152.6$ where $Q_{\min}$ is minimal value of functional (2). For pointed out parameters behavior of empirical time series and model (3) trajectory are presented on figure 3 (two last years values of elements of sample and model trajectory are very close).

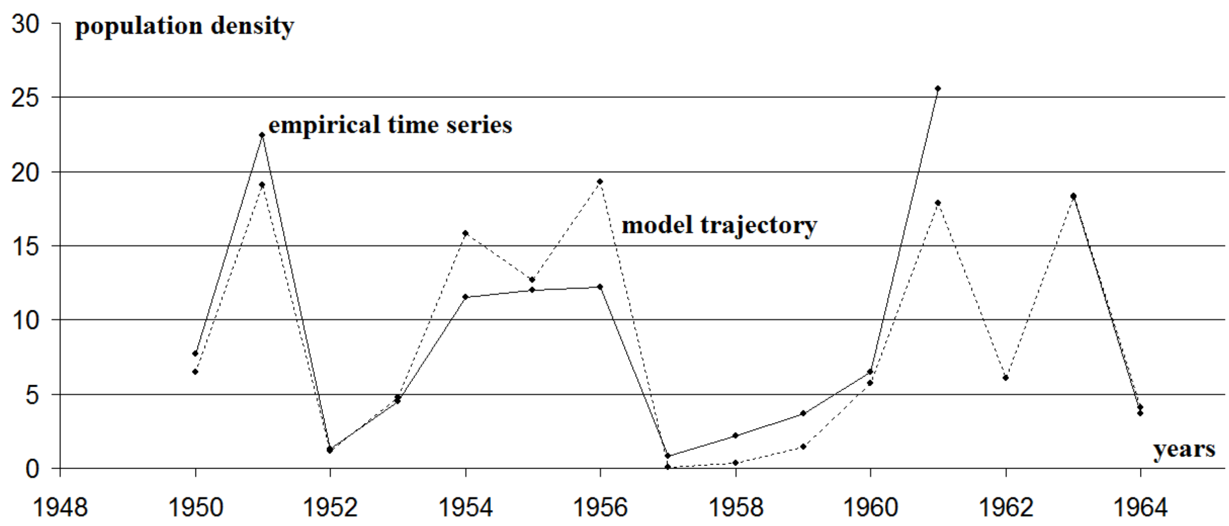

Fig. 3 Empirical dataset (2728, solid line) and model trajectory (broken line) calculated with LSM-estimations of model parameters.

For these estimations of model parameters probability of event that distribution of deviations is Normal is greater than 0.2, $p > 0.2$ (Kolmogorov – Smirnov test and Lilliefors' test); Shapiro – Wilk test showed that this probability is rather big, $p = 0.13045$. Thus, hypothesis about Normality of distribution cannot be rejected even with 10% significance level. For deviations average plus/minus standard error is equal to 0.1152 ± 0.981. Consequently, hypothesis about equivalence of average to zero cannot be rejected.

Durbin – Watson test showed that $d = 1.319$; critical value for 2.5% significance level and one predictor variable is $d_U = 1.18319$, $d_U < d$ (for determination of value of Durbin – Watson test first 12 values of sample were used only). Swed – Eisenhart test showed that probability of sequence of signs of residuals is rather big, $p = 0.762$. Thus, both statistical tests showed that serial correlation is absent in sequence of residuals. It means that within the framework of traditional approach to model parameter estimations we have to conclude that generalized discrete logistic (3) is suitable for fitting of time series 2728.

On Fig. 4 projection of feasible set onto the plane (a,b) is presented. As we can see on this picture, minimum of functional (2) belongs to zone of maximum concentration of marked red points. Additionally, minimum belongs to zone $\{(a,b): ab > 4\}$, and it means that in this case model (3) cannot be used for prediction of population density dynamics.

As we can see on fig. 4, big number of red points belongs to zone $\{(a,b): 1 < ab < 2\}$. It allows concluding that there are no reasons for rejecting hypothesis that population dynamics corresponds to regime of monotonic stabilization at non-zero level. Like in previous case there are no points in the domain $\{(a,b): ab < 1\}$. Thus, observed dynamics doesn't correspond to the regime of population elimination.

Big number of red points belongs to the domain $\{(a,b): 3 < ab < 4\}$. It allows concluding that observed dynamic regime may have cyclic or chaotic nature. Thus, for both first variants (time series 2727 and 2728) we can point out similar dynamic regimes.

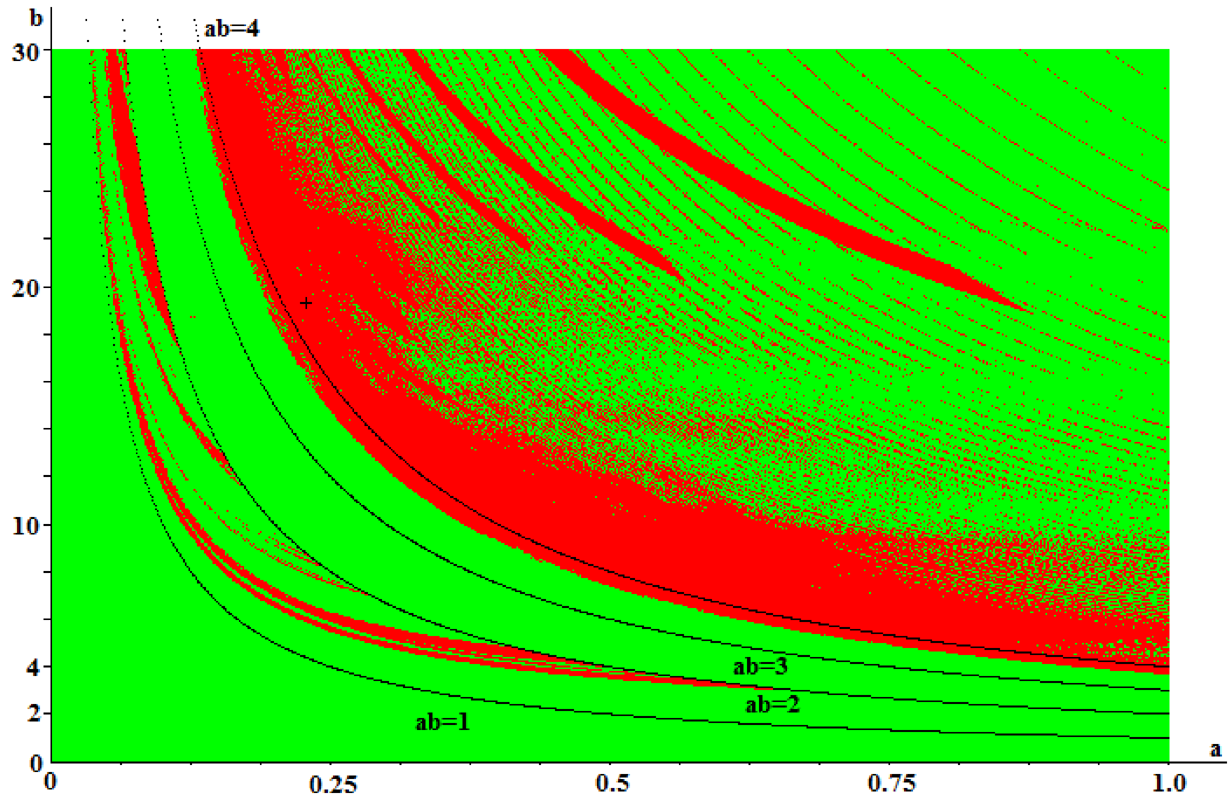

Fig. 4 Projection of feasible set (for 2728) on plane (a, b). Crest corresponds to point of minimum of functional (2). Functions $ab=1$, $ab=2$, $ab=3$, and $ab=4$ are basic bifurcation curves for model (3).

7 Time Series № 2729

For time series 2729 the following LSM-estimations were obtained:: $x_0 = 3.33$, $a = 1.64$, $b = 4.05$, $Q_{\min} = 15.4$.

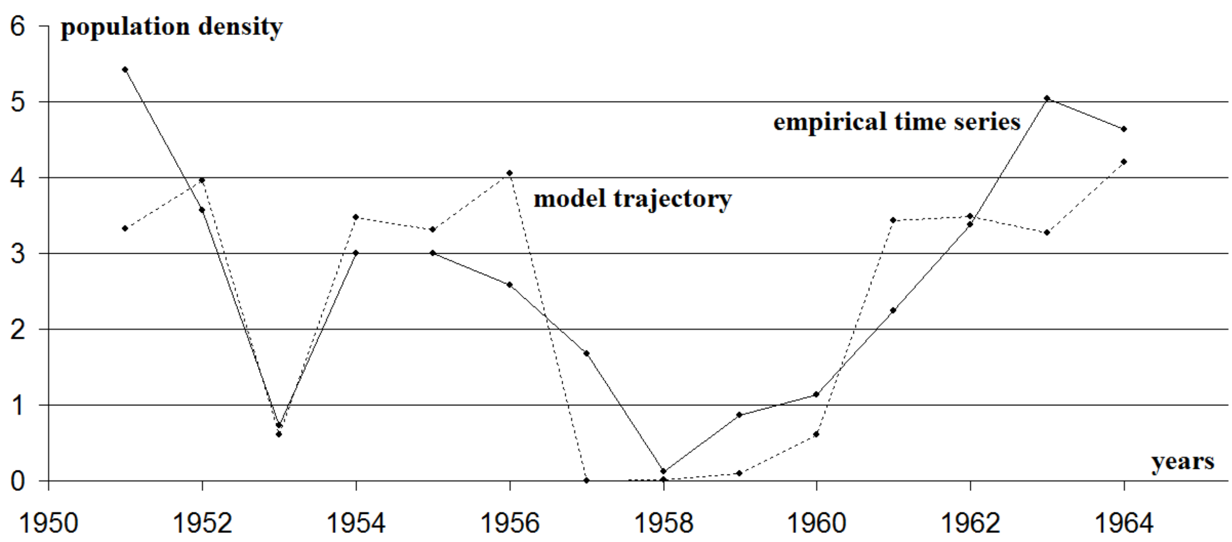

Fig. 5 Empirical dataset (2729, solid line) and model trajectory (broken line) calculated with LSM-estimations of model parameters.

For pointed out parameters behavior of empirical time series and model (3) trajectory are presented on figure 5. As we can see on this picture, model demonstrates similar behavior like initial sample. For these estimations of model parameters probability of event that distribution of deviations is Normal is greater than 0.2, $p > 0.2$ (Kolmogorov – Smirnov test and Lilliefors' test); Shapiro – Wilk test showed that this probability is rather big, $p = 0.68609$. Thus, hypothesis about Normality of distribution cannot be rejected even with 10% significance level. For deviations average plus/minus standard error is equal to -0.254 ± 0.282. Consequently, hypothesis about equivalence of average to zero cannot be rejected (with fixed 5% significance level).

Durbin – Watson test showed that $d = 1.967$. It allows concluding that there are no serial correlations in the sequence of residuals. Swed – Eisenhart test showed that $p = 0.413$: hypothesis about existence of serial correlation in the sequence of residuals must be rejected. Finally, within the framework of traditional approach to model parameter estimations we have to conclude that generalized discrete logistic model (3) is suitable for fitting of time series 2729.

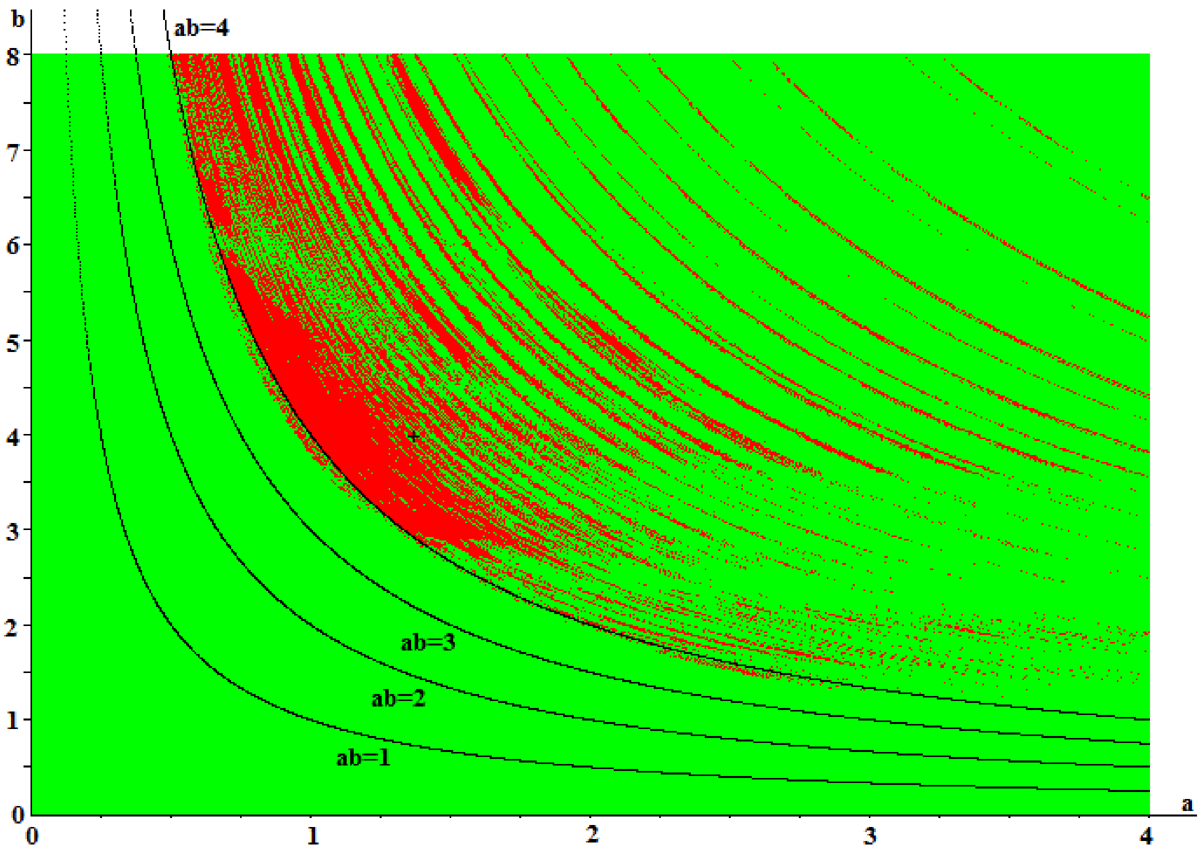

Fig. 6 Projection of feasible set (for 2729) on plane (a,b). Crest corresponds to point of minimum of functional (2). Functions $ab=1$, $ab=2$, $ab=3$, and $ab=4$ are basic bifurcation curves for model (3).

On figure 6 projection of feasible set onto the plane (a,b) is presented. As we can see on this picture, minimum of functional (2) doesn't belong to zone of maximum concentration of marked red points. And, additionally, this minimum belongs to zone $\{(a,b): ab > 4\}$. Like in both previous cases red points can be found in the domain $\{(a,b): 3 < ab < 4\}$: it gives a possibility to find common dynamic regime for all considering time series (taking into account that all time series correspond to dynamics of one and the same pine looper moth and in one and the same place, these samples must correspond to one and the same dynamic regime).

In considering case there are no red points in the domain $\{(a,b): ab \le 3\}$. It allows concluding that regime of asymptotic stabilization at non-zero level doesn't correspond to analyzing time series.

8 Conclusion

In current publication we compared two various approaches to estimation of non-linear model parameters on an example of generalized discrete logistic model (3) and time series on pine looper moth population dynamics (Klomp, 1966) in Netherlands. First of all, least square method (LSM) was used as traditional, widely used, and well-known method. Together with LSM standard statistical methods were used for analyses of properties of sets of deviations between theoretical and empirical datasets (tests for Normality of deviations, tests for absence/existence of serial correlation etc.).

Alternative approach to considering problem is inverse way to LSM. Within the framework of LSM we have to have any loss-function which depends on model parameters. Minimizing of loss-function can be considered as a first step of process, and it gives *best estimations* for model parameters. In other words, it gives one point from a space of model parameters, and final conclusion about suitability of model for fitting of empirical time series we make after application of pointed above statistical criterions to set of deviations obtained for model with these *best estimations*.

It is very important to note that loss-function has no relation to considering biological object or process, no relation to data collection etc. And it is not obligatory to use loss-function as a sum of squared deviations: we can use various modifications of this function. For example, we can summarize absolute values of deviations in any power; we can multiply squared deviations on any positive or non-negative weights etc. In every case we'll get different values of *best estimations of model parameters*.

It allows concluding that LSM is based on logic mistake. Before creation of loss-function, before minimizing of this loss-function we have to determine basic requirements to model and sets of deviations. In Chapter 2 the list of possible basic requirements is presented. These requirements are rather obvious and don't need in special explanations. For example, density function of deviations must be symmetric with respect to origin, and left and right branches of density function must be monotonic and so on. These obvious requirements allow obtaining (after application of respective statistical criterions) *feasible sets* of points in space of model parameters (red points on fig. 2, 4, and 6). Feasible sets contain points when we can observe correspondence (from standpoint of selected requirements to model) between theoretical (model) and empirical datasets.

In all three cases LSM led to results which belong to feasible sets but far from centers of subsets of highest concentration of red points of these feasible sets. Moreover, obtained LSM-estimations belong to feasible sets which correspond to situations when model cannot be used for forecast of population density dynamics. At the same time in all considered situations it is possible to point out red points which can be used for long-term forecast and belong to zone when discrete logistic model (non-generalized model) can be used for fitting of empirical time series.

References

Bard Y. 1974. Nonlinear Parameter Estimation. Academic Press, New York, USA

Bolshev LN, Smirnov NV, 1983. Tables of Mathematical Statistics. Nauka, Moscow, Russia

Draper NR, Smith H, 1998. Applied Regression Analysis. John Wiley and Sons Inc., New York, USA

Hollander M, Wolfe DA, 1973. Nonparametric Statistical Methods. John Wiley and Sons, New York, USA

Klomp H, 1966. The dynamics of a field population of the pine looper, *Bupalus piniarius* L. (Lep., Geom.). Advances in Ecological Research, 3: 207-305

Korn GA, Korn TM, 1973. Mathematical Handbook for scientists and engineers. Nauka, Moscow, Russia

Likesh J, Laga J, 1985. Basic Tables Of Mathematical Statistics. Finance and Statistics, Moscow, Russia

Lilliefors HW, 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. Journal of the American Statististcal Association, 64: 399-402

Nedorezov LV, 2011. Analysis of cyclic fluctuations in larch bud moth populations with discrete-time dynamic models. Biology Bulletin Reviews 72(2): 407-414.

Nedorezov LV, 2012. Chaos and Order in Population Dynamics: Modeling, Analysis, Forecast. LAP Lambert Academic Publishing, Saarbrucken, USA

Nedorezov LV, Sadykova DL, 2010. Analysis of population time series using discrete dynamic models (on an example of green oak leaf roller). Lesovedenie, 2: 14-26

Nedorezov LV, Utyupin YuV, 2011. Continuous-Discrete Models of Population Dynamics: An Analytical Overview. State Public Scientific-Technical Library, Russian Academy of Sciences, Novosibirsk, Russia

Sadykova DL, Nedorezov LV, 2013. Larch bud moth dynamics: can we explain periodicity of population fluctuations by the time lag-dependence in birth rate? Population Dynamics: Analysis, Modelling, Forecast 2(4): 154-181

Shapiro SS, Wilk MB, Chen HJ, 1968. A comparative study of various tests of normality. Journal of the American Statistical Association, 63: 1343-1372.