Basic analysis of population dynamics:
Models and correct statistical analysis of population time series

L.V. Nedorezov

Research Center for Interdisciplinary Environmental Cooperation RAS, Nab. Kutuzova 14, Saint-Petersburg 191187, Russia

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

Abstract

In paper basic approach to analysis of population dynamics is under consideration. Problems of well-known Least Squared Method (LSM) in its application to population time series analysis, is under consideration too. It was demonstrated that LSM is based on a number of unresolvable problems, and it allows concluding that LSM is deadlock way in statistics development. Considering basic approach was applied to several time series and several various models; obtained results were compared with results obtained with LSM.

Content

1. Introduction
2. Least Squared Method is a blind alley
    2.1. Why LSM is blind alley
    2.2. Way out from LSM blind alley
3. General arrangement of population time series analysis
4. Examples
    4.1. Hare-lynx system dynamics
    4.2. Fitting of time series on pine looper moth dynamics by trajectories of Moran – Ricker, discrete logistic and ELP models
5. Conclusion
6. References

1. Introduction

Constructing and analysis of mathematical models of population/ecosystem dynamics are among basic directions of ecological modeling development: knowledge of base properties of

models is very important information we have to have before its use on a practice. It is obvious that we have no reasons to use Kostitzin model (in literature this model has names "Beverton – Holt model", "Skellam model" etc.; Kostitzin, 1937; Beverton, Holt, 1957; Skellam, 1951; Nedorezov, 2012 a; Nedorezov, Utyupin, 2011 a) in a situation when population dynamics has obvious cyclic character: it is known a'priori that final result will be negative.

Other side of this process is following: constructing of new models, modification of old models, and modification of old modifications can be transformed into the end in itself. In other words, this other side of process can be transformed into games for mathematicians with determined topic which have no interests as for ecology as for mathematics.

If we want to escape from these "games for mathematicians" we have to be sure that every time we are talking about real models: we can point out real biological species with dynamics which corresponds to considering model on quantitative level. It can be achieved by the unique way of comparison of theoretical (model) results with empirical/experimental datasets (Isaev et al., 1984, 2001; Tonnang et al., 2009 a, b, 2010, 2012; Nedorezov, Lohr, Sadykova, 2008 and others).

Let's assume that model and real datasets (time series) were selected. And now we have to ask ourselves: what does it mean "to prove that we have good correspondence between dataset and model"? We have *to prove* but not *postulate* it (we constructed a model which takes into account influence of all basic factors, regulators, mechanisms… and it allows us concluding that model is very good for fitting and there is a good correspondence…). There is the only way for this proof: we have to point out a set of model parameters when model demonstrates sufficient approximation of dataset. But what does it mean "*sufficient approximation of dataset*"?

It is rather funny situation: further analysis of real and model datasets is based on faith (or on obvious properties-postulates – but it is the same). Below we'll talk about "good models" which can give sufficient approximation of dataset. First of all, *we believe* that probabilities of equal deviations to negative and positive parts are equal. The second, *we believe* that probabilities of big deviations are smaller of probabilities of small deviations. The third, *we believe* that all deviations must be values of independent stochastic variables. If we cannot prove that deviations are values of independent stochastic variables, it means that within the framework of model we didn't take into account any important factor (regulator, mechanism…) or it was taken into account but not in correct form. In most cases it is assumed that deviations are values of independent stochastic variables with Normal distribution (with zero average and fixed variance; Bard, 1979; Draper, Smith, 1986, 1987; Hudson, 1970; Bolshev, Smirnov, 1983; Lilliefors, 1967; Shapiro, Wilk, Chen, 1968 and many others).

To presented postulates we have to add one more: *we believe* that behavior of theoretical dataset must correspond to behavior of time series. If for every case when time series demonstrates increasing of values model shows its decreasing and vice versa, there are no backgrounds for conclusion about sufficient approximation of real dataset. Hence, *we believe* that in a situation when all pointed out properties for sets of deviations between theoretical and real datasets are observed, it gives us a background for conclusion about suitability of using of model for fitting of considering time series, for *explanation* of observed type of dynamics and so on. Note, that now we are talking about explaining models, in other words, about models which can be considered as laws of population dynamics, which allow providing of computer experiments etc. Models of this type cannot give best forecast: for this purpose it is much better to use simple asymmetric filters which can give good short-term prognosis but can explain nothing. Asymmetric filters are rather popular in ecological modeling (see, for example, Stenseth, Bjornstad, Falck, 1996; Stenseth, Bjornstad, Saitoh, 1998 and others) but these models cannot be used for explanation of population dynamics, cannot be used for providing of computer experiments and so on.

It is real situation when postulates are satisfied for various values of model parameters, and these parameters correspond to various dynamical regimes. We have to ask ourselves: what is right dynamic regime? It is rather typical situation (see, for example, Nedorezova, Nedorezov, 2011, 2012; Nedorezov, 2011 a, b, 2012 a, b and others), and it can be illustrated by following example. Patient came to physician and told him about problems with health. Physician worked with medical expert system, and put into this system initial information: temperature of patient, blood pressure etc. In a result of analysis of initial information expert system presents a set of possible diseases, and for every disease expert system points out respective probability. It is naturally to ask: who can be an author of final diagnosis – expert system or physician? Answer is obvious.

Consequently, for selected model and time series we have to point out estimations of probabilities of realization of one or other dynamic regime for population. This is basic goal of providing of analysis. We can point out estimations of model parameters after presentation of final diagnosis by specialists.

One of basic problems is following: we have to choose a set of statistical criterions which must allow concluding (with strong background) about sufficient or insufficient fitting of time series by model. It is obvious that it is impossible to indicate a set of statistical criterions which must be used in all situations. For every problem we can find a lot of various statistical criterions which can be used for solution of problem. Thus, every investigator must choose criterions on the base of his or her experience.

After selection of statistical criterions we have to check properties of points in interesting (for us) zone of space of model parameters: we have to be sure that it is possible to point out points where statistical criterions are satisfied (for selected levels of significance). In other words, we have to be sure that set of model parameters which allow obtaining of sufficient approximation of time series, isn't empty. If we cannot find such values of parameters it gives us good background for conclusion that model isn't suitable for fitting of considering time series.

As it was noted above, next important step of process of population dynamics analysis is obtaining of estimations of probabilities of realization of one or another dynamic regime. It requires of knowledge of bifurcation surfaces in the same domain of space of model parameters. If we know it, we can model a point with uniform distribution within limits of set where all used statistical criterions demonstrate good results – this procedure will allow estimating of required probabilities.

2. Least Squared Method is a blind alley

Least Squared Method (LSM) is one of the basic methods in statistical analysis of processes of various natures. It was created and developed by famous scientists: J.C.F. Gauss, A.-M. Legendre, P.-S. de Laplace, F.W. Bessel, A.A. Markov, A.N. Kolmogorov. LSM was lucked out to be a method which will allow obtaining famous results. And it is true – with the help of this method a lot of famous results were obtained, and it can be found as in special scientific literature as in various textbooks (Bard, 1979; Draper, Smith, 1986, 1987; Hudson, 1970; Ivchenko, Medvedev, 1984; Lakin, 1990 and many others). Thus, there are no reasons to talk about famous results obtained with the help of LSM one more time. But we have a certain necessity to talk about problems which arise every time we use this method.

2.1. Why LSM is blind alley

Let's assume that we have scalar time series $\{x_k^*\}$, $k = 1,2,...,N$, where N is sample size. Using of this scalar time series we have to estimate values of dynamic model parameters:

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

In (1) $\vec{\alpha}$ is vector of model parameters, $\vec{\alpha} \in A$, A is a space of model parameters; F is non-linear function, x_k is a value of scalar variable at k th time moment, x_0 is initial value of model variable. This problem is typical, for example, for various ecological problems (in particular, when x_k is population size in k th year; Korzukhin, Semevskii, 1992; Nedorezov, Utyupin, 2011 a; Nedorezov, 2011 a, b, 2012 a, b; McCallum, 2000; Moran, 1950; Ricker, 1954; Pielou, 1977 and many others).

If we use canonical way, for the first step we have to construct functional (loss-function), for example, in the following form:

$$Q(\vec{\alpha}) = \sum_{k=1}^{N-1} (x_{k+1}^* - F(x_k^*, \vec{\alpha}))^2 . \tag{2}$$

Every element of expression (2) is a deviation between element of initial sample and respective value obtained with the help of model (1). For obtaining of best estimations of model parameters we have to minimize functional Q (2). This idea is based on our faith: it seems (for us) rather natural that good model with best values of model parameters allows obtaining global minimum of sum of squared deviations between theoretical (model) and empirical/experimental datasets.

Two important questions arise now. The first, what is a relation between functional (2) and considering (biological, chemical, medical, physical…) problem? The second, what is real flavor of functional (2)?

Answer on first question is obvious: functional (2) has no relation to considering problem. It has relation to *our imagination* (about good models, good estimations etc.) only. More precisely, basic phrase (basic idea) must be presented in following form: It seems for us that minimization of functional (2) gives us best values of model parameter estimations; all other values of model parameters are much more badly. Note, that keywords of this phrase are "it seems for us…".

Answer on the second question isn't so obvious. Analysis of literature shows that researches use various types of loss-functions for estimating of model parameters. It is based on obvious idea – *nobody said that we have to use squared deviations* between theory and experiment. Next loss-function looks better than functional (2):

$$Q(\vec{\alpha}) = \sum_{k=1}^{N-1} \left| x_{k+1}^* - F(x_k^*, \vec{\alpha}) \right|^\gamma . \tag{3}$$

In expression (3) γ is any positive number, $\gamma = const > 0$, and it isn't obligatory equivalence $\gamma = 2$. When researches want to take into account an influence of small deviations on estimations of model parameters they can introduce so-called weights into expressions (2) or (3):

$$Q(\vec{\alpha}) = \sum_{k=1}^{N-1} w_k (x_{k+1}^* - F(x_k^*, \vec{\alpha}))^2 . \tag{4}$$

In (4) $w_k = const \geq 0$. Note that these weights can depend on amounts of deviations. We can add to loss-functions (2)-(4) other expressions where modified deviations are also used (but after various non-linear transformations: after logarithmic transformation, double logarithmic transformation etc.), where deviations are calculated between initial sample and model trajectories (global fitting; Wood, 2001 a, b) and so on. After all it becomes obvious that nobody

knows what kind of functional we have to use in one or other situation. Moreover, nobody knows about criterions we can use for selecting of type of minimizing loss-function. And nobody can say about criterions of selection of weights – researches use their own imaginations about process and existing experience.

Thus, after honest answers on two first question we can see that situation with LSM isn't so iridescent like it could seemed at the very beginning. For one and the same problem we can use various minimizing loss-functions – does it mean that we have to have a lot of famous results? If we obtained a famous result for determined loss-function – does it mean that this result will be extraordinary for all other functions? It looks unbelievable…

Let's assume that for functional (2) estimations $\vec{\alpha}^*$ of model parameters were found, and it corresponds to global minimum of Q. On the next step of analysis of correspondence between model and considering datasets we have to provide investigation of properties of set of deviations $\{e_k\}$ where

$$e_k = x_{k+1}^* - F(x_k^*, \vec{\alpha}^*).$$

In most cases it is assumed that deviations $\{e_k\}$ must have Normal distribution with zero average, and in a sequence of residuals serial correlation cannot be observed (Bard, 1979; Draper, Smith, 1986, 1987; Hudson, 1970; Lilliefors, 1967; Shapiro, Wilk, Chen, 1968). If (for selected levels of significance) one of these conditions isn't truthful it gives a background for conclusion that model isn't suitable for fitting of time series. In particular, existing of serial correlation in a sequence of residuals can correspond to situation when in model any important factor (regulator) was not taken into account (or was taken but in incorrect form). Respectively, model needs in modification. If hypothesis about equivalence of average to zero must be rejected it means that there is a regular mistake in process of data collection (if model is correct), or estimations of model parameters are not good (sample isn't good), or selected model doesn't correspond to considering process.

In other words, within the limits of traditional approach we use one point from a space of model parameters $\vec{\alpha}^* \in A$ for creation a conclusion about suitability or unsuitability of model for approximation of datasets. All the more, this point $\vec{\alpha}^*$ was found at minimization of loss-function which has no relation to datasets, model and problem.

In our opinion, requirement about Normality of deviations $\{e_k\}$ is rather strong, and at the same time absolutely incorrect. For example, if elements of initial sample $\{x_k^*\}$ were estimated in grams probability of event that we have an error in several tons or x_k^* is negative for any k cannot be positive at all. But postulating of Normality of distribution for deviations means

that pointed out *events can be observed with positive probabilities*. It is very interesting to look onto researcher who can get negative weight… Thus, apriori assumption about Normality of distribution of deviations is nonnormal.

When we construct model of (1) type we try correctly to take into account influence of various mechanisms and/or processes onto changing of value of variable. At the same time we try to take into account all obvious requirements: if we are talking about population size we don't allow being it negative within the framework of model; we don't allow being it extremely big; we take into account relations/balance between various age groups and so on. And after constructing of model we begin to check a correspondence of model to foolish requirements…

Funny side of a situation (with checking of Normality of deviations) consists in following: everybody knows and understands that nobody talks about Normality. If Null hypothesis (about Normal distribution of deviations) cannot be rejected for selected significance level it means that with a certain probability distribution is close to Normal. But it isn't obligatory that properties which are observed for normally distributed stochastic variables can characterize stochastic variables *with close to Normal distribution*.

It looks natural to check hypothesis about symmetry of deviations $\{e_k\}$ (with respect to origin). For this reason we can use criterions about likeness of two samples. Additionally, we have to check hypothesis about monotonic behavior of branches of density function (monotonic decreasing behavior for positive values of argument, and monotonic increasing behavior for negative values). It corresponds to situations when bigger deviations can be observed with smaller probabilities.

Problems pointed out above underline the necessity in changing of paradigms existing in modern statistics (Orlov, 2004, 2015). Changing of paradigm must allow solution of pointed out problems. Without it (if we will use methods of accomplishment of blind alley) solutions will never be found: blind alley will be a blind alley even after total accomplishment.

2.2. Way out from LSM blind alley

Taking into account all problems with use of LSM pointed out above we can conclude that we have to be careful and awake with all results obtained with use of LSM. LSM is a logic blind alley. And we have to note that method of maximum likelihood is a blind alley too (reasons for this conclusion are the same).

Where is the way out of this blind alley? First of all, it looks rather natural and logic to forget about constructing of any loss-functions of (2)-(4) type. Instead of this we have to scan interesting (for us) domain of a space of model parameters A. For example, it can be provided in

the following manner: for every selected point (it can be selected in a result of stochastic process) of a space of model parameters we have to calculate a set of deviations $\{e_k\}$, and if these deviations have a certain set of properties (all statistical criterions are satisfied) we have to mark this point as belonging to a *feasible set*. It is obvious that *limits and structure of feasible set* depend on levels of significance which are used for testing of deviations. It isn't obligatory for statistical criterions to have one and the same level of significance.

If feasible set is empty it gives us a background for conclusion that considering model isn't suitable for fitting of time series. If feasible set is not empty we have to find (among elements of this set) points with extreme properties: from the stand point of using criterions model with such parameters will have best correspondence to considering datasets. Analysis of real time series showed that in many cases LSM-estimations didn't satisfy to statistical criterions but feasible sets were not empty (Nedorezov, 2013, 2014 a, 2015 a, 2016). Moreover, in some cases LSM-estimations were very close to elements of feasible sets: in such situations LSM-estimations can be modified a little for obtaining better results (Nedorezov, 2011 a, b, 2012 a, b).

It is important to mark it one more time − considering approach is one of possible variants of estimation of model parameters without use of LSM. But considering approach allows to solve one more problem of LSM which appears when we have to use several correlated time series (for example, when we have time series for predators and preys) which we have to use together for estimation of model parameters (for example, for estimation of parameters of Lotka − Volterra model of predator-prey system dynamics, of parasite-host system dynamics etc.; Volterra, 1931; Lotka, 1920, 1925; Ivanchikov, Nedorezov, 2011, 2012). In this case we have to follow to following basic principle: point of a space of model parameters belongs to feasible set if and only if all corresponding sets of deviations are satisfied to selected statistical criterions.

After determination of points of feasible set we have to provide additional investigation of properties of elements of this set. As it was pointed out above, for estimation of model parameters we have to find elements of feasible set which have extreme properties.

But before starting a process of finding of elements with extreme properties we have to understand − what does it mean *point with extreme properties*? If under testing of hypothesis (for example, hypothesis about symmetry of distribution of deviations) we cannot reject Null hypothesis (hypothesis about equivalence of distribution functions for positive deviations e_k and negative deviations taking with sign minus) with significance level in 5% − it doesn't mean that we have to accept Null hypothesis. It means that we have no reasons for rejecting of Null hypothesis − and that's it. For the same hypothesis stronger result is observed if we cannot reject

Null hypothesis with 20% significance level. Finally, if we cannot reject Null hypothesis with 95% significance level it means that *we have to accept this hypothesis*.

3. General arrangement of population time series analysis

As it was noted above, structure of feasible set depends on selected levels of significance for statistical criterions which are used for testing of deviations between theoretical (model) and real datasets. Let's assume that set of significance levels was selected. After that we get a feasible set, and every element of this set has required properties: for every element of feasible set model (1) gives sufficient approximation of dataset. Respectively, we can commit to every element of this set, and in this occasion feasible set plays a role of *confidence domain*.

Note that when we use traditional approach to finding boundaries of confidence domains (Draper, Smith, 1986, 1987) for estimations of model parameters we use Fisher distribution and sections of loss-function at any heights. But as Fisher distribution as sections of loss-function at any heights have no relation to considering problem, to initial sample, to… − natural question arises at this moment: what's a set we obtain after all operations? What's a role it can play in statistical analysis if it has no relation to all basic elements − problem, model, and time series?

In a space of model parameters A we have also a set of bifurcation surfaces which separate various dynamic regimes. If for model (1) initial value of variable x_0 is known (for example, like in experiments provided by G.F. Gause; Gause, 1934) we have one and the same space A for feasible set and bifurcation surfaces. If we have to estimate x_0 that dimension of space for feasible set is equal to $\dim A + 1$. In all situations bifurcation surfaces cut feasible set onto subsets corresponding to various dynamic regimes.

Thus, general analysis of population dynamics contains following basic stages:

1. Selection of mathematical model for approximation of time series.

2. Selection of domain Δ in a space of model parameters where we have to find points of feasible set (which allow obtaining of sufficient approximation by model).

3. Determination of elements of feasible set (in Δ):

3.1. Checking of symmetry of distribution of deviations.

For every stochastic point with uniform distribution in Δ we have to check properties of set of deviations between theoretical (model) and real dataset. Let $\{e_k^+\}$ be a set of positive deviations, and $\{-e_k^-\}$ be a set of negative deviations (with sign minus); $F_1(x)$ is a distribution function for $\{e_k^+\}$, $F_2(x)$ is a distribution function for $\{-e_k^-\}$. It is obvious, deviations $\{e_k\}$ have symmetric distribution (with respect to origin) if for selected significance level we cannot reject Null

hypothesis H_0 : $F_1(x) = F_2(x)$ with alternative hypothesis H_1 : $F_1(x) \neq F_2(x)$. For this purpose Kolmogorov – Smirnov, Mann – Whitney, Lehmann – Rosenblatt, Wald – Wolfowitz and other criterions of homogeneity of two samples can be used (Hollander, Wolfe, 1973; Bard, 1979; Draper, Smith, 1986, 1987; Hudson, 1970; Bolshev, Smirnov, 1983; Likes, Laga, 1985).

3.2. Checking of monotonic behavior of branches of density function can be provided in two different variants. If sample size is rather big then we can check pointed out property for deviations $\{e_k^+\}$ and $\{-e_k^-\}$ separately. If sample size is small then we can check pointed out property for set $\{e_k^+\} \cup \{-e_k^-\}$.

Monotonic decreasing of density function (for positive values of argument) means that bigger values (in sample) must be observed with smaller probabilities. Respectively, if $\{e_k^{+*}\}$ is ordered sample of positive deviations,

$$0 \leq e_1^{+*} \leq e_2^{+*} \leq ... \leq e_m^{+*},$$

where m is sample size (number of positive deviations) then for lengths of intervals

$$[0, e_1^{+*}], \ [e_1^{+*}, e_2^{+*}], \ ... \ , \ [e_{m-1}^{+*}, e_m^{+*}],$$

we have to have the similar order in ideal situation: rank 1 will correspond to shortest interval $[0, e_1^{+*}]$, biggest rank m will correspond to biggest interval $[e_{m-1}^{+*}, e_m^{+*}]$. Ideal case we must compare with real situation which is determined by sample $\{e_k^{+*}\}$. For this reason we have to calculate Spearmen rank correlation coefficient ρ (and/or Kendall correlation coefficient τ), and check Null hypothesis H_0 : $\rho = 0$ with alternative hypothesis H_1 : $\rho > 0$. For selected significance level Null hypothesis must be rejected. Note, we have stronger result in a case when we can reject Null hypothesis with smaller significance level.

3.3. Checking of existence/absence of serial correlation in sequence of residuals.

As it was noted above existence of serial correlation in sequence of residuals means that there is a process (or regulator, or mechanism…), which wasn't taken into account in model (or was taken into account but in incorrect form). In this situation we have to modify model or to choose (construct) another one.

Testing of sequence of residuals on existence/absence of serial correlation we can provide with various methods. In particular, we can analyze behavior of autocorrelation function. Popular Durbin – Watson test cannot be recommended for use because it is based on assumption of Normality of deviations (Draper, Smith, 1986, 1987). It is much better to use Swed – Eisenhart test and test "jumps up – jumps down" (Hollander, Wolfe, 1973; Likes, Laga, 1985).

3.4. "Similar behavior" of model and time series.

Testing of "similar behavior" is one of important elements of analysis of correspondence between model and dataset. If model demonstrates decreasing of values in cases when time series increases and vice versa, it gives us a background for conclusion about unsuitability of model for fitting of dataset. In general case we have to determine a frequency ρ of successful outcomes (number of cases when signs of respective increments for model and time series are equal), and check Null hypothesis H_0 : $\rho = 0.5$ with alternative hypothesis H_1 : $\rho > 0.5$. For "good" model Null hypothesis must be rejected.

But we have to be careful when we use this test. For example, if model demonstrates slow increasing and time series shows fluctuations with big variance near this nonlinear trend (and as a result we can observe alternation of signs of increments) we can get a frequency ρ which is close to 0.5 (and we'll have to conclude that models isn't suitable for fitting of time series) in a case when model corresponds to modeling process.

4. Estimation of probabilities of realization of one or other dynamic regime.

Every element of feasible set can be used as estimations of model parameters: used statistical tests don't allow us making negative conclusion. Bifurcation surfaces are boundaries of sub-domains which correspond to various dynamic regimes. Determination of relations of points of feasible set in respective sub-domains (under pure stochastic search) to total number of points of feasible set allows obtaining estimations of required probabilities.

5. Search of dynamic regimes which correspond to elements of feasible set with extreme properties. Idea of search of these regimes is rather simple: elements of feasible set with extreme properties allow obtaining results of using of model which demonstrate best correspondence to time series.

Search of points with extreme properties can be provided, for example, after *ranking* of used statistical criterions. Let's assume that Kolmogorov – Smirnov test is most important for us. Respectively, we can find (in feasible set) point (or points) when Null hypothesis cannot be rejected with biggest value of significance level. If we have a lot of points with extreme property (from the stand point of Kolmogorov – Smirnov test) and Lehmann – Rosenblatt test is a second in our ranking list, we can continue a procedure of search and find point (or points) when Null hypothesis cannot be rejected with biggest value of significance level (for Lehmann – Rosenblatt test) and so on.

4. Examples

4.1. Hare-lynx system dynamics

One of most interesting (and important from a stand point of development of mathematical ecology) is a problem of approximation of well-known time series collected by Hudson's Bay company (Elton, Nicholson, 1942; Chitty, 1948; Leigh, 1968; Brauer, Castillo-Chavez, 2001; Gilpin, 1973; Gilpin, Ayala, 1973; Rosenberg, 2010; Akcakaya, 1992; Inchausti, Ginzburg, 2002 and many others). In this case we have two correlated time series for population of hares (*Lepus americanus*) and population of lynxes (*Lynx lynx*), and we have to be sure that dynamics of this system can be described by trajectories of well-known Lotka – Volterra model of predator – prey system dynamics (without taking into account of influence of self-regulative mechanisms on system dynamics; Lotka, 1920, 1925; Volterra, 1931):

$$\frac{dx}{dt} = ax - bxy, \quad \frac{dy}{dt} = -cy + dxy. \qquad (5)$$

In (5) $x(t)$ is a number of preys; $y(t)$ is a number of predators at moment t. Coefficient a is Malthusian parameter of population of preys (below we will assume that $a > 0$; if $a \leq 0$ origin is global stable state: system eliminates for all initial values of population sizes). Coefficient c is Malthusian parameter of population of predators. Coefficients b and d are parameters of population interactions. In (5) all coefficients are non-negative, $a, b, c, d = const \geq 0$.

In figure 1 there are trajectories of changing of population sizes of hares and lynxes in time (in thousands of collected skins).
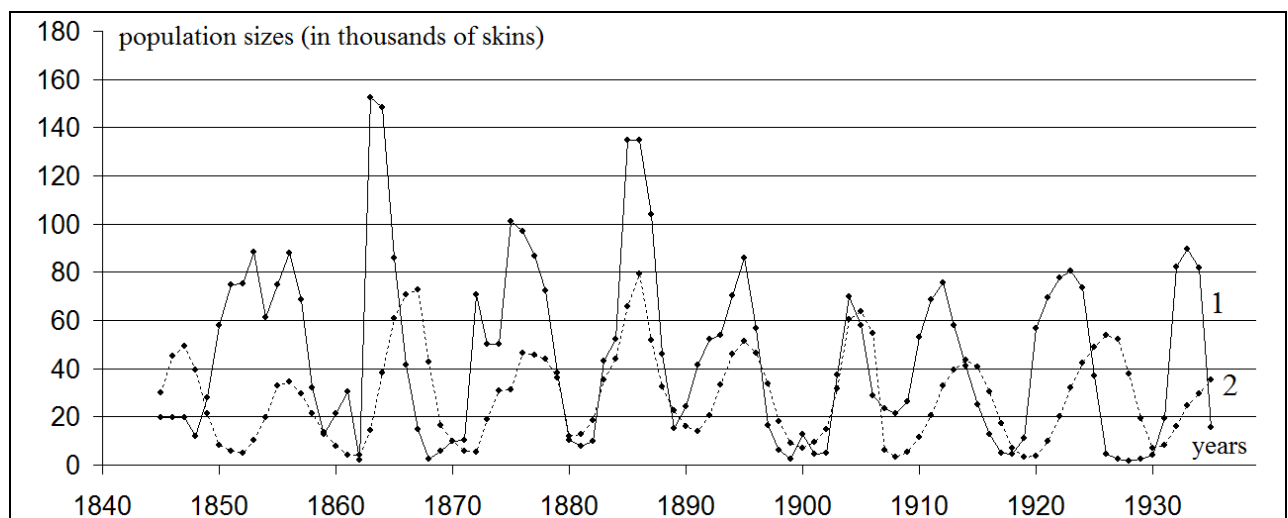

Fig. 1. Dynamics of hares and lynxes in Canada, 1845 – 1935 года. 1 – curve of number of hares changing. 2 – curve of number of lynxes changing (broken line).

M.E. Gilpin (1973) considered a model which contains (5) as a particular case. For estimation of model parameters he followed by methodology which is presented in publications by F. Ayala with co-authors (Ayala, Gilpin, Ehrenfeld, 1973; Gilpin, Ayala, 1973). Use of this approach is based on following procedure: we have to change left sides of equations (for example, of model (5)) on increments of respective populations for one year. After that estimations of parameters can be obtained in a result of minimization of sum of squared deviations. We have to note that within the limits of this approach we don't use original model – we use discrete analog of model which we obtain with Euler scheme (approximation of derivative). Moreover, for approximation of *empirical trajectory* we don't use *trajectory of model* (which is closest to empirical time series) – we use set of values of other nature.

Dimension of space of model parameters is equal to 6 (under conditions pointed out above), and there are no bifurcation surfaces: for all possible values of model parameters non-trivial stationary state is center, and positive part of phase space contains closed curves only. There is no a real time in models of (5) type, thus we can postulate that unit of time in model is equal to one year of real time.

In figures 2 and 3 there are projections of 30000 points of feasible set which were found during a pure stochastic search (with uniform distribution) in Δ: $x_0, y_0 \in [0,220]$, $a \in [0,16]$, $b \in [0,6]$, $d \in [0,3]$, $c \in [0,14]$. For testing of sets of deviations (for both time series) Kolmogorov − Smirnov test and Lehmann − Rosenblatt test were used for checking of symmetry; Spearmen and Kendall rank correlation coefficients were used for testing of monotonic behavior of branches of density functions; Swed − Eisenhart test was used for checking absence/existence of serial correlations in sequences of residuals.
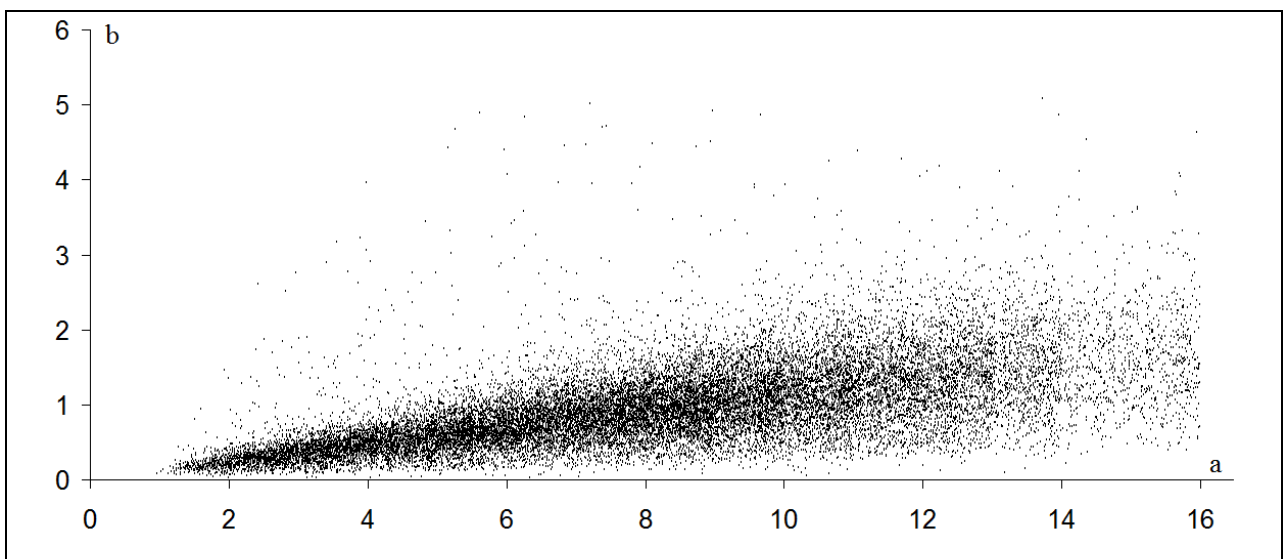

Fig. 2. Projection of feasible set on plane (a,b).

Search of points of feasible set was provided out of limits of set Δ. But it didn't allow obtaining positive results.
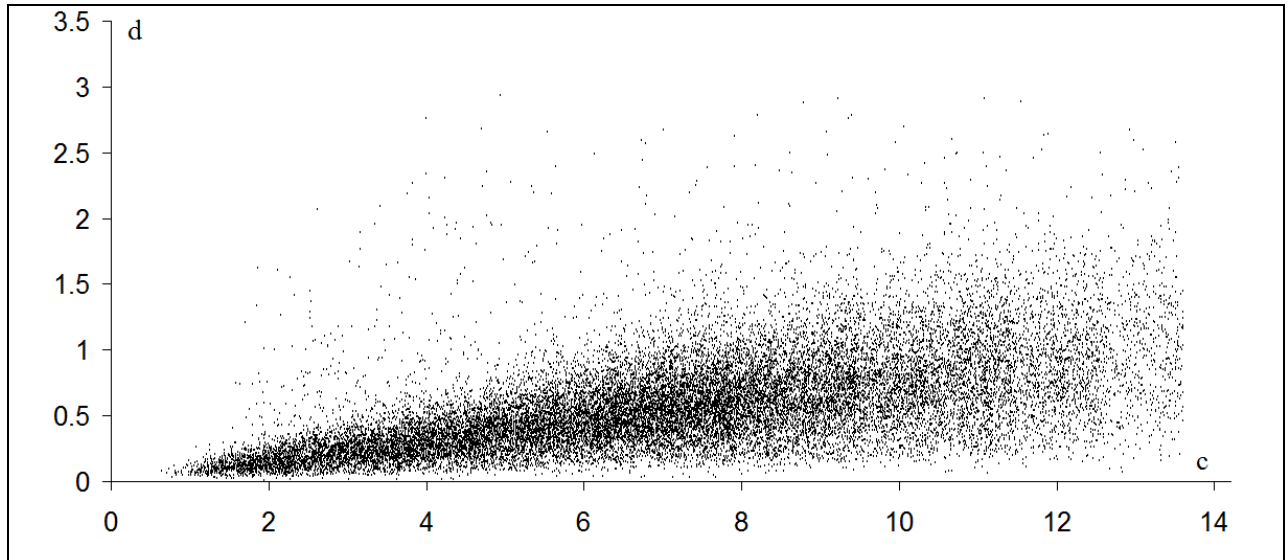

Fig. 3. Projection of feasible set on plane (c, d).

Thus, feasible set (obtained for 5% significance level for all used statistical criterions; fig. 2 and 3) is not empty. Additionally it was assumed that sample sizes for positive and negative deviations must be bigger than 10. Now we can conclude that result presented in publication by M. Gilpin (1973) isn't correct and doesn't correspond to reality: trajectories of model (5) can allow obtaining good fitting of time series without additional assumption that *hares are bloody predators*.

4.2. Fitting of time series on pine looper moth dynamics by trajectories of Moran – Ricker, discrete logistic and ELP models

Datasets on dynamics of pine looper moth (*Bupalus piniarius* L.) in Netherlands (Klomp, 1966) haven't scientifically based explanation of observed dynamic type (Isaev et al., 1984, 2001). For this species we have three correlated time series for egg's densities, larva's densities, and pupa's densities. In numerical format these time series are presented in GPDD, № 2727, № 2728, №2729.

Approximation of these datasets can be provided in two different ways. First, we can choose rather simple model (for example, Moran – Ricker model, discrete logistic model etc.; Moran, 1950; Ricker, 1954), and try to fit every time series by trajectories of this model. It is obvious, good model must demonstrate sufficient approximation for all time series. It looks

unreal when one sample can be approximated by one model, and another sample can be approximated by another one.

Second, we can construct a model which takes into account relations between measured variables (for example, like LPA-model; Costantino et al., 1997; Dennis et al., 2001; Desharnais et al., 2001). In this case we have to organize independent testing of all time series of deviations. If in the first case we can construct any loss-functions of (2)-(4) type, in the second case constructing of loss-functions looks like artificial and senseless activity.

Approximation of time series by trajectories of Moran − Ricker model (Moran, 1950; Ricker, 1954):

$$x_{k+1} = ax_k e^{-bx_k}, \; a,b = const \geq 0, \tag{6}$$

shows that for some of time series feasible set can be very rich (fig. 4), but for another one it can be very poor − feasible set contains several points only. In (6) a is a maximum value of birth rate, b is coefficient of self-regulation.
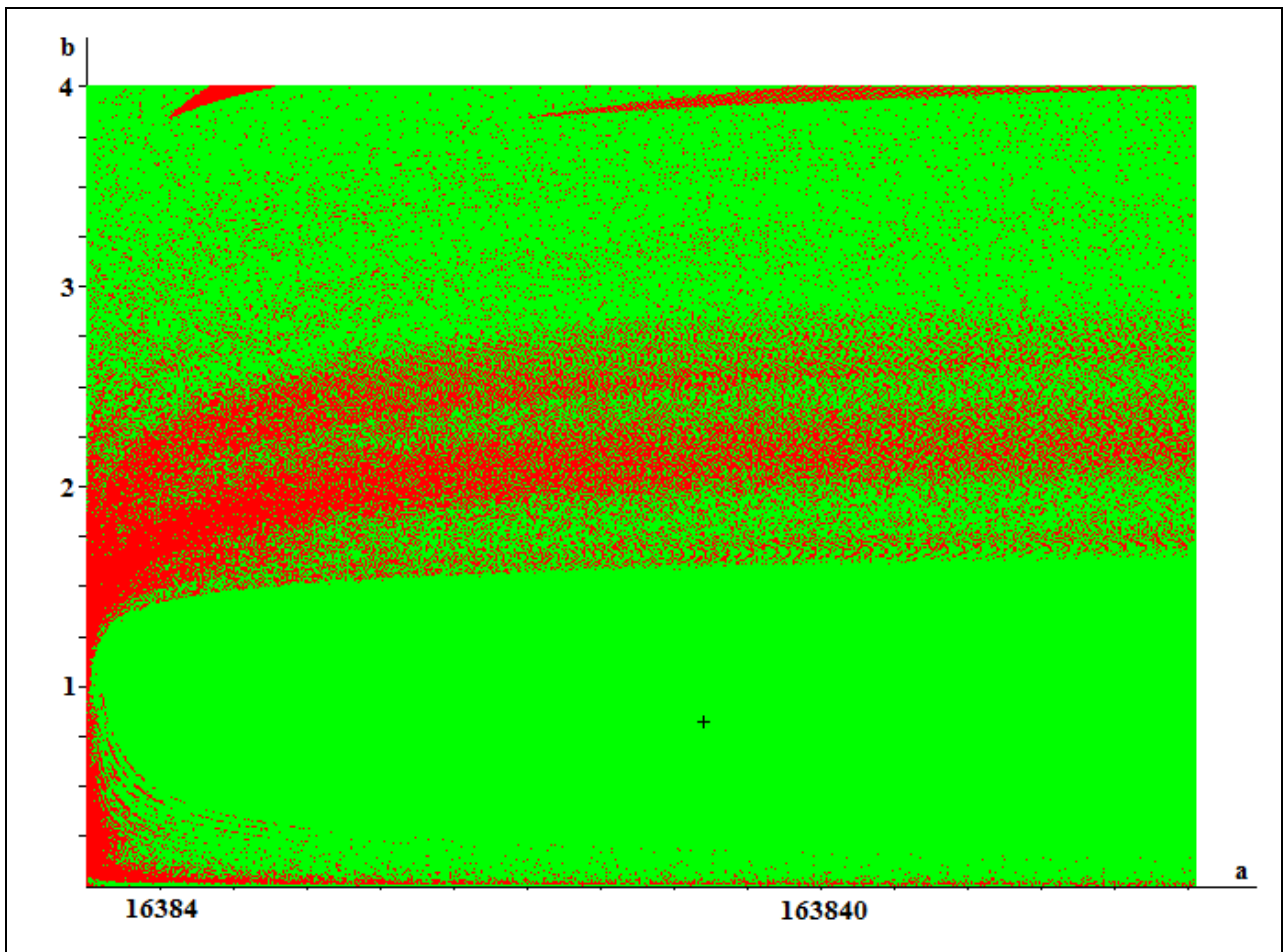

Fig. 4. Projection of feasible set (for time series 2728) onto plane (a,b). Crest corresponds to point of minimum of loss-function which is equal to sum of squared deviations between values of time series and trajectories of model (6). Projections of points of feasible set are marked by red color.

As we can see in figure 4 global fitting of time series 2728 gives results in "non-biological" zone (extremely big value of parameter a). Additionally, set of deviations doesn't satisfy to used statistical criterions. At the same time feasible set isn't empty and contains points in "biological zone".

More interesting results were obtained when generalized logistic model were used for approximation of considering datasets:

$$x_{k+1} = \begin{cases} ax_k(b-x_k), 0 \le x_k \le b, \\ 0, b < x_k. \end{cases}, \quad a,b = const \ge 0. \tag{7}$$

Use of model (7) showed that feasible sets in all cases are rather rich (for time series 2727 feasible set is presented in figure 5). LSM-estimations (obtained with global fitting; Wood, 2001 a, b) are satisfied to statistical criterions. But in all cases LSM-estimations are out of zones of maximum concentration of points of feasible set.
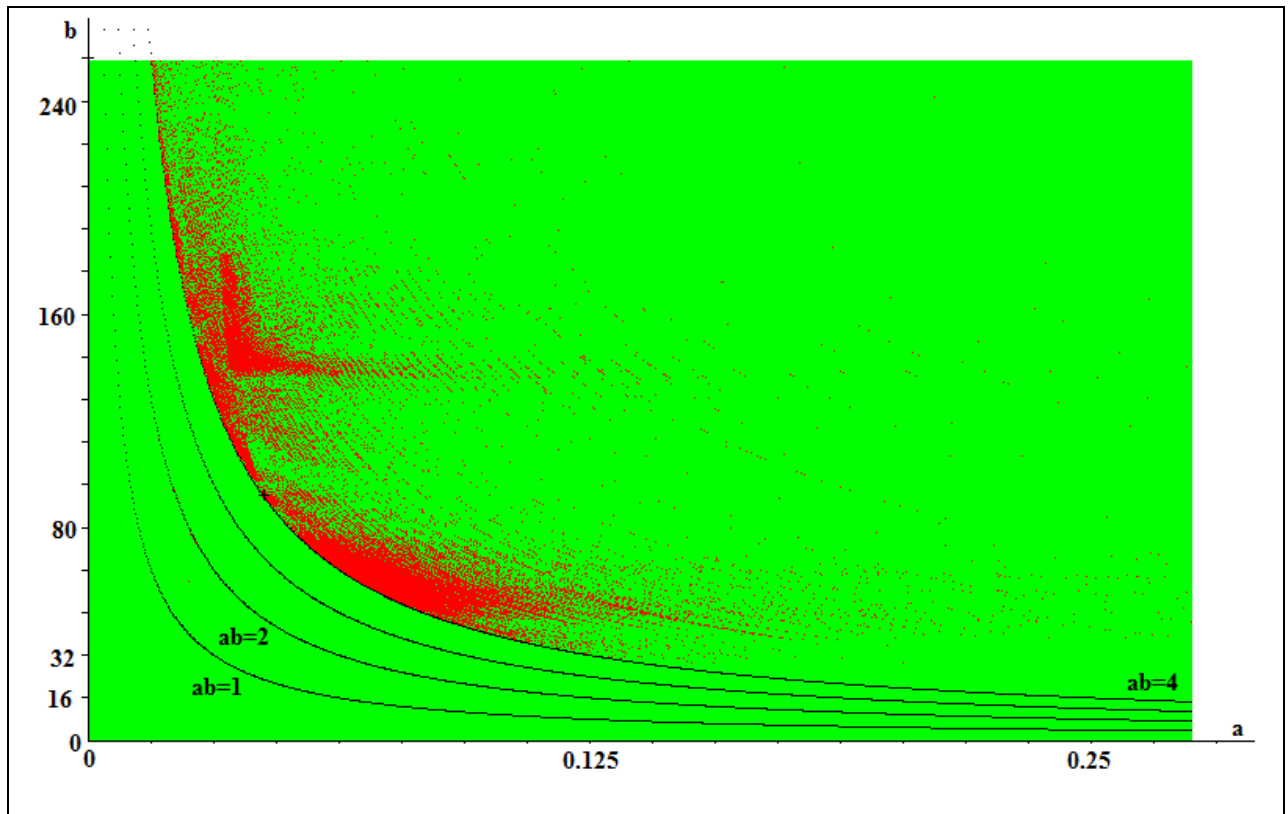

Fig. 5. Projection of feasible set (for time series 2727) onto plane (a,b). Crest corresponds to point of minimum of loss-function which is equal to sum of squared deviations between values of time series and trajectories of model (7) (global fitting). Projections of points of feasible set are marked by red color. $ab=1$, $ab=2$, $ab=3$, $ab=4$ are bifurcation curves of model.

Most interesting results were observed in a case when time series were approximated by trajectories of ELP-model (egg – larvae – pupae) (Nedorezov, 2014 b, 2015 b). In this case three

time series of deviations were tested independently, and testing point belonged to feasible set if and only if all time series satisfied to statistical criterions:

$$E_{k+1} = C\mu_1 P_k,$$

$$L_{k+1} = \mu_2 E_{k+1},$$

$$P_{k+1} = L_{k+1} e^{-\alpha L_{k+1}}. \tag{8}$$

In (8) P_k is density of pupae at time moment k, $k = 0,1,2...$, L_k is density of larvae, and E_k is density of eggs. Functions in right-hand side of equations (8) are following:

$$\mu_1 = \frac{g_1}{1 + g_2 L_k^{g_3}}, \quad C = \frac{c_1}{1 + c_2 L_k^{c_3}}. \tag{9}$$

All parameters in formulas (8)-(9) are non-negative. Function μ_1 is equal to quota of pupae survived during a winter period, $g_1 = const \leq 1$. Function C is equal to average productivity of butterflies. Within the framework of model (8) it is assumed that death rate of pupae during a winter time and productivity of butterflies depend on food conditions of larvae during respective vegetation period. Parameter μ_2 is quota of successful appeared larvae, $\mu_2 = const \leq 1$; α is coefficient of self-regulation (in considering case it is coefficient of intra-population competition for food).

In model (8)-(9) there are eight unknown parameters. To unknown amounts we have to add initial value of density of eggs E_0 (it is sufficient for starting of the model). Thus, using existing time series we have to estimate 9 unknown parameters.

In figure 6 there is a feasible set (presented in projection on plane (c_1, g_1)) obtained with the help of stochastic search (of points of feasible set) with uniform distribution in domain $E_0 \in [0,200]$, $\mu_2 \in [0,1]$, $c_1 \in [0,800]$, $c_2 \in [0,10]$, $c_3 \in [0.5,2.5]$, $g_1 \in [0,1]$, $g_2 \in [0,6]$, $g_3 \in [0.1,4.1]$, $\alpha \in [0,2]$ (and with 5% significance level for all statistical criterions). Curve $c_1 g_1 = 1$ is bifurcation line, boundary of zone of population extinction. 10^6 is total number of points of feasible set in fig. 6. Note, that there are no points in the domain $c_1 g_1 < 1$ (i.e. probability of event that observed fluctuations correspond to regime of asymptotic population extinction is equal to zero). Probability of event that stochastic point (with uniform distribution) belong to feasible set is equal to $8.182 \cdot 10^{-4}$ approximately.

For found points of feasible set it was determined that observed fluctuations correspond to cycle of length 2 is equal to 0.4831, cycle of length 4 is equal to 0.1224, cycle of length 1000 years (or more, or we observe stochastic fluctuations) is equal to 0.2618, cycle of length 1 (asymptotic stabilization at non-zero level) is equal to 0.04313. Probabilities of realization of

other dynamic regimes are much less than pointed out amounts. Thus, we have good background for conclusion that observed regime is cyclic with period in 2 years.
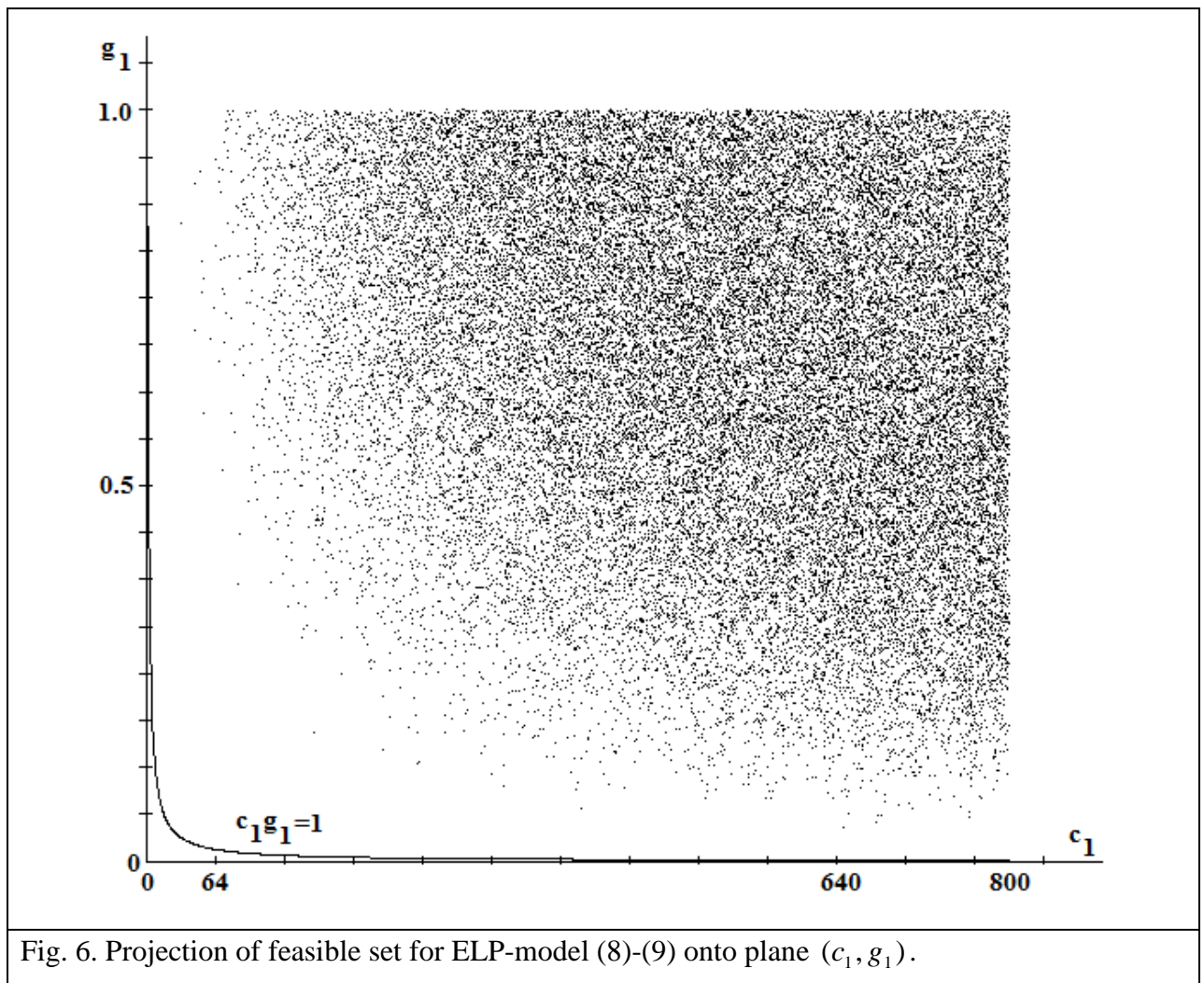

Fig. 6. Projection of feasible set for ELP-model (8)-(9) onto plane (c_1, g_1).

5. Conclusion

Least Squared Method is one of most popular methods of estimation of parameters of ecological models. Within the limits of this approach we can conditionally mark following stages: selection (or construction) of loss-function (in particular, it can be constructed as sum of squared deviations between theoretical and empirical datasets). Loss-function is minimized, and it allows obtaining estimations of model parameters. On the last step of procedure set of deviations (corresponding to estimations of model parameters) is analyzed with use of various statistical criterions. In most cases it is assumed that distribution of deviations must be Normal with zero average, and in sequence of residuals serial correlation cannot be observed.

If for selected significance levels hypothesis about Normal distribution of deviation or hypothesis about equivalence of average to zero must be rejected, it gives a background for conclusion that model isn't suitable for fitting of considering time series. The same background

we have in a case when serial correlation is observed in sequence of residuals. In other words, final conclusion about suitability or non-suitability of model for approximation of time series we make on a base of analysis of trajectory behavior in unique point from space of model parameters. It is not serious background for such a conclusion.

Selection (or construction) of loss-function is one of most serious stages of use of least squared method. Now we have no criterions for selection of type of this function, and in a result of it in modern literature it is possible to find a lot of various modifications (for example, it can be presented as sum of absolute values of deviations, sum of squared deviations with non-negative weights etc.). And all these modifications are used for analysis of biological datasets. Use of different loss-functions lead to obtain different results, and, respectively, to misunderstanding and infidelity in method, results etc.

Freedom in selection of type of loss-function can be explained by following circumstances: this function has no relation to existing datasets, model and considering biological process. It gives a background for conclusion that least squared method is based on logic mistake (and method of maximum likelihood too). Before minimizing of any function we have to determine basic requirements to model, to deviations between theoretical (model) and empirical/experimental datasets; we have to determine – is it possible to find values of model parameters when all requirements are satisfied or not?

One of possible and rather obvious set of requirements is presented in paper. Part of these requirements is well-known: absence of serial correlation in sequences of residuals, symmetry of distribution of deviations with respect to origin, monotonic behavior of branches of density function. Set of points of space of model parameters (where all pointed out requirements are satisfied) was called as *feasible set*. Among points of feasible set points with strongest results can be found – it is naturally to assume that these points we must use as model parameters.

Note that within the framework of this approach we cannot give a guarantee that finally we will have the only solution. But within the framework of least squared method we cannot give same guarantee too when we analyze non-linear ecological models. It is important to mark following idea: within the framework of considered approach we have a possibility to use *in natural way* several time series for parameter estimations, and we don't need to construct any *artificial* loss-functions.

Feasible set has important property: in principle, every element of this set can be used as estimations of model parameters. Statistical criterions cannot give us a background for conclusion that model with respective parameters demonstrates insufficient approximation of datasets. Thus, we can trust to elements of feasible set, and in this occasion this set plays the role of *confident domain*.

If feasible set is divided onto system of sub-sets by bifurcation surfaces we can estimate probabilities of realization (for populations) of one or other dynamic regime. For this reason we can organize a point with uniform distribution on feasible set. If probability of event that point with uniform distribution belongs to fixed sub-set, is bigger than all other probabilities it can be a background for conclusion that this regime is observed for population.

6. References

Akcakaya H.R., 1992. Population cycles of mammals: evidence for a ratio-dependent predation hypothesis// Ecol. Monogr. 62(1): 119-142.

Ayala F.J., Gilpin M.E., Ehrenfeld J.G., 1973. Competition between species: Theoretical models and experimental tests// Theor. Pop. Biol. 4(3): 331-356.

Bard Y., 1979. Nonlinear Parameter Estimation. N.Y.: Academic Press Inc., 349 p.

Beverton R.J., Holt S.J., 1957. On the dynamics of the exploited fish populations, Great Brit. Min. Agr. Fish, Food, Fish. Invest. . Ser. 2-19. P. 1-533.

Bolshev L.N., Smirnov N.V., 1983. Tables of Mathematical Statistics. Moscow: Nauka, 416 p.

Brauer F., Castillo-Chavez C., 2001. Mathematical Models in Population Biology and Epidemiology. N.Y.: Springer. 417 p.

Chitty H., 1948. The snowshoe rabbit inquiry// J. Anim. Ecol. 17: 39-44.

Costantino R.F., Desharnais R.A., Cushing J.M., Dennis B., 1997. Chaotic dynamics in an insect population// Science 275: 389-391.

Dennis B., Desharnais R.A., Cushing J.M., Henson S.M., Costantino R.F., 2001. Estimating chaos and complex dynamics in an insect population// Ecol. Monographs 71(2): 277-303.

Desharnais R.A., Costantino R.F., Cushing J.M., Henson S.M., Dennis B., 2001. Chaos and population control of insect outbreaks// Ecol. Letters 4: 229-235.

Draper N.R., Smith H., 1986. Applied Regression Analysis. Vol. 1. Moscow: Finance and Statistics, 366 p.

Draper N.R., Smith H., 1987. Applied Regression Analysis. Vol. 2. Moscow: Finance and Statistics, 351 p.

Elton C., Nicholson M., 1942. The ten-year cycle in numbers of the lynx in Canada// J. Anim. Ecol. 11(2): 215–244.

Gause G.F., 1934. The Struggle for Existence. Baltimore: Williams and Wilkins, 163 p.

Gilpin M.E., 1973. Do hares eat lynx?// Amer. Nat. 107(957): 727-730.

Gilpin M.E., Ayala F.J., 1973. Global models of growth and competition// Proc. Nat. Acad. Sci. 70: 3590–3593.

Hollander M.D.A., Wolfe D.A., 1973. Nonparametric statistical methods. New York, London, Sydney, Toronto. John Wiley and Sons, Inc., 503 p.

Hudson D.J., 1970. Statistics for physicists. Moscow: Mir, 296 p.

Inchausti P., Ginzburg L.R., 2002. Using the phase shift for assessing the causation of population cycles// Ecological Modelling 152: 89–102.

Isaev A.S., Khlebopros R.G., Nedorezov L.V., Kondakov Yu.P., Kiselev V.V., 1984. Forest Insect Population Dynamics. Novosibirsk: Nauka, 224 p.

Isaev A.S., Khlebopros R.G., Nedorezov L.V., Kondakov Yu.P., Kiselev V.V., Soukhovolsky V.G., 2001. Population Dynamics of Forest Insects. Moscow: Nauka, 347 p.

Ivanchikov P.V., Nedorezov L.V., 2011. About a modification of Beddington − Free − Lawton model of parasite − host system dynamics// Comp. Ecol. Software 1(2): 86-94.

Ivanchikov P.V., Nedorezov L.V., 2012. About a modification of May model of parasite − host system dynamics// Computational Ecology and Software 2(1): 42-52.

Ivchenko G.I., Medvedev Yu.I., 1984. Mathematical statistics. Moscow: Visshaya Shkola, 248 p.

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.

Korzukhin M.D., Semevskii F.N., 1992. Forest Sin-ecology. Saint-Petersburg: Gidrometeoizdat, 192 p.

Kostitzin V.A., 1937. La Biologie Mathematique. Paris: A. Colin. 236 p.

Lakin G.F., 1990. Biometrics. Moscow: Visshaya Shkola, 352 p.

Leigh G., 1968. The ecological role of Volterra's equations // Some of Mathematical Problems of Biology. Proc. Sympos. Math. Soc., Washington: Providence: 1-61.

Likes J., Laga J., 1985. Basic tables of mathematical statistics. Moscow: Finance and statistics, 356 p.

Lilliefors H.W., 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown// J. Amer. Statist. Association 64: 399-402.

Lotka A.I., 1920. Undamped oscillations derived from the law mass action// J. Amer. Chem. Soc. 42(8): 1595-1599.

Lotka A.I., 1925. Elements of physical biology. Baltimore: Williams a. Wilkins, 460 p.

McCallum H., 2000. Population parameters estimation for ecological models. Brisbane: Blackwell Sciences, 348 p.

Moran P.A.P., 1950. Some remarks on animal population dynamics// Biometrics 6(3): 250-258.

Nedorezov L.V., 2011 a. Analysis of some experimental time series by Gause: Application of simple mathematical models// Computational Ecology and Software 1(1): 25-36.

Nedorezov L.V., 2011 b. Analysis of Cyclic Fluctuations in Larch Bud Moth Populations with Discrete-Time Dynamic Models// Biology Bulletin Reviews 72(2): 407-414.

Nedorezov L.V., 2012 a. Chaos and Order in Population Dynamics: Modeling, Analysis, Forecast. LAP Lambert Academic Publishing, Saarbrucken, 352 p.

Nedorezov L.V., 2012 b. Application of Moran – Ricker model for analysis of *Bupalus piniarius* L. population dynamics// Proceedings of the International Academy of Ecology and Environmental Sciences 2(4): 234-245.

Nedorezov L.V., 2013. Temporal variability in model coefficients: *Bupalus piniarius* as a case study// Biology Bulletin Reviews 3(5): 399-405.

Nedorezov L.V., 2014 a. Use of Moran-Ricker model for fitting of larch bud moth time series: structure of feasible set in space of model parameters// Population Dynamics: Analysis, Modelling, Forecast 3(4): 90-92.

Nedorezov L.V., 2014 b. Basic properties of ELP-model// Population Dynamics: Analysis, Modelling, Forecast 3(4): 93-101.

Nedorezov L.V., 2015 a. Application of generalized discrete logistic model for fitting of pine looper moth time series: Feasible sets and estimations of model parameters// Proceedings of the International Academy of Ecology and Environmental Sciences 5(1): 38-48.

Nedorezov L.V., 2015 b. Non-traditional approach to estimation of parameters of ecological models (on examples of larch bud moth and pine looper moth time series)// Siberian J. of Forest Science, N 1: 72-84.

Nedorezov L.V., 2016. Application of nonlinear model of population dynamics with phase structure to analysis of pine looper moth time series// Proc. of the International Academy of Ecology and Environmental Sciences 6(1): 1-12.

Nedorezov L.V., Lohr B.L., Sadykova D.L. 2008. Assessing the importance of self-regulating mechanisms in diamondback moth population dynamics: Application of discrete mathematical models// J. Theor. Biol. 254: 587–593.

Nedorezov L.V., Utyupin Yu.V., 2011 a. Continuous-Discrete Models of Population Dynamics: An Analytical Overview. State Public Scientific-Technical Library, Novosibirsk, 234 p.

Nedorezova B.N., Nedorezov L.V., 2011. Pine looper population dynamics: forecast with discrete time mathematical models// Euro-Asian Entomological Journal 10(3): 280-288.

Nedorezova B.N., Nedorezov L.V., 2012. Pine looper moth population dynamics in Netherlands: Prognosis with generalized logistic model// Proceedings of the International Academy of Ecology and Environmental Sciences, 2(2): 70-83.

Orlov A.I., 2004. Applied Statistics. Moscow: Ekzamen, 656 p.

Orlov A.I., 2015. New paradigm in mathematical methods in research// Zavodskaya laboratoriya. Diagnostika materialov 81(7): 5-5.

Pielou E.C., 1977. Mathematical Ecology. N.Y.: John Wiley and Sons, 385 p.

Ricker W.E., 1954. Stock and recruitment// J. Fish. Res. board of Canada 11(5): 559-623.

Rosenberg G.S., 2010. Some comments to publication by M. Gilpin "Do hares eat lynx?"// Samarskaya Luka: Prob. Region. i Glob. Ekologii 19(3): 180-186.

Shapiro S.S., Wilk M.B., Chen H.J. 1968. A comparative study of various tests of normality. J. of the American Statistical Association 63: 1343-1372.

Skellam J.G., 1951. Random dispersal in theoretical populations// Biometrika 38: 196-218.

Stenseth N.C., Bjornstad O.N., Falck W., 1996. Is spacing behaviour coupled with predation causing the microtine density cycle? A synthesis of current process-oriented and pattern-oriented studies// Proc. R. Soc. Lond. B 263: 1423-1435.

Stenseth N.C., Bjornstad O.N., Saitoh T., 1998. Seasonal forcing on the dynamics of *Clethrionomys rufocanus*: Modelling geographic gradients in population dynamics// Res. on Pop. Ecol. 40: 85-94.

Tonnang H., Nedorezov L.V., Owino J., Ochanda H., Löhr B. 2009 a. Evaluation of discrete host –parasitoid models for diamondback moth and *Diadegma semiclausum* field time population density series// Ecological Modelling 220: 1735-1744

Tonnang H., Nedorezov L.V., Ochanda H., Owino J., Lohr B. 2009 b. Assessing the impact of biological control of *Plutella xylostella* through the application of Lotka – Volterra model. Ecological Modelling 220: 60-70.

Tonnang H., Nedorezov L.V., Owino J., Ochanda H., Löhr B. 2010. Host–parasitoid population density prediction using artificial neural networks: diamondback moth and its natural enemies. Agricultural and Forest Entomology 12(3): 233-242.

Tonnang H., Löhr B., Nedorezov L.V. 2012. Theoretical Study of the Effects of Rainfall on the Population Abundance of Diamondback Moth, *Plutella xylostella*. Population Dynamics: Analysis, Modelling, Forecast 1(1): 32-46.

Volterra V., 1931. Lecons sur la theorie mathematique de la lutte pour la vie. Paris: Gauthiers-Villars, 214 p.

Wood S.N. 2001 a. Partially specified ecological models. Ecological Monographs, 71: 1-25.

Wood S.N., 2001 b. Minimizing model fitting objectives that contain spurious local minima by bootstrap restarting// Biometrics 57: 240-244.