About an approach to population periodic dynamics analysis
(on an example of larch bud moth fluctuations)

L.V. Nedorezov

University of Nova Gorica, Vipavska Cesta 13, Nova Gorica SI-5000, Slovenia
e-mail: lev.nedorezov@ung.si

Abstract
In current publication a preliminary statistical method of population dynamics analysis is considered. Method was applied to analysis of well-known time series of larch bud moth (Zeiraphera diniana Gn.) fluctuations, and showed that observed fluctuations don’t correspond to strong 2-, 3-,..., 9-year cycle, which can be generated by one-dimensional discrete models.

Keywords larch bud moth fluctuations, time series, preliminary statistical analysis, approximation

Introduction
Search of suitable mathematical model and estimation of model parameters using empirical datasets are the main elements of population dynamics analysis (McCallum, 2000; Nedorezov, 1986, 2010, 2011a; Nedorezov, Lohr, Sadykova, 2008; Sibly et al., 2005; Tonnang et al., 2009, 2010; Turchin, 2003; Wood, 2001 and many others). Without finding of suitable model (or without constructing of new suitable model) it is impossible to prepare strong scientifically-based forecasts of pest population changing and optimal methods of its management. But up to current moment there are no criterions, which can help in finding suitable model before comparison of theoretical and empirical results (Isaev et al., 1984, 2001).

In a result of it preliminary and respective statistical analysis, which can help in creation of plausible hypothesis about the character of population fluctuations, can play important, key role in choosing of mathematical models. This preliminary analysis must include as standard characteristics of population time series (like average level of population size, range of sample etc.) as creation of generalized pictures (phase portraits), which describe basic features of

It is possible to point out one more variant of preliminary statistical analysis which seems to be very important. In particular, without choosing of any mathematical model it is possible to check a hypothesis about correspondence of observed fluctuations to cyclic dynamics of any fixed size $n$ ($n = 1, 2, 3...$). Respectively, all deviations from coordinates of $n$-cycle can be explained by the influence of external stochastic factors, by used methods of data collection etc. In other words, before choosing of model we can try to determine a dynamical regime which is observed in natural conditions. More precisely, we can try to find a dynamical regime with the following property: modern statistical methods don’t allow us concluding that considering regime doesn’t correspond to observed fluctuations.

Necessity of such preliminary statistical analysis can also be explained by the following assumptions. It is known, that for estimation of model parameters under the use of global fitting and empirical time series, researches use initial parts of model trajectories, and don’t use parts of trajectories which correspond to stabilized regime of population fluctuations (McCallum, 2000; Nedorezov, 2010, 2011a, 2012a; Nedorezov, Sadykova, 2005, 2008, 2010; Tonnang et al., 2009, 2010, 2012; Turchin, 2003; Wood, 2001 and many others). Approximation of empirical time series by initial parts of model trajectories is correct if we analyze fluctuations of invasive species. In such situations we cannot talk about asymptotic stabilizations of population fluctuations. But if we analyze dynamics of species which exist in local habitat (where datasets were collected) during long time period, the use of initial parts of model trajectories for fitting of empirical time series needs in additional explanation.

Algorithm
Let $x_1, x_2, ..., x_N$ be a time series of considering hypothetical population. Time step is equal to one year, thus $x_k$ is a population size (or density) at $k$th year. First of all, we have to solve the following question: what kind of datasets we have now? If, for example, we analyze time series presented in book by G.F. Gause (1934), it is obvious, that for every trajectory we can point out initial part (it can be exponential phase of population growth), mid part of trajectory (where we can observe growth of influence of self-regulative mechanisms on population size), and stabilized behavior (fluctuations near any stable level). In such a situation we have a good
background for application of initial parts of trajectories for fitting of experimental datasets (Nedorezov, 2011b, 2012b).

But in the case when we analyze insect population dynamics in locations where insects live thousands and thousands years, we haven’t background with the same properties. In these situations we observe stabilized dynamical regime. Thus, for the estimation of model parameters we have to minimize, for example, the sum of squared deviations of real datasets from coordinates of asymptotically stable attractors.

Let’s consider a situation when we have to fit a time series by a trajectory of stabilized regime. There are two possible ways for doing it. The first way is following. We have to choose a model, to fix initial values of model parameters, and to get asymptotic trajectory. If we have a discrete model (model with discrete time), we have to start this model: first 1000 values of population size generated by model must be skipped, and the next 1000 values must be saved. Taking into account that $N << 1000$ for biggest part of existing time series on fluctuations of insect populations, 1000 values (generated by model) will be enough for estimation of value of minimizing functional form.

Starting with first element of saved trajectory we summarize squared deviations between theoretical and empirical trajectories. The same procedure must be repeated starting from the second value of saved trajectory and so on (up to $1000 - N$). For the set of calculated sums we can obtain minimum one – the last value will be the value of minimizing functional form for selected values of model parameters. This procedure must be continued for determination of global minimum of minimized functional form.

The second way to solution of considering problem is following. On the first step of process we must determine a dynamical regime which is realized for population (it is hypothesis we have to check analyzing datasets). For example, we can start with assumption that observed fluctuations of population size correspond to cycle of the length two: $ababab\ldots$ Let’s also assume that minimizing functional form is equal to sum of deviations squared. In this case functional form can be presented in the following form:

$$\sum_k (a_k - a)^2 + \sum_k (b_k - b)^2 \rightarrow \min_{a,b}.$$  \hspace{1cm} (1)

From (1) we get the following estimations for coordinates of 2-cycle:

$$a = \frac{1}{N} \sum_k x_k, \quad b = \frac{1}{N} \sum_k x_k.$$  \hspace{1cm} (2)
where $N^* + N'' = N^*$, and $N^* = N''$ or $N^* = N'' + 1$. After estimations (2) of coordinates of 2-cycle we have to check hypothesis that observed regime is 2-cycle: more precisely, we have to analyze two sequences $x_1 - a$, $x_3 - a$ ..., and $x_2 - b$, $x_4 - b$ ... and to show that arithmetic averages are equal to zeros, distribution functions for both sets are symmetric functions, and there are no serial correlation in both sequences.

If all used tests show that there are no reasons for rejecting of the respective Null hypotheses, we have to start the second stage of process – we can start the process of selection of mathematical model. It is obvious, if observed changing of population size corresponds to 2-cycle there are no reasons for consideration of the Skellam model or Kostitzin model as a basic model – in both models there are the regimes of asymptotic stabilizations of population size at any levels only for all values of model parameters (Skellam, 1951; Kostitzin, 1937; Beverton, Holt, 1957). In this situation it is much better to use Moran – Ricker model or discrete logistic model which contain a lot of various dynamical regimes (Moran, 1950; Ricker, 1954; Isaev et al., 1984, 2001; Bazykin, 1985; Nedorezov L.V. 1986, 1997; Svirezhev, 1987). Let’s assume that we decided to choose Moran – Ricker model for the description of population size dynamics:

$$y_{k+1} = Ay_k e^{-\alpha y_k},$$

where $A, \alpha = const > 0$, $y_k$ is population size at time moment $k$. If we observe 2-cycle for model (3) it means that following relations must be truthful for coefficients:

$$a = Abe^{-\alpha b}, b = Aae^{-\alpha a}.$$  

Values of model (3) parameters must be obtained as solutions of this system of non-linear algebraic equations. After logarithmic transformation of this system of algebraic equations we get new linear system for variables $\alpha$ and $w = \ln A$:

$$\ln a - \ln b = w - ab, \ln b - \ln a = w - aa.$$  

For $b \neq a$ basic determinant of this linear system doesn’t equal to zero, and, respectively, solutions of this system exist and unique:

$$\alpha = \frac{2(\ln a - \ln b)}{a - b}, w = \ln a - \ln b + b \frac{2(\ln a - \ln b)}{a - b}.$$  

After obtaining of model parameter estimations we have to check a stability of 2-cycle: if it isn’t asymptotically stable model cannot give us a sufficient approximation of datasets. For checking of stability of 2-cycle we have to calculate eigenvalue of Jacobi matrix for $F(F(\cdot))$ of model (3).
for given parameters, where $F(F(\cdot))$ is second iteration of function $F$ in right-hand side of equation.

Let’s consider more difficult situation when we check hypothesis about realization of the 3-cycle $abcabc…$ for considering time series. In this situation coordinates of 3-cycle can be determined with formulas like (2): in first sum there are elements of initial sample with numbers 1, 4, 7…; in second sum there are elements of initial sample with numbers 2, 5, 8…; in third sum there are elements with numbers 3, 6, 9…Like in the previous case after the estimation of cycle coordinates it is important to check three sequences of deviations with respective numbers. If all used statistical criterions didn’t allow rejecting respective Null hypotheses for all sequences, after that model parameters can be estimated.

Let’s also assume that Moran – Ricker model (3) is the best one for fitting. Thus, model coefficients must satisfy to the following algebraic equations:

$$ a = Ace^{-ac}, \ b = Aae^{-aa}, \ c = Abe^{-ab}. $$

Consequently, after logarithmic transformation we have three linear equations for two unknown variables (non-correct mathematical problem). Estimations of model parameters can be obtained, for example, minimizing the following functional form:

$$ Q(a, b) = |a - Ace^{-ac}| + |b - Aae^{-aa}| + |c - Abe^{-ab}|. $$

After obtaining of model parameter estimations sequence of deviations between theoretical and empirical time series must be tested using statistical criterions.

This approach to identification of population dynamics regime has the following important problem: length of cycle must be much less than sample size. If we assume that for obtaining good (confidence) estimations of coordinates of cycle we have to have about ten real values, then even for datasets by F. Schwerdtfeger (1944, 1968; there are about 60 values in 4 time series of forest insect populations) it will be problematic to check cycles of the length 6 or more.

Within the limits of this approach every time we have to give answers on sequence of questions: can we say that observed changing of population size in time is pure stochastic fluctuation near stationary level? Can we conclude that observed fluctuations are two-years, three-years and so on cyclic processes?

Let’s assume that we have a hypothesis that considering changing of population size in time is cycle of the length 2, $ababab…$. Following the steps of algorithm we estimate coordinates
of this cycle (formulas (1) and (2)) and obtain two sequences of deviations between empirical values and estimated coordinates of cycle. First of all, we have to be sure those estimations of cycle coordinates are confidently different. It can be checked with the help of Student test (but before using this test we have to check the Normality of the sets of these deviations). For cycles of bigger lengths we can use analysis of variance. If there is not a confident difference between averages (2) (for selected significance level) we haven’t a background for conclusion about existence of 2-cyclic fluctuations. If we have confident differences between averages, we have a background for continuation of analysis of considering situation. The second, we have to check the hypothesis about the relation between sample variations: for smaller coordinate of cycle we have to have smaller (or equal) sample variation. But this last requirement cannot be strong, and cannot be realized for sufficient big values of population size (Nedorezov, 2012 a, c). The third, after two successive steps we have to check hypotheses about absence/existence of serial correlation (Draper, Smith, 1986, 1987).

Datasets

Regular observations of larch bud moth fluctuations had been started in Swiss Alps (in Upper Engadine Valley) in 1949 (Auer, 1977; Baltensweiler, Fischlin, 1988). Used in current publication time series can be free downloaded in Internet (NERC Centre for Population Biology, Imperial College (1999) The Global Population Dynamics Database, N 1407). Population densities are presented in units “number of larvae per kilogram of branches”.

In GPDD it is pointed out that data were collected in Upper Engadine Valley on 1800 m above the sea level (optimal zone for species; Isaev et al., 1984, 2001). Total sample size is 38 values (from 1949 to 1986).

Results

In table there are the estimations of cycle’s coordinates (under the assumptions that one or other attractor is realized for larch bud moth fluctuations), and respective values of minimizing functional form \( Q \) (it is a sum of squared deviations of real values from the estimated coordinates of cycle).

A big decreasing of value of minimizing functional form is observed for the case when it is assumed the fluctuations of moth corresponds to 9-cycle. If we assume that length of cycle is less than 9, there are no strong changing of the value of minimizing functional form \( Q \) (table).
Table

<table>
<thead>
<tr>
<th></th>
<th>1*</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>92.66</td>
<td>82.06</td>
<td>104.63</td>
<td>79.24</td>
<td>151.06</td>
<td>86.15</td>
<td>83.17</td>
<td>141.81</td>
<td>0.032</td>
</tr>
<tr>
<td>2</td>
<td>103.25</td>
<td>45.47</td>
<td>119.71</td>
<td>27.83</td>
<td>56.57</td>
<td>69.03</td>
<td>130.04</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>130.81</td>
<td>85.2</td>
<td>44.32</td>
<td>126.83</td>
<td>51.1</td>
<td>21.36</td>
<td>0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>84.96</td>
<td>100.72</td>
<td>126.19</td>
<td>83.23</td>
<td>2.93</td>
<td>2.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>147.18</td>
<td>32.52</td>
<td>136.11</td>
<td>16.66</td>
<td>36.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>134.78</td>
<td>171.82</td>
<td>109.37</td>
<td>387.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>69.06</td>
<td>165.01</td>
<td>337.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>187.5</td>
<td>111.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>4.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>899181</td>
<td>894916</td>
<td>850904</td>
<td>889028</td>
<td>798326</td>
<td>843668</td>
<td>840931</td>
<td>727251</td>
<td>114293</td>
</tr>
</tbody>
</table>

*1-cycle is stationary level

In table there are the estimations of cycle’s coordinates (under the assumptions that one or other attractor is realized for larch bud moth fluctuations), and respective values of minimizing functional form $Q$ (it is a sum of squared deviations of real values from the estimated coordinates of cycle).

A big decreasing of value of minimizing functional form is observed for the case when it is assumed the fluctuations of moth corresponds to 9-cycle. If we assume that length of cycle is less than 9, there are no strong changing of the value of minimizing functional form $Q$ (table).

It is obvious, that increasing of the length of cycle must lead to decreasing of the value of minimizing functional form $Q$ (as we can see from the table, process of decreasing can be non-monotonic). If length of cycle is equal to 38 (in considering situation it is equal to sample size), $Q = 0$. Thus, strong decreasing of minimizing functional form $Q$ while crossing from 8-cycle to 9-cycle doesn’t give us conclusion that 9-cycle gives us the best approximation, and this cycle is really observed for larch bud moth. This changing can be a “natural” decreasing of functional form $Q$.

If we assume that considering time series is stochastic fluctuations near one stable level, we have to check Normality of deviations and existing/absence of serial correlation. Kolmogorov
Smirnov test gives \( d = 0.35991 \), and probability \( p < 0.01 \) (this is the probability of event that distribution of deviations is Normal); Lilliefors test shows \( p < 0.01 \); Shapiro – Wilk test: \( W = 0.6436 \) and \( p < 10^{-5} \) (Lilliefors, 1967; Shapiro, Wilk, Chen, 1968; Bolshev, Smirnov, 1983). Durbin – Watson criteria (Draper, Smith, 1986, 1987) is equal to 1.0804, and it means that in sequence of residuals the negative serial correlation is observed (for sample size 38 and one predictor critical values of this criterion for 5% significance level are following: \( d_L = 1.43 \) and \( d_U = 1.54 \); and \( d_L = 1.23 \) and \( d_U = 1.33 \) for 1% significance level). Thus, with 1% significance level our hypothesis that considering sample is stochastic fluctuations near stationary level must be rejected.

If we assume that observed fluctuations correspond to 2-cycle \( \text{abab} \ldots \), we have to divide initial sample on two sub-samples \( x_1, x_2 \ldots \) and \( x_3, x_4 \ldots \), and check the hypothesis \( H_0 : a = b \) against alternative hypothesis \( H_1 : a \neq b \). If with rather big significance level \( H_0 \) cannot be rejected we have not a background for conclusion that 2-cycle is realized for population. For considering situation (for first sub-sample) the following results were obtained: Kolmogorov – Smirnov test – \( d = 0.37399 \), \( p < 0.01 \); Lilliefors test – \( p < 0.01 \); Shapiro – Wilk test – \( W = 0.63368 \), \( p = 10^{-5} \). For the second sub-sample obtained results are following: Kolmogorov – Smirnov test – \( d = 0.37171 \), \( p < 0.01 \); Lilliefors test – \( p < 0.01 \); Shapiro – Wilk test – \( W = 0.65546 \), \( p = 2 \cdot 10^{-5} \). Thus, with 1% significance level hypotheses about Normality of two sub-samples must be rejected, and we haven’t a background for application of parametric statistical criterions for checking of the hypothesis \( H_0 \).

Taking into account that shift of the sample on any constant doesn’t lead to changing of amounts of the tests on Normality, rejecting of the hypothesis about Normality of the sample \( x_1, x_3, \ldots \) means that we have to reject the hypothesis about Normality of the sample \( x_1 - a, x_3 - a, \ldots \). Thus, we obtain a background for conclusion that our hypothesis about the correspondence of observed fluctuations to 2-cycle is untenable. It is important to note that for every considering situation (from 3-cycle up to 9-cycle) we had sub-samples with non-Normal distribution: more precisely, hypothesis about Normality could be rejected with 1% significance level. Analysis of correspondence of considering fluctuations to 9-cycle showed that for one of nine sub-samples hypothesis about Normality can be rejected with 1% significance level; for
biggest part of sub-samples this hypothesis cannot be rejected even with 20% significance level. Analysis of correspondence of considering fluctuations to 8-cycle showed that for two sub-samples hypothesis about Normality cannot be rejected even with 20% significance level, and for two sub-samples hypothesis about Normality can be rejected with 1% significance level.

Obtained results show that we haven’t a background for conclusion that observed fluctuations of larch bud moth correspond to strong 2-cycle, 3-cycle,…, 9-cycle, which can be generated by one-dimensional models of (3) type. This observed cycle has more complicated nature. But taking into account that statistics isn’t the probative science (at every time we have a probability that presented result is mistaken) we have to check other properties of sets of deviations.

For 9-cycle Kruskal – Wallis test is following: $H = 31.83104, \ p = 0.0001$. Consequently, we have confident difference between cycle coordinates (with high significance level we have to reject hypothesis that respective sub-samples have one and the same distribution functions). For all other cases (for cycles of the lengths 3, 4,…, 8) Kruskal – Wallis test has rather small values with $p > 0.3$. Respectively, we have no reasons for the rejecting hypothesis that sub-samples have the same distribution functions.

Thus, there are no reasons for conclusion that the length of observed cycle is less than 9. For analysis of correspondence of observed fluctuations to 9-cycle we have to show that we have confident difference between all 9 sub-samples. For pairwise comparison Kolmogorov – Smirnov test, Wilcoxon – Mann – Whitney test, and Wald – Wolfovitz test were used. The total number of considered cases is equal to 36. In 20 cases all three tests showed that for 5% significance level sub-samples are different. In 10 cases all three tests showed that there are no reasons for rejecting of the respective Null hypotheses. In 6 cases Wilcoxon – Mann – Whitney test showed that sub-samples are different (hypotheses about equivalence of distribution functions must be rejected), at the same time two other tests showed that Null hypotheses cannot be rejected.

Finally, it give us additional reasons for conclusion that larch bud moth fluctuations don’t correspond to 9-cycle. Absence of confident differences between coordinates of 9-cycle can be explained by small sub-sample sizes (in every sub-sample we have 4 or 5 values only). It also can be also explained by more complicated nature of observed cycle. It is possible to point out some other reasons but, anyway, in all situations we have reasons to say that observed cycle is close to 9-year cycle, and we haven’t reasons for saying this cycle is close to 8-year cycle.
Discussion

Difference in results which were obtained using various types of approximation of real datasets – fitting by initial parts of model trajectories and fitting by the systems of coordinates of asymptotically stable attractors – demonstrates the level of influence of stochastic factors on population dynamics. Comparison of results of approximation of larch bud moth fluctuation obtained by two these different ways (Nedorezov, 2011a, 2012 a) show that stochastic factors have strong influence on population dynamics. This conclusion cannot be the final one, because in most cases used for fitting models didn’t give us a sufficient approximations of time series.

After all provided calculations we can conclude that observed fluctuations are close to 9-year cycle. Obtained results show also that we haven’t a background for conclusion that observed fluctuations of larch bud moth correspond to strong 2-cycle, 3-cycle,…, 9-cycle, which can be generated by one-dimensional models of (3) type. This observed cycle has more complicated nature. Respectively, for approximation of considered datasets we have to use more complicated models.

References


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


