

УДК 57.087:595.782

НЕТРАДИЦИОННЫЙ ПОДХОД К ОЦЕНКЕ ПАРАМЕТРОВ ЭКОЛОГИЧЕСКИХ МОДЕЛЕЙ (НА ПРИМЕРАХ ДИНАМИКИ ЧИСЛЕННОСТИ СЕРОЙ ЛИСТВЕННИЧНОЙ ЛИСТОВЕРТКИ И СОСНОВОЙ ПЯДЕНИЦЫ)

© 2015 г. Л. В. Недорезов

Центр междисциплинарных исследований по проблемам окружающей среды РАН
191187, Санкт-Петербург, наб. Кутузова, 14

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

Поступила в редакцию 01.03.2015 г.

Рассматривается подход к оценке параметров экологических моделей, не требующий построения функций потерь или правдоподобия. Суть метода заключается в том, что первоначально формулируются требования к модели, к отклонениям эмпирических данных от теоретических (модельных), а затем с использованием соответствующих статистических методов определяются допустимые множества в пространстве параметров модели (где все требования к модели выполняются). В допустимых множествах находятся элементы, для которых выполняются наиболее *сильные* результаты. Построение допустимых множеств для модели Морана–Риккера с запаздыванием при использовании данных по динамике численности серой лиственничной листовертки *Zeiraphera diniana* Gn. и модели ELP для данных по динамике численности сосновой пяденицы *Bupalus piniarius* L. показывает, что имеются значительные расхождения с результатами, полученными с помощью метода наименьших квадратов. Указываются значения параметров, когда модели дают удовлетворительное описание популяционной динамики. Для временного ряда динамики численности серой лиственничной листовертки получено, что допустимые множества не пусты для модели Морана–Риккера без запаздывания и с запаздыванием в один год. В то же время проведенный анализ свойств элементов допустимых множеств показал, что среди них нет таких, которые бы соответствовали сложившимся биологическим представлениям о динамике листовертки. Полученные МНК-оценки позволили при учете запаздывания в один год получить режим с циклом длины 18 с двумя максимумами. Однако при этом все статистические критерии дали отрицательный результат. Полученные результаты показывают, что запаздывание в действии саморегуляторных механизмов составляет более одного года. Дан способ одновременной оценки параметров модели динамики численности сосновой пяденицы для связанных временных рядов. Проведенный анализ показывает, что есть все основания утверждать, что наблюдаемый режим – циклический с длиной цикла в два года.

Ключевые слова: динамика численности популяций, временные ряды, математическая модель, оценка параметров, серая лиственничная листовертка, сосновая пяденица.

DOI: 10.15372/SJFS20150307

ВВЕДЕНИЕ

Для оценок значений параметров экологических моделей используются разные подходы: от экспертных до байесовского (Бард, 1979; Боровков, 1984; Дрейпер, Смит, 1986, 1987; Лемешко и др., 2011; Wood, 2001a, b). Для одной и той же модели при ее использовании для аппроксимации одного и того же временного ряда можно получить разные оценки: экспертные, полученные с помощью мето-

дов наименьших квадратов и максимального правдоподобия и т. д. Вопрос о том, какие лучше, не ставится, как в некоторых случаях не ставится вопрос об анализе соответствия теоретических (модельных) и эмпирических данных.

Пусть требуется оценить значения параметров следующей модели с дискретным временем:

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

В (1) x_k – численность (или плотность) популяции в момент времени k , $k = 0, 1, 2, \dots$; $\vec{\alpha}$ – вектор параметров модели; F – неотрицательная функция при неотрицательных значениях x_k и допустимых значениях параметров. В начальный момент времени $k = 0$ численность популяции равна x_0 , что также является неизвестной величиной, требующей оценки по исходной выборке.

Пусть также дана выборка $\{x_k^*\}$, $k = 0, 1, \dots, N$, – временной ряд значений численности популяции; $N + 1$ – размер выборки. По этой выборке $\{x_k^*\}$ необходимо оценить значения параметров $\vec{\alpha}$ и x_0 .

При использовании метода наименьших квадратов (Бард, 1979; Боровков, 1984) сначала выбирается функция потерь (это может быть сумма квадратов отклонений теоретических значений от эмпирических), минимизация которой дает МНК-оценки параметров. Если временной ряд аппроксимируется траекториями модели (глобальное приближение (Wood, 2001a, b)), функция потерь имеет вид

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

В (2) $\{x_k(\vec{\alpha}, x_0)\}$ – траектория модели (1), полученная при фиксированных значениях $\vec{\alpha}$ и x_0 . После нахождения минимума функции потерь (2)

$$Q(\vec{\alpha}^{**}, x_0^{**}) = \min_{\vec{\alpha}, x_0} \left(\sum_{k=0}^N (x_k(\vec{\alpha}, x_0) - x_k^*)^2 \right) \quad (3)$$

проводится анализ отклонений $\{e_k\}$:

$$e_k = x_k(\vec{\alpha}^{**}, x_0^{**}) - x_k^*. \quad (4)$$

Обычно предполагается, что отклонения $\{e_k\}$ – значения независимых случайных величин, имеющих нормальное распределение с нулевым средним. В соответствии с этими предположениями выбираются статистические критерии: в частности, для проверки нормальности могут быть выбраны критерии Колмогорова–Смирнова, Лиллифорса, Шапиро–Вилка (Большев, Смирнов, 1983; Lilliefors, 1967; Shapiro et al., 1968), для проверки независимости – критерий Дарбина–Ватсона и тест Сведа–Эйзенхарта (Дрейпер, Смит, 1986, 1987; Холлендер, Вулф, 1983; Ликеш, Ляга, 1985).

Если в последовательности отклонений обнаруживается сериальная корреляция, то это служит основанием для заключения о том, что модель не учитывает влияния какого-то механизма или фактора и нуждается в модификации. Иными словами, заключение о пригодности или непригодности модели для аппроксимации данных делается на основе анализа отклонений, которые соответствуют одной-единственной точке пространства параметров модели $(\vec{\alpha}^{**}, x_0^{**})$, удовлетворяющей условию (3). Тот же самый вывод имеет место при отклонении гипотезы о нормальности распределения отклонений (4) или о равенстве среднего нулю. Именно это представляется необоснованным: какими бы исключительными свойствами ни обладала точка $(\vec{\alpha}^{**}, x_0^{**})$ пространства параметров, она никоим образом не отражает свойств других точек данного пространства.

Кроме этого, возникает целый ряд вопросов, в частности: на чем именно основано предположение о нормальности отклонений (4). В действительности никакого обоснования этому предположению нет. Более того, оно изначально неверно (точнее, абсурдно): если, к примеру, оценивается известными методами локальная численность популяции, предположение о нормальности отклонений означает, что с *положительной вероятностью* может быть обнаружено отрицательное количество особей. Поэтому проверку на нормальность отклонений следует заменить проверкой симметрии распределения отклонений относительно нуля и монотонности поведения ветвей плотности распределения (Недорезов, 2012).

Далее, на чем основан выбор минимизируемой функции потерь (2), (3)? Никаких реальных оснований для выбора вида функции потерь нет, и с одинаковым успехом можно выбирать и суммы квадратов отклонений (2), и суммы квадратов отклонений с весами, и суммы абсолютных величин в каких-либо положительных степенях и с какими-либо положительными весами. Свобода выбора обусловлена тем, что эта функция никак не связана ни с исследуемым биологическим объектом, ни с имеющимися данными, ни с моделью. Также ниоткуда не следует, что модель (1) – даже если она является законом изменения числен-

ности популяции – должна обеспечивать минимум некоей неотрицательной функции.

Таким образом, можно утверждать, что в вопросе оценки параметров моделей *телега поставлена впереди лошади*. Конечно, подобная конструкция тоже может двигаться, существовать, но целесообразно поменять их местами. Сначала определить критерии, *которым должна удовлетворять модель*, затем найти точки пространства параметров, при которых модель удовлетворяет выбранным критериям (в дальнейшем множество таких точек будем называть *допустимым*). На следующем шаге можно использовать функции потерь (2) на допустимом множестве. Это не является обязательным, можно использовать и другие характеристики случайных множеств (коим является допустимое множество).

Существует еще одна весьма важная и типичная проблема в оценке параметров экологических моделей – одновременного использования нескольких связанных временных рядов, поскольку всегда помимо численности популяции фиксируется зараженность паразитами, плодовитость бабочек и многие другие показатели. Конечно, нужно иметь математическую модель, которая бы позволяла учитывать при оценке параметров наличие дополнительных временных рядов. Помимо этого нужно дать ответ на вопрос, как именно следует объединить «зараженность», «численность», «длину хвои» и прочее в одной-единственной функции потерь. Ниже рассматривается данная проблема и описывается подход к оценке параметров сложных моделей, не требующий построения каких-либо функций потерь.

Рассматривается альтернативный подход к оценке параметров модели, который основан на первоначальном нахождении допустимых множеств и лишен недостатков, указанных для метода наименьших квадратов. Апробация подхода проводится на известных данных по колебаниям численности серой листовичной листовертки *Zeiraphera diniana* Gn. (Baltensweiler, 1964, 1978) и сосновой пяденицы *Bupalus piniarius* L. (Klomp, 1966).

Все требования к модели можно условно разбить на следующие три группы:

1. Отклонения теоретических и эмпирических данных должны иметь симметричную относительно нуля плотность распределения

с монотонно изменяющимися ветвями в положительной и отрицательной областях (монотонно убывающей в положительной области и монотонно возрастающей ветвью в отрицательной области). Это требование существенно более слабое, чем требование нормальности у совокупности отклонений, и многие распределения удовлетворяют этому ограничению.

Пусть $\{e_k^+\}$ – множество положительных отклонений (4) и $\{-e_k^-\}$ – множество отрицательных отклонений, взятых со знаком минус. Симметричность относительно нуля означает, что при выбранном уровне значимости гипотеза о том, что выборки $\{e_k^+\}$ и $\{-e_k^-\}$ имеют одно и то же распределение, не может быть отклонена. Для проверки симметричности использовали критерии Колмогорова–Смирнова, Лемана–Розенблатта и Манна–Уитни (Большев, Смирнов, 1983; Холлендер, Вулф, 1983; Ликеш, Ляга, 1985).

Для проверки монотонности поведения ветвей плотности распределения использовали коэффициент корреляции Спирмена (Большев, Смирнов, 1983). Пусть $\{e_k^{*+}\}$ – упорядоченная по возрастанию выборка $\{e_k^+\}$: $e_1^{*+} < e_2^{*+} < \dots$. При монотонном снижении плотности распределения в идеальном случае длины интервалов $[0, e_1^{*+}]$, $[e_1^{*+}, e_2^{*+}]$, ... также должны быть упорядочены по возрастанию (и могут быть ранжированы 1, 2, ...). Пусть ρ – ранговый коэффициент корреляции Спирмена. Сравнивая ранги длин интервалов для выборки $\{e_k^{*+}\}$ с идеальным случаем, нулевая гипотеза $H_0: \rho = 0$ (при альтернативной гипотезе $H_1: \rho > 0$ и выбранном уровне значимости) должна быть отклонена.

2. В последовательности отклонений не должно быть сериальной корреляции. Для этих целей использовались тест серий Сведы–Эйзенхарта (Дрейпер, Смит, 1986, 1987) и тест серий «скачков вверх – скачков вниз» (Ликеш, Ляга, 1985).

3. Поведение модели должно быть *сходным* с поведением выборки. Под этим понимается следующее: у нас нет оснований для утверждения, что модель позволяет получить хорошую аппроксимацию данных, если во всех случаях, когда значения выборки снижаются, модель «показывает» рост значений, и

наоборот. Оснований также нет, когда таких фактов много.

В данной ситуации необходимо оценить долю случаев «увеличение значения во временном ряду – увеличение значения на траектории модели» и «снижение значения во временном ряду – снижение значения на траектории модели» среди всех возможных вариантов (при соответствующих фиксированных значениях параметров модели). Пусть q – доля таких случаев. При «хорошем» совпадении в поведении модели и исходной выборки при фиксированном уровне значимости нулевая гипотеза $H_0 : q = 0.5$ (при альтернативной гипотезе $H_1 : q > 0.5$) должна быть отклонена.

Данное требование к траекториям модели является очень сильным фильтром: в некоторых случаях не удавалось найти точки в пространстве параметров модели, принадлежащие допустимому множеству (не исключено, что при подобной фильтрации допустимое множество оказывалось пустым). Именно поэтому в отдельных случаях требование 3 не учитывали.

Указанные критерии использовали для решения следующих задач: во-первых, для построения допустимых множеств (при фиксированных значениях уровня значимости), во-вторых, для поиска точек в пространстве параметров, для которых выполняются наиболее сильные утверждения (в рамках этих же критериев). К примеру, если гипотеза о симметрии не отклоняется при 5%-м уровне значимости, то это еще не является основанием для принятия нулевой гипотезы. Более сильный результат – если та же гипотеза не отклоняется при 10%-м уровне значимости. Но мы должны принять нулевую гипотезу, если она не может быть отклонена с 95%-м уровнем значимости.

Изменение уровня значимости (для разных критериев по-разному) позволяет найти одну–две точки в пространстве параметров, которые и используются в качестве оценок параметров модели. Одна из базовых идей подхода состоит в следующем: в качестве оценок параметров модели следует выбирать такие, для которых выполняются наиболее сильные результаты при анализе отклонений. Другая базовая идея подхода такова: помимо анализа «критических» точек (для которых наблюдаются наиболее сильные результаты) резуль-

татом исследования допустимых множеств должно стать распределение (оценки вероятностей) возможных динамических режимов.

При оценке вероятностей реализации у популяции того или иного динамического режима использовали следующую процедуру: на множестве точек, принадлежащих допустимому множеству, разыгрывали равномерное распределение и для каждого полученного значения определяли асимптотический режим. Для большого числа «разыгрываний» случайной точки оценивали известными методами меру подобласти, соответствующую какому-либо динамическому режиму (Михайлов, 1974; Ермаков, 1975), что и позволило оценить вероятность попадания в эту подобласть при равномерном распределении.

ДИНАМИКА ЧИСЛЕННОСТИ СЕРОЙ ЛИСТВЕННИЧНОЙ ЛИСТОВЕРТКИ

Рассмотрим временной ряд по колебаниям численности серой лиственничной листовертки *Zeiraphera diniana* Gn. в Швейцарских Альпах (NERC Centre for Population Biology, Imperial College (1999); the Global Population Dynamics Database, N. 1407) (Baltensweiler, 1964, 1978). Анализ данных, проведенный в работах А. С. Исаева и Р. Г. Хлебопроста (Исаев, Хлебопрос, 1973, 1977; Исаев и др., 1984, 2001), показал, что листовертку следует отнести к эруптивным открытоживущим видам, способным реализовать перманентную вспышку массового размножения.

Вне зоны оптимального обитания динамика вида может протекать по продромальному типу с сохранением определенной периодичности (Исаев и др., 1984, 2001), поэтому, прежде чем аппроксимировать временные ряды с помощью моделей, имеющих сложно организованную структуру фазового пространства (Исаев и др., 1982), следует использовать более простые математические модели, имеющие, например, только одно нетривиальное равновесие в фазовом пространстве. При этом важно, чтобы модель учитывала влияние запаздывания в действии саморегуляторных механизмов на динамику численности популяции (Исаев, Хлебопрос, 1973, 1977; Исаев и др., 1974; Недорезов, 2007). В частности, для решения подобных задач представляется весь-

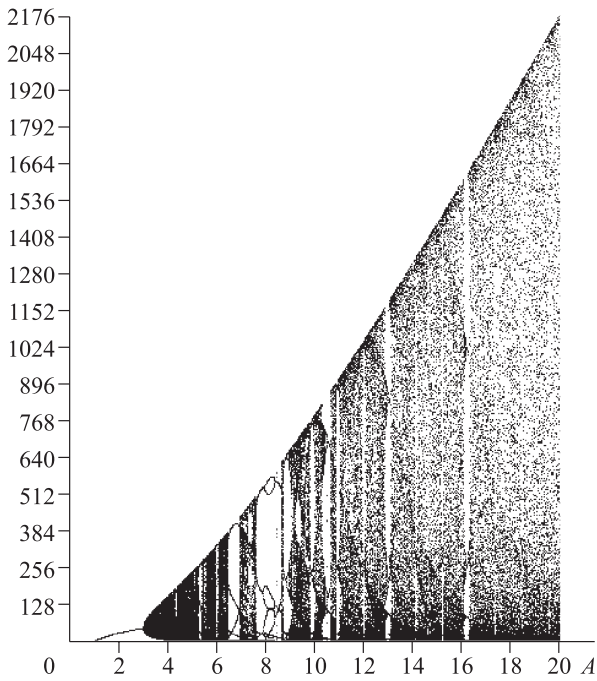


Рис. 1. Бифуркационная диаграмма для модели (5) при $m = 1$, $a_0 = 2.25 \cdot 10^{-3}$, $a_1 = 0.02214$.

ма удобной модель Морана–Риккера с запаздыванием (Moran, 1950; Ricker, 1954):

$$x_{k+1} = Ax_k \exp \left\{ - \sum_{j=0}^m a_j x_{k-j} \right\}, \quad m = 0, 1. \quad (5)$$

В модели (5) неотрицательны все параметры – $A, a_0, a_1 = \text{const} \geq 0$ и начальные значения численностей – $x_0^0, x_1^0 \geq 0$. Модель обладает богатым набором динамических режимов: на рис. 1 представлена бифуркационная диаграмма для $m = 1$, $a_0 = 2.205 \cdot 10^{-3}$, $a_1 = 0.02214$, на которой видны некоторые из возможных режимов (Sadykova, Nedorezov, 2013). Кроме этого, для некоторых временных рядов по колебаниям численности серой лиственничной листовертки получены хорошие результаты аппроксимации с помощью модели (5) (McCallum, 2000) при запаздывании в один год $m = 1$.

Вариант: модель Морана–Риккера без запаздывания, $m = 0$. При $m = 0$ в традиционном подходе для этого случая получены следующие результаты (Nedorezov, Sadykova, 2015): $Q_{\min} = 465\,435.7$, $x_0 = 53.44$, $A = 517.0$, $a_0 = 0.1168$, где Q_{\min} – минимальное значение функционала (2), (3). Разыгрывая значения случайной точки, равномерно распределенной в прямоугольнике $[0, 100] \times [0, 1000] \times [0, 1]$, получаем возможность оценить структуру допустимого множества (для всех статистических

критериев выбран один уровень значимости 5 %). На рис. 2 представлена проекция этого множества на плоскость (A, a_0) . Как видно из представленного рисунка, МНК-оценка лежит вне области наибольшей концентрации точек допустимого множества и сама точка не принадлежит этому множеству.

При МНК-оценках параметров модели (5) анализ отклонений $\{e_k\}$ (4) показывает, что вероятность того, что выборка имеет нормальное распределение, $p < 0.01$ (тесты Колмогорова–Смирнова и Лиллифорса); по тесту Шапиро–Вилка эта вероятность $p < 10^{-5}$ (Большев, Смирнов, 1983; Lilliefors, 1967; Shapiro et al., 1968). С 5%-м уровнем значимости тесты Вальда–Вольфовица ($p = 0.002507$) и Колмогорова–Смирнова ($p < 0.025$) позволяют отклонить гипотезу о симметрии распределения остатков. Получаем, что в рамках традиционного подхода модель Морана–Риккера следует признать непригодной для аппроксимации данных.

Замечание. Отмеченные (черным цветом) точки на рис. 2 соответствуют случаям, когда для фиксированных значений параметров A и a_0 находится по крайней мере одно начальное значение x_0^0 такое, что полученная при данных значениях траектория модели (5) удовлетворяет всем перечисленным выше статистическим критериям. Если точка никак не отмечена, то это означает, что какой-либо из критериев «не работает» (например, при достаточно малом числе положительных или отрицательных отклонений) или какой-то один критерий показывает отрицательный результат (при отсут-

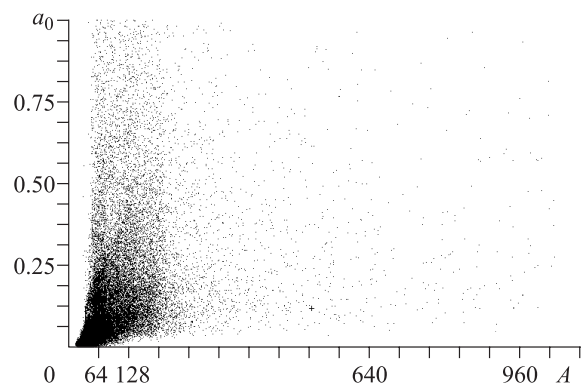


Рис. 2. Проекция допустимой области для модели (5) при $m = 1$ на плоскость (A, a_0) для временного ряда по колебаниям численности серой лиственничной листовертки (GPDD № 1407). Крестом отмечено место минимума функционала (2), (3).

ствии симметрии распределения отклонений, при наличии сериальной корреляции и т. д.). Крестом отмечена МНК-оценка параметров модели.

Как видно из рис. 2, наибольшая концентрация точек наблюдается в области (приблизительно) $18.5 \leq A \leq 70$, $a_0 \leq 0.05$. Таким образом, на основании проведенных расчетов можно сделать следующий вывод: модель Морана–Риккера (5) пригодна для описания динамики численности серой лиственничной листовертки.

Точка пространства параметров $x_0^0 = 50.6028$, $A = 29.0773$, $a_0 = 0.0263$ обладает следующими характеристиками: значение теста Колмогорова–Смирнова равно 0.2832 (гипотеза о симметрии отклонений не может быть отклонена с 99.9996%-м уровнем значимости), значение теста Лемана–Розенблатта равно 0.01864 (гипотеза не отклоняется с 99.7%-м уровнем значимости), значение теста Манна–Уитни равно 0.03026 (гипотеза не отклоняется с 96.8%-м уровнем значимости). Таким образом, все три используемых теста показывают, что нулевая гипотеза *должна быть принята*.

Нулевая гипотеза о равенстве нулю коэффициента ранговой корреляции Спирмена должна быть отклонена, поэтому в данном случае уровень значимости следует уменьшать для получения более сильного результата. Критическое значение коэффициента корреляции Спирмена равно 0.513 при объеме выборки 38 и 0.1%-м уровне значимости; для рассматриваемой точки пространства параметров значение равно 0.63913. Таким образом, гипотеза о монотонном поведении ветвей плотности вероятностей не может быть отклонена.

Тест серий показывает, что гипотеза об отсутствии сериальной корреляции в последовательности отклонений не может быть отвергнута с 25.4%-м уровнем значимости. Значение теста «скачков вверх – скачков вниз» равно 25 при критических значениях 19 и 31 (при 5%-м уровне значимости). При 20%-м уровне значимости критические значения равны 21 и 29, что также не позволяет отклонить гипотезу об отсутствии сериальной корреляции.

Значения автокорреляционной функции при указанных значениях параметров (значения функции вычислялись по 20 000 значений

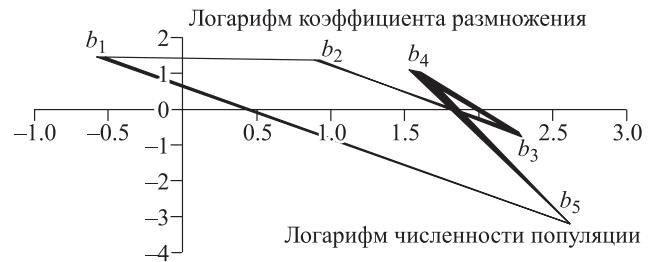


Рис. 3. Асимптотический режим для модели Морана–Риккера без запаздывания при $x_0^0 = 50.6028$, $A = 29.0773$, $a_0 = 0.0263$. Точки b_1, \dots, b_5 – элементы 5-цикла.

после 200 000 шагов работы модели «вхолостую») лежат в интервале $[-0.14807, 0.999838]$. Причем каждое пятое значение функции больше 0.999. Таким образом, асимптотический режим является нечетким циклом длины 5 (рис. 3): точки b_1, \dots, b_5 – элементы цикла представлены на плоскости $(\log_{10} x, \log_{10} y)$, где x – плотность популяции, y – коэффициент размножения (отношение плотностей двух смежных генераций).

При МНК-оценках автокорреляционная функция быстро снижается, и в дальнейшем ее значения не выходят за границы интервала $[-0.043, 0.0867]$ (Nedorezov, Sadykova, 2015). Подобное поведение автокорреляционной функции означает, что наблюдаемый режим является циклическим (с циклом в 1000 лет или более) или хаотическим. Как уже отмечено, для множества отклонений при этих оценках различные статистические тесты не выполняются.

Пусть ξ – целочисленная случайная величина, равная длине цикла асимптотически устойчивого режима, который реализуется для случайно выбранной точки из допустимого множества. Тогда $P\{\xi = k\}$ равна доле соответствующего динамического режима среди всех режимов допустимого множества. Распределение ξ может служить основанием для заключения о динамическом режиме, который реализуется у популяции. Для рассматриваемого случая получено, что вероятность события $\{\xi > 1000\}$ примерно равна 0.924. Вероятности остальных событий меньше: в частности, $P\{\xi = 5\} \approx 0.0206$, $P\{\xi = 8\} \approx 0.0015$. Таким образом, ни режимы, соответствующие наиболее сильным утверждениям (относительно симметрии), ни наиболее вероятные режимы, ни МНК-оценки не дают режима, который бы соответствовал сложившимся пред-

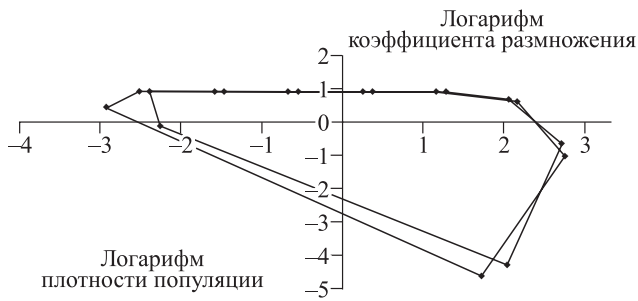


Рис. 4. Асимптотический режим для модели Морана–Риккера с запаздыванием при МНК-оценках параметров (цикл длины 18).

ставлениям о динамике листовёртки (Исаев и др., 1984, 2001; Auer, 1977; Baltensweiler, 1964, 1978). В этом смысле модель (5) при $t = 0$ непригодна для описания динамики серой листовёртки.

Вариант: модель Морана–Риккера с $t = 1$. При традиционном подходе (2), (3) для этого случая получены следующие результаты (Nedorezov, Sadykova, 2015): $Q_{\min} = 154\,148.4$, $x_0^0 = 4.148 \cdot 10^{-14}$, $x_1^0 = 0.1274$, $A = 8.3$, $a_0 = 2.205 \cdot 10^{-3}$, $a_1 = 0.02214$. При данных параметрах асимптотический режим – цикл длины 18 (с близкими значениями координат, что вполне можно принять за цикл длины 9; рис. 4), что согласуется с мнением биологов (Auer, 1977; Baltensweiler, 1964, 1978). Анализ отклонений показывает, что с уровнем значимости, меньшим 1 %, гипотеза о нормальности должна быть отклонена. Тест Сведа–Эйзенхарта указывает на наличие сериальной корреляции.

Разыгрывая значения случайной точки, равномерно распределенной в прямоугольнике $\Delta = [0, 100] \times [0, 100] \times [0, 1000] \times [0, 1] \times [0, 1]$, получаем возможность оценить структуру допустимого множества. Картина получается сходной с той, которая представлена на рис. 2 (очевидно, что допустимое множество для модели Морана–Риккера с $t = 0$ является подмножеством нового). При этом новых областей сгущения точек на плоскости (A, a_0) не наблюдается.

Для 40 000 точек, принадлежащих пересечению допустимого множества с Δ (вероятность обнаружить точку допустимого множества при разыгрывании равномерного распределения в Δ примерно равна $5.3356 \cdot 10^{-5}$), получено, что точка с координатами $x_0^0 = 44.705^4$, $x_1^0 = 53.333$, $A = 719.01$,

$a_0 = 0.0318$, $a_1 = 0.0017$ обладает наилучшими показателями (по симметрии). Для этой точки имеем следующие характеристики: значение теста Колмогорова–Смирнова равно 0.4425 (гипотеза о симметрии отклонений не отклоняется с 98.7411%-м уровнем значимости), значение теста Лемана–Розенблатта равно 0.03305 (93.315 %), значение теста Манна–Уитни равно 0.0908 (92.04 %). Таким образом, нулевая гипотеза *должна быть принята*.

Нулевая гипотеза о равенстве нулю коэффициента ранговой корреляции Спирмена должна быть отклонена при 0.1%-м уровне значимости: для рассматриваемой точки пространства параметров значение равно 0.74483, что больше критического 0.513. Тест серий показывает, что гипотеза об отсутствии сериальной корреляции в последовательности отклонений не может быть отклонена с 67.44%-м уровнем значимости. Значение теста «скачков вверх – скачков вниз» равно 21 при критических значениях 19 и 31 (при 5%-м уровне значимости) – нулевая гипотеза об отсутствии сериальной корреляции не отклоняется. Однако при 20%-м уровне значимости критические значения равны 21 и 29, что свидетельствует о наличии оснований для отклонения гипотезы об отсутствии сериальной корреляции.

Значения автокорреляционной функции при указанных параметрах (значения функции вычислялись по 20 000 значений после 200 000 шагов работы модели «вхолостую») лежат в интервале $[-0.033; 0.076]$. Динамика численности носит нерегулярный характер: в течение какого-то времени наблюдается монотонный рост численности, одномоментное снижение плотности (нет значений, соответствующих фазе разреживания и депрессии), после чего снова наблюдается монотонный рост. Множество точек асимптотически устойчивого режима на плоскости $(\log_{10}x, \log_{10}y)$ представлено на рис. 5.

Сходное поведение автокорреляционной функции, а также близкая структура асимптотического устойчивого режима динамики популяции наблюдаются при минимизации вероятности отклонения гипотезы об отсутствии сериальной корреляции и при минимизации сумм квадратов отклонений. Во всех указанных случаях асимптотический режим не является циклическим (с циклом меньше 1000 лет)

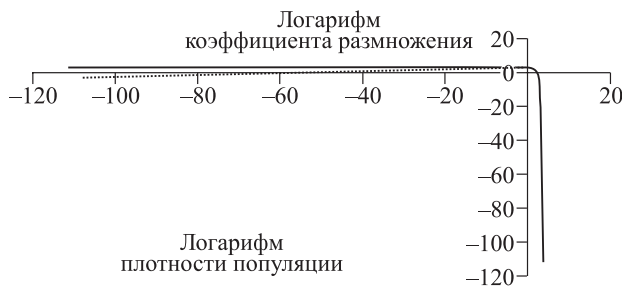


Рис. 5. Асимптотический режим для модели Морана-Риккера с запаздыванием.

и не обладает чертами, характерными для перманентной вспышки массового размножения (Исаев и др., 1984, 2001).

Анализ 40 000 точек допустимого множества показывает также, что среди них есть такие, которые соответствуют циклам от 1 (режим стабилизации на положительном уровне или незначительные флуктуации около ненулевого уровня с дисперсией, меньшей 10^{-40}) до 7, а также циклы длиной 11, 15, 16, 19, 27, 30, 56. Циклы длиной 8 или 9 не выявлены. Циклы с длинами, кратными 8 и 9 (16, 27, 56), представляют собой изменение численности с одним максимумом и «растянутой» фазой нарастания численности, что не соответствует наблюдаемым колебаниям численности. Таким образом, в допустимой области не выявлено режимов, сходных с наблюдаемыми в природной обстановке; близкий режим выявлен при использовании МНК, но отклонения теоретических и эмпирических данных не удовлетворяют всем статистическим критериям.

Для рассматриваемого случая получено, что $P\{\xi = 1\} \approx 0.0164$, $P\{\xi = 2\} \approx 0.0131$. Остальные вероятности для $\xi \leq 1000$ существенно меньше. Наибольшая вероятность равна 0.9546 и наблюдается для события $\{\xi > 1000\}$ — хаотического режима или цикла в 1000 лет или более.

Таким образом, в обоих рассмотренных случаях ($m = 0$ и $m = 1$) не удается получить динамические режимы (ни с помощью МНК, ни с помощью рассматриваемого нетрадиционного подхода), которые бы удовлетворяли и сложившимся представлениям о динамике серой лиственничной листовертки, и совокупности статистических тестов. Следовательно, в модели (5) необходимо учитывать запаздывание за большее число лет.

ДИНАМИКА ЧИСЛЕННОСТИ СОСНОВОЙ ПЯДЕНИЦЫ

Используемые в настоящем разделе работы данные (Klomp, 1966) по колебаниям численности сосновой пяденицы имеются в свободном доступе в базе данных в Интернете (NERC Centre for Population Biology, Imperial College (1999) The Global Population Dynamics Database, № 2727, 2728 и 2729). Эти данные в первом случае (временной ряд № 2727) представлены в единицах «среднее число отложенных яиц на квадратный метр», во втором (ряд № 2728) — «среднее число гусениц на квадратный метр», в третьем (ряд № 2729) — «средняя плотность куколок на квадратный метр». В первом случае объем выборки равен 15 (первое значение получено в 1950 г.), во втором — 14 (первое значение также получено в 1950 г., но данные по плотности гусениц за 1962 г. отсутствуют), в третьем случае имеем 14 значений (первое получено в 1951 г.).

Поскольку данные собирали в одном месте, имеем «связанные» временные ряды, непосредственно относящиеся к одному и тому же популяционному процессу. Поэтому прежде всего необходимо иметь модель, которая бы позволяла использовать одновременно все три временных ряда для оценки параметров и обладала богатым набором динамических режимов (Недорезов, 2010, 2011). Именно для такого случая разработана ELP-модель (Nedorezov, 2014):

$$\begin{aligned} E_{k+1} &= C\mu_1 P_k, \\ L_{k+1} &= \mu_2 E_{k+1}, \\ P_{k+1} &= L_{k+1} e^{-dL_{k+1}}. \end{aligned} \quad (6)$$

В модели (6) P_k — плотность куколок в момент времени k , $k = 0, 1, 2, \dots$, L_k — плотность гусениц, E_k — плотность отложенных яиц. Функции, стоящие в правой части системы уравнений (6), имеют вид

$$\mu_1 = \frac{g_1}{1 + g_2 L_k^{g_3}}, \quad C = \frac{c_1}{1 + c_2 L_k^{c_3}}. \quad (7)$$

Все параметры в формулах (6) и (7) неотрицательны. Функция μ_1 равна доле куколок, выживших за зимний период, $g_1 = \text{const} \leq 1$. Функция C равна средней плодовитости бабочек. В модели (6) предполагается, что и смерт-

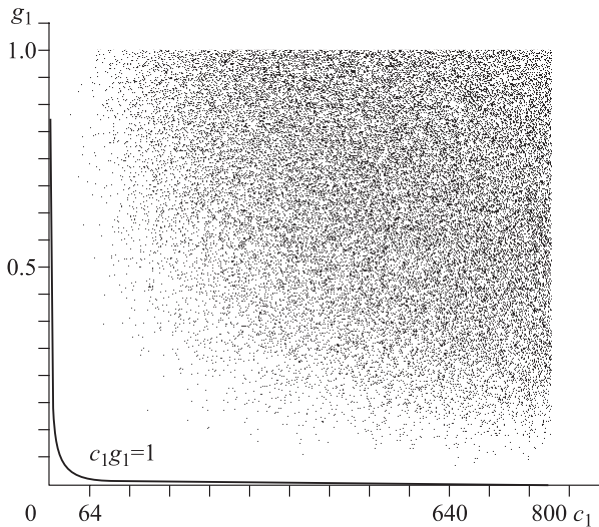


Рис. 6. Проекция допустимой области для ELP-модели (6) на плоскость (c_1, g_1) .

ность куколок за зимний период, и плодовитость бабочек зависят от условий питания гусениц в соответствующий год. Параметр μ_2 – коэффициент успешного отрождения гусениц, $\mu_2 = \text{const} \leq 1$; α – коэффициент саморегуляции (в данном случае – коэффициент внутривидовой конкуренции за корм). Заметим, что в модели (6) и (7) при допустимых значениях параметров траектории продолжаемы по времени, что важно для построения прогнозов динамики численности (Недорезов, 2010, 2011).

В модели (6) имеется 8 неизвестных параметров. К ним следует добавить начальное значение плотности яиц E_0 (этого достаточно для «запуска» модели). Таким образом, по имеющимся временным рядам следует оценить значения девяти неизвестных показателей. При оценке значений параметров модели использовался следующий *основной принцип*: точка девятимерного пространства параметров считалась принадлежащей допустимому множеству, если для *каждого* временного ряда статистические критерии показывали требуемый результат.

При разыгрывании равномерного распределения в прямоугольнике $E_1 \in [0, 200]$, $\mu_2 \in [0, 1]$, $c_1 \in [0, 800]$, $c_2 \in [0, 10]$, $c_3 \in [0.5, 2.5]$, $g_1 \in [0, 1]$, $g_2 \in [0, 6]$, $g_3 \in [0.1, 4.1]$, $\alpha \in [0, 2]$ получено допустимое множество (при 5%-м уровне значимости), представленное в проекции на плоскость (c_1, g_1) на рис. 6. $c_1 g_1 = 1$ – бифуркационная кривая, граница области

вырождения популяции. Всего получено 10^6 точек допустимого множества. Ни одно из значений не попало в область $c_1 g_1 < 1$ (рис. 6). Вероятность нахождения точки допустимого множества примерно равна $8.182 \cdot 10^{-4}$.

Для случая, когда нулевая гипотеза о симметрии распределения отклонений по критерию Колмогорова–Смирнова не может быть отклонена с 95%-м уровнем значимости (для трех временных рядов), выявлены две точки в допустимом множестве. В первом случае $\mu_2 = 0.1162$, $c_1 = 350.34$, $c_2 = 5.31$, $c_3 = 1.41$, $g_1 = 0.92$, $g_2 = 4.0$, $g_3 = 3.27$, $\alpha = 0.021$, $E_1 = 195.76$. Расчеты значений автокорреляционной функции (после 10^6 шагов работы модели «вхолостую») показывают, что каждое значение функции с номером $3k$, $k = 1, 2, \dots$, превосходит величину 0.965. Это дает основание полагать, что в данном случае имеет место нечеткий цикл длины 3. На рис. 7 представлены 30 значений данного цикла по изменению значений плотности яиц и гусениц.

Во втором случае $\mu_2 = 0.1285$, $c_1 = 213.96$, $c_2 = 0.91$, $c_3 = 1.08$, $g_1 = 0.506$, $g_2 = 0.446$, $g_3 = 1.983$, $\alpha = 0.105$, $E_1 = 182.26$. Асимптотический режим – цикл длины два с координатами $a = 35.49$ и $b = 5.35$ для плотности яиц, $a = 4.56$ и $b = 0.69$ для плотности гусениц и $a = 2.83$ и $b = 0.64$ для плотности куколок.

Поскольку рассматриваемые временные ряды коротки, сказать что-либо о сходстве полученных результатов с представлениями биологов о популяционной динамике в данном локальном местообитании не представляется возможным, тем более что идентификация типа динамики численности по этим временным рядам встречает определенные трудности (Исаев и др., 1984, 2001; Kendall et al., 2005). Полученные результаты позволяют высказать гипотезу, что динамика вида в данном локаль-

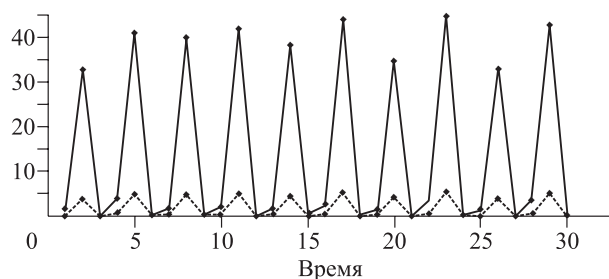


Рис. 7. Динамика плотности яиц (сплошная линия) и плотности гусениц (штриховая линия) во времени в стационарном режиме.

ном месте может соответствовать циклу не-большой длины. Иными словами, динамика сосновой пяденицы соответствует продромальному типу.

На это же указывают результаты оценки вероятностей для различных режимов. Для рассматриваемого случая получены следующие оценки для вероятностей ξ (оценки даны по результатам анализа 32 000 точек допустимого множества): наибольшая вероятность $P\{\xi = 2\} \approx 0.4831$, $P\{\xi = 4\} \approx 0.1224$. Вероятность того, что наблюдаемый режим имеет цикл в 1000 лет или больше (или наблюдается хаотический режим), примерно равна 0.2618. Вероятность $P\{\xi = 1\} \approx 0.04313$, остальные вероятности существенно меньше. Таким образом, есть основания полагать, что наблюдаемый режим является циклическим с периодом в 2 года.

ЗАКЛЮЧЕНИЕ

Метод наименьших квадратов, видимо, наиболее популярный для оценки параметров экологических моделей. В рамках этого метода можно условно выделить следующие этапы: сначала выбирается (или строится) функция потерь (в частности, в виде суммы квадратов отклонений теоретических и эмпирических данных). Эта функция минимизируется, что позволяет получить некие оценки параметров модели. Наконец, на последнем шаге проводится статистический анализ совокупности отклонений значений модели, полученных при найденных оценках параметров, от имеющихся эмпирических данных. Это вызывает определенные нарекания со стороны специалистов (см., например, Лемешко и др., 2011), поскольку с использованием одного и того же массива данных оцениваются значения параметров модели и проводится проверка свойств отклонений. В качестве альтернативы предлагается по одной части временного ряда оценивать значения параметров модели, а проверять свойства отклонений – по другой части. Однако для коротких (в подавляющем большинстве) экологических временных рядов подобная альтернатива неприемлема.

Если в результате анализа совокупности отклонений должна быть отклонена гипотеза либо об их нормальном распределении, либо об отсутствии сериальной корреляции в по-

следовательности остатков, то в таком случае модель признается непригодной для аппроксимации данных. Иными словами, заключение о пригодности или непригодности модели для аппроксимации данных делается на основе анализа поведения траектории в одной-единственной точке пространства параметров.

Важный этап в использовании метода наименьших квадратов (МНК) – выбор вида функции потерь. В настоящее время никаких критериев выбора вида данной функции не существует, и отчасти именно поэтому в литературе можно найти огромное число модификаций этой функции (например, в виде суммы модулей отклонений, суммы квадратов отклонений с какими-либо неотрицательными весами и т. д.), используемых в анализе биологических данных. Использование каждой из подобных модификаций приводит не только к различным результатам, но и к непониманию того, каким именно результатам следует отдать предпочтение.

Свобода в выборе вида функции потерь обусловлена тем обстоятельством, что она никак не связана ни с имеющимися данными, ни с моделью, ни с анализируемым биологическим процессом. Это позволяет предположить, что в основе метода наименьших квадратов лежит логическая ошибка (как, впрочем, и в основе метода максимального правдоподобия). Прежде чем минимизировать некую функцию, нужно определить требования к модели, к отклонениям теоретических от эмпирических данных и выяснить, существуют ли вообще такие значения параметров, при которых эти требования выполняются. Не исключено, что множество таких точек в пространстве параметров пусто.

Один из таких возможных и очевидных наборов требований представлен в работе. Частично требования совпадают с общепринятыми: отсутствие сериальной корреляции в последовательности отклонений, симметрия относительно нуля распределения отклонений, монотонность ветвей функции плотности (последние два использовались вместо требования нормальности отклонений). Множество точек пространства параметров, для которых указанные требования выполняются, названо допустимым. Когда допустимое множество не пусто, то вновь не требуется строить какую-либо функцию потерь – среди точек

допустимого множества можно найти такие, для которых наблюдаются наиболее сильные результаты (если гипотеза о симметрии не отклоняется с 95%-м уровнем значимости, то это более сильный результат, чем когда та же гипотеза не отклоняется с 5%-м уровнем значимости). Именно такие точки пространства параметров и следует использовать в качестве оценок.

Заметим, что при таком подходе к оценке параметров также нельзя гарантировать единственность решения, как и при МНК, когда речь идет о нелинейных экологических моделях. Однако важно другое: при таком подходе появляется возможность *естественным* образом использовать несколько временных рядов одновременно для оценки параметров модели и не строить какие-либо *искусственные* функции потерь.

Допустимое множество обладает важным свойством: любой его элемент можно использовать в качестве оценок параметров модели. Используемые статистические критерии не дают оснований для заключения о том, что модель при соответствующих значениях параметров дает неудовлетворительную аппроксимацию данных. В таком случае важными характеристиками являются меры подмножеств допустимого множества, которые соответствуют тому или иному динамическому режиму. Меры этих подмножеств пропорциональны вероятностям «попадания» в них при разыгрывании равномерно распределенной случайной точки на допустимом множестве. Если вероятность, соответствующая какому-либо режиму, существенно больше других вероятностей, то это вполне может служить основанием для утверждения о том, что именно этот режим реализуется у популяции.

Для временного ряда по колебаниям численности серой листовенничной листовертки получено, что допустимые множества (для модели Морана–Риккера без запаздывания и с запаздыванием в один год) не пусты. В то же время проведенный анализ свойств элементов допустимых множеств показал, что среди них нет таких, которые бы соответствовали сложившимся биологическим представлениям о динамике листовертки. Полученные МНК-оценки при учете запаздывания в один год позволили получить такой режим (цикл длины 18 с двумя максимумами), однако при этом все

статистические критерии дали отрицательный результат. Полученные результаты показывают, что запаздывание в действии саморегуляторных механизмов больше одного года, что требует использования модели с показателем $m \geq 2$.

Для связанных временных рядов по колебаниям численности сосновой пяденицы в работе продемонстрировано, как именно эти ряды могут быть одновременно учтены при оценке параметров модели. Проведенный анализ показывает, что есть все основания для утверждения, что наблюдаемый режим — циклический с длиной цикла в два года.

СПИСОК ЛИТЕРАТУРЫ

- Бард Й.* Нелинейное оценивание параметров. М.: Статистика, 1979. 349 с.
- Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. М.: Наука, 1983. 416 с.
- Боровков А. А.* Математическая статистика. М.: Наука, 1984. 472 с.
- Дрейпер Н., Смит Г.* Прикладной регрессионный анализ. Т. 1. М.: Финансы и статистика, 1986. 366 с.
- Дрейпер Н., Смит Г.* Прикладной регрессионный анализ. Т. 2. М.: Финансы и статистика, 1987. 351 с.
- Ермаков С. М.* Метод Монте-Карло и смежные вопросы. М.: Наука, 1975. 472 с.
- Исаев А. С., Недорезов Л. В., Хлебопрос Р. Г.* Влияние трансформации корма на динамику численности насекомых // Консортивные связи дерева и дендрофильных насекомых. Новосибирск: Наука. Сиб. отд-ние, 1982. С. 67–75.
- Исаев А. С., Хлебопрос Р. Г.* Принцип стабильности в динамике численности лесных насекомых // Докл. АН СССР. 1973. Т. 208. № 1. С. 225–228.
- Исаев А. С., Хлебопрос Р. Г.* Эффекты запаздывания в регуляции численности лесных насекомых // Докл. АН СССР. 1977. Т. 232. № 6. С. 1448–1451.
- Исаев А. С., Хлебопрос Р. Г., Кондаков Ю. П.* Закономерности динамики численности лесных насекомых // Лесоведение. 1974. № 3. С. 27–42.
- Исаев А. С., Хлебопрос Р. Г., Недорезов Л. В., Киселев В. В., Кондаков Ю. П.* Динамика

- численности лесных насекомых. Новосибирск: Наука. Сиб. отд-ние, 1984. 224 с.
- Исаев А. С., Хлебопрос Р. Г., Недорезов Л. В., Киселев В. В., Кондаков Ю. П., Суховольский В. Г. Популяционная динамика лесных насекомых. М.: Наука, 2001. 347 с.
- Лемешко Б. Ю., Лемешко С. Б., Постовалов С. Н., Чимитова Е. В. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход. Новосибирск: Изд-во НГТУ, 2011. 888 с.
- Ликеш И., Ляга Й. Основные таблицы математической статистики. М.: Финансы и статистика, 1985. 356 с.
- Михайлов Г. А. Некоторые вопросы теории методов Монте-Карло. Новосибирск: Наука. Сиб. отд-ние, 1974. 143 с.
- Недорезов Л. В. Влияние качества корма и паразитизма на циклические колебания серой листовёртки // Евразийский энтомолог. журн. Т. 6. № 2. 2007. С. 229–244.
- Недорезов Л. В. Анализ динамики численности сосновой пяденицы с помощью дискретных математических моделей // Математическая биология и биоинформатика. 2010. Т. 5. № 2. С. 114–123.
- Недорезов Л. В. Динамика численности сосновой пяденицы: прогноз с помощью дискретных математических моделей // Евразийский энтомолог. журн. 2011. Т. 10. № 3. С. 280–288.
- Недорезов Л. В. Хаос и порядок в популяционной динамике: моделирование, анализ, прогноз. Саарбрюкен: Lambert Acad. Publ., 2012. 352 с.
- Холлендер М., Вулф Д. Непараметрические методы статистики. М.: Финансы и статистика, 1983. 520 с.
- Auer C. Dynamik von Larchenwicklerpopulationen langs des Alpenbogens // Mitteilungen der Eidgenossischen Anstalt fiir forstliches Versuchswesen. 1977. V. 53. P. 71–105.
- Baltensweiler W. Zeiraphera griseana Hubner (Lepidoptera, Tortricidae) in the European Alps. A contribution to the problem of cycles // Can. Entomol. 1964. V. 96. N. 5. P. 792–800.
- Baltensweiler W. Ursache oder Wirkung? Huhn oder Ei? // Mitteilungen der Schweizerischen Entomologischen Gesellschaft. 1978. V. 51. P. 261–267.
- Kendall B. E., Ellner S. P., McCauley E., Wood S. N., Briggs C. J., Murdoch W. W., Turchin P. Population cycles in the pine looper moth: dynamical tests of mechanistic hypotheses // Ecol. Monogr. 2005. V. 75. N. 2. P. 259–276.
- Klomp H. The dynamics of a field population of the pine looper *Bupalus piniarius* L. (Lep., Geom.) // Adv. Ecol. Res. 1966. V. 3. P. 207–305.
- Lilliefors H. W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown // J. Amer. Statist. Ass. 1967. V. 64. P. 399–402.
- McCallum H. Population parameters estimation for ecological models. Brisbane: Blackwell Sci. Ltd., 2000. 224 p.
- Moran P. A. P. Some remarks on animal population dynamics // Biometrics. 1950. V. 6. N. 3. P. 250–258.
- Nedorezov L. V. Basic properties of ELP-model // Population Dynamics: Analysis, Modelling, Forecast. 2014. V. 3. N. 4. P. 93–101.
- Nedorezov L. V., Sadykova D. L. Dynamics of larch bud moth populations: application of Moran–Ricker models with time lags // Ecol. Model. 2015. V. 297. P. 26–32.
- Ricker W. E. Stock and recruitment // J. Fish. Res. Board Can. 1954. V. 11. N. 5. P. 559–623.
- Sadykova D. L., Nedorezov L. V. Larch bud moth dynamics: can we explain periodicity of population fluctuations by the time lag dependence in birth rate? // Population Dynamics: Analysis, Modelling, Forecast. 2013. V. 2. N. 4. P. 154–181.
- Shapiro S. S., Wilk M. B., Chen H. J. A comparative study of various tests of normality // J. Amer. Stat. Ass. 1968. V. 63. P. 1343–1372.
- Wood S. N. Minimizing model fitting objectives that contain spurious local minima by bootstrap restarting // Biometrics. 2001a. V. 57. P. 240–244.
- Wood S. N. Partially specified ecological models // Ecol. Monogr. 2001b. V. 71. P. 1–25.

Non-Traditional Approach to Estimation of Ecological Models Parameters (On Examples of Larch Bud Moth and Pine Looper Moth Population Dynamics)

L. V. Nedorezov

*Center for Interdisciplinary Studies of Environmental Problems, Russian Academy of Sciences
Nabereznaya Kutuzova, 14, Saint-Petersburg, 191187 Russian Federation*

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

In current publication, a non-traditional approach to estimation of parameters of ecological models is considered. Within the framework of this approach it is not necessary to construct loss-function or likelihood function. The basic idea of method is as follows: first of all, basic requirements to model and to deviate between theoretical (model) and empirical datasets must be formulated. After that, respective statistical criteria must be selected, and with the help of these statistical criteria, a *feasible set* in space of model parameters must be determined (where all statistical criterions demonstrate required results). Within the limits of this feasible set must be found elements with *strongest properties*. These elements can be used as estimations of model parameters. Constructions of feasible sets for Moran – Ricker model with and without time lag and ELP-model were demonstrated on examples of larch bud moth (*Zeiraphera diniana* Gn.) and pine looper moth (*Bupalus piniarius* L.) time series. Results of approximation are discussed. For a time series of population dynamics of larch budworm it was found that feasible sets are not empty for the model Moran – Ricker without delay and with delay of one year. At the same time, the analysis of properties of elements of admissible sets showed that among them there are no properties that match the prevailing biological understanding of the dynamics of the *Zeiraphera diniana* Gn. The resulting estimation allowed taking into account the delay of one year to receive treatment with a cycle length of 18 with two maxima. However, almost all statistical tests gave negative results. These results indicate that the delay in the self-regulation mechanisms of action of more than one year. Given way to simultaneously estimate the parameters of the population dynamics model of pine looper for related time series. The analysis shows that there is every reason for the claim that the observed mode – cyclic short cycle length of two years.

Keywords: *population dynamics, time series, mathematical model, parameter estimation, larch bud moth, pine looper moth.*

How to cite: *Nedorezov L. V. Non-traditional approach to estimation of ecological models parameters (on examples of larch bud moth and pine looper moth population dynamics) // Sibirskij Lesnoj Zurnal (Siberian Journal of Forest Science). 2015. N. 3: 70–82 (in Russian with English abstract).*