

В.М. СКРОБАЛА

Український державний лісотехнічний університет,
вул. О.Кобилянської, 1, 79005, м. Львів

**МОДЕЛЮВАННЯ БІОЛОГІЧНИХ ПРОЦЕСІВ МЕТОДАМИ
БАГАТОВИМІРНОЇ ОПТИМІЗАЦІЇ**

ключові слова: біологічні процеси, математичне моделювання, багатовимірна оптимізація

key words: biological processes, mathematical modeling, multivariate optimization

V.M. SKROBALA

**MODELING OF BIOLOGICAL PROCESSES BY THE METHODS OF
MULTIVARIATE OPTIMIZATION**

Ukrainian State University of Forestry and Wood Technology
1 Kobylyanska str., Lviv, 79005, Ukraine

It was proposed algorithm, which foresee a using of multivariate square interpolation-extrapolation for the purposes of mathematical modeling of biological processes. It was analyzed preferences and defects of proposed approach in comparison with a method of least squares.

Сучасний етап розвитку біології базується на тісній і багатогранній взаємодії її методів з експериментальними й теоретичними методами інших наук, зокрема, з широким використанням ідей і методів математики [2, 3, 6, 7, 9]. Математика все більше перетворюється у звичайний інструмент досліджень, потребу у використанні якого відчувають спеціалісти різних галузей знань. У той же час існуючий арсенал математичних методів не завжди задовільняє потреби біології [2]. Пояснюється це тим, що біологія має справу з об'єктами й процесами виняткової складності, значно складнішими, ніж ті, що є предметом досліджень у фізиці [1, 2, 11, 12]. Тому окремі спроби використання неадекватних математичних методів для розв'язання біологічних проблем призвели до математичного формалізму [2]. З іншого боку, рівень підготовки біологів у галузі математики не завжди достатньо високий для того, щоби коректно використовувати існуючі або створювати нові математичні підходи до вирішення біологічних задач [2, 10].

Багатофакторність та нелінійний зв'язок між параметрами створюють чималі труднощі у математичному моделюванні біологічних процесів [6, 9]. Немає універсального способу для вирішення проблеми вибору функціональної залежності [6, 8]. Інколи аналіз графічного зображення наявних даних, а також розуміння суті механізму дослі-

джуваного процесу допомагають вибрати певний вид аналітичної залежності. Але, як правило, проблему вибору останньої доводиться вирішувати методом спроб і помилок, використовуючи як критерій суму квадратів відхилень фактичних і теоретичних значень функції [2, 6, 8, 9, 10]:

$$F = \sum_{i=1}^n [y_i - f(x_{1i}, x_{2i}, \dots, x_{ji}, a_1, a_2, \dots, a_k)]^2 = \min, \quad (1)$$

де F – сума квадратів відхилень; y_i – фактичне значення функції; f – теоретичне значення функції; x – незалежна змінна; j – кількість аргументів; a – коефіцієнт рівняння, який підлягає визначенню; k – кількість невідомих коефіцієнтів рівняння; n – кількість спостережень.

Для вирішення задачі (1) методом найменших квадратів потрібно обчислити часткові похідні функції F за коефіцієнтами a_1, a_2, \dots, a_k , прирівняти їх до нуля та розв'язати отриману систему k лінійних рівнянь [2, 6, 8, 9]. Складність цього завдання для біолога-дослідника полягає в тому, що для вибору іншої цільової функції F задачу (1) доводиться вирішувати багатократно, витрачаючи багато часу на зміну алгоритму комп'ютерної програми. Метод найменших квадратів стає непридатним у випадку вибору нелінійної за коефіцієнтами функції [8, 10]. Так, для простої нелінійної функції виду:

$$y = a_1 + a_2 \exp(a_3 x) \quad (2)$$

доведеться розв'язувати систему нелінійних рівнянь [10]:

$$\begin{cases} na_1 + a_2 \sum \exp(a_3 x_i) = \sum y_i \\ a_1 \sum \exp(a_3 x_i) + a_2 \sum \exp(2a_3 x_i) = \sum y_i \exp(a_3 x_i) \\ a_1 \sum x_i \exp(a_3 x_i) + a_2 \sum x_i \exp(2a_3 x_i) = \sum y_i x_i \exp(a_3 x_i) \end{cases} \quad (3)$$

У цьому випадку рекомендується використовувати ітеративні методи, зокрема, метод Ньютона-Рафсона [4, 8]. Проте, для більшості спеціалістів-біологів це завдання може виявитися непосильним. У науковій літературі, присвяченій математичному аналізу біологічних даних, рідко можна знайти рекомендації, як адаптувати статистичні методи лінійної регресії для нелінійної за параметрами функції. Окремі поради щодо поєднання ітерацій з методом найменших квадратів наведені в роботах [6, 10]. Так, для залежності (2) можна припустити, що параметр a_3 відомий, і необхідно знайти тільки a_1 і a_2 . У цьому випадку залежність (2) є лінійною за невідомими параметрами a_1 і a_2 , і для їх визначення можна використати метод найменших квадратів.

Зрозуміло, що параметр a_3 не є відомим, тому потрібно багаторазово змінювати його наближене значення, щоби досягти необхідного мінімуму суми квадратів відхилень. Такий підхід використовувався нами для побудови математичних моделей продуктивних запасів ґрунтової вологи та випаровування в умовах зелених насаджень м.Львова [13, 14]. Недолік цього методу полягає у збільшенні затрат часу на складання алгоритму й програмування порівняно з безпосереднім часом комп'ютерних обчислень.

Для спеціаліста-біолога вигідніше менше витратити часу на складання алгоритмів, якщо є можливість адекватно перекласти значний обсяг роботи на машину за умови однакової точності й повноти результатів. Досягти вказаної ситуації можна шляхом спрощення алгоритму за рахунок збільшення кількості обчислювальних операцій. Відносно апроксимації залежностей це означає використання такого методу, де коефіцієнти цільової функції знаходяться шляхом послідовного поліпшення з урахуванням критерію (1) або величини помилки.

У ряді математичних програм багатовимірного статистичного аналізу для визначення параметрів нелінійної моделі регресії реалізований метод, що не потребує визначення часткових похідних – так званий DUD-метод [5, 15]. Запозичивши ідею цього методу, а також відомі алгоритми багатовимірної оптимізації [4, 5, 10], наше завдання полягало в конструюванні такої програми, користування якою вимагає мінімально необхідного обсягу математичних знань і мінімальних навичок програмування. Створюючи програму, ми враховували той факт, що на практиці часто постає задача пошуку екстремумів (максимумів або мінімумів) деякої цільової функції $F(x_1, x_2, \dots, x_k)$ к змінних x_i . Така функція описує $(k+1)$ -вимірну поверхню. У найпростішому випадку функція $F(x)$ одного аргументу описує деяку криву на площині. Складність нашого завдання полягає в тому, що цільова функція F задана у вигляді суми (1), яку потрібно визначити багаторазово. Проте висока швидкодія сучасних комп'ютерів дозволяє досить просто вирішити цю проблему. Реалізація запропонованого методу в простому випадку функції з однією змінною $F(x)$ здійснюється за таким алгоритмом [4]:

1. Задати початкове наближене значення x_1 для мінімуму й визначити два суміжні значення аргумента $F(x)$:

$$x_0 = x_1 - h, \quad x_2 = x_1 + h,$$

де h – півінтервал інтерполяції-екстраполяції.

2. Обчислити три значення $F(x)$: $F(x_0)=F_0$, $F(x_1)=F_1$, $F(x_2)=F_2$.

3. Аналітично визначити положення мінімуму, замінивши функцію $F(x)$ на проміжку $x_{1\pm}h$ квадратичною параболою $y(x)=x^2+bx+c$

$$x_{\min} = -\frac{b}{2c} = \frac{F_0(2x_1 + h) - 4F_1x_1 + F_2(2x_1 - h)}{2(F_0 - 2F_1 + F_2)}, \quad (4)$$

1. Перевірити виконання умови $|x_{\min} - x_1| < \varepsilon$. Якщо умова не виконується, задати $x_{\min} = x_1$ і повернутися до п.1. Якщо умова виконується, вважати x_{\min} -знайденим із заданою похибкою ε , визначити $F(x_{\min})$ і зупинити обчислення.

У випадку функції багатьох змінних $F(x_1, x_2, \dots, x_k)$ після знаходження точки мінімуму за координатою x_1 потрібно перейти до знаходження точки мінімуму за координатою x_2 і т.д. з перевіркою умови одночасно для k змінних. Для розв'язання задачі (1) потрібно тільки врахувати, що невідомими змінними у нас є коефіцієнти цільової функції a_1, a_2, \dots, a_k , у той час як аргументи x_j є відомими. З математичної точки зору, суть пропонованого методу зводиться до розв'язання системи n рівнянь із k невідомими, причому $n \gg k$ (n – кількість спостережень; k – кількість коефіцієнтів аналітично заданої функції), тоді як у методі найменших квадратів кількість рівнянь у системі завжди дорівнює k . Для розв'язування задачі (1) можна використати й інші алгоритми багатовимірної оптимізації, зокрема, координатного спуску і спірального координатного спуску [4, 10].

Для роботи з програмою потрібно виконати такі пункти:

1. Задати масиви значень аргументів x_1, x_2, \dots, x_k і функції y .
2. Задати цільову функцію у вигляді виразу (1).
3. Задати масив наближених значень коефіцієнтів цільової функції $A(k)$, значення похибки ε та інтервалу інтерполяції h .

У випадку зміни цільової функції потрібно уточнити тільки пункти 2 і 3. Програма дозволяє використовувати для апроксимації залежностей математичні моделі високого рівня складності, в тому числі нелінійні за параметрами.

Досвід роботи з програмою дозволив виявити низку її недоліків:

а) результати обчислень є наближеними, вони залежать від вхідних параметрів ε, h і масиву $A(k)$;

б) визначений мінімум може виявитися локальним, а не глобальним; результати слід уточнити, змінивши вхідні параметри $\varepsilon, h, A(k)$;

в) у процесі обчислень необхідно стежити за проміжними значеннями цільової функції (суми квадратів відхилень), які повинні систематично зменшуватися; у випадку циклічної зміни значень функції обчислення відбуваються безкінечно довго, потрібно зупинити виконання програми й задати інші вхідні параметри $\varepsilon, h, A(k)$;

г) у процесі обчислень трапляються випадки ділення на машинний нуль (division by zero) і переповнення (overflow). Машина, як правило, легко справляється з цією проблемою. У протилежному випадку слід

зупинити виконання програми й задати інші вхідні параметри;

д) за великої кількості спостережень і незалежних змінних x_j обчислення доцільно починати, задавши високі значення похибки ϵ та інтервалу пошуку h ; отримані значення коефіцієнтів треба уточнити, зменшивши вхідні параметри ϵ , h .

Слід зауважити, що робота над створенням математичної моделі полягає, переважно, у спостереженні за ходом обчислень і внесенні оперативних змін у виконання програми. Готових рецептів (яку цільову функцію і вхідні параметри використовувати) немає. Постійний пошук оптимальних рішень та орієнтація на фізичний зміст досліджуваного явища є найкращими гарантіями правильності розв'язку. Автор статті сподівається, що запропонований ним метод математичного моделювання на основі багатовимірної інтерполяції-екстраполяції стане в пригоді багатьом біологам-дослідникам.

ЛІТЕРАТУРА

1. Арманд Д.Л. Наука о ландшафте (основы теории и логико-математические методы). – М.: Мысль, 1975. – 288 с.
2. Владимирский Б.М. Математические методы в биологии. – Ростов: Изд-во Ростов. ун-та, 1983. – 304 с.
3. Горя В.С. Алгоритмы математической обработки результатов исследований. – Кишинев: Штиинца, 1978. – 118 с.
4. Дьяконов В.П. Справочник по алгоритмам и программам на языке бейсик для персональных ЭВМ. – М.: Наука, 1989. – 240 с.
5. Енюков И.С. Методы, алгоритмы, программы многомерного статистического анализа. – М.: Финансы и статистика, 1986. – 232 с.
6. Зайцев Г.Н. Математика в экспериментальной ботанике. – М.: Наука, 1990. – 296 с.
7. Лиена И.Я. Математические методы в биологических исследованиях. Факторный и компонентный анализы. – Рига: Изд-во Латв. гос. ун-та, 1980. – 104 с.
8. Львовский Е.Н. Статистические методы построения эмпирических формул. – М.: Высшая школа, 1988. – 239 с.
9. Плохинский Н.А. Математические методы в биологии. – М.: Изд-во Моск. ун-та, 1978. – 265 с.
10. Поллард Дж. Справочник по вычислительным методам статистики. – М.: Финансы и статистика, 1982. – 344 с.
11. Сачок Г.И., Цуркова Т.Ф. Математико-картографическое моделирование природных условий Белоруссии. – Мн.: Наука и техника, 1984. – 252 с.
12. Скробала В.М., Данилик Р.М. Моделювання біогеоценотичного покриву урбанізованих територій методами багатовимірного статистичного аналізу // Екологічний стрес і адаптація в біологічних системах. – Тернопіль: Вид-во Терноп. держ. пед. ун-ту, 1998. – С.137-139.
13. Скробала В.М. Антропогенна трансформація водного режиму ґрунтів (на прикладі м.Львова) // Наук. вісник Львівського УкрДІТУ. – Вип. 9.1. – 1998. – С.67-74.
14. Скробала В.М. Визначення вологозабезпеченості ґрунту лісових насаджень статистичними методами // Наук. вісник Львівського УкрДІТУ. – Вип. 9.10. – 1999. – С.190-193.
15. Ralson M.L., Jennrich R.I. DUD, a Derivative-Free Algorithm for Nonlinear least Squares. – Technometrics. – 1978. – 20, № 1. – P.7-14.