

І.О. Калініна, О.П. Гожий, О.П. Нечахін, С.І. Шиян

## СИНТЕЗ ПАРАМЕТРІВ НЕЛІНІЙНОЇ ПРОГНОЗНОЇ МОДЕЛІ ЗА ДОПОМОГОЮ ГЕНЕТИЧНОГО АЛГОРИТМУ

*Анотація.* В статті розглядається визначення та оцінка параметрів нелінійної регресійної моделі за допомогою генетичного алгоритму. Оцінка параметрів, є різновидом задачі оптимізації, що дає змогу її розв'язувати за допомогою стохастичних алгоритмів, це представляє можливість використання таких алгоритмів, щоб відповідати моделям нелінійної регресії. Було досліджено набір даних *trees*, який відображає нелінійну залежність між ознаками. В наборі міститься 1200 спостережень і три змінних: ID дерева; вік дерева; об'єм деревини. Була проаналізована залежність об'єму деревини (*Vol*) від віку дерева (*Age*). Увага на локації дерев дозволило підібрати логістичну функцію Річардса в якості функціональної залежності, нелінійною як за параметрами, так і за змінною. Для оцінки параметрів функції Річардса було використано генетичний алгоритм. Були обчислені коефіцієнти нелінійних моделей для випадково обраних локацій дерев. Для оцінки якості моделей використано коефіцієнт детермінації  $R^2$ .

*Ключові слова:* Генетичний алгоритм, метод найменших квадратів, нелінійна модель, функція Річардса, параметри моделі.

### Вступ

Статистичне моделювання має важливе значення для підтримки фундаментальних і наукових гіпотез в теоретичних дослідженнях, а також процесів моделювання реальних процесів на практиці. Однією з головних проблем, яка виникає при побудові моделей є врахування нелінійностей, які притаманні процесу, який моделюється. Моделі з нелінійностями, або нелінійні моделі мають велику кількість прикладних застосувань у фізичних, біологічних та соціальних науках, бізнесі, інженерії, економіці та управлінні [1-3]. На сьогоднішній день визначена та успішно застосована велика кількість нелінійних моделей для вирішення різних завдань. Однак існує велика кількість ситуацій, які не можуть бути нелінійно змодельовані через те, що ситуації можуть бути складними або математично, чи статистично важкорозв'язуємі. Нелінійна поведінка деяких відносин ускладнює завдання мінімізації втрат конкретної моделі, яка

може бути додатково ускладнена математичними характеристиками прикладної моделі.

Однім із засобів подолання таких проблем є застосування нелінійних регресійних моделей. Нелінійна регресійна модель — це нелінійна функція з принаймні однією незалежною змінною та однією залежною змінною [2]. Основним завданням для побудови нелінійної регресійної моделі є визначення та оцінка параметрів. Оцінка параметрів, по суті, є типом задачі оптимізації, що дає змогу її розв'язувати за допомогою стохастичних алгоритмів, це представляє можливість використання таких алгоритмів, щоб відповідати моделям нелінійної регресії. Однак такі алгоритми мають обмеження щодо стабільності та стійкості їх результатів [1]. Прикладом є генетичний алгоритм (ГА).

Генетичні алгоритми є евристичними методами пошуку, які емулюють природну генетику та дарвінівські принципи еволюційного відбору [4]. Генетичні алгоритми використовують для вирішення широкого кола завдань, таких як дискретна оптимізація, багатокритерійна оптимізація, пошук і генерація нових рішень, а також для вирішення задач нелінійної оптимізації [5]. Він складається з набору потужних пошукових засобів евристики [6] і придатний для прогнозування нелінійних параметрів, завдяки своїй основній характеристиці отримання кількох рішень з кожною ітерацією [7]. Враховуючи це, та обмеження традиційних нелінійних алгоритмів оптимізації, таких як алгоритм Левенберга–Марквардта, коли початкові припущення параметрів мають бути визначеними та призводити до збіжності оцінок параметрів, метаевристики, то генетичний алгоритм може бути використано, для того щоб обрахувати параметри нелінійних моделей.

### **Постановка проблеми**

Метою даної роботи є дослідження процесу визначення та оцінки параметрів нелінійних прогнозних моделей за допомогою генетичного алгоритму.

### **Аналіз останніх досліджень**

Поряд із численними успішними застосуваннями в широкому ряді дисциплін [4-6], генетичні алгоритми були запропоновані для вирішення завдань регресійного аналізу. Відповідні дослідження можна розділити на три групи: ГА для оцінки параметрів регресійної моделі, ГА які ідентифікують регресійну модель для даної проблеми, і ГА які використовують як для оцінки, так і для ідентифікації регресійної моделі.

До першої групи можна віднести наступні наукові публікації. В роботі [5] оцінено параметри лінійної та двох нелінійних регресійних моделей за допо-



Для аналізу характеру зав'язків аналізується графік залежності об'єму деревини (Vol) від віку дерева (Age) (рис. 2а).

При візуальному аналізі виявлено наявність додаткових ознак, які впливають на об'єм деревини. Для подальшої роботи обираються лише дані для дерев з певним місцем розташування, наприклад, ID = 1.3.11 (рис. 2б). В результаті вибірка має 12 спостережень та двома змінними (рис.3).

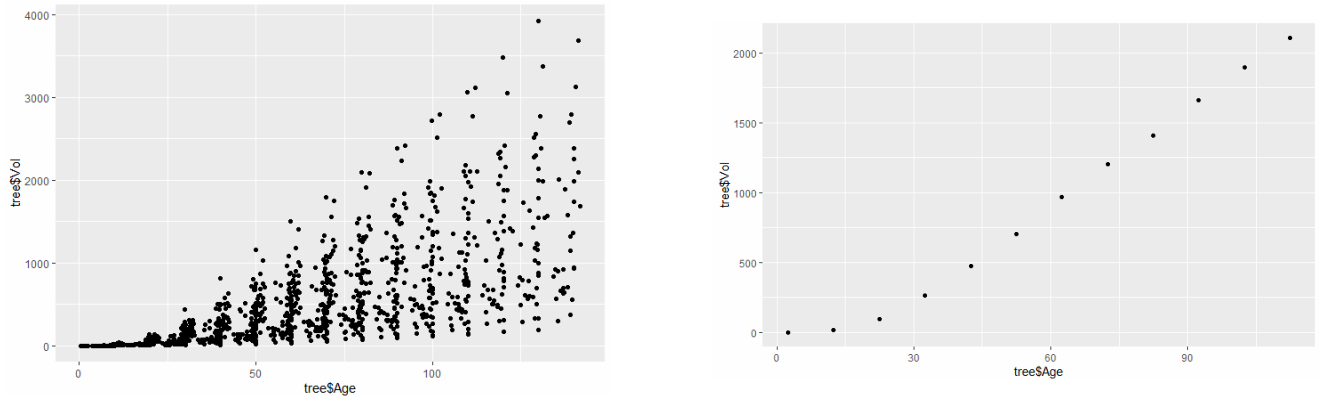


Рисунок 2 –Графіки залежності об'єму деревини (Vol) від віку дерева (Age)  
а) для повного набору даних; б) для вибірки ID = 1.3.11

```

ROWS: 12
Columns: 2
$ Age <dbl> 2.44, 12.44, 22.44, 32.44, 42.44, 52.44, 62.44, 72.44, 82.44, 92.44~
$ Vol <dbl> 2.2, 20.0, 93.0, 262.0, 476.0, 705.0, 967.0, 1203.0, 1409.0, 1659.0~
    
```

Рисунок 3 – Вибірка з набору даних trees

Залежність об'єму деревини (y) від його віку дерева (x) описується за допомогою логістичної функції Річардса [5]:

$$y = a(1 - e^{-bx})^c.$$

Наступним кроком буде застосування методу найменших квадратів (МНК). Параметри якого входять в нелінійну модель:

$$a, b, c = \arg \min_{a, b, c} \sum_{i=1}^{12} \left( Vol_i - a(1 - e^{-bx})^c \right)^2.$$

Мінімізація суми квадратів помилок вирішується, за допомогою генетичного алгоритму. В середовищі програмування R генетичний алгоритм реалізується за допомогою функції ga, яка максимізує fitness-функцію. Вона задана задалегідь.

Так як поставлено завдання мінімізації, то скористаємося наступним фактом:

$$\arg \min_{\Theta} (\Theta) = \arg \min_{\Theta} (-\Theta).$$

Для цього необхідно задати функцію Річардса. Вона буде використана для

fitness-функцію, яка вказана в аргументах функції *ga*. Задається функція *fit* суми квадратів помилок. Функція при роботі виводить в консоль значення *fit* на кожній ітерації алгоритму (рис. 4).

```
iter = 1 | Mean = -71956884.8 | Best = -643228.9
iter = 2 | Mean = -63488550.80 | Best = -60568.38
iter = 3 | Mean = -54229145.71 | Best = -16069.03
iter = 4 | Mean = -49497174.27 | Best = -16069.03
iter = 5 | Mean = -42940214.660 | Best = -4380.708
iter = 6 | Mean = -38641933.25 | Best = -4024.16
iter = 7 | Mean = -30580733.91 | Best = -4024.16
iter = 8 | Mean = -23764633.39 | Best = -4024.16
iter = 9 | Mean = -15770500.00 | Best = -4024.16
iter = 10 | Mean = -11832425.11 | Best = -4024.16
```

Рисунок 4 – Значення *fit* на кожній ітерації алгоритму

Результат роботи генетичного алгоритму залежить від того, яким чином налаштовані його параметри. При реалізації генетичного алгоритму були встановлені наступні параметри:

- *Popsizе*=500 – розмір популяції;
- *Maxiter*=5000 – максимальне число ітерацій, після якого робота генетичного алгоритму припиняється;
- *Run*=200 – число послідовних поколінь без будь-якого покращення в значенні *fitness*-функції перед зупиненням алгоритму.

Для операції селекції використана випадкова схема. При операції мутації виконується зміна випадково обраного біта. Вид операції кросоверу – одноточковий. Кросовер використовується для 80% хромосом кожного покоління.

Кінцевим результатом застосування генетичного алгоритму є підбір параметрів функції Річардса. На рисунку 5а представлено результат роботи генетичного алгоритму (підбор параметрів функції).

В результаті підбору параметрів отримано наступне рівняння регресії:

$$y = 3597.809(1 - e^{-b \cdot 0.015 \cdot x})^{2.780}.$$

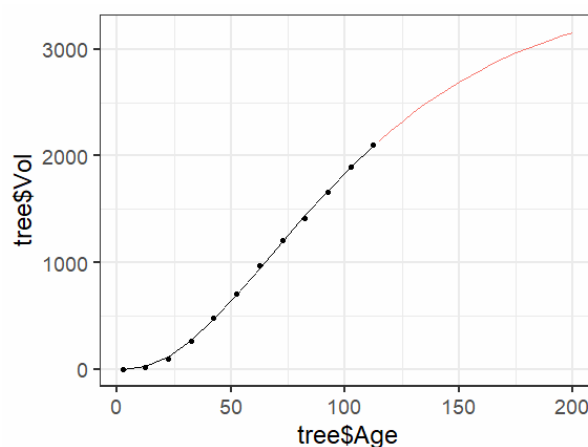
Обчислене значення коефіцієнту детермінації  $R^2 = 0.9996$  свідчить про високу якість моделі. На основі дослідженої моделі отримуємо прогноз об'єму деревини від віку дерева. Точковий графік вихідних даних, модельні значення по функції Річардса (гладка чорна лінія) та прогноз об'єму деревини від віку дерева до 200 років (гладка червона лінія) представлено на рисунку 5б.

-- Genetic Algorithm -----

```
GA settings:
Type           = real-valued
Population size = 500
Number of generations = 5000
Elitism        = 25
Crossover probability = 0.8
Mutation probability = 0.1
Search domain =
  a b c
lower 3000 0 2
upper 4000 1 4
```

```
GA results:
Iterations           = 503
Fitness function value = -2774.012
Solution =
```

```
      a          b          c
[1,] 3597.809 0.01541044 2.780185
```



а)

б)

Рисунок 5 – а) результати підбору параметрів функції;  
б) графік вихідних даних та отриманої функції Річардса з прогнозом об'єму деревини від віку дерева до 200 років

На рисунку 6 представлено зміни значень *fitness* - функції від числа ітерацій генетичного алгоритму.

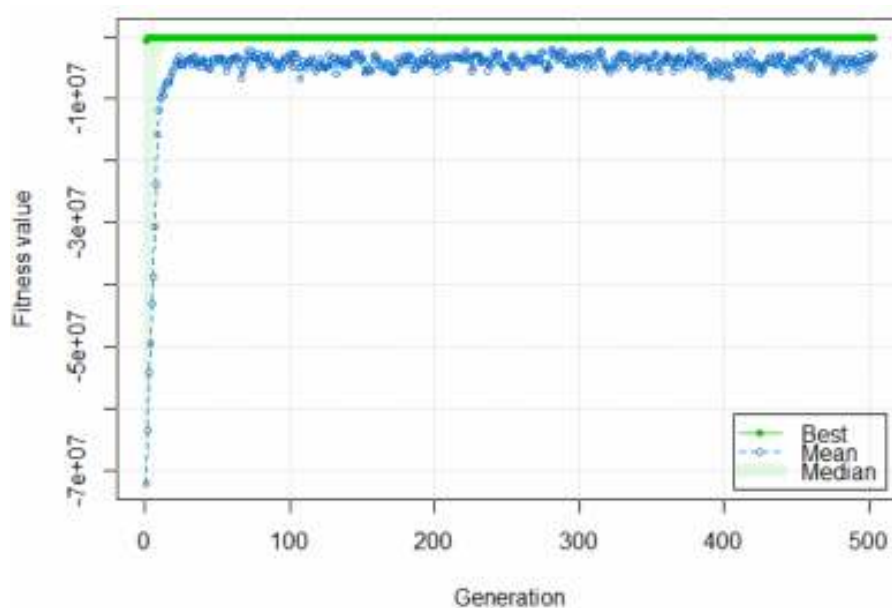


Рисунок 6 – Зміна значень *fitness*-функції від числа ітерацій

Параметри нелінійної функції Річардса для різних вибірок з набору даних

№	ID	$n$	$a$	$b$	$c$	$R^2$
1	1.3.11	12	3597,809	0,015	2,780	0,999
2	2.4.7	13	3999,838	0,014	3,367	0,995
3	1.1.3	14	4000,000	0,013	2,624	0,998
4	5.7.21	13	3000,418	0,002	2,000	0,993
5	3.1.12	14	3346,793	0,010	2,822	0,999
6	5.4.10	13	3001,146	0,010	3,226	0,999
7	4.7.15	11	3999,981	0,007	2,837	0,992
8	4.5.37	12	3000,001	0,005	2,192	0,999
Весь набір даних		1200	3999,999	0,007	2,078	0,497

Для аналізу змін параметрів нелінійної функції береться випадково різні вибірки даних із генеральної сукупності за допомогою ознаки ID. Параметри нелінійної функції Річардса представлено в таблиці 1. Для оцінки точності моделі в таблиці 1 також представлено відповідні значення коефіцієнту детермінації  $R^2$ .

### Висновки

В роботі представлено дослідження пошуку параметрів нелінійної регресії за допомогою генетичного алгоритму. Було досліджено набір даних «trees», що відображає складний нелінійний процес. Для аналізу змін параметрів нелінійної функції, були використані випадкові вибірки даних із генеральної сукупності. Головна увага в наборі даних була сфокусована на локації дерев. Це дозволило підібрати логістичну функцію Річардса в якості функціональної залежності. Вона є нелінійною як за параметрами, так і за змінною. Для оцінки параметрів функції Річардса було використано генетичний алгоритм. Високе значення коефіцієнту детермінації  $R^2$  підтвердило якість моделей на всіх досліджуваних вибірках з набору даних.

### ЛІТЕРАТУРА

1. Бідюк П.І., Гожий О.П. Ймовірно-статистичні методи моделювання і прогнозування: [монографія] . – Миколаїв: Вид-во ЧДУ ім. Петра Могили, 2014. – 440 с.

2. De Gooijer J.G. Elements of Nonlinear Time Series Analysis and Forecasting. Springer Series in Statistics. Library of Congress Control Number: 2017935720. 2017, 620 p. DOI 10.1007/978-3-319-43252-6/
3. Bidyuk P., Gozhyj A., Kalinina I., Vysotska V. (2020) Methods for Forecasting Nonlinear Non-stationary Processes in Machine Learning. Communications in Computer and Information Science. – Volume 1158. – Springer, Cham, 2020. – pp. 470-485. – Print ISBN978-3-030-61655-7, Online ISBN978-3-030-61656-4. DOI: [https://doi.org/10.1007/978-3-030-61656-4\\_32](https://doi.org/10.1007/978-3-030-61656-4_32).
4. Gen M., Cheng R. (2000), Genetic Algorithms and Engineering Optimization (New York: John Wiley & Sons Inc).
5. Kapanoglu M., Koc I.O., Erdogmus S., Genetic algorithms in parameter estimation for nonlinear regression models: an experimental approach, Journal of Statistical Computation and Simulation, 77(10):851-867, (2007).
6. Haj Seyed Hadi M.R., Gonzalez-Andujar J.L. Comparison of Fitting Weed Seedling Emergence Models with Nonlinear Regression and Genetic Algorithm. Comput. Electron. Agric. 2009, 65, pp. 19–25.
7. Mitra S., Mitra A., Kundu D. Genetic algorithm and M-estimator based robust sequential estimation of parameters of nonlinear sinusoidal signals. Communications in Nonlinear Science and Numerical Simulation. Vol. 16, Issue 7, 2011, pp. 2796-2809. <https://doi.org/10.1016/j.cnsns.2010.10.005>.
8. Paterakis E., Petridis V., Kehagias A. Genetic algorithm in parameter estimation of nonlinear dynamic systems. International Conference on Parallel Problem Solving from Nature, 2006, 171 Accesses, 2 Citations, pp. 1008–1017. DOI: 10.1007/BFb0056942.
9. Monti C.A.U., Oliveira R.M., Roise J.P., Scolforo H.F., Gomide L.R. Hybrid Method for Fitting Nonlinear Height–Diameter Functions. Journals “Forests”, Vol. 13, Issue 11, 2022. <https://doi.org/10.3390/f13111783>.
10. Leichter, Y., Sethares, W.A. Nonlinear Parameter Estimation via the Genetic Algorithm. IEEE Trans. Signal Process. 1994, 42, pp. 927–935.
11. Chatterjee, S., Laudato M., Lynch, L.A. Genetic Algorithms and Their Statistical Applications: An Introduction. Comput. Stat. Data Anal. 1996, 22, pp. 633–651.
12. Jin Y.-F., Yin Z.-Y., Shen S.-L., Hicher, P.-Y. Selection of Sand Models and Identification of Parameters Using an Enhanced Genetic Algorithm. Int. J. Numer. Anal. Methods Geomech. 2016, 40, pp. 1219–1240.
13. VanderNoot T.J. Abrahams I. The Use of Genetic Algorithms in the Non-Linear Regression of Immittance Data. J. Electroanal. Chem. 1998, 448, pp. 17–23.

14. Lacerda T.H.S., Miranda E.N., E Lopes I.L., Fonseca G.R., França L.C., de Jesus França L.C., Gomide L.R. Feature Selection by Genetic Algorithm in Nonlinear Taper Model. Can. J. For. Res. 2022, 52, pp. 769–779.

15. Zhou X., Wang J., (2005) A genetic method of LAD estimation for models with censored data. Computational Statistics and Data Analysis, 48(3), pp. 451–466.

Received 01.03.2023.

Accepted 03.03.2023.

***Synthesis of the parameters of a nonlinear predictive model  
using a genetic algorithm***

*The article deals with the definition and estimation of the parameters of a nonlinear regression model using a genetic algorithm. Parameter estimation is a type of optimization problem that can be solved using stochastic algorithms, which presents the possibility of using such algorithms to fit nonlinear regression models. The trees data set, which displays the non-linear relationship between traits, was investigated. The nonlinear method of least squares (NLS) was used to find the coefficients of the regression equation using a genetic algorithm on the trees data set belonging to the spuRs library. The set contains 1200 observations and three variables: tree ID; age of the tree; volume of wood. Visual analysis revealed the presence of additional signs that affect the volume of wood. Only data for trees with a certain location are selected for further work. As a result, the sample has 12 observations and two variables. The dependence of the volume of wood (Vol) on the age of the tree (Age) was analyzed. Attention to tree locations made it possible to choose the Richards logistic function as a functional dependence, nonlinear both in terms of parameters and variables. A genetic algorithm was used to estimate the parameters of the Richards function. The result of the work of the genetic algorithm depends on how its parameters are configured. When implementing the genetic algorithm, the following parameters were set: population size; the maximum number of iterations after which the work of the genetic algorithm stops; the number of consecutive generations without any improvement in the value of the fitness function before stopping the algorithm. A random scheme was used for the selection operation. A mutation operation involves changing a randomly selected bit. The type of crossover operation is one-point. Coefficients of nonlinear models were calculated for randomly selected tree locations. The coefficient of determination  $R^2$  was used to assess the quality of the models.*

**Калініна Ірина Олександрівна** – доцент кафедри інтелектуальних інформаційних систем Чорноморського національного університету ім. Петра Могили, к.т.н., доцент

**Гожий Олександр Петрович** – професор кафедри інтелектуальних інформаційних систем Чорноморського національного університету ім. Петра Могили, д.т.н., професор

**Нечахін Владіслав Володимирович** – аспірант кафедри інтелектуальних інформаційних систем Чорноморського національного університету ім. Петра Могили.

**Шиян Сергій Іванович** – аспірант кафедри інтелектуальних інформаційних систем Чорноморського національного університету ім. Петра Могили.

**Kalinina Iryna Oleksandrivna** - Associate Professor of the Department of Intellectual Information Systems of the Black Sea National University named after Petra Mohyly, Ph.D., associate professor

**Gozhyj Oleksandr Petrovych** - professor of the Department of Intellectual Information Systems of the Black Sea National University named after Petra Mohyly, Doc.of Ing.Sc., professor

**Nechahin Vladislav Volodymyrovych** - PhD student of the Department of Intellectual Information Systems of the Black Sea National University named after Petra Mohyly.

**Shiyan Serhii Ivanovych** – PhD student of the Department of Intellectual Information Systems of the Black Sea National University named after Petera Mohyly.