УДК 004.428.4: 519.657

Л.П. ВАКАЛ

ПРОГРАМНІ ЗАСОБИ ДЛЯ НАБЛИЖЕННЯ ДОСЛІДНИХ ДАНИХ ЕМПІРИЧНИМИ ФОРМУЛАМИ

L. Vakal

SOFTWARE FOR EXPERIMENTAL DATA APPROXIMATION BY EMPIRIC FORMULA

Methods and software for experimental data approximation by empiric formulas are reviewed. Bundled software for empiric formulas construction which has some advantages is described .

Рассматриваются методы и программные средства для приближения данных измерений эмпирическими формулами. Описывается программный комплекс построения эмпирических формул, который обладает рядом преимуществ.

Розглядаються методи та програмні засоби для наближення даних вимірювань емпіричними формулами. Описується програмний комплекс побудови емпіричних формул, який має ряд переваг.

Вступ. При математичній обробці даних спостережень, дослідів або експериментів часто виникає задача знаходження функціональної залежності між фізичними величинами x та y за результатами вимірювань, поданими у вигляді таблиць відповідних значень (x_i, y_i), $i = \overline{1, n}$. Через обмежену кількість значень та наявність випадкових помилок вимірювань (так званого шуму в експерименті) однозначно відтворити функціональну залежність неможливо. Йдеться лише про побудову формули, яка достатньо добре наближала досліджувану залежність у діапазоні значень $x_1 \le x \le x_n$.

Для розв'язання цієї задачі можуть успішно застосовуватися емпіричні формули, які дають досить точне наближення дослідних даних, є відносно простими і дозволяють згладжувати похибки вимірювань. Зауважимо, що на відміну від емпіричних використання інтерполяційних формул не дозволить розв'язати задачу належним чином, оскільки при інтерполяції зберігається увесь шум [1].

Велика кількість формул, що застосовуються в науці та техніці, є емпіричними. Це, наприклад, формули залежності опору металевого стрижня від температури в фізиці; показника політропи стиснення від кутової швидкості валу в техніці; швидкості потоку рідини, що проходить через шар піни, від об'єму рідини в піні в хімії та ін. [2–4].

В історії науки багато прикладів того, як отримання вдалої емпіричної формули стало першою, дуже важливою віхою на шляху до великих наукових відкриттів. Класичним прикладом цього факту є відкриття закону випромінювання Планка на основі побудови трьох емпіричних формул.

Методи побудови емпіричних формул. Задача побудови емпіричної формули для наближеного подання експериментальних даних (x_i, y_i), $i = \overline{1, n}$, полягає у знаходженні формули (функції)

$$y = f(x;\ a_0, a_1, \ldots, a_m), \tag{1}$$

значення якої при $x = x_i$ якомога менше відрізнялися від значень y_i. Невідомі величини a_i називаються параметрами емпіричної формули. Найпростішою емпіричною формулою з двома параметрами є лінійна функція $y = a_1 x + a_0$, а з трьома параметрами – квадратична функція $y = a_2 x^2 + a_1 x + a_0$.

Процес побудови емпіричної формули складається з вибору типу формули та знаходження найкращих значень її параметрів.

Вибір типу формули – головне, що визначає результативність і точність емпіричної формули. До цього часу немає загального методу для визначення найбільш підходящого типу емпіричної формули для подання тих або інших даних. Іноді його можна визначити на основі теоретичних міркувань щодо характеру досліджуваної залежності. Якщо ж характер залежності невідомий, то функціональний вигляд емпіричної формули, взагалі кажучи, може бути довільним. На практиці перевага віддається більш простим формулам (з 2–3 параметрами), що мають достатню точність. Оскільки фізичні процеси здебільшого мають нелінійний характер, то для більш адекватного їх представлення доцільно вибирати нелінійні функції.

Найбільш поширений підхід, коли тип формули вибирають шляхом порівняння кривої, побудованої за експериментальними даними, з графіками елементарних функцій. При порівнянні графіків можливі різні складні ситуації, зокрема, коли експериментальна крива схожа на графіки одразу декількох функцій або, навпаки, мало схожа на типовий графік функції, яка цілком відповідає її характеру. У першій ситуації слід вибирати такий тип формули, якій найбільше відповідає фізичній суті явища. В іншій ситуації причин її виникнення може бути декілька: невдалий вибір масштабу координатних осей, відповідність експериментальної кривої тільки деякій частині типового графіку функції (наприклад, експериментальна крива є опуклою кривою, що не має вираженого максимуму, разом з тим їй може відповідати рівняння параболи [5]). Такий підхід до вибору типу формули є досить суб'єктивним, вдалий результат багато в чому залежить від досвіду та інтуїції дослідника. Існують також спеціальні прийоми для перевірки можливості виразити експериментальну криву даним типом формули (вирівнювання, спрямлення [1, 2]).

Щоб не залежати від суб'єктивних чинників та автоматизувати сам процес визначення типу формули, запропоновано алгоритм [6], в якому тип емпіричної формули вибирається з набору найпоширеніших функцій (набір містить більше

20 формул) на основі певних аналітичних критеріїв. В алгоритмі для кожного типу формули спочатку виконуються відповідні перетворення даних, які дозволяють посилити в них роль лінійної або квадратичної компоненти, а далі перевіряються необхідні та достатні умови наявності лінійної чи квадратичної залежності. Формула, для якої відповідний критерій виконується точніше, і є найбільш підходящою.

Після вибору типу емпіричної формули необхідно знайти найкращі значення її параметрів a_i. Під терміном "найкращі" розуміємо такі значення, за яких деяка визначена міра відхилення експериментальних даних y_i від значень, обчислених за емпіричною формулою, буде мінімальною. Міра відхилення описує критерій (спосіб) наближення і визначає метод знаходження параметрів. Далі коротко проаналізуємо критерії та методи знаходження найкращих значень параметрів емпіричної формули, які найчастіше застосовуються на практиці.

Так у методі середніх параметри a_i знаходять з умови рівності нулю алгебраїчної суми $\sum_{i=1}^{n} \Delta_i = 0$, де відхилення Δ_i визначаються формулою

$$\Delta_i = y_i - f(x_i; a_0, a_1, \ldots, a_m). \tag{2}$$

У методі найменших квадратів для визначення параметрів a_i застосовується критерій квадратичного наближення

$$\sum_{i=1}^{n} \Delta_i^2 \Rightarrow \min. \tag{3}$$

Цей метод має ту перевагу перед методом середніх, що якщо сума (3) мала, то й відхилення Δ_i також є малими за модулем. Для методу середніх, де береться алгебраїчна сума відхилень, такого висновку зробити не можна. Тому для більш точних обчислень краще використовувати метод найменших квадратів. Слід додати, що замість формули (3) точність наближення методом найменших квадратів іноді зручніше оцінювати за середньоквадратичним критерієм

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n} \Delta_i^2} \Rightarrow \min. \tag{4}$$

Якщо похибки експериментальних даних мають нормальний закон розподілу, то критерій (3) можна розглядати як статистичний**,** а отримані за методом найменших квадратів значення a_i є найбільш ймовірними.

При лінійному входженні параметрів в емпіричну формулу їх значення знаходяться за методом найменших квадратів досить просто з нормальної системи лінійних рівнянь. Якщо ж параметри входять у формулу нелінійно, то система рівнянь для їх визначення також є нелінійною і знаходження її розв'язку звичайно пов'язано з громіздкими обчисленнями. Крім того, нелінійна система часто може мати декілька розв'язків. Якщо при цьому початкові значення параметрів (методи знаходження розв'язків нелінійної системи є ітераційними) вибрані невдало, то це може привести до незадовільного результату наближення емпіричною формулою. Тому намагаються звести нелінійний випадок до лінійного, наприклад відповідним перетворенням формули із заміною основних змінних x та y на нові [1, 5]. Однак варто пам'ятати, що перетворення формули порушує принцип найменших квадратів (3), тобто формула, яка задовольняє умові мінімуму в нових змінних не буде задовольняти такій умові після повернення до основних змінних x і y.

Найкращі значення параметрів a_i емпіричної формули можна визначати також за критерієм найкращого чебишовського наближення

$$\max_{1 \le i \le n} |\Delta_i| \Rightarrow \min . \qquad (5)$$

Використання для побудови емпіричної формули принципу мінімакса (5) гарантує, що модулі відхилень Δ_i не перевищать заданої похибки наближення на всьому відрізку $x_1 \le x \le x_n$. Чебишовський критерій доцільно застосовувати, зокрема, для обчислень з високою точністю та в задачах ущільнення результатів вимірювань, що містять методичну похибку у відомих межах (це можуть бути, наприклад, масиви геофізичних чи сейсмічних вимірювань [7]). Методи чебишовського наближення табличних залежностей лінійними формулами розроблені досить добре й успішно використовуються на практиці [8]. Методи нелінійного наближення є досить громіздкими, хоча наближення деякими типами нелінійних формул вдається знайти за спрощеними методами [9].

Останнім часом підвищився інтерес до визначення параметрів формули з умови мінімуму суми модулів відхилень

$$\sum_{i=1}^{n} |\Delta_i| \Rightarrow \min . \qquad (6)$$

Огляд існуючих програмних засобів для наближення дослідних даних. Практично в усіх відомих універсальних пакетах програм, таких як Mathcad, Maple, Mathematica та ін., реалізовано наближення дослідних даних за методом найменших квадратів, хоча й у різному обсязі.

Так у пакеті Mathcad існує можливість наближення довільними лінійними формулами (функція *linfit*) та наближення типовими нелінійними формулами з 2–3 параметрами, наприклад $ae^{bx} + c$, $a/(1 + be^{cx})$, $a\ln(x + b) + c$, $ax^b + c$, $a\ln x + b$, $a + bx$, $a\sin(x + b) + c$, що реалізується функціями *expfit, lgsfit, logfit, pwrfit, lnfit, medfit* та *sinfit* відповідно [10]. При роботі з функціями *expfit, lgsfit, pwrfit* та іншими необхідно задавати початкові значення невідомих параметрів для розв'язання системи нелінійних рівнянь ітераційним методом. Оскільки нелінійна система часто має декілька розв'язків, то невдалі початкові значення можуть привести до незадовільного результату наближення. Крім того, у пакеті Mathcad є функція *genfit*, яка теоретично дозволяє виконувати наближення більш складними нелінійними виразами. Але для багатьох користувачів пакета ця функція є «китайською грамотою» з непрозорим способом виклику [11]. Як аргументи вона вимагає, окрім самої формули, також її часткові похідні по шуканих параметрах.

Наближення лінійними відносно невідомих параметрів формулами реалізовано також в інших математичних пакетах – Maple (команда *leastsquare*), Mathematica (функції *FindFit, Fit*) та MATLAB (функція *polyfit* для наближення поліномом) [12, 13]. Крім того, для наближення деякими простими нелінійними формулами у системі Mathematica передбачено функцію *NonlinearFit*, а в пакеті MATLAB – функцію *lsqcurvefit*.

Слід зауважити, що задача знаходження формул, які найкраще описують експериментальні дані, виникає також у регресійному аналізі. У цьому випадку апроксимація даних здійснюється з урахуванням їх статистичних параметрів, оскільки вони є результатом вимірювань процесів або явищ, статистичних за своєю природою або на високому рівні шумів. Види регресії називаються відповідно до типу апроксимуючої формули, і для знаходження їх параметрів, як правило, застосовується метод найменших квадратів. Тому в ряді випадків для визначення параметрів емпіричної формули можна скористатися програмними засобами для побудови регресій.

Що стосується наближення дослідних даних за чебишовським критерієм (5), то жоден з вищеназваних універсальних математичних пакетів їх не виконує. Для цього необхідно скористатися спеціалізованими програмними засобами. Наприклад, наближення лінійними емпіричними формулами (поліномами та ін.) виконуються в деяких бібліотеках програм з чисельного аналізу [14, 15]. При цьому користувачу необхідно додатково задавати початкові значення або іншу інформацію для роботи програм. Знайти параметри нелінійних формул за критерієм (5) можна тільки за допомогою спеціалізованих пакетів програм для чебишовського наближення, наприклад [7, 16].

З наведеного огляду існуючих програмних засобів можна зробити висновок, що найкраще в них реалізовано наближення за критерієм методу найменших квадратів (3). Однак і в цьому випадку при визначенні параметрів нелінійних формул не завжди вдається отримати бажаний результат. Програмні засоби для знаходження наближень дослідних даних емпіричними формулами за іншими критеріями є, в основному, спеціальними і доступ до них пов’язаний для користувача з різного роду труднощами (організаційними, фінансовими та ін.).

Отже, актуальною є розробка програмних засобів для наближення експериментальних даних емпіричними формулами з використанням різних критеріїв наближення, з можливістю програмного визначення найбільш підходящого типу формули та графічними представленням отриманих результатів.

Програмний комплекс побудови емпіричних формул. Цей комплекс призначений для наближеного подання дослідних даних за допомогою найбільш поширених типів емпіричних формул з 1–3 параметрами, перелік яких наведено в табл. 1.

Програмний комплекс розроблений у системі програмування Delphi і орієнтований на користувачів “нематематиків”. Він зручний та інтуїтивно зрозумілий у використанні.

ТАБЛИЦЯ 1. Перелік типів формул для наближення дослідних даних

№ п/п	Число параметрів	Формула	№ п/п	Число параметрів	Формула
1	1	$y = ax$	11	3	$y = ax^2+bx+c$
2	2	$y = ax+b$	12	3	$y=1/(a+bx+cx^2)$
3	2	$y = ax^b$	13	3	$y = a+b/x+c/x^2$
4	2	$y = ae^{bx}$	14	3	$y=x^2/(ax^2+bx+c)$
5	2	$y = a\ln x+b$	15	3	$y = x/(a+bx+cx^2)$
6	2	$y = 1/(a+be^{-x})$	16	3	$y=(cx^2+bx+a)/x$
7	2	$y = 1/(ax+b)$	17	3	$y=a+b\ln x+c\ln^2 x$
8	2	$y = x/(ax+b)$	18	3	$y=a\,e^{bx+cx^2}$
9	2	$y = (ax+b)/x$	19	3	$y=a\,e^{bx}+c$
10	2	$y = \ln(ax+b)$			

У програмному комплексі реалізовано 6 критеріїв знаходження найкращих значень параметрів формули, за яких вона реалізує такі способи наближення:

1) найкраще абсолютне середньоквадратичне наближення (4);

2) найкраще відносне середньоквадратичне наближення

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\Delta_i}{y_i}\right)^2} \Rightarrow \min\ ;$$

3) найкраще абсолютне чебишовське наближення (5);

4) найкраще відносне чебишовське наближення

$$\max_{1\le i\le n}\left|\frac{\Delta_i}{y_i}\right| \Rightarrow \min\ ;$$

5) найкраще абсолютне наближення сумою модулів відхилень (6);

6) найкраще відносне наближення сумою модулів відхилень

$$\sum_{i=1}^{n}\left|\frac{\Delta_i}{y_i}\right| \Rightarrow \min\ .$$

За допомогою цього програмного комплексу можна розв’язати такі задачі.

Перша задача – знаходження найкращих (за відповідним критерієм) значень параметрів формули заданого типу. Тип формули вибирається із запропонованих списків відповідно до числа параметрів або задається самостійно (для цього потрібно у списках формул з 2 і 3 параметрами вибрати позицію з назвою «Інша формула»).

Друга задача – побудова емпіричної формули, яка найкраще (за певним критерієм) серед усіх формул із заданим числом параметрів наближає дослідні дані.

Алгоритм розв’язання першої задачі, який реалізовано у програмному комплексі, детально описано в [17]. Обчислювальна схема алгоритму в загальних рисах є такою:

– виконується перетворення даних (своє для кожного типу формули), яке дозволяє перейти для перетворених даних від нелінійної формули до лінійної функції (у випадку двох параметрів) або квадратичної функції (у випадку трьох параметрів);

– методом найменших квадратів знаходять коефіцієнти лінійної або квадратичної функцій;

– через знайдені коефіцієнти за формулами зворотного перерахунку обчислюються початкові значення параметрів нелінійної емпіричної формули;

– способом пошуку на решітці [18] ці значення уточнюються з тим, щоб визначити найкращі згідно із вибраним критерієм значення.

Якщо користувач сам задає тип емпіричної формули, а не вибирає із запропонованих списків, то йому необхідно вказати початкові значення параметрів самостійно. Однак навіть за невдалого вибору початкових значень бажане наближення буде отримано, хоча за більш тривалий проміжок часу.

При розв'язанні другої задачі вищеописаний алгоритм модифікується. Спочатку знаходяться наближення даних усіма типами формул із заданим числом параметрів, а потім серед них вибирається найкраща формула, яка забезпечує найменшу відповідно до вибраного критерію похибку наближення [17].

Результатами розв'язання першої і другої задач є значення параметрів a_i, найбільш підходящий тип формули $f(x;\, a_0, a_1, \ldots, a_m)$ (тільки для другої задачі), похибка наближення за вибраним критерієм та значення $f(x_i)$. Крім того, для візуалізації якості отриманого наближення виводяться графіки експериментальної кривої та емпіричної формули.

Програмний комплекс включає також довідкову систему, яка містить основні відомості щодо наближення експериментальних даних емпіричними формулами, правил роботи з комплексом та ін.

Приклади використання програмного комплексу. Розглянемо два приклади застосування описаного комплексу для побудови емпіричних формул. У прикладі 1 розв'язується перша задача, а у прикладі 2 для тих самих дослідних даних розв'язується друга задача.

ТАБЛИЦЯ 2. Залежність швидкості різання V (м/хв) від площі поперечного перерізу стружки F(мм^2)

i	F_i	V_i	i	F_i	V_i
1	1,1	25,0	8	7,0	10,0
2	1,4	22,7	9	10,0	8,2
3	1,7	22,1	10	12,8	6,7
4	2,1	19,8	11	16,5	5,6
5	2,6	17,0	12	20,8	5,0
6	4,7	12,3	13	40,6	3,5
7	6,1	10,7			

<u>Приклад 1</u>. Необхідно знайти емпіричну формулу для наближеного подання залежності швидкості різання хромонікелевої сталі від площі поперечного перерізу стружки за дослідними даними з табл. 2 [3, с. 421].

З теорії відомо, що залежність швидкості різання металу від поперечного перерізу стружки подається степеневою функцією

$$V = a \cdot F^{-1/\varepsilon}. \qquad (7)$$

Отже, тип емпіричної формули відомий, залишається тільки знайти найкращі значення її параметрів, наприклад, способом найкращого абсолютного середньоквадратичного наближення. Для цього необхідно в діалоговому вікні задати вхідну інформацію (рис. 1). Якщо дослідні дані вводяться з файлу, то вибір файлу здійснюється у спеціальному вікні, яке відкривається після вибору відповідної радіокнопки.

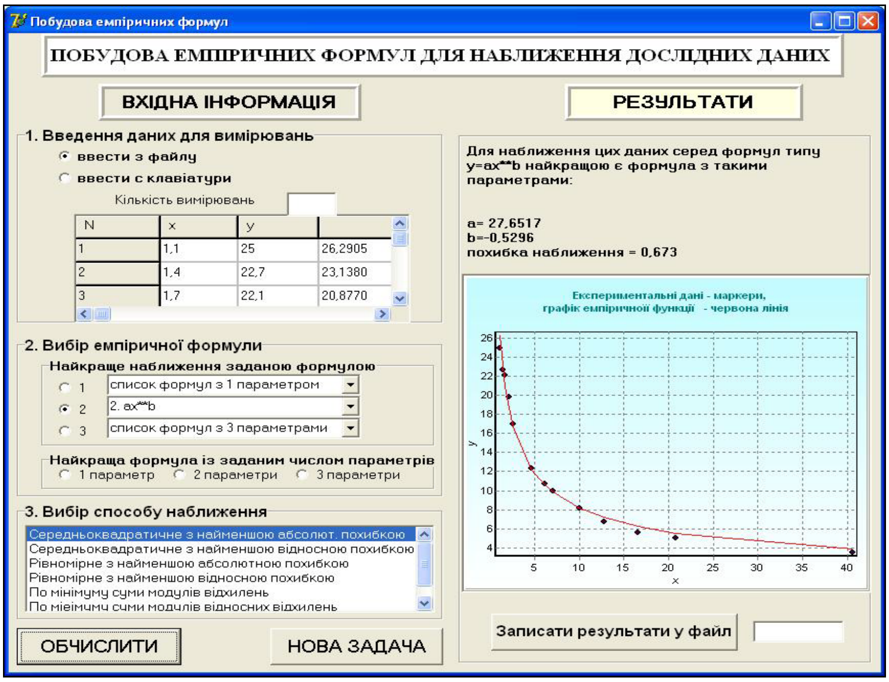

РИС. 1. Вхідна інформація та результати обчислень для прикладу 1

Після натиснення кнопки «Обчислити» на панелі результатів з'являється повідомлення (див. рис. 1), що найкращою є формула з параметрами $a = 27{,}652$, $b = -0{,}530$ (відповідно в формулі (7) $\varepsilon = -1/b$) і середньоквадратичною похибкою 0,673. Отже, шукана емпірична формула має вигляд $V = 27{,}652 \cdot F^{-0{,}530}$. Крім того, третій стовпчик таблиці з вхідними даними заповнюється значеннями швидкості різання, обчисленими за отриманою формулою, та будуються графіки емпіричної (лінія) та експериментальної (точки) залежностей. Варто додати, що емпірична формула вигляду $V = 28,753 \cdot F^{-0,565}$, знайдена за методом найменших квадратів [3], наближає дані з табл. 2 з більшою середньоквадратичною похибкою, яка дорівнює $0,802$.

Приклад 2. Необхідно знайти емпіричну формулу для наближення залежності швидкості різання сталі від площі поперечного перерізу стружки , заданої даними з табл. 2 (як у прикладі 1), за умови, що тип емпіричної формули невідомий.

Отже, необхідно знайти формулу, яка найкраще за вибраним критерієм серед усіх формул з 2 параметрами наближає дослідні дані. На рис. 2 можна побачити установки, які слід зробити у діалоговому вікні програмного комплексу для цього прикладу. В результаті отримаємо повідомлення, що найкращою є формула степеневої залежності $y = a \cdot x^b$ з параметрами $a = 28,617$, $b = -0,564$, тобто емпірична формула вигляду $V = 28,617 \cdot F^{-0,564}$, яка наближає дані з табл. 2 з відносною похибкою 4 %.

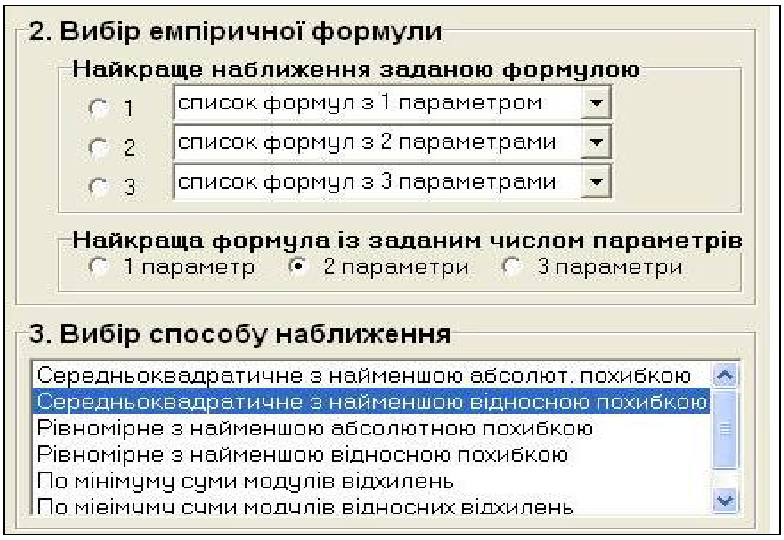

РИС. 2. Фрагмент діалогового вікна з установками для прикладу 2

Висновки. Розроблений програмний комплекс дозволяє розв'язувати важливу задачу наближеного подання дослідних даних за допомогою емпіричних формул. До його переваг слід зарахувати реалізацію декількох способів наближення, програмне визначення найбільш підходящого типу формули, а також графічне представлення результатів, що дозволяє візуально пересвідчитись в якості наближення дослідних даних побудованою емпіричною формулою. Програмний комплекс тестувався на багатьох задачах обробки дослідних даних з різних галузей науки та техніки і показав високу ефективність.

1. *Демидович Б.П., Марон И.А., Шувалова Э.З.* Численные методы анализа. – М.: Наука, 1967. – 368 с.
2. *Веденяпин Г.В.* Общая методика экспериментального исследования опытных данных. – М.: Колос, 1967. – 159 с.
3. *Гутер Р.С., Овчинский Б.В.* Элементы численного анализа и математической обработки результатов опыта. – М.: Наука, 1970. – 432 с.
4. *Батунер Л.П., Позин М.Е.* Математические методы в химической технике. – Л.: Химия, 1968. – 823 с.
5. *Турчак Л.И.* Основы численных методов. – М.: Наука, 1987. – 320 с.
6. *Вакал Л.П.* Про один підхід до автоматизації процесу вибору типу емпіричної формули // Интеллектуальные информационно-аналитические системы и комплексы. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2000. – С. 53–62.
7. *Малачівський П.С., Монцібович Б.Р.* Програмне забезпечення задач рівномірної апроксимації дослідних даних // Математические методы в компьютерных системах. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 1996. – С. 38–43.
8. *Ремез Е.Я.* Основы численных методов чебышевского приближения. – К.: Наук. думка, 1969. – 623 с.
9. *Попов Б.А., Теслер Г.С.* Приближение функций для технических приложений. – Киев: Наук. думка, 1980. – 352 с.
10. *Дьяконов В.* Mathcad 8/2000: специальный справочник. – СПб.: Питер, 2000. – 592 с.
11. http://twt.mpei.ac.ru/ochkov/Mathcad_2001/index.htm#_Toc506643752.
12. *Говорухин В., Цибулин Б.* Компьютер в математическом исследовании: Maple, MATLAB, LaTeX. – СПб.: Питер, 2001. – 624 с.
13. http://documents.wolfram.com.
14. http://wwwasdoc.web.cern.ch/wwwasdoc/shortwrupsdir/e222/top.html.
15. http://www.srcc.msu.su/num_anal/lib_na/cat/cat922.htm.
16. *Каленчук-Порханова А.А., Вакал Л.П.* Пакет программ аппроксимации функций // Комп'ютерні засоби, мережі та системи. – 2008. – № 7. – С. 32–38.
17. *Вакал Л.П.* Построение наилучших нелинейных эмпирических формул // Праці Міжнар. симп. "Питання оптимізації обчислень ПОО-XXXV", Кацивелі, 2009. – Київ: Ин-т кібернетики ім. В.М. Глушкова НАН України, 2009. – С. 62–65.
18. http://www.wasm.ru/article.php?article=approx.

Отримано 17.09.2009