УДК 004:519.216

Л. Я. Нич, Р. М. Камінський
Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

ВИЗНАЧЕННЯ ПОКАЗНИКА ГЕРСТА ЗА ДОПОМОГО ФРАКТАЛЬНОЇ РОЗМІРНОСТІ, ОБЧИСЛЕНОЇ КЛІТИНКОВИМ МЕТОДОМ НА ПРИКЛАДІ КОРОТКИХ ЧАСОВИХ РЯДІВ

Розроблений алгоритм реалізації трьох варіантів клітинкового методу для визначення фрактальної розмірності в табличному процесорі Ms Excel. Для десяти коротких індивідуальних часових рядів отримані значення показників Герста, визначені цим алгоритмом, на підставі рівняння D=2-H. Також наведені значення показника Герста, отримані загальними методами. Результати проведених досліджень підтверджують, що точне значення фрактальної розмірності та показника Герста можна отримати, використовуючи третій варіант цього алгоритму.

Ключові слова: фрактальна розмірність, показник Герста, клітинковий метод, метод нормованого розмаху, фрактальний аналіз.

The three versions of box counting algorithm for determining the fractal dimension of a spreadsheet Ms Excel is implemented. Values of Hearst exponent have been obtained for ten short individual time series, by the designed algorithm, based on the equation D = 2-H. The values of Hearst exponent received by means of general methods are given. The research results prove that the exact value of the fractal dimension and Hearst exponent can be obtained using the third version of the algorithm.

Key words: fractal dimension, Hearst exponent, box counting algorithm, R/S analysis, fractal analysis.

Вступ

Для всебічного розвитку інформаційних технологій обробки та аналізу даних в останні десятиліття щодо обробки та моделювання часових рядів широко використовують методи нелінійної динаміки, зокрема фрактальний аналіз. Основною задачею фрактального аналізу є визначення фрактальної розмірності D та показника Герста H часових рядів, поданих послідовністю значень показника $y(t_i)$ фіксованих в еквідистантні моменти часу t_i, між якими існує простий зв'язок – $D=2-H$. Проте, безпосереднє обчислення цих показників має певні особливості, пов'язані з їхнім визначенням.

Постановка проблеми

Фрактальна розмірність характеризує заповнюваність досліджуваним об'єктом простору свого вкладення, наприклад, для графіка одновимірного часового ряду простором вкладення є поверхня – двовимірна декартова система координат. Вона є одним з центральних понять фрактальної геометрії і характеризує міру структурованості об'єкта, його самоподібність. За [1], множина X називається фрактальною, якщо її розмірність Хаусдорфа $D(X)$ не є цілим числом. Розмірність Хаусдорфа за означенням показує, яку кількість n «куль» діаметром ε потрібно для того, щоб в кожній такій кулі знаходився хоча б один елемент множини X, тобто кількість таких куль відповідає співвідношенню $n(\varepsilon)\approx 1/\varepsilon^D$, коли $\varepsilon\to 0$, звідки значення розмірності обчислюється так:

$$D=\lim_{\varepsilon\to 0}\frac{\log n(\varepsilon)}{\log\left(\frac{1}{\varepsilon}\right)}. \quad (1)$$

Оскільки часовий ряд $y(t)$ подається графіком на площині, то простором його вкладення є поверхня – декартова система координат. Точки, що відповідають рівням ряду $y(t)$, для кращої його візуалізації послідовно з'єднують відрізками. Фрактальну розмірність D графіка такого ряду можна визначити за допомогою *клітинкового* методу, суть якого полягає в тому, що замість куль використовують набір прямокутних сіток з різними розмірами клітинок переважно у формі квадрата. Для кожної сітки підраховують кількість клітинок, що покривають хоча б одну точку ламаної лінії графіка. Використовуючи формулу (1) для кожного значення кількості клітинок за даного їхнього розміру, будують у логарифмічному масштабі графік залежності кількості $n(\varepsilon_i)$ від розміру клітинки ε_i, де $i = 1, 2, \ldots$ відповідає характеру сітки в даному наборі. Далі, для цієї залежності будують графік функції $log\, n(\varepsilon_i) = f\left(log\left(\varepsilon_i\right)\right)$, тобто в логарифмічному масштабі. Відзначені на графіку точки апроксимують прямою лінією методом найменших квадратів. Отримане рівняння прямої має такий вигляд:

$$log\ y = -k\ log\ x + b$$

де x та y є логарифмічними координатами, а додатне значення параметра k відповідає значенню фрактальної розмірності D даного часового ряду $y(t)$.

Використання клітинкового методу для визначення фрактальної розмірності клітинкового методу має певні особливості, суть яких полягає ось у чому. За теорією Хаусдорфа кількість клітинок має бути цілим числом. Безпосередній підрахунок кількості клітинок буде завжди завищеним, оскільки локалізація екстремальних значень $y(t)$ в крайніх клітинках по вертикалі є випадковою. У результаті розмах $y_{max} - y_{min}$ залежно від локалізації значень $y_{\max}$ та $y_{\min}$ може займати в одних випадках n клітинок, а при деякому вертикальному зміщенні сітки вже $n-1$ клітинку. Крім того, знаючи розмір клітинки ε і різницю $y_{max} - y_{min}$, можемо знайти оптимальне значення клітинок $\dfrac{y_{max} - y_{min}}{\varepsilon}$, яке переважно є дробовим числом. З математичного погляду, це не суперечить використанню формули (1), а, оскільки використовується сума таких значень, то похибка не перевищує розміру клітинки. Інакше кажучи, в кожному з цих випадків будемо мати різні значення фрактальної розмірності для одного і того ж часового ряду.

Показник Герста характеризує відношення сили тренду (як фактора детермінованості ряду) до рівня шуму (фактора випадковості). Г. Герст показав, що більшість природних явищ: річкові стоки, добові температури, опади, сонячні плями тощо відповідають зміщеному випадковому блуканню, тобто описуються трендом з накладеним на нього шумом. Крім того, за допомогою показника Герста можна виявляти в емпіричних даних різної природи такі властивості: зміну самого показника в часі, наявність періодичних і неперіодичних циклів, схильність процесу до трендів. Визначають показник Герста логарифмуванням емпіричної формули

$$\frac{R}{S} = \left(\frac{n}{2}\right)^H, \tag{2}$$

де R – розмах кумулятивного ряду різниць $\sum_{i=1}^{n}\left(y(t_i) - y_{середнє}\right)$, S – середньоквадратичне значення n рівнів часового ряду $y(t)$, H – значення показника Герста.

Значення H визначають з рівняння (2), логарифмуючи його.

Існує декілька способів визначення цього показника, проте найпоширенішими серед них є описані [3, 4], проте жоден з цих емпіричних способів не дає фактичного в сенсі об'єктивного значення показника Герста, через різні припущення, умови, попередні обробки та перетворення початкових даних. У результаті дані про один і той самий об'єкт дослідження, які наводять різні дослідники, будуть відрізнятися між собою, а висновки будуть хибними.

Отже, наявна важлива і актуальна *проблема*: забезпечення об'єктивної відповідності між показником Герста H і фрактальною розмірністю D в сенсі наведеної Б. Мандельбродтом рівності

$$D = 2 - H, \tag{3}$$

шляхом вдосконалення або модифікації існуючих способів визначення одного з цих показників.

Формулювання мети

Враховуючи той факт, для забезпечення (3) достатньо визначити один з показників, причому визначення значення D має більш конкретний і простий алгоритм, можна сформулювати ***мету*** дослідження, а саме: «враховуючи властивості фрактальної розмірності вдосконалити клітинковий метод її визначення та експериментально порівняти значення показників Герста, отриманих як за прийнятими алгоритмами, так і за варіантами, власне клітинкового методу».

Аналіз останніх досліджень та публікацій

Відомо доволі багато публікацій стосовно застосування клітинкового методу для обчислення фрактальної розмірності та визначення показника Герста. Суть фрактального аналізу часових рядів часто розуміють по-різному. Фрактальний аналіз в одних випадках зводиться до визначення фрактальної розмірності, в інших – до визначення показника Герста або до застосування методу нормованого розмаху. Одні надають перевагу знаходженню фрактальної розмірності, другі – визначенню показника Герста, треті – процедурі нормованого розмаху. У [5] обговорюються методи фрактального аналізу: фрактальної розмірності клітинковим методом, кореляційному інтегралу, нормованому розмаху та їх реалізації на сигнальних процесорах. У [6] розкриваються недоліки клітинкового методу визначення фрактальної розмірності: частина клітинок залишається неповністю заповненою графіком, а також складність підрахунку клітинок, коли їх розмір $\varepsilon \to 0$. У [7] автор подає два методи розрахунку фрактальної розмірності часового ряду, а саме: R/S аналізу та сегментно-варіаційного методу. Розмірність мінімального покриття в [8] подається як нова фрактальна характеристика, виражена індексом фрактальності. Тут як і в попередніх двох роботах розглядається поверхня, яка покриває графік часового ряду, розділений суміжними вертикальними прямокутниками однакової ширини з висотами, що дорівнюють різниці між максимальним і мінімальним рівнями, локалізованими в даному прямокутнику. Автори доводять, що при зменшенні ширини значення покриття збігається зі значенням фрактальної розмірності. У [9] запропонований модифікований алгоритм обчислення оцінки фрактальної розмірності. Його перевагами є нижчі обчислювальні вимоги порівняно з класичним алгоритмом клітинкового методу, який базується на цілочисельній арифметиці, що дозволяє його реалізацію сигнальними процесорами з цілими числами або програмованими логічними схемами, проте він має доволі повільне зближення з точним значенням розмірності Хаусдорфа. У [10] на основі аналізу отриманих результатів – порівняння значення фрактальної розмірності броунівського руху зроблено висновок про те, що найбільш точним серед розглянутих методів є метод, який використовує показник Герста. Показник Герста, як вказується в [11], є показником самоподібності часових рядів. У додаток до стандартних методів визначення показника Герста, основаних на формулі Герста, розглянуті ще два методи, що використовують вейвлети, один з яких показав найкращий збіг зі значенням цього показника. Оцінено показник Герста в [12] за методом нормованого розмаху з метою вивчення розподілу його властивостей на підставі генерування часових рядів з нормальним розподілом рівнів. Показано, що в цьому випадку розподіл оцінок показника Герста також є нормальним. У [13] показник Герста оцінюється за допомогою R/S аналізу і, як міра стійкості статистичного явища, використовується для визначення таких фрактальних характеристик: фрактальної розмірності, кореляційного параметра, спектрального та фрактального показників. У [14] на прикладі часового ряду цін досліджена залежність показника Герста цього ряду від довжини вибірки та часу. Запропоновано декілька варіантів узагальнення цього показника на нееквідистантні часові ряди. Побудовані розподіли цього показника для вибірок різної довжини і показано, що вони є стаціонарними в межах точності, з якою визначені відповідні емпіричні ймовірності і наближаються до нормальних розподілів. У [15] на прикладі часових рядів екологічного моніторингу визначається показник Герста і вказується на те, що він нетривіально залежить від довжини часового ряду, а також від властивостей множини моментів часу, на яку відображається досліджуваний часовий ряд. Крім того, пояснюється чому значення показника Герста природних процесів групуються навколо значень 0.72–0.73. У [16] наведено алгоритм розрахунку показника Герста методом нормованого розмаху, а також дані рекомендації щодо вибору кроку між значеннями рівнів досліджуваного часового ряду та його довжини. Порівняння методів визначення фрактальної розмірності наведено в [17] і зроблені такі висновки: зіставлення значень фрактальної розмірності та показника Герста узгоджуються одне з одним, але відрізняються точністю, крім того, вказується, що клітинковий метод дає точніше значення фрактальної розмірності часового ряду, ніж її визначення через показник Герста.

Аналіз отриманих наукових результатів

Організація дослідження полягає у теоретичному обґрунтуванні використаних методів та проведенні розрахунків запропонованими та відомими алгоритмами. Суть дослідження полягає в тому, щоб порівняти значення фрактальної розмірності і показника Герста, визначені наведеними алгоритмами в сенсі рівності (3). Річ у тому, що значення показника Герста визначають за співвідношенням (2), ігноруючи той факт, що в даному виразі є дві невідомих: сам показник Герста H і коефіцієнт пропорційності α, який у цій формулі дорівнює 2, тобто $\alpha = 2$. Його значення вибрано емпірично, а тому значення H, визначені різними способами, будуть відрізнятися. У цьому сенсі значення розмірності D, визначене клітинковим методом, практично не містить емпірики і має строго визначену послідовність операцій. Єдиним зауваженням є подання кількості клітинок цілим числом. Проте формула (1) не вимагає подання $n(\varepsilon)$ лише цілими числами, а тому в цій роботі розглядаються результати, отримані трьома її варіантами. Для визначення значення показника Герста використані відомі широко використовувані методики, наведені в [3, 4].

Варіанти клітинкової розмірності. Зміст цих варіантів описаний вище, нагадаємо лише, що: перший варіант дає ціле, але завищене значення кількості клітинок, другий дає мінімальне ціле значення кількості клітинок, третій – оптимальне значення кількості, але переважно виражене дробовим числом. Наведемо алгоритми для їх реалізації засобами табличного процесора Ms Excel. Вибір саме Ms Excel обґрунтований тим, що:

по-перше, безпосередня і наочна робота з даними «номер – значення»;

по-друге, великий вибір математичних та статистичних формул;

по-третє, можливість складання спеціальних формул та швидка побудова графіків і діаграм.

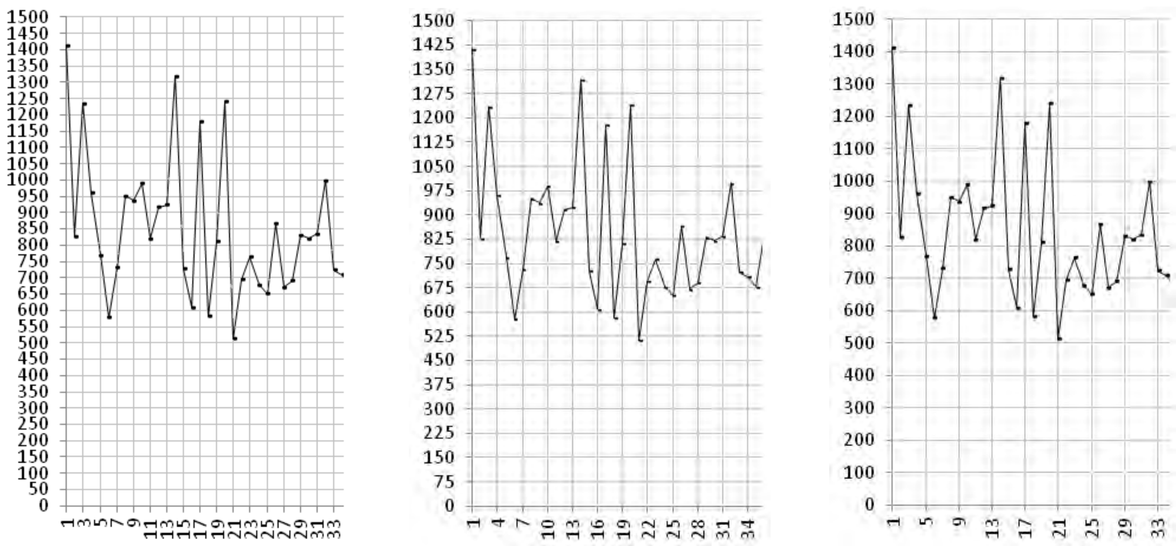
по-четверте, він доступний практично в будь-якій комп’ютерній системі з офісним пакетом.

Природа експериментальних даних. Використані в дослідженні дані є індивідуальними часовими рядами, отримані в експериментальних дослідженнях і представляють результат індивідуальної інтелектуальної операторської діяльності, яка пов’язана з опрацюванням візуальної інформації, наданої на моніторі у формі послідовності зображень. Зображення мають однорідне тло з випадково локалізованими малорозмірними, в сенсі [2], зображеннями об’єктів заданого класу, а в якості операторів задіяні молоді люди із числа студентів. Суть опрацювання полягає у виявленні на кожному зображенні об’єкт уваги та прийняти рішення про належність цього об’єкта заданому класу. Час з моменту експозиції зображення до моменту прийняття рішення вважається часом опрацювання. Послідовність значень цього часу послідовності опрацьованих зображень і подається значеннями рівнів індивідуального часового ряду, тобто експериментальними даними є часові ряди часу опрацювання кожного з зображень. Усім операторам надається одна і та сама, і з тими самими параметрами подання, послідовність зображень. Особливістю цих рядів є те, що вони, як показала попередня обробка, мають значну дисперсію, і асиметричний, одномодальний та зрізаний в області малих значень розподіл Вейбула. Вже перший погляд на їхнє графічне відображення дає підстави вважати їх випадковими, і в принципі стаціонарними, процесами. Проте, переважно для всіх рядів спостерігається на початку ряду існування деякого тренду, можливо пов’язаного з періодом адаптації, який практично далі зникає. Другим моментом є те, що значення рівнів, за винятком одного або кількох, які різко виділяються за величиною, але не є викидами, хаотично локалізовані в межах деякого коридору в межах стандартного відхилення. Така ситуація, асоціюється з самоподібністю часового ряду і в певному сенсі стимулює застосування до цих рядів фрактального аналізу. Фрактальний аналіз належить до нових сучасних методів обробки часових рядів і повною мірою є орієнтований на використання комп'ютерних технологій. Особливістю цих методів є те, що з їхньою допомогою виявляються закономірності в динаміці показників за наявності випадковостей.

Реалізація клітинкового алгоритму для варіанта 1. Безпосередній (вручну) підрахунок клітинок є вельми рутинною і монотонною процедурою, не захищеною від помилок визначення кінців відрізків часового ряду в клітинках. Суть алгоритму клітинкової розмірності полягає в тому, що на графік часового ряду накладають набір сіток з клітинками розміром $\varepsilon_x \times \varepsilon_y$, причому візуально варто забезпечити для них квадратну форму, тобто $\varepsilon_x \cong \varepsilon_y$, як зображено на рис. 1.

У цьому сенсі, необхідно забезпечити вимогу: розмір клітинок ε не повинен бути меншим за найменшу відстань між двома сусідніми рівнями часового ряду, а тому шкала значень номерів вертикальних ліній сітки на осі абсцис для включення двох значень рівнів має позначки 1, 3, 5, …., для трьох значень – 1, 4, 7, … , для чотирьох відповідно, 1, 5, 9 , … і т.д.

Розміри клітинок підбираємо так, щоб вони виглядали квадратами і вкладалися між поділками шкали ординат і, таким чином приводимо розмір клітинки до одиниць значень рівнів часового ряду. На рис. 2 зображені два фрагменти часового ряду, для яких обчислення кількості клітинок можна подати в такий спосіб.



а *б* *в*

Рис. 1. Набір сіток, накладених на графічне зображення часового ряду. Розмір клітинок такий: а – 50 ms і об'єднує два рівні; б – 75 ms і об'єднує три рівні; в – 100 ms і об'єднує 4 рівні

Підраховують кількість клітинок так. Оскільки розмір клітинок ε по осі ординат є відомий і відповідає відстані між двома сусідніми номерами по осі абсцис, а шкала ординат починається з нуля, то значення кількості цілих клітинок на інтервалі абсциси, що дорівнює також ε, дорівнюватиме значенню $k_{max} = \frac{y_{max}}{\varepsilon}$, заокругленому до найближчого більшого цілого. Натомість, кількість цілих клітинок, що знаходяться нижче від клітинки зі значенням y_{min}, визначається значенням $k_{min} = \frac{y_{min}}{\varepsilon}$, заокругленим до найближчого найменшого цілого. Наприклад, якщо $k = 7.23$ чи $k = 7.68$, то для k_{max} кількість клітинок як ціле число має бути $k = 8$, а для k_{min} відповідно $k = 7$. Отже, кількість клітинок у даному стовпці дорівнює різниці заокруглених значень $k_{max} - k_{min}$. У такий спосіб загальна кількість клітинок $n(\varepsilon_i)$ кожної сітки, в яких локалізоване графічне зображення часового ряду

$$n(\varepsilon_i) = \sum_{i=1}^{n/\varepsilon_i} \left(k_{i_{max}} - k_{i_{min}}\right). \tag{4}$$

Якщо в такому стовпці більше ніж два значення, то розглядається лише їх розмах, тобто максимальне і мінімальне значення. Така процедура не порушує принципу або змісту клітинкової розмірності, оскільки не порушує регулярність сітки вздовж абсцис та не змінює розміру самих клітинок, лише уточнює їхню кількість.

Недоліком цього варіанта є те, що отримана кількість є завищеною, оскільки враховуються клітинки, в яких знаходиться навіть одна точка часового ряду, тобто клітинки з екстремальними значеннями можуть бути мало заповненими, а враховуються повністю.

Реалізація клітинкового алгоритму за варіантом 2. Цей недолік легко усунути, якщо враховувати заповнення клітинок з екстремальними значеннями в такий спосіб. Нехай, ε – розмір клітинки, а y_{max} і y_{min} – екстремальні значення в інтервалі $(\varepsilon_i, \varepsilon_{i+1})$, тоді заокруглене до найближчого більшого значення частки $k = \dfrac{y_{max} - y_{min}}{\varepsilon}$ дає значення кількості заповнених клітинок у цьому стовпці. Загальна сума клітинок для кожної сітки (ε_i) визначається як

$$n(\varepsilon_i) = \sum_{i=1}^{n/\varepsilon_i} \left(\frac{y_{i_{max}} - y_{i_{min}}}{\varepsilon_i}\right). \tag{5}$$

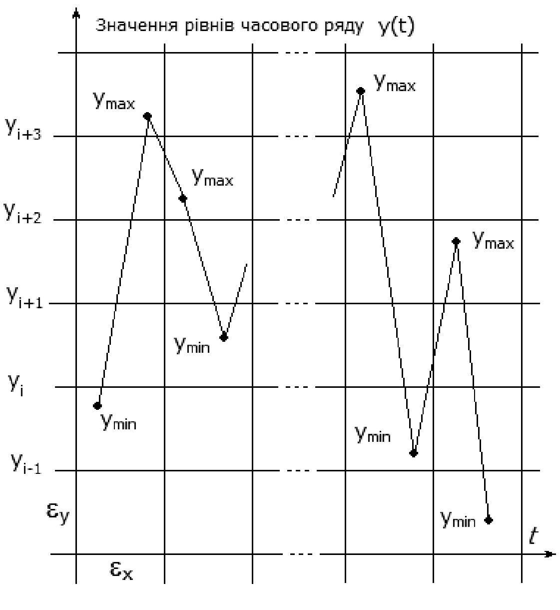


Рис. 2. Фрагменти положення екстремальних значень рівнів

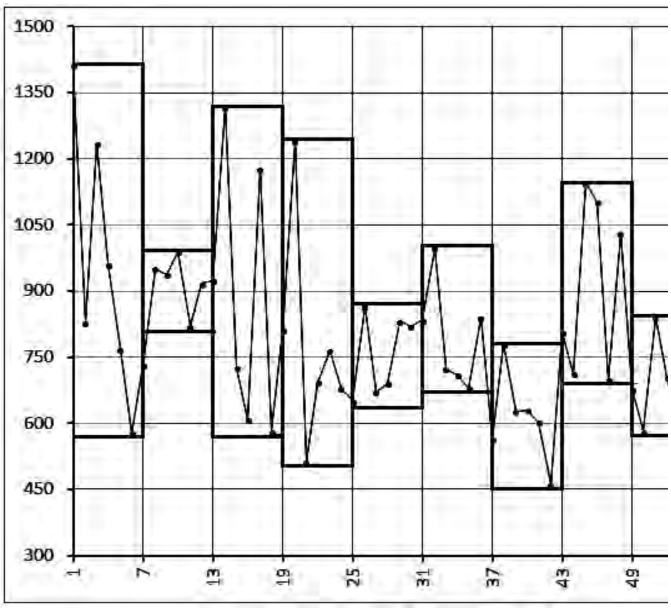


Рис. 3. Ілюстрація методу найменшого накриття

Кількість клітинок у цьому варіанті подається цілим числом, проте не всі клітинки є повністю заповнені, тобто на кожен стовпець з основою ε_i залишається, майже завжди, одна незаповнена повністю клітинка. Очевидно, сума таких незаповнених клітинок дає деяке ціле число клітинок плюс одну неповну, яку в результаті заокруглення додаємо до загальної суми як одиницю. Тому, якщо йдеться саме про ціле число клітинок, то цей варіант дає їх точну кількість.

Реалізація клітинкового алгоритму для варіанта 3. Якщо значення k_i кількості клітинок вертикального стовпця не заокруглювати до найближчого цілого і прийняти

$$k_i = \frac{y_{i_{max}} - y_{i_{min}}}{\varepsilon_i} \tag{6}$$

як дробове число, то отримаємо кількість клітинок у дійсних числах, тобто має місце співвідношення (6), але без заокруглення. Інакше кажучи, для *i*-го стовпця матимемо ціле число повністю заповнених клітинок і одну неповну, що відповідає дробовій частині частки (6). Оскільки для формули (1) немає значення чи кількість клітинок $n(\varepsilon_i)$ подана цілим числом чи дробовим і це не змінює її суті, а тому загальна кількість заповнених клітинок також буде дійсним числом і відповідатиме мінімальному покриттю даного часового ряду в «одиницях» вимірювання – клітинках.

Наведені варіанти клітинкового методу обчислення фрактальної розмірності не заперечують змісту топологічної розмірності Хаусдорфа.

Визначення показника Герста за алгоритмом [3]. Нехай $y(t) = y_1(t), y_2(t), \dots, y_n(t)$, де n – кількість рівнів, а $t,\ t \in T$ – множина моментів часу, в які заміряні значення деякого показника, що характеризує динаміку явища – опрацювання людиною-оператором послідовності зображень, наданих на моніторі в даному досліджені. Вважаємо, що заміряють через однакові проміжки часу, а сама їх послідовність утворює еквідистантний часовий ряд.

Першим кроком є визначення середнього значення $\bar{y}_n$ і середньоквадратичного відхилення $\sigma_y = S$. Далі, для кожного рівня обчислюють його відхилення від середнього $y_i(t) - \bar{y}_n$ і утворюють з послідовності відхилень, шляхом послідовного накопичення суми, кумулятивний ряд за формулою

$$Z = \sum_{i=1}^{n} \left(y_i(t) - \bar{y}_n \right). \tag{7}$$

Для кумулятивного ряду (7) визначають розмах R за такою формулою:

$$R = max\, Z - min\, Z\ . \tag{8}$$

Використовуючи емпіричну формулу $\frac{R}{S} = \left(\frac{n}{2}\right)^H$, визначають показник Герста логарифмуванням цього виразу

$$H = \frac{log(R/S)}{log(n/2)}\ . \tag{9}$$

Зауважимо, що показник Герста для значної більшості природних процесів є більш або менш симетрично розподілений навколо середнього значення 0.73 зі стандартним відхиленням приблизно 0.09.

Визначення показника Герста за алгоритмом [4]. Цей метод відрізняється від попереднього тим, що для усунення тренду здійснюють таке перетворення: беруть логарифм відношення наступного значення до поточного, тобто утворюють новий ряд $Z = (z_1, z_2, \dots, z_{n-1})$ з рівнями $z_i = log\left(y_{i+1}(t) / y_i(t)\right)$, де $i = 1, 2, \dots, (n-1)$.

Першим етапом у цьому варіанті є розбиття вихідного ряду на k суміжних інтервали довжиною $m_1, m_2, \dots, m_k$ так, щоб за величиною $m \times k = Z$, причому вказується на вимогу $m_{k+1} = m_k + \Delta$, де Δ – величина такого інтервалу і, крім того, має бути забезпечена ще й така вимога: максимальне значення довжини інтервалу $m_k \le \left[\frac{n}{2}\right]$, тобто максимальна величина інтервалу m_k має забезпечити поділ часового ряду Z довжиною $n-1$ на два інтервали однакової довжини (з однаковою кількістю рівнів).

У [4] наведене емпіричне правило, за яким: довжина інтервалу є $m_k \geq 10$, а кількість інтервалів $k \geq 5$. Крім того, пропонується обирати лише такі значення m_k, які є дільниками числа $n-1$, тобто кількість рівнів у інтервалі m_k має бути кратна ($n-1$).

Очевидно, для коротких часових рядів обсягом 100 – 300 рівнів це правило можна забезпечити, але забезпечити кратність для величин інтервалів практично є неможливим.

На другому етапі для кожного інтервалу визначають середнє значення $\bar{z}_k$ та стандартне відхилення S_k. Третім етапом є те, що для кожного інтервалу $z^k(m_k)$ створюється ряд z_k^* накопичених відхилень як

$$z_k^* = \sum_{j=1}^{m_k} \left(z_j^k - \bar{z}_k \right),$$

де $j = 1, 2, \ldots, m_k$.

Інакше кажучи, від кожного рівня z_j^k в інтервалі $z^k(m_k)$ віднімається середнє значення $\bar{z}_k$, і це відхилення додається до суми попередніх відхилень, тобто послідовність таких сум є елементами кумулятивного ряду z_k^* для даного інтервалу $z^k(m_k)$.

На четвертому етапі визначають розмах елементів кумулятивного ряду z_k^* за формулою

$$R_k = max\, z_{jk}^* - min\, z_{jk}^*, \qquad \forall j = \overline{1, m_k}.$$

На п'ятому етапі нормують значення розмаху діленням R_k / S_k і обчислюють середнє зі всіх k нормованих розмахів

$$\frac{R}{S} = \frac{1}{k} \sum_{t=1}^{k} \left(\frac{R_k}{S_k} \right).$$

У шостому етапі для кожного k обчислюється логарифм довжини інтервалу m_k, який відповідає абсцисі $log(m_k)$, і логарифм нормованого розмаху $log(R/S)$, який є ординатою точки на площині для побудови графіка регресії $log(R/S) = f[log(m_k)]$ в подвійних логарифмічних координатах.

Потім початковий ряд $Z = \{z_i\}$, $i = 1, 2, \ldots, (n-1)$ знову розбивається на декілька інтервалів, але з іншою кількістю даних m і процедуру повторюють.

Стосовно цього алгоритму можна вказати на його емпіричність, оскільки точного математичного обґрунтування вибору кількості підінтервалів сьогодні не виявлено. Не зрозуміло також скільки таких ітерацій має бути. Тому в цій роботі показник Герста визначений для однієї ітерації, з забезпеченням вимоги – мінімальна величина підінтервалу m не менша, ніж 10.

Опис алгоритму в Excel. Обчислення фрактальної розмірності проілюстровано на такому прикладі. Часовий ряд $y = f(x)$ є еквідистантний, де y – час опрацювання людиною-оператором зображення з об'єктом уваги, наданого на моніторі, вимірюваний у мілісекундах, а x – значення моментів реального часу, в які людина-оператор приймає рішення про те, що виявлений, на зображенні наданому на моніторі, об'єкт відповідає заданому класу. Оскільки фактичні значення x не використовуються, їх замінено простою зростаючою нумерацією. Значення y змінюються в діапазоні 431 – 1413 мілісекунд, є випадковими величинами з асиметричним, одномодальним розподілом, зрізаним в області малих значень.

Для обчислення фрактальної розмірності в середовищі Excel вводимо дані.

1. Відкриваємо нову Книгу і з файла вводимо дані часового ряду: номер значення і саме значення, починаючи з стовпців відповідно **A4** і **B4**. Перших три рядки для зручності залишаємо для заголовків.

2. Вибираємо параметри сітки. Будуємо графік даного часового ряду разом з основними лініями сітки по обох координатах. Мінімальне значення розміру клітинки вздовж абсцис $\varepsilon = 2$, тобто шкала основних ліній сітки вздовж абсциси є такою: 1, 3, 5, … . Вибираємо шкалу для основних ліній сітки вздовж ординати так, щоб клітинки мали форму квадрата. Для прикладу на рис. 1 наведені графіки з кроками шкали ординат $\varepsilon = 50$, $\varepsilon = 75$ і $\varepsilon = 100$ тобто $y_{i+1} - y_i = 50$, 75 та 100. Значення кроку вздовж абсцис варто вибирати за кількістю рівнів, що охоплюються клітинками сіток, причому необхідно забезпечити вздовж абсцис збіжність відстаней між рівнями в клітинці і розмірами клітинки.

Очевидно, що для досягнення клітинками форми квадрата можна незначно маніпулювати, розтягуючи чи стискаючи в обох напрямках сам графік. Важливим є те, щоб була збережена відповідність між розмірами клітинки та кроками шкали, що своєю чергою забезпечить візуальну перевірку обчислення фрактальної розмірності.

3. Вводимо в клітинку **D4** таку формулу для першого варіанта:

=ROUNDUP(MAX(**B4**:**B5**)/50;0)-INT((MIN(**B4**:**B5**)/50), (9)

де перший доданок визначає кількість повних клітинок, що охоплює максимальний рівень y_{max}, а другий – кількість клітинок, що лежать нижче від клітинки, в якій знаходиться значення y_{min}.

Для другого варіанта формула матиме такий вигляд:

=INT((MAX(**B4:B5**)-MIN(**B4:B5**))/50), (10)

а для третього, відповідно

=(MAX(**B4:B5**)-MIN(**B4:B5**))/50. (11)

4. Після введення формули натискаємо **Enter** і далі, шляхом ***автозаповнення*** заповнюємо значеннями стовпець **D** до кінця часового ряду. У цьому стовпці кожне друге значення є невідповідним, оскільки стосується значень, розділених лінією сітки (нас цікавить лише результат, обчислений для значень ряду, що лежать між вертикальними лініями сітки). Тому очищаємо кожну другу клітинку як показано на рис. 4.

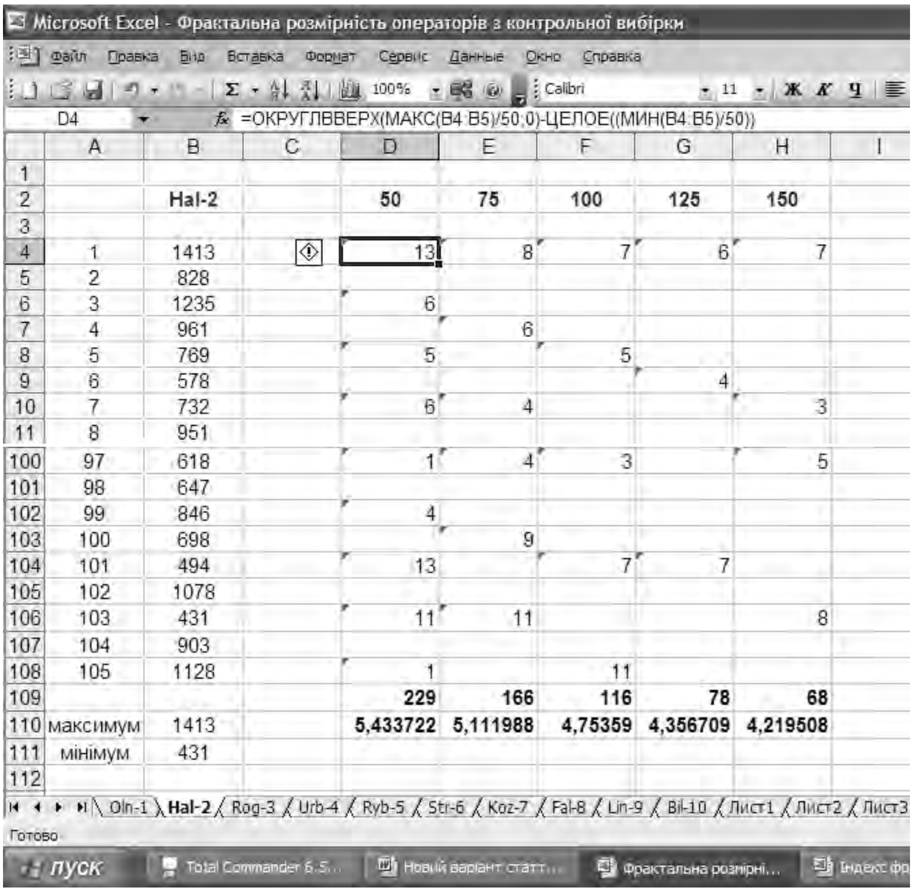
Інакше кажучи, в результаті автозаповнення ми отримуємо усі можливі комбінації обчислення значень: (**B4**:**B5**), (**B5**:**B6**), (**B6**:**B7**) і т.д., а нам потрібні лише результати для таких інтервалів: (**B4**:**B5**), (**B6**:**B7**), (**B8**:**B9**) тощо, бо ці значення знаходяться між двома сусідніми лініями шкали осі абсцис. Для значень, що знаходяться в інтервалах (**B5**:**B6**), (**B7**:**B8**) і т.д., одне значення від другого буде відокремлене лінією осі абсцис.

Далі підсумовуємо значення стовпця **D**.

На рис. 4 у рядку № 154 показані результати підрахунку кількості клітинок, а в рядку № 155 наведені логарифми цих величин. Цифри 50, 75, 100, 125 і 150 визначають розміри сіток за ординатами.

Для стовпця **D** таблиці кожна друга клітинка є порожньою, оскільки формула охоплює лише два рівні, для яких значення кількості клітинок становить 7, і саме в клітинках **D4**, **D6**, **D8** і т.д. ця формула вказує їх правдиву кількість. Для стовпця **E** формула враховує три значення – **B4**:**B6**, з яких лише перше є правильним, а тому очищаються клітинки **E5** і **E6**, тобто залишається лише перше значення з інтервалу, вказаного у формулі, а решта стирається, оскільки в цьому випадку легко знайти суму клітинок сітки, зображену на рис. 4 у клітинці **D154**.

5. Цей крок аналогічний крокам 2 – 4, але за інших параметрів. Копіюємо графік (рис. 1, *а*) і змінюємо значення шкал координат: для абсциси $\varepsilon_x = 3$ і для ординати $\varepsilon_y = 75$, в результаті графік матиме вигляд як на рис. 1, *б*.



	A	B	C	D	E	F	G	H	I
1									
2		Hal-2		50	75	100	125	150	
3									
4	1	1413		13	8	7	6	7	
5	2	828							
6	3	1235		6					
7	4	961			6				
8	5	769		5		5			
9	6	578					4		
10	7	732		6	4			3	
11	8	951							
100	97	618		1	4	3		5	
101	98	647							
102	99	846		4					
103	100	698			9				
104	101	494		13		7	7		
105	102	1078							
106	103	431		11	11			8	
107	104	903							
108	105	1128		1		11			
109				229	166	116	78	68	
110	максимум	1413		5,433722	5,111988	4,75359	4,356709	4,219508	
111	мінімум	431							
112									

Рис. 4. Зображення робочого аркуша книги Ексель з даними для обчислення фрактальної розмірності за першим варіантом

6. Копіюємо формулу (9) з клітинки D4 в клітинку E4, активізуємо її і змінюємо інтервал **B4**:**B5** на **B4**:**B6**, параметр 50 на 75, тобто:

=ROUNDUP(MAX(**B4**:**B6**)/75;0)-INT((MIN(**B4**:**B6**)/75),

аналогічно і для формул (10) і (11) відповідно

=INT((МАКС(**B4:B6**)-MIN(**B4:B6**))/75),

=(MAX(**B4:B6**)-MIN(**B4:B6**))/75.

Потім натискаємо **Enter** і, далі, за допомогою автозаповнення заповнюємо стовпець **E** до кінця ряду. Оскільки між вертикальними лініями осей має бути три значення ряду (**B4**:**B6**), тому кожне друге і третє значення треба відкинути, бо в них входять значення, розділені вертикальною лінією осі, тобто істинними результатами є кожне перше з трьох значень, що й зображено на рис. 4 у стовпці E.

7. Визначаємо суму клітинок за даними стовпця **E**.

8. Аналогічно здійснюємо кроки 2 – 4 для інтервалів значень (**B4**:**B7**), (**B4**:**B8**), (**B4**:**B9**), змінюючи відповідно значення шкал графіків $\varepsilon_x = 4, 5, 6$ і $\varepsilon_y = 100, 125, 150$ та редагуючи відповідно формулу (9).

9. Після введення формули натискаємо **Enter** і за допомогою ***автозаповнення*** заповнюємо значеннями відповідний наступний стовпець до кінця часового ряду і очищаємо для кожного інтервалу значень всі $\varepsilon - 1$ клітинок крім першої.

10. Формуємо за отриманими даними табл. 1 результатів, в яку записуємо такі дані: ε, $N(\varepsilon)$, $log(\varepsilon)$, $log\,N(\varepsilon)$. В якості розміру клітинки ε можна вводити як 2, 3, … , 6, так і 50, 75, … , 150, оскільки кожна з цих послідовностей є лінійною, а самі значення – незалежними аргументами.

Таблиця 1

Допоміжні результати визначення фрактальної розмірності

№ з/п	Розмір клітинки ε	Кількість клітинок $N(\varepsilon)$	Розмір клітинки $log(\varepsilon)$	Кількість клітинок $\log N(\varepsilon)$
1	2	229	0,693	5,434
2	3	166	1,098	5,112
3	4	116	1,386	4,753
4	5	78	1,609	4,357
5	6	68	1,792	4,219

11. Будується графік, зображений на рис. 5, у подвійних логарифмічних координатах, значення яких наведені в четвертому і п'ятому стовпцях табл. 1.

Апроксимація прямою лінією послідовності точок методом найменших квадратів подає таке рівняння для апроксимованої лінії:

$$\log(y) = -1.1861 \cdot log(x) + 6{,}3118, \tag{12}$$

причому висока якість апроксимації підтверджується близьким до одиниці коефіцієнтом детермінації $R^2 = 0{,}9852$.

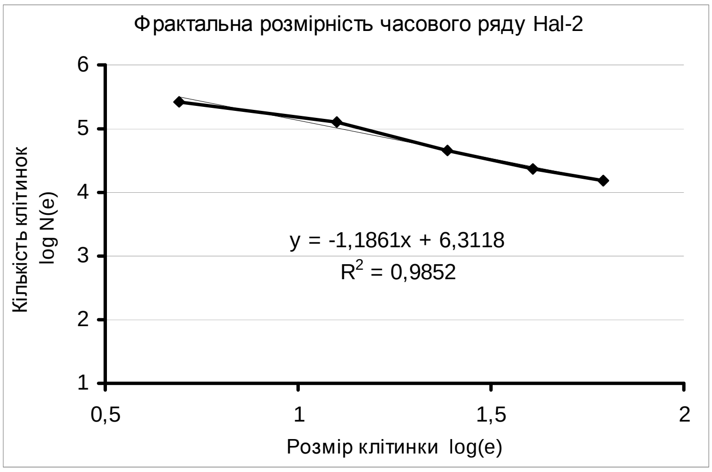


Рис. 5. Регресія кількості клітинок у логарифмічній системі координат

Таблиця 2

Зведена таблиця значень показника Герста, отриманих різними методами

Оператори	Показник Герста за загальним алгоритмом Хаусдорфа	Показник Герста за точною кількістю клітинок	Показник Герста за мінімумом покриття в клітинках	Показник Герста за методикою, наведеною в [3]	Показник Герста за методикою, наведеною в [4]
1	2	3	4	5	6
1	0.859	0.801	0.781	0,677	0,706
2	0.836	0.738	0.701	0,746	0,675
3	0.895	0.815	0.789	0,739	0,680
4	0.799	0.756	0.732	0,698	0,679
5	0.764	0.725	0.670	0,575	0,667
6	0.875	0.840	0.796	0,721	0,669
7	0.841	0.807	0.729	0,637	0,716
8	0.831	0.831	0.759	0,663	0,670
9	0.843	0.796	0.783	0,493	0,685
10	0.767	0.761	0.724	0,672	0,636

Додатне значення кутового коефіцієнта рівняння (12) відповідає значенню фрактальної розмірності D, тобто фрактальна розмірність даного часового ряду є $D = 1{,}1861$.

Результатом дослідження є таблиця значень показника Герста, отриманих п'ятьма методами: трьома варіантами клітинкового методу визначення фрактальної розмірності, використовуючи рівняння (3) та методами, наведеним у [3], і дещо модифікованому авторами методі, наведеному в [4]. Модифікація здійснена тому, що у відповідній літературі зустрічаються багато емпіричних варіантів щодо поділу часового ряду на інтервали та знаходження самого значення показника. Підсумовуючи, варто відзначити, що характеристики нелінійних динамічних властивостей часових рядів, отримані фрактальним аналізом, можуть бути лише першим кроком щодо розуміння поведінки досліджуваної системи. Тому лише застосування комплексного системного підходу може допомогти під час прогнозування, прийняття стратегічних рішень і планування довготривалої політики.

Висновки та перспективи подальших наукових розвідок

Проведені дослідження визначення фрактальної розмірності клітинковим методом показали, що використання з цією метою табличного процесора Ms Excel не лише забезпечує її визначення, але завдяки відмові від цілочисельного подання кількості клітинок, що покривають цей часовий ряд, дає її точне значення. Алгоритм є доволі простий у користуванні і практично може бути застосований для часових рядів з довільною кількістю рівнів.

Визначення показника Герста здійснюється за рахунок рівняння $D = 2 - H$, тобто показник Герста визначається за допомогою фрактальної розмірності, а не навпаки, як рекомендується в багатьох публікаціях. Наведені в четвертому стовпці значення показника Герста добре узгоджуються з зауваженням самого Герста, наведеному в [3, с. 154], а саме: значення показника H більш або менш симетрично розподілені навколо середнього значення 0.73 з стандартним відхиленням, що дорівнює приблизно 0.09.

1. Мандельброт Б. Фрактальнная геометрия природы / Мандельброт Б.– М., 2002.– 656 с. 2. Камінський Р. М. Властивості та характеристики малорозмірних об'єктів в тестових зображеннях систем розпізнавання та обробки зорових образів / Р. М. Камінський // Вісник Нац. ун-ту «Львівська політехніка». – 2003. – № 496: Комп'ютерна інженерія та інформаційні технології. – С. 153–162. 3. Федер Е. Фракталы / Е. Федер. – М.: Мир, 1991. – 254 с. 4. Петерс Э. Фрактальный анализ финансовых рынков: Применение теории хаоса в инвестициях и экономике / Петерс Э. – М.: Интернет-Трейдинг, 2004. – 304 с. 5. Борисов В. Д. Метод фрактального анализа временных рядов / Борисов В. Д., Садовой Г. С. // Автометрия. – 2000. – № 6. – С. 10–19. 6. Zwolankowska M. Metoda segmentowo-wariacyjna. Nowa propozycja liczenia wymiaru fraktalnego / Zwolankowska M. // Przegląd Statystyczny. – 2000. – R. 47, z. 1–2. – S. 209–224. 7. orzeszko W. Wymiar fraktalny szeregów czasowych a ryzyko inwestowania / orzeszko W. // Acta Universitatis Nicolai Copernici. Ser. Nauki Humanistyczno-Społeczne. Ekonomia. – 2010. – Z. 397. – S. 57–70. 8. Дубовиков М. М., Крянев А. В., Старченко Н. В. Размерность минимального покрытия и локальный анализ фрактальных временных рядов // Вестник РУДН , Серия Прикладная и компьютерная математика. – 2004. – Т. 3. – № 1. – С. 30–44. 9. Panek D. On fractal dimension estimation / Panek D., Kropik P., Predota A. // Przeglad Elektrotechnizny (Electrical Review). – 2011. – R. 87. – № 5. – P. 120–122. 10. Анализ методов определения фрактальной размерности / Курдюков В. И., Остапчук А. К., Овсянников В. Е., Рогов Е. Ю. // Вестник Кузбасского государственного технического университета. – 2008. – Вып. № 5. – С. 46–49. 11. Racine R. Estimating the Hurst exponent / Roman Racine // MOSAIC Group: Bachelor thesis, April 14, 2011 y., Zürich. – Zürich, 2011. – P. 1–30. 12. Malhar K. Fractal Analysis of Time Series and Distribution Properties of Hurst Exponent / Malhar K., Ferry B.B. // Journal of Mathematical Sciences & Mathematics Education. – 2011. – Vol. 5, No. 1. – P. 8–19. – Режим доступу: http://www.msme.us/2011-1-2.pdf. 13. Бельков Д.В. Статистический анализ сетевого трафика / Бельков Д.В., Едемская Е.Н., Незамова Л.В. // Наукові праці ДонНТУ. Серія Інформатика, кібернетика та обчислювальна техніка. – 2011. – Вип. 13(185). – С. 66–75. 14. Распределения показателя Херста нестационарного маркированного временного ряда / Кириллов Д. С., Короб О. В., Митин Н. А., Орлов Ю. Н., Плешаков Р. В. // Препринт ИПМ им. М. В. Келдыша. – 2013. – № 11. – 16 с. – Режим доступу: http://library.keldysh.ru/preprint.asp?id=2013-11. 15. Калуш Ю. А. Показатель Хёрста и его скрытые свойства / Калуш Ю. А., Логинов В. М. // Сибирский журнал индустриальной математики. – 2002. – Том 5, № 4(12): Октябрь-декабрь. – С. 29–37. 16. Чепелева М. С. Прогнозирование в управлении потенциально опасным объектом / Чепелева М.С., Ткалич С. А. Чепелев С. А. // Научный журнал КубГАУ. – 2011. – № 74(10). – С. 1–12. – Режим доступу: http://ej.kubagro.ru/2011/10/pdf/23.pdf. 17. Кривоносова Е. К. Сравнение фрактальних характеристик временных рядов экономических показателей / Кривоносова Е. К., Первадчук В. П., Кривоносова Е. А. // Современные проблемы науки и образования. – 2014. – № 6. – Режим доступу: http://www.science-education.ru/120-15974.