

УДК 504.501

## **МОДЕЛЮВАННЯ ЗАБРУДНЕННЯ АТМОСФЕРИ ТА ПОВЕРХНЕВИХ ВОД В УКРАЇНІ**

*П.І. Бідюк, д-р техн. наук, проф.  
(Інститут прикладного системного аналізу НТУУ «КПІ»);*

*О.І. Савенков, д-р техн. наук, проф.  
(Національна академія управління);*

*О.М. Трофимчук, д-р техн. наук, проф.  
(Інститут телекомунікацій і глобального  
інформаційного простору НАН України)*

*В роботі досліджується проблема екологічного моделювання з використанням реальних статистичних даних. Наведено результати моделювання екологічних процесів на основі статистичних даних на основі методів регресійного аналізу та методу групового урахування аргументів. Результати розрахунків, аналізу та моделювання свідчать про адекватність отриманих моделей реальним процесам.*

*В работе исследуется проблема экологического моделирования с использованием фактических статистических данных. Приведены результаты моделирования экологических процессов на основе статистических данных с использованием методов регрессионного анализа и метода группового учета аргументов. Результаты расчетов, анализа и моделирования свидетельствуют об адекватности полученных моделей реальным процессам.*

*The work is dedicated to the problem of ecological modeling using actual statistical data. The results of ecological processes modeling are provided on the basis of statistical data using the methods of regression analysis and the group method for data handling. The results of computing achieved show an adequacy of the models constructed to real life processes.*

У зв'язку з постійним зростанням навантаження на довкілля та проголошеним світовою спільнотою курсом на сталий розвиток очевидна висока роль саме екологічного моделювання, як способу передбачення впливу наслідків економічної діяльності на довкілля.

© П.І. Бідюк, О.І. Савенков, О.М. Трофимчук, 2011

Враховуючи складне теоретичне підґрунтя багатьох методів моделювання та складність застосування результатів теоретичних досліджень на практиці, складається тенденція до застосування не «найкращого» методу високої складності, а декількох методів моделювання з подальшим порівнянням та можливою корекцією результатів. За такої постановки задачі перевага надається простішим з точки зору технічної реалізації методам, які дають прийнятну якість прогнозу. Чим більше методів використовується для побудови моделі досліджуваної системи, тим більшою вважається ймовірність визначення природної закономірності, відповідно до якої формуються значення досліджуваних параметрів [1–4].

Виходячи з описаних тенденцій та нинішнього рівня розвитку методів прогнозування за мету даної роботи були поставлені задачі дослідити та проаналізувати можливості традиційних методів обробки статистичної інформації та розробити методику їх адаптації до реальних задач з урахуванням особливостей досліджуваних процесів.

### **Постановка завдання**

Проаналізувати задачі, пов'язані з прогнозуванням розвитку екологічних процесів, представлених часовими рядами. Дослідити існуючі методи прогнозування часових рядів з урахуванням особливостей конкретних процесів на основі фактичних даних. Модифікувати методику моделювання часових рядів для врахування нелінійності та наявності сезонних тенденцій у процесі. Здійснити порівняльний аналіз використаних методів прогнозування на реальних даних. Зробити висновки щодо можливостей подальшого застосування використаних методів.

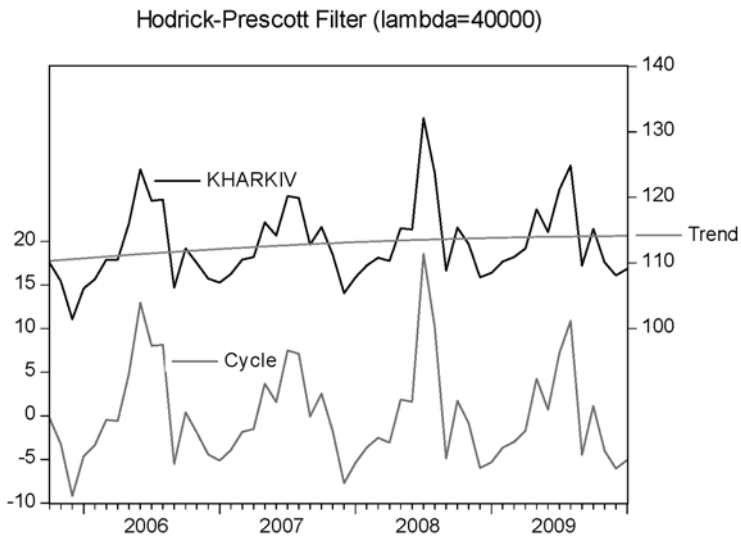
### **Моделювання та прогнозування вмісту забруднюючих речовин у атмосферному повітрі і поверхневих водах в Україні.**

Характерною особливістю досліджуваних даних є невелика кількість їх значень. Це зумовлено тим, що накопичення та збереження даних загальнонаціонального обов'язкового екологічного моніторингу в Україні було впроваджено лише у середині 2005 р. У ході досліджень проаналізовано динаміку зміни показників стану довкілля у масштабах всієї України або басейну головних українських річок і в конкретних місцях, зокрема у водах р. Уда Харківської області.

### **Аналіз даних**

Для перевірки можливості застосування сучасних методів аналізу часових рядів до реальних даних екологічного моніторингу на території України необхідно визначити, чи є досліджувані процеси стаціонарними. Для цього необхідно перевірити статистичні дані на наявність тренду. Для виявлення можливого лінійного тренду скористаємось відомим пакетом прикладних програм Eviews. Виділення тренду здійснимо за допомогою фільтра Ходріка-Прескотта [2, 5]. Результати аналізу наведено на рис. 1.

Як видно з графіка на рис. 1, вміст завислих речовин у поверхневих водах даної ріки має яскраво виражений сезонний характер, що може бути зумовленим як коливаннями екологічних показників (середньомісячна температура, рівень опадів, тощо), так і циклічністю роботи індустрії, зосередженої у даному регіоні. Крім того, тестування ряду даних свідчить про те, що в досліджуваній період спостерігається незначне підвищення середнього значення даного ряду.



*Рис. 1. Результати застосування фільтра Ходріка-Прескотта для даних стосовно забруднення річки Уда.*

Отже, доцільно розглянути цей ряд двома різними способами: як стаціонарного та як ряду з лінійним трендом (для дослідження тенденцій розвитку процесу у довгостроковій перспективі).

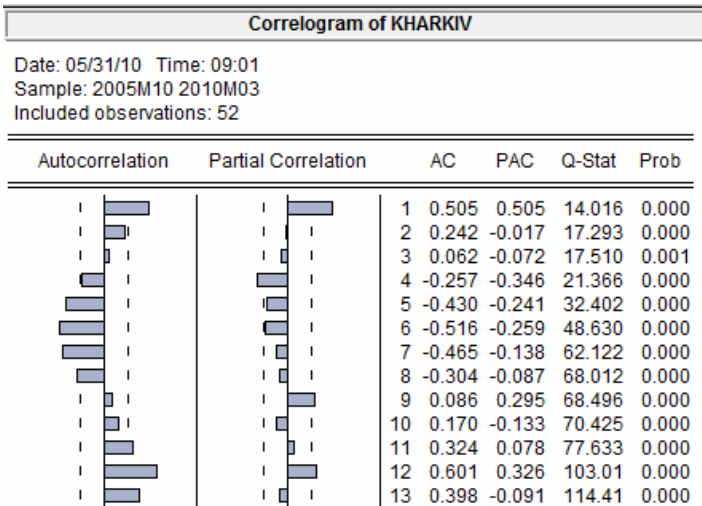
### Застосування авторегресійних моделей

Розглянемо можливість опису динаміки досліджуваних процесів забруднення за допомогою моделі авторегресії з ковзним середнім (АРКС). АРКС моделі вимагають менше витрат часу для побудови, проте при цьому вони забезпечують високу ступінь адекватності досліджуваному процесу, що робить їх придатними для подальшого аналізу [2, 5].

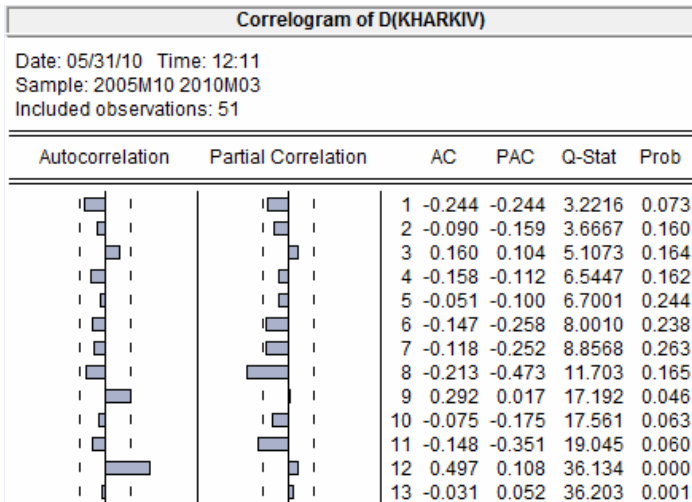
Проаналізуємо динаміку вмісту завислих речовин у воді ріки Уди Харківської області за допомогою рівнянь АРКС. Дослідимо глибину взаємозв'язків між розділеними в часі значеннями досліджуваного числового ряду за допомогою автокореляційної функції (АКФ) та часткової АКФ (табл. 1 і 2).

Аналіз АКФ для перших різниць свідчить про те, що побудову авторегресійних моделей слід починати з моделей нижчих порядків, які часто мають високу адекватність процесу і забезпечу-

Таблиця 1 — АКФ та ЧАКФ для значень часового ряду



Таблиця 2 — АКФ та ЧАКФ для першої різниці



ють прийнятну якість прогнозу. Також високі значення АКФ з 12-м лагом та ЧАКФ з 11-м лагом свідчать про яскраво виражену сезонність вмісту завислих речовин у воді, що спричиняється режимом коливань екологічного режиму та сезонністю основної маси викидів.

Таблиця 3 — Результати оцінювання моделі AP(1)

Метод найменших квадратів

Використано спостережень: 51 після корегування

Моделю AP (1):  $KNARKIV = C(1) + C(2) * KNARKIV(-1)$

	Оцінки коэф-тів	Станд. похибка	t-Статистика	Імовірність
C(1)	56.08706	14.37563	3.901538	0.0003
C(2)	0.504339	0.127112	3.967670	0.0003
R2	0.254969	Середнє значення змінної		113.0458
Скорегований R2	0.238773	Станд. відхилення змінної		6.004961
Станд. похибка регресії	5.239229	Критерій Акаїке		6.190999
Сума квадр. помилок	1262.678	Критерій Шварца		6.268966
		Стат. Дарбіна-Уотсона		1.974551

Отримана модель АР(1):

$$\begin{aligned} kharkiv(k) &= a_0 + a_1 kharkiv(k-1) + e(k) = \\ &= 56,087 + 0,5043 kharkiv(k-1) + e(k), \end{aligned}$$

де  $e(k)$  — залишки (похибки) моделі. Три основних статистичних характеристики якості моделі:

$$R^2=0,255; J=\text{СКП}=1262,7; DW=1,975.$$

Як видно з наведених характеристик, коефіцієнт детермінації має низьке значення (0,255), що свідчить про низьку інформативність моделі у порівнянні з вихідними даними. Статистика Дарбіна-Уотсона (1,975) наближається до свого найкращого значення, що свідчить про незначну автокореляцію залишків моделі. Загалом значення критеріїв свідчать про низьку адекватність моделі, тобто її структура вимагає уточнення. Характеристики якості прогнозу (за навчальною вибіркою) для даної моделі:

$$\text{СеКП} = 3,2313; \text{САП} = 3,0359; \text{САПП} = 2,7813; U = 0,0146$$

Значення коефіцієнта Тейла (0,0146) свідчить про загальну прийнятність моделі для прогнозування. Проте попередній аналіз критеріїв якості моделі свідчить про потребу уточнення структури даної моделі.

Проаналізувавши АКФ та ЧАКФ даного часового ряду, робимо висновок, що наступним регресором має бути значення з 4-м лагом, і, відповідно, наступна модель будується за допомогою АР (4). Для цієї моделі досліджувані статистичні характеристики набувають таких значень:

$$R^2=0,315; J=\text{СКП}=1042,4; DW=2,239.$$

Зміна значень коефіцієнта детермінації та суми квадратів похибок у порівнянні з попередньою моделлю загалом свідчать про покращення якості структури моделі та збільшення рівня її адекватності реальному процесові. Характеристики якості прогнозу для даної моделі:

$$\begin{aligned} \text{СекП} &= 4,8424; \text{САП} = 4,7986; \\ \text{САПП} &= 4,4028; \text{U} = 0,0217. \end{aligned}$$

Порівняння характеристик статичного однокрокового прогнозу даної моделі з характеристиками попередньої моделі загалом свідчить про погіршення прогностичних якостей моделі. Враховуючи загальне покращення адекватності даної моделі, її слід розширити додаванням наступних регресорів. Серед побудованих моделей найкращою виявилась АР(12).

**Таблиця 4 – Результати оцінювання моделі АР(12)**

Метод найменших квадратів

Використано спостережень: 40 після корегування

$$\begin{aligned} \text{KHARKIV} &= \text{C}(1) + \text{C}(2) * \text{KHARKIV}(-1) + \text{C}(3) * \text{KHARKIV}(-4) + \text{C}(4) * \text{KHARKIV}(-5) + \\ &+ \text{C}(5) * \text{KHARKIV}(-6) + \text{C}(6) * \text{KHARKIV}(-7) + \text{C}(7) * \text{KHARKIV}(-9) + \\ &+ \text{C}(8) * \text{KHARKIV}(-11) + \text{C}(9) * \text{KHARKIV}(-12) \end{aligned}$$

	Оцінки коєф-тів	Станд. похибка	t-Статистика	Імовір- ність
С(1)	71.11624	56.86668	1.250578	0.2204
С(2)	0.101609	0.128976	0.787819	0.4368
С(3)	-0.029107	0.125142	-0.232588	0.8176
С(4)	-0.051087	0.134768	-0.379076	0.7072
С(5)	-0.061572	0.134237	-0.458681	0.6497
С(6)	-0.194706	0.126841	-1.535036	0.1349
С(7)	0.123075	0.121163	1.015774	0.3176
С(8)	-0.104351	0.130789	-0.797860	0.4310
С(9)	0.591723	0.133933	4.418048	0.0001
$R^2$	0.705197	Середнє значення змінної		113.1150
Скорегований $R^2$	0.629119	Станд. відхилення змінної		5.577820
Станд. похибка регресії	3.396891	Критерій Акаїке		5.478706
Сума квадр. помилок	357.7050	Критерій Шварца		5.858704
		Стат. Дарбіна-Уотсона		2.046050

Статистичні характеристики даної моделі:

$$R^2=0,705; J=\text{СКП}=357,7; DW=2,046.$$

Характеристики якості статичного однокорокового прогнозу:

$$\begin{aligned} \text{СеКП} &= 2,9904; \text{САП} = 1,977; \\ \text{САПП} &= 1,709; U = 0,0132. \end{aligned}$$

Враховуючи циклічність екологічних процесів та сезонність дії основних забруднюючих факторів, додавати до моделей нові регресори з більшим лагом недоцільно. Побудуємо спільний графік реальних даних та адекватної їм моделі  $AR(12)$ :

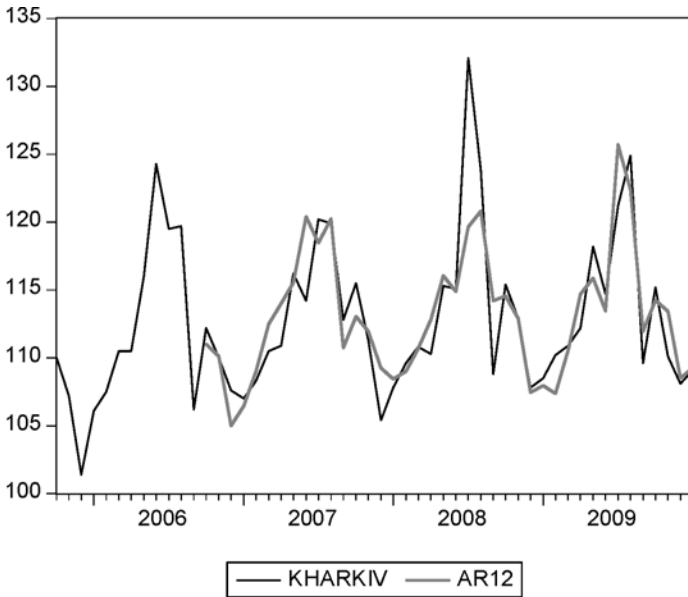


Рис. 2. Спільний графік динаміки початкового ряду та моделі  $AR(12)$ .

Як видно з побудованого прогнозу, модель  $AR(12)$  в цілому адекватно відображає тенденцію подальшого розвитку процесу.



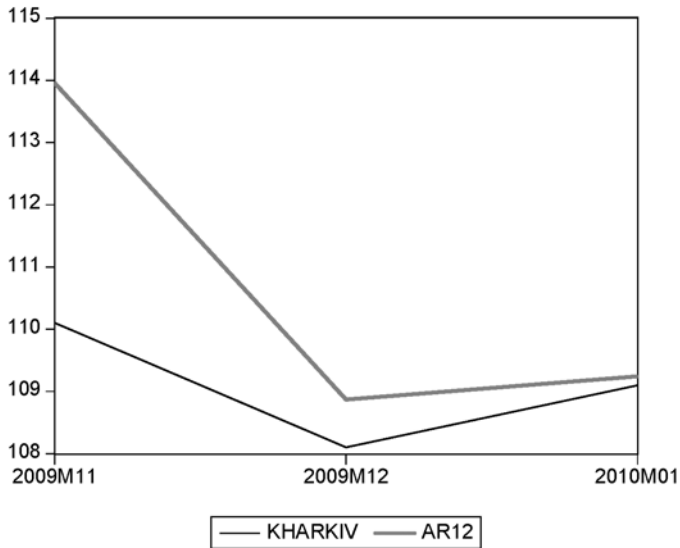


Рис. 3. Прогноз на три кроки вперед і реальне значення.

### Виділення тренду для довгострокового прогнозування

Враховуючи результати тесту Ходріка-Прескота, на довгих проміжках часу спостерігається деяке зростання середнього значення досліджуваного часового ряду. Це дозволяє зробити висновок, що для покращення якості прогнозу на довгий проміжок часу необхідно включити в модель час як додатковий регресор. Таблиця 5 містить результати опису вибраного процесу моделями різних порядків.

Як показує зведена таблиця характеристик побудованих моделей, модель АРКС(12, 6) з трендом характеризується найкращим коефіцієнтом детермінації та суми квадратів помилок, проте саме для неї найгіршим є значення статистики Дарбіна-Уотсона, отже, занадто точна апроксимація ряду на такому короткому фрагменті призводить до прояву явища автокореляції залишків.

Оцінимо тепер якість прогнозів, що дозволяють нам побудувати досліджені моделі.

Дослідження переносу забруднюючих речовин регресійним методом.

Таблиця 5 – Характеристики адекватності та якості прогнозу

Модель	Характеристики моделі			Характеристики прогнозу			
	$R^2$	SSE	DW	СеКП	САП	САПП	Коеф-т Тейла
AP(1)	0,255	1262,7	1,975	3,2313	3,0359	2,7813	0,0146
AP(4)	0,315	1042,4	2,239	4,8424	4,7986	4,4028	0,0217
AP(5)	0,356	958,5	2,2276	4,516	3,302	2,869	0,0200
AP(6)	0,441	826,8	2,221	4,239	3,114	2,708	0,0187
AP(7)	0,511	719,0	2,157	3,998	2,856	2,480	0,0175
AP(8)	0,540	673,2	1,831	3,911	2,696	2,342	0,0172
AP(9)	0,497	675,8	2,136	3,964	2,818	2,450	0,0175
AP(10)	0,468	693,5	2,104	4,0633	2,864	2,450	0,0180
AP(11)	0,469	668,9	1,988	4,0392	2,874	2,500	0,0179
AP(12)	0,705	357,7	2,046	2,9904	1,977	1,709	0,0132
AP(12)+c*t	0.707902	262.2	1,700	2,8615	2,060	1,869	0,0126
APKC(12,6)+c*t	0.9687	22.8	3.368	2.9192	3.225	2.919	0.0190

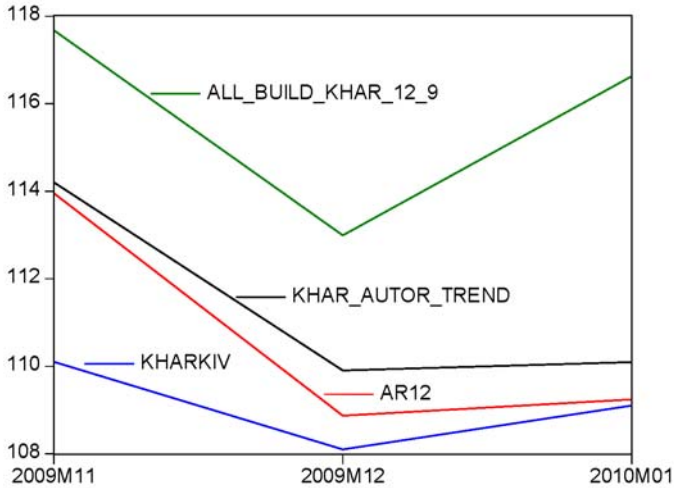


Рис. 4. Результати побудови динамічного трикрокового прогнозу для регресійних моделей.

Традиційними методами очищення річкових вод є використання відстійників. Ефективність схеми парної регресії дослідимо на основі двох споріднених рядів: вмісту завислих речовин у воді річки Уди Харківської області після одного та після двох відстійників. Обчислимо коефіцієнт кореляції між цими двома показниками.

**Таблиця 6 — Порівняльна характеристика моделей, побудованих на базі парної регресії**

Модель	Характеристики моделі			Характеристики прогнозу			
	R <sup>2</sup>	SSE	DW	СєКП	САП	САПП	Коеф-т Тейла
Модель 1	0,334	48,232	2,218	0,698	0,500	4,760	0,032
Модель 2	0,358	46,282	2,416	0,821	0,620	5,890	0,037
Модель 3	0,338	43,471	2,309	0,840	0,648	6,153	0,038
Модель 4	0,506	28,766	2,803	0,924	0,696	6,615	0,042
Модель 5	0,883	6,656	1,864	0,355	0,347	3,255	0,016
Модель 6	0,473	37,273	1,982	0,502	0,480	4,513	0,0229
Модель 7	0,867	7,402	2,056	0,335	0,333	3,104	0,0154

За загальною сукупністю критеріїв якості моделей можна зробити висновок, що найбільш адекватними в плані моделювання закономірностей перебігу процесів та ефективного прогнозування подальших значень ряду мають бути моделі під номерами 5 та 7. Побудуємо на їх основі прогноз та порівняємо його з реальними значеннями ряду та прогнозом за моделлю 1 (випадок парної регресії без використання лагів).

Найближчою за значеннями абсолютних відхилень до реальних даних є модель під номером 5.

### **Побудова моделей за допомогою методу групового урахування аргументів (МГУА)**

Враховуючи невелику довжину досліджуваних вибірок, невисоку точність проведених вимірювань та зашумленість аналізованих даних, доцільно скористатись алгоритмами МГУА для побудови моделей на основі цих даних [3, 6].

Виконано моделювання часового ряду, що відображає вміст завислих речовин у воді ріки Уди Харківської області після 2-го відстійника.

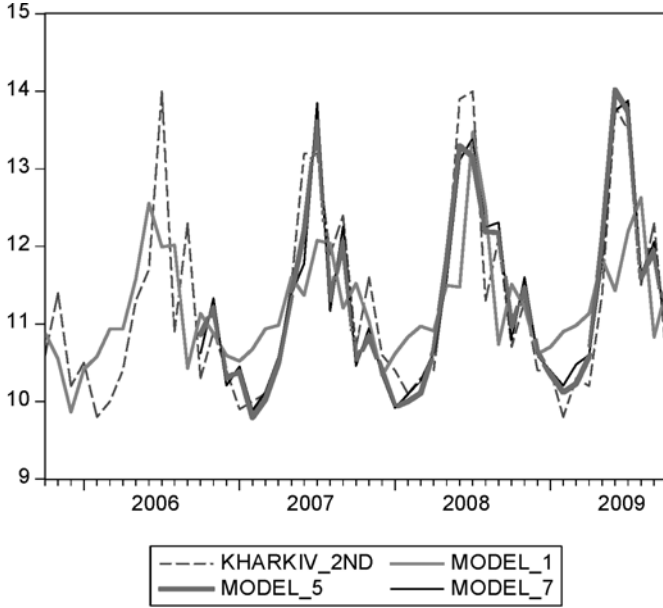


Рис. 5. Моделі часового ряду, побудовані на базі парної регресії.

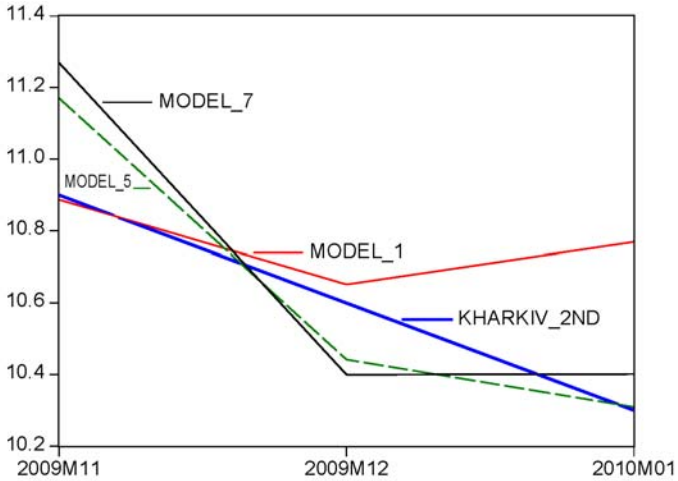
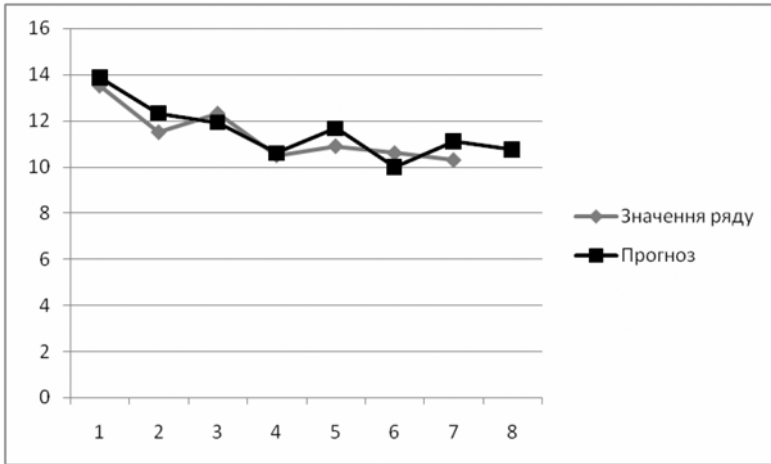


Рис. 6. Прогнози, побудовані на базі моделей парної регресії.

1. Базовий клас моделей — лінійні. Критерій регулярності — корінь з середньоквадратичної похибки. Довжина навчальної вибірки — 40 значень, довжина прогнозу — 8 значень.

**Таблиця 7 — Моделі, генеровані алгоритмом МГУА у випадку лінійного базового класу моделей**

Ранг моделі	Модель у явному вигляді	Значення критерію
1	$0,9882*a(k-1)$	0,0011835
2	$0,5821*a(k-1)+0,4101*a(k-2)$	0,0018721
3	$0,8993*a(k-1)+0,08929*a(k-2)$	0,0018835
4	$0,7796*a(k-1)+0,2135*a(k-3)$	0,001954
5	11,28	0,001962



*Рис. 7. Реальні та прогнозовані за МГУА значення числового ряду.*

1. Базовий клас моделей — частково лінійні. Критерій регулярності — корінь з середньоквадратичної помилки. Довжина навчальної вибірки — 40 значень, довжина прогнозу — 8 значень.

**Таблиця 8 — Моделі, згенеровані алгоритмом МГУА у випадку частково лінійного базового класу моделей**

Ранг моделі	Модель у явному вигляді	Значення критерію
1	$-0,04269 * a(k-1) * a(k-3) + 1,478 * a(k-1)$	0,0010235
2	$-34,91 - 0,3032 * a(k-1) * a(k-2) - 0,3488 * a(k-1) * a(k-3) + 7,975 * a(k-1) + 0,3094 * a(k-2) * a(k-3)$	0,0010799
3	$0,3702 * a(k-1) * a(k-2) - 0,07736 * a(k-1) * a(k-3) + 5,768 * a(k-1) + 0,3794 * a(k-2) * a(k-3) - 4,002 * a(k-3)$	0,0011364
4	$-0,4217 * a(k-1) * a(k-2) + 5,492 * a(k-1) + 0,3568 * a(k-2) * a(k-3) + 0,9116 * a(k-2) - 4,674 * a(k-3)$	0,001954
5	$0,9882 * a(k-1)$	0,0011835

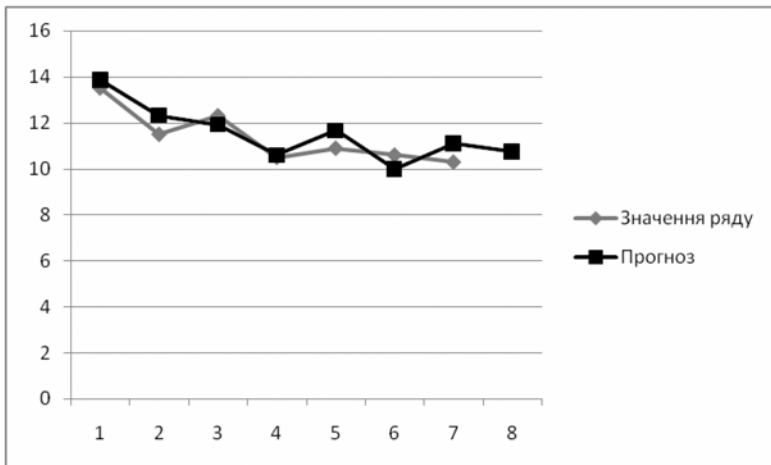
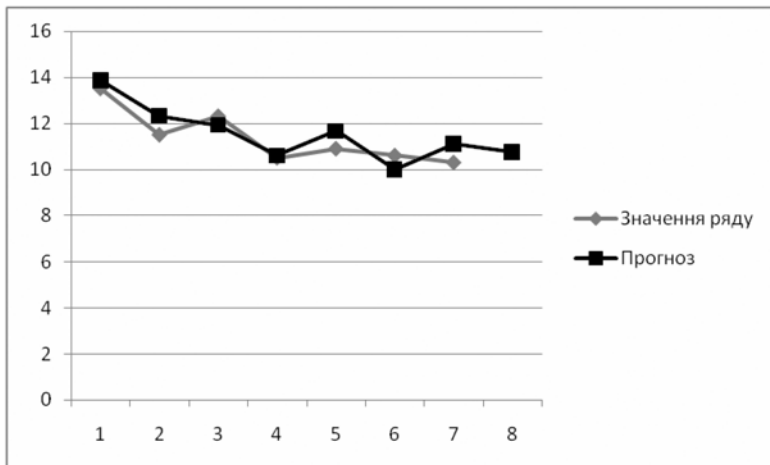


Рис. 8. Результати моделювання значень часового ряду за допомогою МГУА, базовий клас моделей — лінійні.

1. Базовий клас моделей — квадратичні. Критерій регулярності — корінь з середньоквадратичної помилки. Довжина навчальної вибірки — 40 значень, довжина прогнозу — 8 значень.

**Таблиця 9 – Моделі, генеровані алгоритмом МГУА у випадку нелінійного базового класу моделей**

Ранг моделі	Модель у явному вигляді	Значення критерію
1	$-74,27 - 0,3508 * a(k-1)^2 - 0,5222 * a(k-1) * a(k-3) + 14,55 * a(k-1) + 0,3557 * a(k-2) + 0,2279 * a(k-2)^2$	0,0009605
2	$-72,29 - 0,363 * a(k-1)^2 - 0,4944 * a(k-1) * a(k-3) + 14,53 * a(k-1) + 0,03 * a(k-1) * a(k-3) + 0,1996 * a(k-3)^2$	0,0096053
3	$-72,73 - 0,363 * a(k-1)^2 - 0,505 * a(k-1) * a(k-3) + 14,63 * a(k-1) + 0,0154 * a(k-2)^2 + 0,2192 * a(k-3)^2$	0,0009607
4	$-73,1 - 0,3661 * a(k-1)^2 - 0,031 * a(k-1) * a(k-2) - 0,536 * a(k-1) * a(k-3) + 0,2348 * a(k-3)^2$	0,0009657
5	$-59,26 - 0,3689 * a(k-1)^2 - 0,04354 * a(k-1) * a(k-3) + 14,02 * a(k-1) + 0,3969 * a(k-2) - 2,171 * a(k-3) + 0,2756 * a(k-3)^2$	0,0009872



*Рис. 9. Значення ряду та прогноз, згенерований МГУА для нелінійних базових моделей.*

### Розробка інформаційно-аналітичної системи оцінювання екологічної ситуації

В зв'язку з тим, що традиційні методи регресійного аналізу здатні в цілому адекватно відображати перебіг екологічних процесів, проте занадто точна апроксимація може призвести до погіршення прогностичних якостей моделі, прийнято рішення розробити інформаційно-аналітичну систему для моделювання зашумлених даних за невеликими вибірками. Як окрема складова, до системи включено блок перевірки лінійності структури моделі, та блок поетапного процесу побудови моделі на основі комплексного аналізу традиційними методами моделювання. Ілюстрації функціонування системи наведені на рис. 10—12.

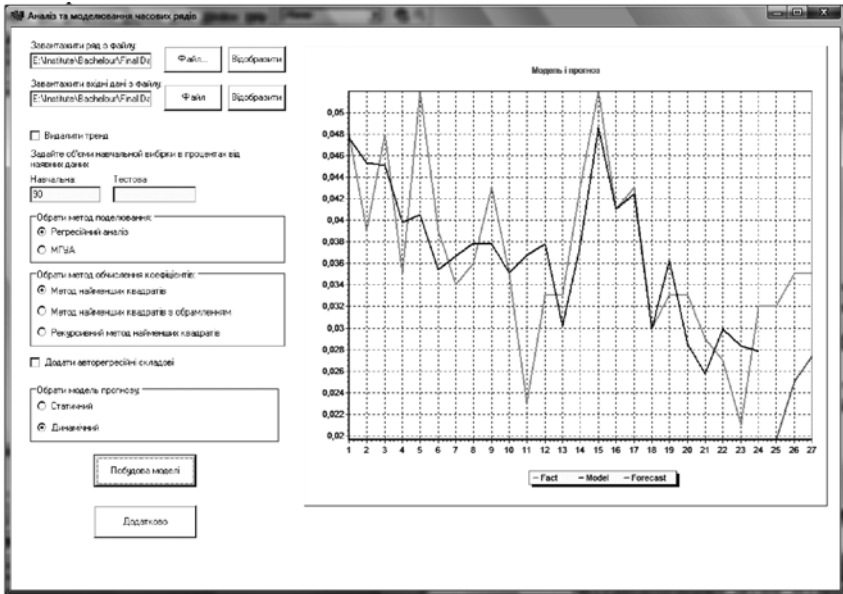


Рис. 10. Результати побудови моделі на основі множинної регресії.



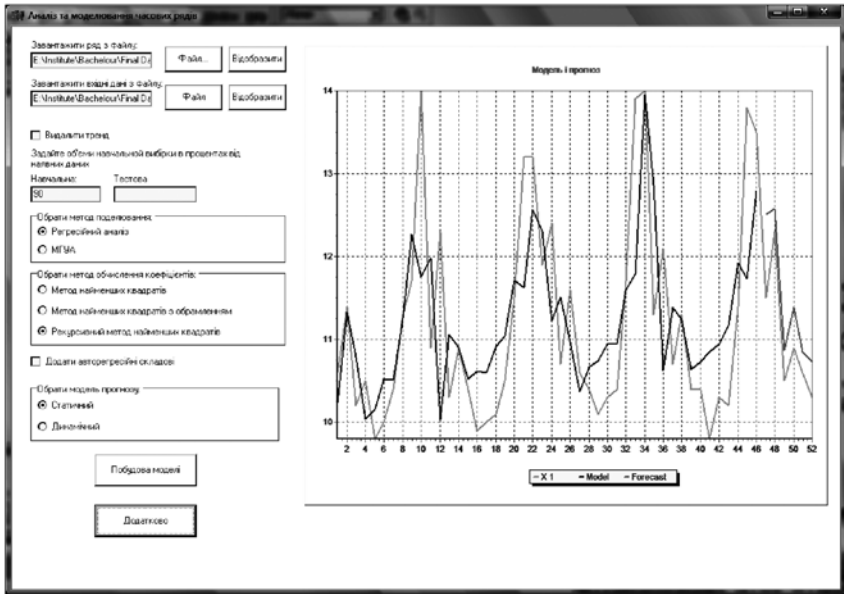


Рис. 11. Результати побудови моделі та прогнозу на основі парної регресії.

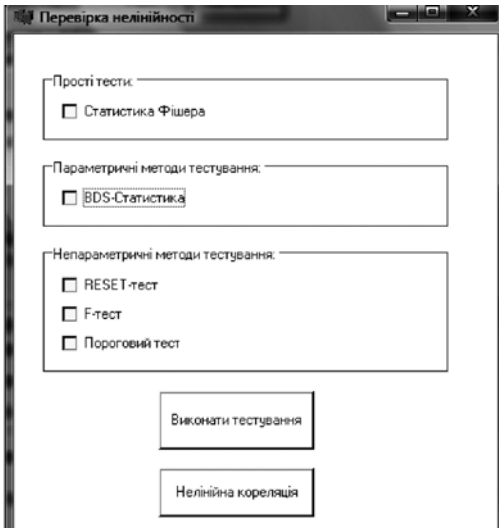


Рис. 12. Модуль тестування нелінійності.

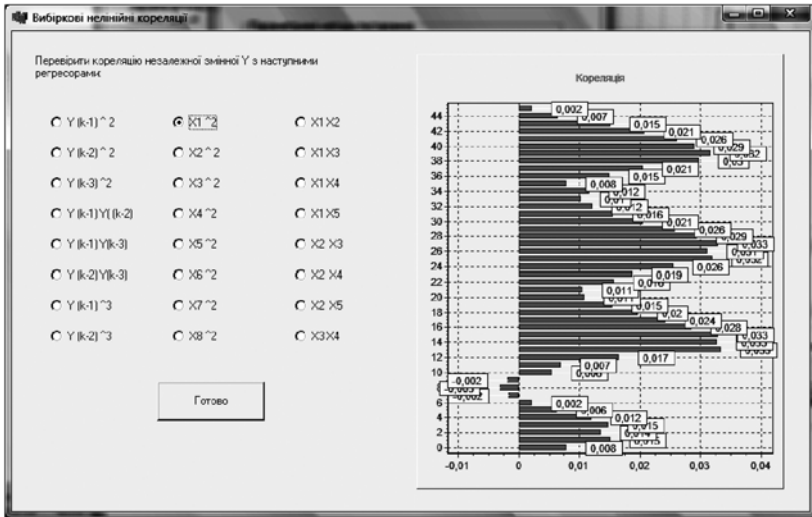


Рис. 13. Блок перевірки вибіркової нелінійної кореляції.

### Модифікація методів регресійного аналізу для ряду з вираженою сезонністю

Практично всі фактори, що впливають на рівень забруднення довкілля (а саме — сільське господарство, промислова активність, використання транспортних засобів населенням, тощо), підвладні яскраво вираженому сезонному ефектові, який виражається у регулярних підйомах та спадах активності залежно від пори року. Загалом це відображається на рівні забруднення довкілля у вигляді яскраво виражених тенденцій до сезонних коливань. Попередньо виконаний аналіз реальних даних за допомогою авторегресії та з урахуванням виділення тренду довів високу корельованість елементів числового ряду з періодичністю в 12 вимірів (високий рівень адекватності моделі з включенням 12-го лагу).

Для моделювання яскраво вираженої сезонності досліджуваних параметрів модифікована традиційна методика регресійного аналізу. А саме: впроваджено мультиплікативні сезонні зрізи, значення яких ітеративно уточнюються на кожному періоді, застосовано елементи завадостійкого моделювання (зокрема — МГУА, який добре апроксимує структуру моделі на коротких вибірках) на

кожному сезонному періоді з ітеративним уточненням значень коефіцієнтів на нових періодах, комбінування різних підходів на певних етапах обробки інформації та побудови моделі.

### **Висновки**

Виконано аналіз часових рядів, що відображають динаміку вмісту забруднюючих речовин в поверхневих водах річок України. Аналіз даних, представлених часовими рядами, виявив яскраво виражену сезонність перебігу процесів, що вимагає введення додаткових методів аналізу та уточнення значення вектора параметрів. На основі виконаної апроксимації побудовано моделі, які дають прийнятну точність прогнозування. Також використано алгоритм МГУА, розглянуто апроксимацію структури часового ряду, обчислено значення прогнозів, згенерованих алгоритмом МГУА для досліджуваних часових рядів.

На основі отриманих результатів та виконаного аналізу переваг та недоліків досліджуваних класичних методів представлено інформаційно-аналітичну систему для аналізу та моделювання перебігу екологічних процесів. Головною інновацією в даній системі є комплексне тестування досліджуваного часового ряду на лінійність.

\* \* \*

1. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов: учебное пособие / Ю. П. Лукашин. — М.: Финансы и статистика, 2003. — 416 с.
2. Дрейпер Н. Прикладной регрессионный анализ / в 2-х томах. Н. Дрейпер, Г. Смит; пер. с англ. Ю. П. Адлер, В. Г. Горский. — М.: Финансы и статистика, 1986. — Т. 1. — 366 с.; Т. 2. — 351 с.
3. Ивахненко А. Г. Долгосрочное прогнозирование и управление сложными системами / А. Г. Ивахненко. — К.: Техника, 1975. — 312 с.
4. Берлянд М. Е. Прогноз и регулирование загрязнения атмосферы. [Текст] / М. Е. Берлянд. — М.: Гидрометеиздат, 1985. — 272 с.
5. Себер Дж. Линейный регрессионный анализ / Себер Дж., пер. с англ. В. П. Носко; под ред. М. Б. Малютова. — М.: Мир, 1980. — 456 с.
6. Зайченко Ю.П. Основи проектування інтелектуальних систем / Ю. П. Зайченко. — К.: Слово, 2003. — 351 с.

*Отримано: 28.02.2011 р.*