

# МЕТОДИ І АЛГОРИТМИ СУЧАСНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

УДК 519.004.942

О. Кожухівська

Черкаський державний технологічний університет,  
кафедра інформатики та інформаційної безпеки

## РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ ФІНАНСОВО-ЕКОНОМІЧНИХ ПРОЦЕСІВ

© Кожухівська О., 2013

Для прийняття об'єктивних обґрунтованих рішень у всіх сферах діяльності необхідно створювати інформаційні системи підтримки прийняття рішень на основі статистичних даних і експертних оцінок та з використанням відповідних математичних моделей та методів. Запропоновано процедуру побудови інформаційної системи підтримки прийняття рішень за методологією системного аналізу на основі сучасних методів і моделей прогнозування, яка надає можливість оцінювати та прогнозувати фінансово-економічні та інші процеси в умовах впливу збурень довільних типів та різної природи. Показано можливості врахування статистичних і параметричних невизначеностей завдяки використанню ймовірнісно-статистичних методів обробки даних. На всіх етапах процесу обробки даних використовуються відповідні критерії якості, які забезпечують високу якість даних, моделей і прогнозів. Наведено приклади побудови моделей різних типів для вибраних макроекономічних і демографічних процесів та їх застосування для розв'язання задачі середньо- та короткострокового прогнозування. Виконано порівняльний аналіз результатів прогнозування, отриманих різними методами.

**Ключові слова:** система підтримки прийняття рішень, системний підхід, фінансово-економічні процеси, моделювання і прогнозування за статистичними даними, порівняльний аналіз.

To make objective substantiated decisions in all spheres of activities it is necessary to construct information decision support systems on the basis of statistical data and expert estimates with application of appropriate mathematical models and procedures. Using the system analysis methodology, a procedure for constructing information decision support system on the basis of modern models and forecasting techniques is developed, that provides a possibility for estimation and forecasting financial, economic and other types of processes. This can be done in conditions of influence of arbitrary disturbances of various origin. Also an attention is paid to the possibility of taking into consideration of statistical and parametric uncertainties due to application of appropriate probabilistic and statistical data processing techniques. All stages of data processing are controlled with the use of appropriate quality criteria that provide high quality of data, models and forecast estimates. Examples of application of the system constructed are provided touching upon solving the problems of short- and medium-term forecasting for selected macroeconomic and demographic processes. A comparative analysis of forecasting results received with alternative techniques is performed.

**Key words:** decision support system, system approach, financial and economic processes, modeling with statistical data, comparative analysis.

### Вступ

Прийняття коректних рішень надає гарантію успішного подальшого розвитку подій у будь-якій сфері діяльності, незалежно від типу чи характеру рішень, враховуючи рішення особистого

характеру. Для успішного розв'язання цієї задачі створюються інформаційні системи підтримки прийняття рішень (ІСППР), які ґрунтуються на сучасних принципах і підходах до прийняття рішень, а також методах і технологіях обробки даних та експертних оцінок. ІСППР набувають все більшого поширення практично в усіх галузях діяльності людини завдяки відносній простоті їх архітектури, цілеспрямованому функціональному наповненню, повноті інструментарію, необхідного для аналізу даних, та забезпеченню можливості прискореного генерування і вибору альтернатив. Їх використовують з метою розв'язання задач математичного моделювання процесів різної природи, прогнозування їх розвитку, оптимізації рішень управлінського чи іншого характеру, розроблення альтернативних стратегій розвитку підприємств, підтримки прийняття особистих рішень тощо. Поширеними типами задач, які розв'язують за допомогою ІСППР, є коротко- та середньострокове прогнозування розвитку фінансово-економічних процесів, процесів у екології, оскільки оцінки прогнозів необхідні для прийняття обґрунтованих рішень [1–4]. Проектування ІСППР ґрунтується на використанні альтернативних математичних моделей, процедурах прогнозування та оптимізації, множинах критеріїв для аналізу якості обробки даних та експертних оцінок на всіх етапах формування альтернатив і логічних правилах вибору кращої альтернативи. У системах такого типу альтернативні методи, моделі й критерії аналізу якості передбачають на кожному етапі обробки даних та експертних оцінок.

Математичні моделі досліджуваних процесів можна розробляти до проектування та реалізації системи або у процесі використання ІСППР на основі нових статистичних даних і експертних оцінок. Такий підхід дає змогу адаптувати систему до можливих змін структури і параметрів досліджуваних процесів та об'єктів і застосовувати кращі моделі, наявні на поточний момент часу. Крім того, ІСППР, яка проектується і реалізується із застосуванням сучасних принципів системного аналізу, надає змогу враховувати невизначеності різних типів і природи. Все це сприяє підвищенню якості створюваних математичних моделей, обчислюваних оцінок прогнозів та формуванню рішень, які приймають на їх основі [3, 4].

У роботі розглянуто побудову ІСППР для розв'язання задач моделювання і прогнозування фінансово-економічних процесів на основі статистичних даних. Можливості створеної системи ілюструються прикладами моделювання і прогнозування сучасних макроекономічних та фінансових процесів.

### **Постановка задачі**

Ставиться задача проектування та реалізації інформаційної СППР для розв'язання задач моделювання і прогнозування процесів довільної природи на основі методики моделювання часових рядів та обчислення оцінок прогнозів на основі побудованих моделей. Процеси можуть бути стаціонарними і нестаціонарними, а також містити нелінійності стосовно змінних. ІСППР повинна ґрунтуватись на альтернативних процедурах оцінювання параметрів моделей та обчислення оцінок коротко- і середньострокових прогнозів. Для аналізу якості даних, моделей і оцінок прогнозів необхідно застосувати відповідні множини статистичних критеріїв якості.

### **Принципи проектування СППР**

Проектування сучасних інформаційних СППР ґрунтується на принципах системного аналізу [5], які забезпечують створення сучасної високоорганізованої системи, придатної для ефективного практичного використання. Ці принципи наведено нижче:

1. Відкрита модульна архітектура побудови системи, яка забезпечує можливість її розвитку з розширенням кількості функцій, методів обробки даних, моделювання, прогнозування, прийняття рішень та представлення результатів.

Цей важливий принцип дає змогу користувачеві системи модифікувати ІСППР, вводячи нові методи обробки даних, оцінювання параметрів моделей, прогнозування, структурної і параметричної оптимізації (процесів моделювання і прийняття рішень), управління тощо. Він реалізується завдяки коректному вибору інструментальної платформи для реалізації системи та використанню сучасних методів модульного програмування.

2. ІСППР високого рівня повинна мати *властивість адаптування* до користувача та навчатись на основі досвіду, який з'являється у процесі її функціонування. Наприклад, у реальному часі можна так виконувати адаптацію функціонування системи до користувача та класу задач, що розв'язуються:

- аналізувати час реакції користувача і відповідно до цього параметра корегувати динаміку діалогу користувач – комп'ютер;
- накопичувати ефективні (щодо досягнення необхідної точності обчислень та витрат часу) алгоритми розв'язування задач визначених класів і автоматизувати процес їх вибору для знаходження розв'язку конкретної задачі, тобто пропонувати альтернативи користувачу;
- система повинна адаптуватись до бажаних для користувача типів представлення результатів обчислень і надавати їх можливу інтерпретацію;
- СППР автоматично визначає тип досліджуваного процесу, застосовуючи спеціальні тести на стаціонарність та лінійність, і пропонує користувачеві модель або декілька типів альтернативних моделей для опису досліджуваного процесу;
- СППР накопичує знання стосовно типів процесів та кращих моделей, що їх описують, зберігаючи поточні результати обчислень у базі знань і даних (БЗД) з метою їх повторного ретроспективного аналізу і застосування.

3. Система має, як правило, *ієрархічну архітектуру* обчислювального процесу, яка узгоджується з методами та принципами прийняття рішень, які застосовують особи, що приймають рішення (ОПР), у повсякденному житті і які використовуються у багатьох реальних системах керування та прийняття рішень.

Ієрархічність представлення проблем притаманна системному підходу до розв'язування складних задач; це природний спосіб аналізу складних проблем та створення відповідних систем менеджменту. Як приклади, можна навести ієрархічну структуру *системи освіти, систем керування технологічними процесами, ієрархічний процес планування у бізнесі та на виробництві, систему ціноутворення* на вироблену продукцію тощо.

4. ІСППР призначена для швидкісної, за можливості, паралельної обробки даних, яка забезпечує її адекватну реакцію на запити користувача. Прийнятним часом очікування результату обчислень є 3–4 секунди [2, 6]; якщо час очікування перевищує цю межу, то користувач повинен отримувати відповідну інформацію про поточний стан системи.

Паралельна обробка даних необхідна, наприклад, під час реалізації процедур статистичного моделювання досліджуваних процесів, використання складних багатовимірних моделей і методу Монте-Карло для оцінювання параметрів таких математичних моделей. Паралельно виконується обробка даних в ієрархічних системах керування технологічними процесами та технічними системами, наприклад, при керуванні рухомими об'єктами.

5. Як правило, під час розв'язання більшості прикладних задач виникають невизначеності різних типів і природи. СППР повинна надавати можливість *моделювати та враховувати ці невизначеності* з метою підвищення якості рішень.

Прикладами можуть бути *структурні, статистичні і параметричні невизначеності*, які виникають у процесі математичного моделювання процесів різної природи. Їх можна врахувати, наприклад, введенням випадкових процесів, описаних статистичними розподілами з відповідними параметрами, нечіткими множинами, ймовірнісними оцінками змінних, параметрів і ситуацій, інтервальними оцінками параметрів тощо.

*Структурні невизначеності* пов'язані, передусім, з проблемами оцінювання структури математичної моделі – порядок процесу (і моделі) може змінюватись у часі; неточне знання лагу (або лаг змінюється у часі); недостатня інформація стосовно нелінійності процесу, нестационарні збурення з розподілами змішаного типу.

*Статистичні невизначеності* стосуються проблем визначення типу розподілу випадкових величин, статистичних характеристик цих розподілів, наявності екстремальних значень, які потребують додаткової обробки, заповнення пропусків даних, видалення шумових складових тощо.

Прикладом підходу до врахування статистичних невизначеностей є модель процесу в просторі станів та оптимальне оцінювання стану за допомогою фільтра Калмана або іншого належного інструменту. Для того, щоб застосувати фільтр Калмана, необхідно побудувати модель і представити її у просторі станів [7, 8]:

$$\mathbf{x}(k) = \mathbf{F} \mathbf{x}(k-1) + \mathbf{B} \mathbf{u}(k-1) + \mathbf{w}(k), \quad (1)$$

$$\mathbf{z}(k) = \mathbf{H} \mathbf{x}(k) + \mathbf{v}(k), \quad (2)$$

де  $\mathbf{x}(k)$  – вектор стану процесу;  $\mathbf{F}$  – матриця динаміки процесу (перехідна матриця станів);  $\mathbf{B}$  – матриця коефіцієнтів керування;  $\mathbf{u}$  – вектор керуючих впливів;  $\mathbf{W}$  – вектор випадкових зовнішніх збурень;  $\mathbf{z}$  – вектор вимірів;  $\mathbf{H}$  – матриця (коефіцієнтів) вимірів;  $\mathbf{V}$  – вектор шумів вимірів;  $k = 0, 1, 2, \dots$  – дискретний час (номер виміру), який зв'язується з неперервним часом  $t$  через період дискретизації вимірів  $T_s$ :  $t = k T_s$ . Отриману модель використовують надалі в алгоритмі оптимальної фільтрації. До виразів (1), (2) легко приводяться всі дискретні моделі і диференціальні рівняння, наприклад, авторегресія та авторегресія з ковзним середнім [6].

Основне рівняння процедури оптимальної фільтрації даних має вигляд:

$$\hat{\mathbf{x}}(k, k) = \mathbf{F} \hat{\mathbf{x}}(k-1, k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H} \mathbf{F} \hat{\mathbf{x}}(k-1, k-1)],$$

де  $\hat{\mathbf{x}}(k, k)$  – оптимальна оцінка стану процесу у момент часу  $k$ , обчислена на основі вимірів на момент  $k$ , включно;  $\mathbf{K}(k)$  – оптимальний коефіцієнт фільтра, який знаходять за умови мінімізації математичного сподівання суми квадратів похибок оцінок вектора стану:  $\min_K E \{[\mathbf{x}(k) - \hat{\mathbf{x}}(k)]^T [\mathbf{x}(k) - \hat{\mathbf{x}}(k)]\}$ .

З наведеного вище рівняння оптимальної фільтрації випливає, що оптимальна оцінка стану складається з однокрокового прогнозу та додаткової складової (інновації), яка визначається різницею між значенням останнього (поточного) виміру та оцінкою прогнозу. У процесі оцінювання стану процесу фільтр дає можливість врахувати збурення стану  $\mathbf{w}(k)$  і шум вимірів  $\mathbf{v}(k)$ . Основною вимогою успішного застосування фільтра для обробки даних є адекватна модель процесу та коректне визначення оцінок коваріаційних матриць збурень стану і шумів вимірів. Оскільки фільтр розроблено для лінійних систем, то кращі результати оцінювання можна очікувати саме для систем цього класу. Нелінійні системи необхідно належно лінеаризувати. Оцінки коваріацій збурень стану і шумів вимірів можна отримати також за допомогою адаптивного варіанта цього фільтра [7, 8].

*Параметричні невизначеності* зумовлені такими труднощами: оцінювання параметрів статистичної чи математичної моделі виконується на основі неповних зашумлених даних або даних з екстремальними значеннями; наявна мультиколінеарність вимірів; занадто короткі вибірки, погана обумовленість матриць вимірів – відношення максимального власного числа до мінімального перевищує 10.

Для подолання більшості типів невизначеностей необхідно виконувати [8]:

- коректну попередню обробку даних з використанням належної множини критеріїв якості;
- цифрову й оптимальну фільтрацію спостережень;
- модель повинна бути адаптивною (за структурою і параметрами), тобто такою, щоб можна було оцінювати її параметри повторно у процесі використання;
- у випадку розв'язання задачі прогнозування для одного процесу необхідно використовувати моделі ідеологічно різних типів і комбінувати отримані за їх допомогою оцінки прогнозів;
- для прийняття рішень необхідно використовувати взаємодоповнювальні методи, наприклад, такі: оптимізаційні процедури, експертне оцінювання, нечітку та нейронечітку логіку, дерева рішень, байєсівські мережі тощо;
- для побудови ІСППР також необхідно застосовувати принципи системного аналізу (ієрархічності процесів прийняття рішень, оптимальності обчислювальних процедур, повноти аналізу даних та експертних оцінок, багатокритеріальності, альтернативності рішень).

*Системний аналіз* – це методологія аналізу та математичного опису процесів та об'єктів різної природи, які функціонують в умовах наявності невизначеностей, з метою виявлення та описання зв'язків між змінними одного та різних рівнів ієрархії з кінцевою метою прийняття оптимальних або раціональних рішень на основі виконаного аналізу.

Методологія системного аналізу ґрунтується на таких принципах:

- досліджувані процеси мають ієрархічну структуру, в якій між рівнями ієрархії існують прямі та зворотні зв'язки;
- математичний опис проходження процесів, пов'язаний з аналізом невизначеностей різної природи (походження) та характеру;
- невизначеності можуть мати структурний, статистичний (ймовірнісний), параметричний та інший характер;
- для пошуку рішень стосовно прогнозування напрямів проходження процесів та керування ними застосовують теорію оптимальних рішень: субоптимальні та раціональні методи;
- за необхідності застосовується багатокритеріальна оптимізація;
- пошук альтернативних рішень виконується за множиною взаємодоповнювальних методів, а вибір кращого рішення з множини можливих – за допомогою множини критеріїв.

Успішним інструментом реалізації процедури системного аналізу може бути інформаційна СППР, яка повинна містити такі елементи: базу знань і даних (БЗД); процедури попередньої обробки даних; критерії якості результатів обробки даних на кожному етапі обчислень (попередня обробка даних, моделювання, прогнозування, формування альтернативних рішень, вибір кращої альтернативи); процедури формування висновку (прийняття рішень); процедури (правила) вибору кращого рішення з множини обчислених альтернатив за допомогою множини критеріїв.

Сформулюємо вимоги користувача та функціональні вимоги до ІСППР для моделювання і прогнозування процесів довільної природи.

### **Вимоги користувача до системи**

Розглянемо можливий варіант формулювання вимог користувача, що стосуються: типу комп'ютера й операційної системи, для якої створюється система; можливості введення даних з клавіатури, жорсткого диска, гнучкого диска, компакт-диска, локальної та глобальної мереж; перегляду введених даних у цифровому та графічному вигляді або у вигляді діаграми; адаптивності інтерфейсу; наявності можливості вибирати метод прогнозування; оцінювання точності отримуваних оцінок прогнозів під час прогнозування на задану кількість кроків; можливості доповнювати та редагувати дані; можливості зберігати результати прогнозування для подальшої роботи.

**Вибір методів прогнозування.** Для прогнозування процесів різної природи вибрано такі методи: експоненціальне згладжування; на основі лінійних різницевих рівнянь типу авторегресії з ковзним середнім; методи ковзного середнього; адаптивний фільтр Калмана; метод подібних траєкторій (МПТ).

**Критерії вибору кращих моделей з множини оцінених кандидатів.** Для вибору скористаємося такими статистичними критеріями якості регресійних моделей [6]:

- коефіцієнт детермінації як міра інформативності моделі:

$$R^2 = \frac{\text{Var}[\hat{y}]}{\text{Var}[y]}$$

- сума квадратів похибок моделі:  $SE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$ ;

- інформаційний критерій Акайке:  $IKA = N \cdot \ln\left[\sum_{k=1}^N e^2(k)\right] + 2n$ , де  $N$  – довжина вибірки даних;  $n$  – кількість параметрів моделі;

- статистика Дарбіна–Уотсона для визначення ступеня корельованості похибок моделі:

$$DW = \frac{\sum_{k=2}^N [e(k) - e(k-1)]^2}{\sum_{k=1}^N e^2(k)} ;$$

- середньоквадратична похибка:  $RSME = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$ .

**Критерії для оцінювання якості прогнозів.** Використаємо такі формальні критерії для оцінювання якості прогнозів за моделями:

- середня похибка:  $ME = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)$ ;

- середня похибка у відсотках:  $MPE = \frac{1}{N} \sum_{i=1}^N \frac{(y_i - \hat{y}_i)}{y_i} \cdot 100\%$ ;

- середня абсолютна похибка у відсотках:

- $MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{|y_i|} \cdot 100\%$ ;

- коефіцієнт нерівності Тейла:  $U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^T (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^T (y_i)^2} + \sqrt{\frac{1}{N} \sum_{i=1}^T (\hat{y}_i)^2}}$ .

**Функціональні вимоги до системи.** Функціональні вимоги стосуються безпосередньо тих конкретних функцій, які виконуватиме система. Для створюваної системи (архітектура системи представлена на рис. 1, а функціональна схема – на рис. 2) можна сформулювати такі функціональні вимоги:

- ✓ введення даних (часового ряду) вручну з клавіатури та з диска;
- ✓ редагування введених даних;
- ✓ застосування попередньої обробки даних (нормування, фільтрація);
- ✓ візуальне дослідження і аналіз початкових даних;
- ✓ автоматичне обчислення описових статистик для початкових даних;
- ✓ аналіз початкових даних на наявність тренду і гетероскедастичності;
- ✓ вибір методів прогнозування;
- ✓ оцінювання параметрів математичних моделей;
- ✓ обчислення автокореляційної функції процесу;
- ✓ застосування декількох критеріїв для оцінювання якості моделі;
- ✓ автоматичне оцінювання якості моделі;
- ✓ застосування декількох критеріїв для оцінювання якості прогнозу;
- ✓ автоматичне оцінювання якості прогнозу;
- ✓ відображення результатів у різному вигляді (цифровому і графічному);
- ✓ експорт результатів роботи програми в Microsoft Office Excel.

**Короткий опис функцій ІСППР.** *Введення даних з файла* – відкриття файла вихідних даних, занесення даних у таблицю і побудова графіків.

*Редагування даних* – реалізація можливості роботи з даними, а саме: додавання нових даних, зміна та видалення вже наявних даних, збереження відредагованих даних у файл та оновлення даних системи (заміна даних на відредаговані).



Рис. 1. Архітектура ІСППР

**Аналіз даних на наявність тренду** – виконання аналізу даних на наявність тренду за допомогою критеріїв серій та наявність одиничних коренів.

**Обчислення описових статистик** – обчислення таких характеристик даних, як кількість вимірів, максимальне і мінімальне значення вибірки, математичне очікування, стандартне відхилення, дисперсія, коефіцієнт асиметрії, ексцес, статистика Жак-Бера.

**Модифікація даних** – попередня обробка даних (нормування, логарифмування, цифрова фільтрація, заповнення пропусків).

**Підбір  $\alpha$**  – тобто автоматичний вибір параметра моделі експоненціального згладжування з кроком 0,05 за критерієм мінімального середнього квадрата похибки однокрокового прогнозу.

**Побудова автокореляційної функції (АКФ)** – побудова графіка АКФ і запис її числових значень у таблицю.

**Вибір порядку авторегресії (АР)** – визначення порядку авторегресійної складової (від 1 до 30).

**Вибір порядку ковзного середнього (КС)** – визначення порядку ковзного середнього (від 0 до 30).

**Оцінювання коефіцієнтів моделі авторегресії з ковзним середнім (АРКС)** – здійснюється за рекурсивним методом найменших квадратів (РМНК) за наявними даними і параметрами побудови моделі, що задаються.

**Оцінювання параметрів нелінійних моделей** (стосовно змінних) – метод Монте-Карло для марковських ланцюгів (МКМЛ).

**Вибір кількості періодів** – визначення кількості періодів, за якими обчислюватиметься ковзне середнє (від 3 до 7).

**Задавання кількості сусідніх точок** – визначення кількості точок, за якими буде знайдена найближча траєкторія (від 3 до 7).

**Задавання потужності тестової вибірки** – визначення розміру вибірки, яка буде використовуватися для перевірки моделі.

**Вибір горизонту прогнозування** – визначення кількості прогнозованих значень (3, 5 або 10).

**Оцінювання прогнозу** – обчислення прогнозних значень за знайденими або заданими параметрами і заданим горизонтом прогнозування.

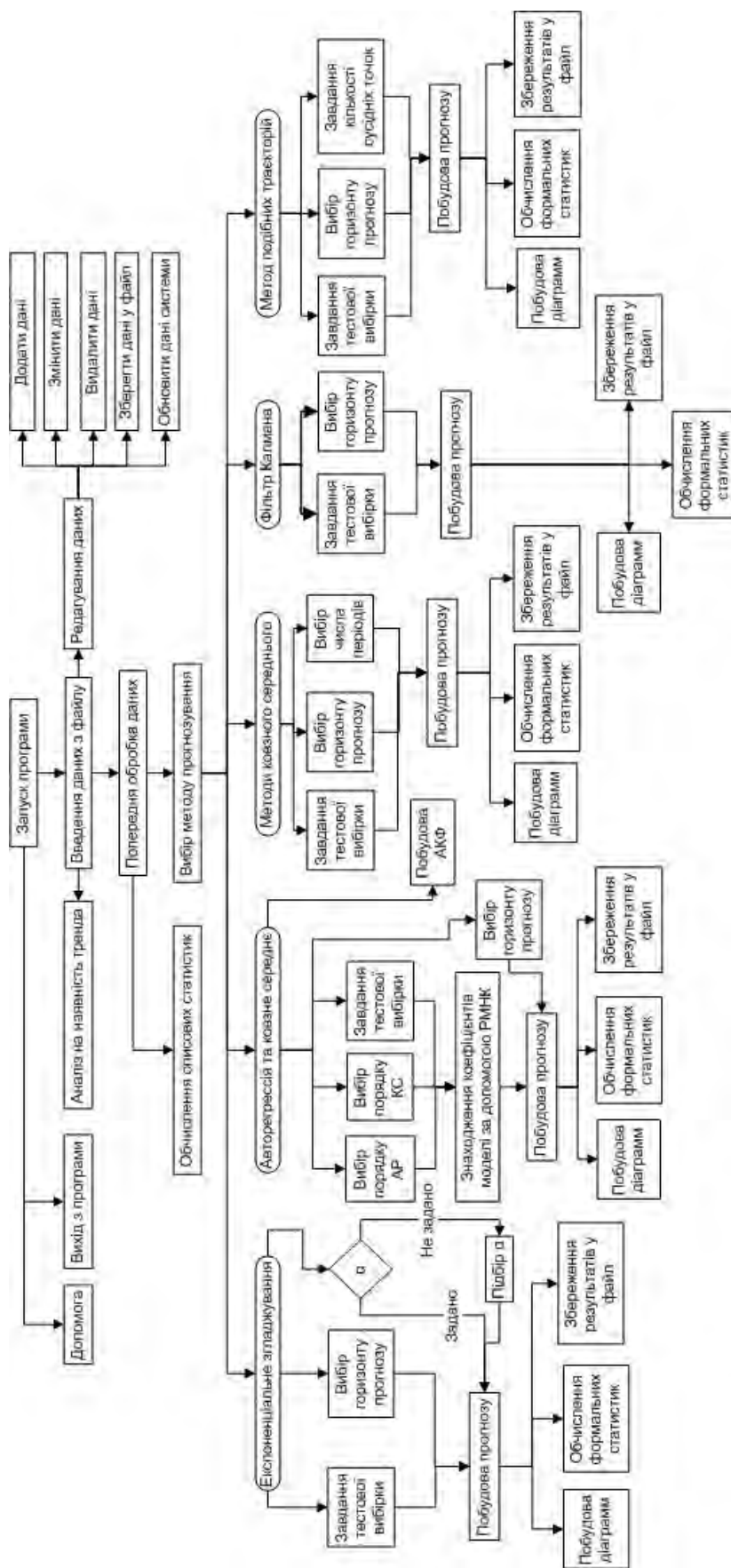


Рис. 2. Функціональна схема системи



**Побудова діаграм** – побудова графіків, що відображають початкові дані, результати прогнозування і похибки прогнозу.

**Збереження результатів у файл** – збереження результатів прогнозування (числових даних та діаграм) у файл Microsoft Office Excel.

**Обчислення формальних статистик** – обчислення критеріїв якості моделі (коефіцієнт детермінації, інформаційний критерій Акайке, статистика Дарбіна–Уотсона) і якості прогнозу (середня квадратична похибка, середня похибка, середня похибка у відсотках, середня абсолютна похибка у відсотках, сума квадратів похибок і коефіцієнт нерівності Тейла).

**Допомога** – коротке керівництво для користувача.

**Робота з програмою.** Зовнішній вигляд головного вікна програми з відкритим часовим рядом зображено на рис. 3. Для завантаження часового ряду треба вибрати файл, що містить необхідний часовий ряд. Файл “\*.txt” повинен містити по одній даті і по одному елементу ряду в кожному рядку. За необхідності попередньої обробки даних вибирають відповідний пункт “Нормування” для нормування ряду (рис. 4). Для прогнозування часового ряду слід вибрати один з пунктів “Статистичні методи” або “Інші методи” (рис. 5). Відкривається відповідне вікно: експоненціальне згладжування; авторегресія з ковзним середнім (АРКС); методи ковзного середнього; фільтр Калмана; метод подібних траєкторій і методи на основі байєсівського аналізу даних [9 – 11].

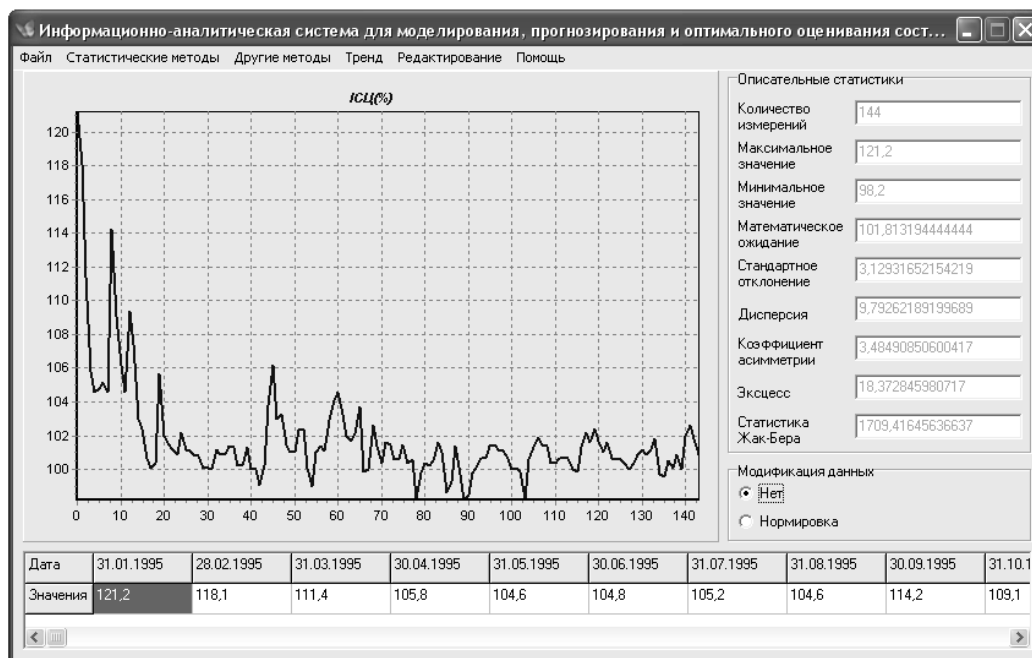


Рис. 3. Зовнішній вигляд програми з відкритим часовим рядом

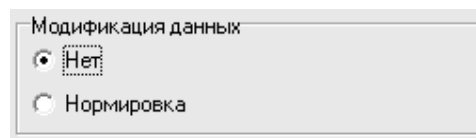


Рис. 4. Попередня обробка даних

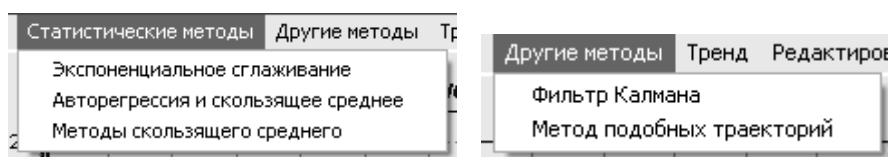


Рис. 5. Меню wyboru методів прогнозування

**Прогнозування процесів різної природи.** За допомогою створеної ІСПП досліджено макроекономічні показники, грошову масу та демографічну ситуацію у нашій країні. Для моделювання макроекономічних показників вибрано валовий внутрішній продукт (ВВП) та індекс споживчих цін (ІСЦ). Валовий внутрішній продукт – сума доданих вартостей, створених за певний період всіма виробниками на території країни (або вартість всіх кінцевих товарів або послуг, вироблених в економіці за певний проміжок часу). ІСЦ характеризує зміни у часі загального рівня цін на товари та послуги, які купує населення для невиробничого споживання. Він є показником зміни вартості фіксованого набору споживчих товарів та послуг у поточному періоді до його вартості у базисному періоді. ІСЦ – це найважливіший показник, який характеризує інфляційні процеси в економіці країни і використовується для вирішення багатьох питань державної політики, аналізу і прогнозу цінових процесів в економіці, перегляду розмірів грошових доходів та мінімальних соціальних гарантій населення, вирішення правових спорів, перерахунку показників системи національних рахунків у постійні ціни. Для виконання аналізу демографічної ситуації використано чисельність наявного населення. До наявного належить населення, яке перебуває на момент перепису на певній території, враховуючи тих, що проживають тимчасово. В результаті тестування встановлено, що усі згадані вище процеси містять тренд. Надалі побудуємо прогнози для кожного ряду даних за такими методами.

**Прогнозування за різницевиими рівняннями типу АРКС.** Як модель на основі різницевиими рівнянь вибрана авторегресійна модель (АР). Виконано серії експериментів із моделювання часових рядів, що містять інформацію про ВВП, ІСЦ і наявне населення України. Побудовано моделі різного порядку і виконано дослідження залежності точності прогнозування від порядку моделі. Для ВВП найкращою за статистичними показниками вибрано модель АР(5), для ІСЦ – АР(5) і для наявного населення України – АР(4). Наведемо рівняння побудованих моделей:

➤ ВВП:

$$\text{ВВП}(k)=0,7409193824*\text{ВВП}(k-1)+0,1522832846*\text{ВВП}(k-2)+0,09375652451* \text{ВВП}(k-3)+0,1176123754*\text{ВВП}(k-4)-0,07951539751*\text{ВВП}(k-5)$$

➤ ІСЦ:

$$\text{ІСЦ}(k)=24,20010469+0,6349467277*\text{ІСЦ}(k-1)-0,07223892767*\text{ІСЦ}(k-2)+0,07143236572*\text{ІСЦ}(k-3)+0,1780720866*\text{ІСЦ}(k-4)-0,05120774325*\text{ІСЦ}(k-5)$$

➤ Наявне населення України (ННУ):

$$\text{ННУ}(k)=1,588507937*\text{ННУ}(k-1)-0,3361088231*\text{ННУ}(k-2)-0,6126302522*\text{ННУ}(k-3)+0,3599296135*\text{ННУ}(k-4).$$

Побудуємо прогнози за зазначеними моделями для кожного ряду даних на десяти останніх точках вибірок. На рис. 6, 7 показано графіки значень ВВП і наявного населення України, а також їх прогнози значення на десять періодів.

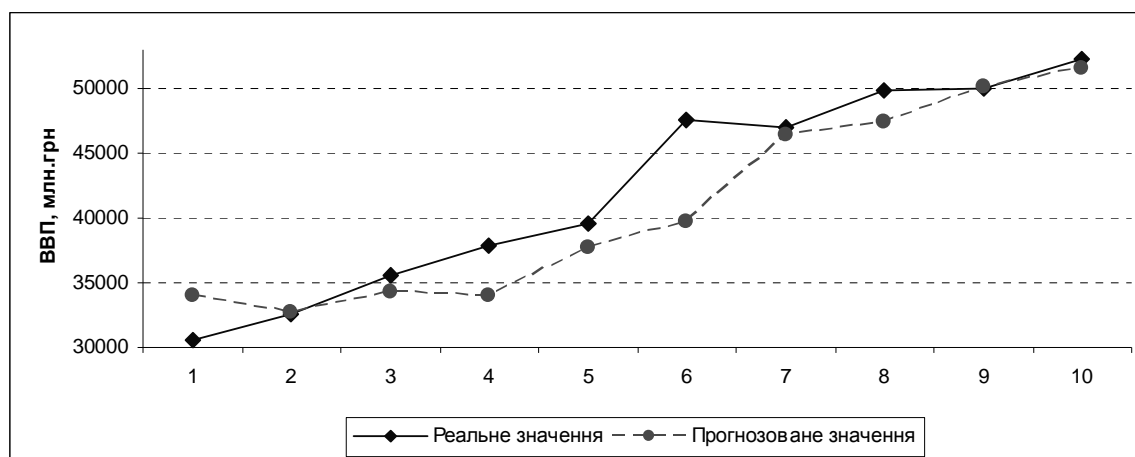


Рис. 6. Результати прогнозування ВВП

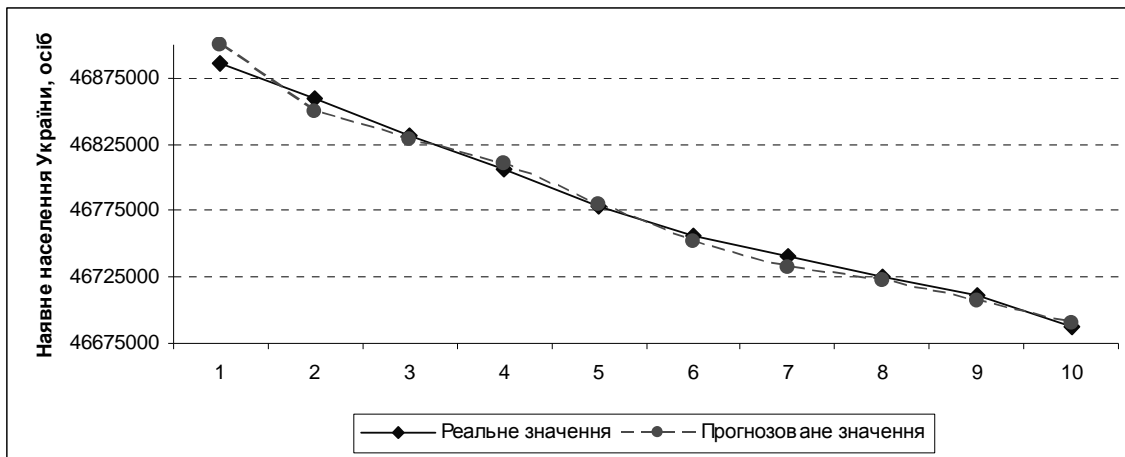


Рис. 7. Результати прогнозування наявного населення України

Також за допомогою розробленої програми виконано дослідження стосовно встановлення взаємозв'язку вибраних фінансово-економічних та соціальних показників.

**Прогнозування за фільтром Калмана.** Побудуємо для кожного ряду даних прогнози для десяти останніх значень вибірок; графіки фактичних і прогнозованих значень наведено на рис. 8, 9.

**Порівняльний аналіз побудованих моделей.** Перевірку отриманих моделей на адекватність виконано за значеннями статистичних характеристик, розглянутих вище. Характеристики адекватності побудованих моделей для часових рядів ВВП, ІСЦ і наявного населення України наведено у табл. 1 – 3 відповідно.

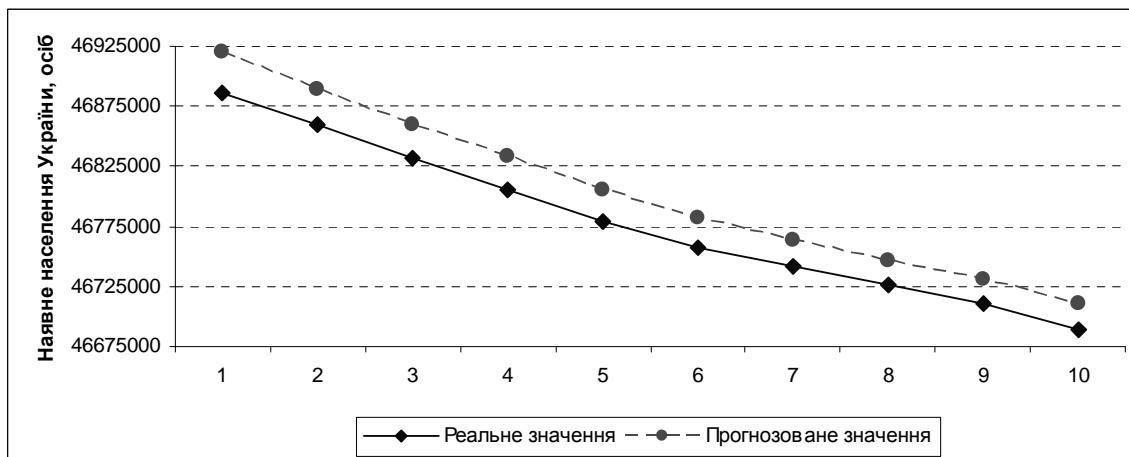


Рис. 8. Результати прогнозування наявного населення України

Таблиця 1

**Модель ВВП**

Показник	$R^2$	$SE$	$DW$	$U$
AP	0,826593593	98759541,12	1,238126047	0,03723826
EЗ	<b>0,995534464</b>	<b>2543240,989</b>	1,109246098	<b>0,005892091</b>
SMA	0,956105795	24998912	1,413651492	0,018636925
ЕМА	0,921965854	44442512,87	1,413651492	0,024956829
WMA	0,980491465	11110627,56	1,413651492	0,012370786
TMA	0,824423182	99995648	1,413651492	0,037757477
ФК	0,957263433	24339606,98	0,680009799	0,018385732
МПТ	0,62871582	211456173,8	0,199439752	0,055273709
МГУА	0,886644756	64558813,33	<b>1,512293054</b>	0,029721953
НМГУА	0,989657681	5890224,672	2,713155231	0,008924331

Таблиця 2

## Модель ІСЦ

Показник	$R^2$	$SE$	$DW$	$U$
АР	0,31916943	9,683033205	1,65285686	0,004877535
ЕЗ	<b>0,971976628</b>	<b>0,26815565</b>	1,630881556	<b>0,000812113</b>
SMA	0,669505695	3,1625	1,846640316	0,002788478
EMA	0,412454535	5,622222557	1,846640316	0,003717689
WMA	0,853113642	1,405555556	1,846640316	0,001859123
TMA	0,200543421	7,65	3,053594771	0,005575662
ФК	0,828420136	1,641847718	1,349777419	0,002009261
МПТ	0,188197851	7,768134766	1,065256827	0,009291999
МГУА	0,920060918	0,764937076	<b>2,173943823</b>	0,001371148
НМГУА	0,892549244	1,028196288	2,888909757	0,001590142

Таблиця 3

## Модель наявного населення України

Показник	$R^2$	$SE$	$DW$	$U$
АР	0,988536499	448175271,4	1,665935293	0,000071557
ЕЗ	0,994864899	200761128,7	0,03794962	0,000047890
SMA	0,958708967	1614307841	0,060106437	0,000135788
EMA	0,926593715	2869880776	0,060106437	0,000181043
WMA	0,98164843	717470151,3	0,060106437	0,000090529
TMA	0,834835868	6457231362	0,060106437	0,000271541
ФК	0,828039131	6722955520	0,006390825	0,000277067
МПТ	0,828448576	6706947941	0,083415473	0,000276859
МГУА	0,995623094	171118842,3	<b>2,183814099</b>	0,000044215
НМГУА	<b>0,999042371</b>	<b>37439333,73</b>	2,419789137	<b>0,000020682</b>

На підставі отриманих результатів аналізу адекватності побудованих моделей можна зробити такі висновки.

1. Модель високого рівня інформативності для всіх часових рядів одержано за експоненціальним згладжуванням:  $R^2 = 0,972 \div 0,997$ . Також високі характеристики отримано із застосуванням нечіткого методу групового урахування аргументів (НМГУА):  $R^2 = 0,892 \div 0,999$  і для фільтра Калмана  $R^2 = 0,828 \div 0,964$  (рис. 9). Найгірші показники отримано у разі застосування методу подібних траєкторій ( $R^2 = 0,188 \div 0,943$ ). Можна констатувати, що НМГУА здебільшого краще описує модель, ніж МГУА. Фільтр Калмана дає кращі результати, ніж методи ковзного середнього.

2. Можна упорядкувати всі методи за збільшенням коефіцієнта Тейла так (рис. 10): експоненціальне згладжування; нечіткий метод групового урахування аргументів; зважене ковзне середнє; фільтр Калмана; просте ковзне середнє; експоненціальне ковзне середнє; метод групового урахування аргументів; модель авторегресії; трикутне ковзне середнє; метод подібних траєкторій.

**Аналіз якості прогнозування.** Важливим моментом процесу прогнозування є об'єктивне визначення якості отриманого прогнозу за статистичними критеріями, наведеними нижче (табл. 4 – 6).

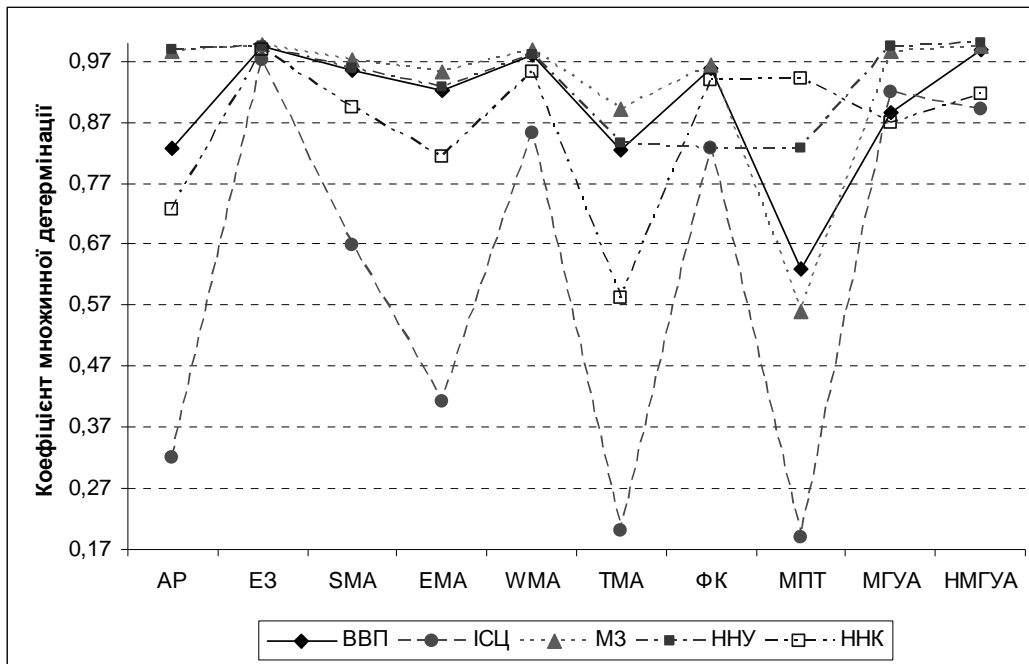


Рис. 9. Значення коефіцієнта множинної детермінації

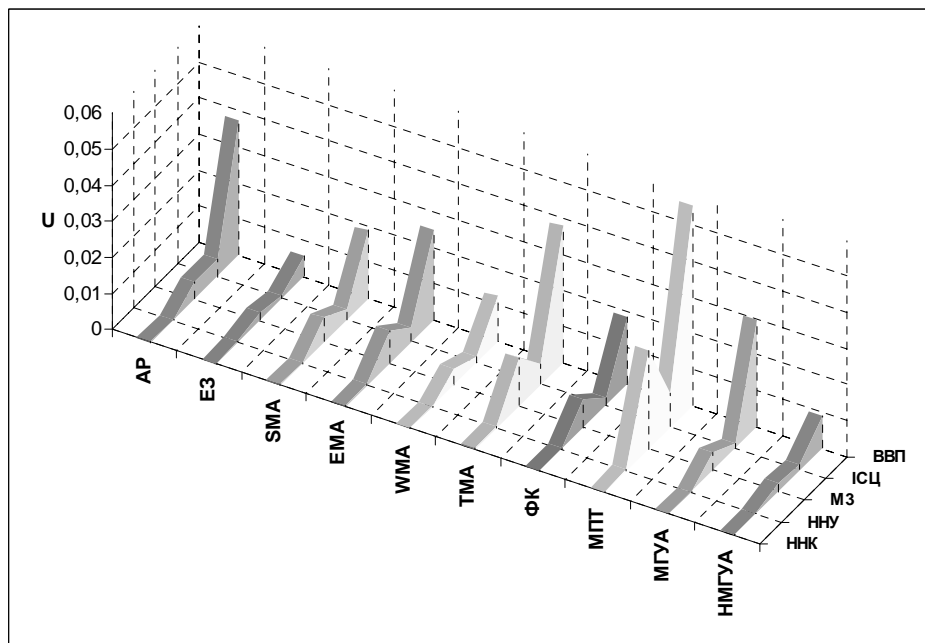


Рис. 10. Значення коефіцієнта нерівності Тейла

Таблиця 4

Аналіз точності прогнозування ВВП

Показник	<i>RSME</i>	<i>MAPE</i>	<i>ME</i>
AP	3142,603079	5,383522022	1489,257446
<b>E3</b>	<b>504,305561</b>	<b>0,938163236</b>	350,8504026
SMA	1581,104424	2,787741957	1123,4
EMA	2108,139295	3,716989387	1497,866711
WMA	1054,069616	1,858494638	748,9333333

Продовження табл. 4

1	2	3	4
<b>ТМА</b>	3162,208848	5,575483914	2246,8
<b>ФК</b>	1560,115604	3,14245718	1011,797626
<b>МПТ</b>	4598,436405	7,517585257	2305,55
<b>МГУА</b>	2540,842642	4,48227187	214,418784
<b>НМГУА</b>	767,4779913	1,433851672	<b>-66,1373368</b>

Таблиця 5

**Аналіз точності прогнозування ІСЦ**

Показник	<i>RSME</i>	<i>MAPE</i>	<i>ME</i>
<b>АР</b>	0,984024045	0,813557916	-0,11943
<b>ЕЗ</b>	<b>0,163754588</b>	<b>0,14118422</b>	<b>-0,011541997</b>
<b>SMA</b>	0,562361094	0,470905704	-0,045
<b>ЕМА</b>	0,749814814	0,62787429	-0,06
<b>WMA</b>	0,374907396	0,313937136	-0,03
<b>ТМА</b>	1,124722188	0,941811407	-0,09
<b>ФК</b>	0,405197201	0,349896649	-0,037326483
<b>МПТ</b>	1,889904563	1,755261675	-1,7640625
<b>МГУА</b>	0,276574958	0,24194038	-0,082284
<b>НМГУА</b>	0,320655	0,243713468	-0,022803682

Таблиця 6

**Аналіз точності прогнозування наявного населення України**

Показник	<i>RSME</i>	<i>MAPE</i>	<i>ME</i>
<b>АР</b>	6694,589393	0,011689812	1352,036706
<b>ЕЗ</b>	4480,637551	0,009148788	-4281,201547
<b>SMA</b>	12705,54147	0,025774964	-12061,5
<b>ЕМА</b>	16940,72247	0,034366620	-16082,0005
<b>WMA</b>	8470,36098	0,017183309	-8041
<b>ТМА</b>	25411,08294	0,051549928	-24123
<b>ФК</b>	25928,66275	0,054710077	-25597,6531
<b>МПТ</b>	25897,77585	0,042705635	15463,17813
<b>МГУА</b>	4136,65133	0,006254529	228,321855
<b>НМГУА</b>	<b>1934,924643</b>	<b>0,003754804</b>	<b>31,61441168</b>

Якщо максимальне значення середньої абсолютної похибки у процентах використати як критерій точності прогнозу, то можна упорядкувати усі методи за збільшенням критерію так, як показано на рис. 11.

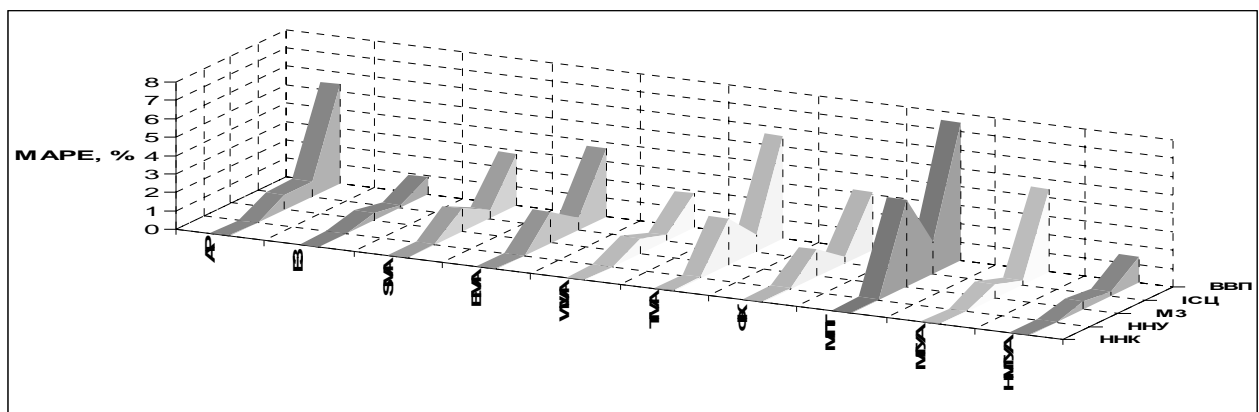


Рис. 11. Значення середньої абсолютної похибки, %

## Висновки

Розроблена і програмно реалізована інформаційна СППР для моделювання, прогнозування та оптимального оцінювання стану процесів різної природи. В системі реалізовано такі методи оцінювання параметрів прогнозувальних моделей: метод найменших квадратів, рекурсивний МНК, метод максимальної правдоподібності і метод Монте-Карло для марковських ланцюгів.

Для опису динаміки вибраних процесів використано такі моделі: модель авторегресії з ковзним середнім, експоненціальне згладжування, просте ковзне середнє, зважене ковзне середнє, експоненціальне ковзне середнє, трикутне ковзне середнє, моделі у просторі станів з фільтром Калмана, метод подібних траєкторій і деякі ймовірнісні моделі байєсівського типу.

Для виконання порівняльного аналізу додатково використано результати прогнозування за чітким та нечітким методами групового урахування аргументів. Для аналізу якості побудованих моделей застосовано такі критерії якості: коефіцієнт детермінації, сума квадратів похибок, коефіцієнт нерівності Тейла та статистика Дарбіна–Уотсона. Моделі найвищого ступеня адекватності отримано із застосуванням експоненціального згладжування, нечіткого методу групового урахування аргументів та зваженого ковзного середнього.

Для аналізу точності прогнозів використано такі критерії: стандартне відхилення залишків (середньоквадратична похибка), середня абсолютна похибка в процентах та середня похибка. Усі обчислені прогнози виявилися прийнятними за точністю; найкращі прогнози отримано з використанням експоненціального згладжування, нечіткого методу групового урахування аргументів, зваженого ковзного середнього та простого ковзного середнього. Дещо гірші за точністю прогнози одержано за допомогою фільтра Калмана, експоненціального ковзного середнього, МГУА, авторегресійної моделі, трикутного ковзного середнього та методу подібних траєкторій. За моделями ймовірнісного типу оцінено ймовірності попадання оцінок у вибрані інтервали.

1. Grunig R. *Successful decision-making* / R. Grunig, R. Kuhn . – Berlin: Springer, 2005. – 236 p.
2. Holsapple C.W. *Decision support systems* / C.W. Holsapple, A. B. Winston . – St. Paul (USA): West Publishing Company, 1996. – 850 p.
3. Figueira J. *Multiple criteria decision analysis* / J. Figueira, S. Greco, M. Ehrgott. – Berlin: Springer, 2005. – 1085 p.
4. Бідюк П.І. *Проектування систем підтримки прийняття рішень* / П.І. Бідюк, О.П. Гожий, Л.О. Коршевнюк. – Київ: Вид-во ЧДТУ ім. Петра Могили, 2012. – 380 с.
5. Згуровський М.З. *Системний аналіз* / М.З. Згуровський, Н.Д. Панкратова. – Київ: Наукова думка, 2011. – 727 с.
6. Бідюк П.І. *Аналіз часових рядів* / П.І. Бідюк, В.Д. Романенко, О.Л. Тимощук. – Київ: Політехніка, 2012. – 520 с.
7. Згуровський М.З. *Аналітичні методи калмановської фільтрації* / М.З. Згуровський, В.М. Подладчиков. – Київ: Наукова думка, 1995. – 293 с.
8. Згуровський М.З. *Аналіз та керування великими космічними конструкціями* / М.З. Згуровський, П.І. Бідюк. – Київ: Наукова думка, 1995. – 450 с.
9. Rossi P.E. *Bayesian statistics and marketing* / P.E. Rossi, G.M. Allenby, R. McCulloch. – New York: John Wiley & Sons, Ltd, 2005. – 348 p.
10. Robert C. *The Bayesian choice* / C. Robert. – Berlin: Springer, 2007. – 617 p.
11. *Probabilistic networks and expert systems* / [Cowell R.G., Dawid A.Ph., Lauritzen S.L., Spiegelhalter D.J.]. – Berlin: Springer, 1999. – 323 p.