

УДК 004.94

С.В. Голуб

*Черкаський національний університет імені Богдана
Хмельницького*

**ЗНИЖЕННЯ ВПЛИВУ ХАРАКТЕРУ РОЗПОДІЛУ ЗНАЧЕНЬ
ВХІДНИХ ДАНИХ НА ЯКІСТЬ МОДЕЛЕЙ ОБ'ЄКТІВ
СОЦІОЕКОЛОГІЧНОГО МОНІТОРИНГУ**

Представлені результати досліджень із підвищення якості результатів моделювання за рахунок підвищення інформативності масиву вхідних даних без проведення додаткових досліджень. Запропоновано новий перелік характеристик масиву вхідних даних, які можуть бути використані як додаткові параметри моделювання. Зниження критерію регулярності результатів моделювання доводить ефективність запропонованого методу.

Ключові слова: *інформативність масиву вхідних даних, перетворення інформації, соціоекологічний моніторинг.*

Представлены результаты исследований по повышению качества результатов моделирования за счет повышения информативности массива входных данных без проведения дополнительных исследований. Предложен новый перечень характеристик массива входных данных, которые могут быть использованы как дополнительные параметры моделирования. Снижение критерия регулярности результатов моделирования доказывает эффективность предложенного метода.

Ключевые слова: *информативность массива входных данных, преобразование информации, социоэкологический мониторинг.*

Research results to improve the quality of simulation results by increasing the information content of an array of input data without conducting further research were presented. A new list of characteristics of an array of input data that can be used as additional parameters of the simulation was proposed. Reduced regularity criterion of simulation results proves the effectiveness of the proposed method.

Keywords: *informative array of input data, transform information, Socio-Ecological Monitoring.*

Вступ. Основою ієрархічних систем багаторівневого перетворення інформації, які використовуються в технологіях моніторингу довкілля є моделі об'єктів [1]. Вони використовуються в якості локальних алгоритмів перетворення інформації (АПІ) у випадку, коли значення критерію їх якості не гірше наперед визначеного.

Структура підсистеми перетворення інформації представлена на рис. 1.

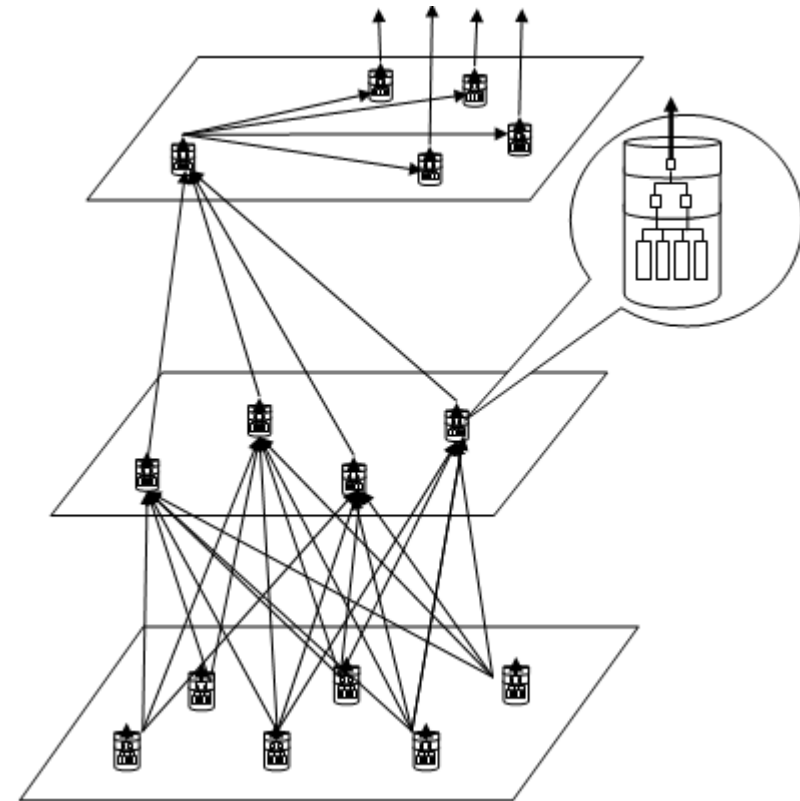


Рис. 1. Структура підсистеми перетворення інформації

Вона являє собою ієрархічне поєднання моделей об'єктів довкілля відповідних рівнів. Формування цієї структури відбувається за методом висхідного синтезу локальних алгоритмів перетворення інформації [2]. Результати моделювання об'єктів нижчого рівня подаються на вхід кожної моделі верхнього рівня за принципом «Всі до одного». Синтез моделей об'єктів верхнього рівня моніторингу відбувається шляхом вирішення задачі ідентифікації функціональної залежності моделюваної характеристики цього об'єкта від вихідних сигналів моделей попереднього рівня. Вирішення цієї задачі відбувається за багаторядним алгоритмом Методу групового врахування аргументів (МГУА) [3]

Технологічність процесу синтезу моделей забезпечується тим, що без значних переналаштувань, синтезатор (підсистема синтезу АПІ) створює моделі об'єктів для кожного з рівнів перетворення інформації. За умови зміни інформативності масиву вхідних даних якість моделей забезпечується достатньою різноманітністю синтезатора.

Під різноманітністю засобу синтезу моделі розуміється його здатність здобувати інформацію із масиву вхідних даних та адекватно відображати властивості об'єкта у своїй структурі за рахунок формування ефективних зв'язків між параметрами [1]. Ця властивість визначається потужністю методів синтезу моделей та конструктивними особливостями самого синтезатора.

Зміна властивостей вхідних даних приводить до зниження адекватності моделей та необхідності їх пересинтезування. Зниження адекватності моделі при фіксованій різноманітності засобів їх синтезу пов'язано із зниженням інформативності вхідних даних. Це може бути пов'язано із зміною стану об'єкта, який не відображено в базі даних або використанням характеристик об'єкта, які несуть суміщену інформацію [4].

Вважається [4], що для кожного алгоритму синтезу моделей існує мінімальний рівень інформативності вхідних даних, який дозволяє забезпечити адекватність цих моделей. Цей мінімальний рівень зветься межею інформаційної достатності (МІД) [4]. На даний час МІД оцінюється якісно за результатами випробування моделей. Установлено, що МІД вхідних даних для синтезу адекватних регресійних моделей, із застосуванням методу найменших квадратів, вищий, ніж при використанні МГУА або нейромереж для перетворення цих же даних. У [5] подані результати досліджень, які дозволяють оцінити МІД кількісно.

Але в будь-якому випадку при зниженні інформативності масиву вхідних даних достатньою умовою для забезпечення якості синтезованих моделей є здобування із існуючих даних додаткової інформації. Цього досягають застосуванням процедури обробки даних перед синтезом моделі.

Одним із ефективних способів підвищення інформативності масиву вхідних даних без проведення додаткових спостережень є використання характеристик самого масиву. Інформативність збільшується шляхом збільшення кількості параметрів, які несуть інформацію, яка раніше не використовувалась. Додаткові параметри отримують шляхом перетворення вигляду чисельних характеристик об'єктів масиву вхідних даних. Для цього застосовують побудову

взаємодій шляхом взаємного перемноження значень кількох параметрів масиву вхідних даних [6, с. 259]. До подібних заходів можна віднести також методику формування додаткових параметрів шляхом їх логарифмування та піднесення до степеня.

Постановка задачі. Метою даної роботи є підвищення інформативності масиву вхідних даних при зміні властивостей цього масиву, з метою синтезу адекватних моделей об'єктів в автоматизованих системах багаторівневого моніторингу.

Розв'язувалась задача відтворення функціональної залежності

$$Y = f(x_1; x_2; \dots, x_n), \quad (1)$$

$$S_x = \sqrt{\frac{\sum_{i=1}^m (y_i^c - y_i^{c*})^2}{m}} \rightarrow \min, \quad (2)$$

де $Y = \{y_1, y_2, \dots, y_m\}$ – вектор результатів моделювання; c – послідовність даних, які не використовувались в процесі створення моделей; $x_1; x_2; \dots, x_n$ – масив вхідних даних; n – кількість характеристик об'єкта моніторингу; m – кількість спостережень екзаменаційного масиву даних; y_i^c, y_i^{c*} – значення модельованого параметра, розраховане та отримане експериментально відповідно; S_x – цільова функція.

Необхідно визначити перелік характеристик масиву вхідних даних, які при їх використанні в процесі синтезу моделі забезпечили б значення цільової функції S_x менше, порівнянням із значенням, отриманим без використання цих додаткових параметрів.

Результати досліджень. Відомо, що характер розподілу значень параметрів вхідного масиву даних впливає на результати синтезу моделей. Критерії рангової кореляції рекомендується застосовувати [6] у випадку, коли розподіл значень характеристик об'єкта не відповідає нормальному закону. Застосування зовнішнього критерію якості у процесі селекції моделей за алгоритмами МГУА дозволяє знизити залежність якості синтезованих моделей від наявності нормального закону розподілу вхідних даних. Але, безумовно, залежність якості моделі від впливу неврахованих факторів залишається.

Використання коефіцієнтів рангової кореляції для здобування інформації про зв'язки між показниками масиву вхідних даних мало на меті використання додаткового ресурсу для покращення характеристик синтезованих моделей. Передбачалось, що оскільки поки що неможливо врахувати вплив невідомих факторів довкілля, які викликають «зашумленість» вхідних даних, застосування критеріїв

рангової кореляції дозволить знизити їх вплив на якість синтезованих моделей.

Була запропонована гіпотеза про те, що підвищити інформативність масиву вхідних даних можливо за рахунок використання додаткових параметрів, що несуть інформацію про зв'язки між характеристиками об'єкта моніторингу і їх порядок розрахунку не залежить від відповідності розподілу значень цих характеристик нормальному закону.

Для експериментальної перевірки цієї гіпотези проведено дослідження залежності результатів індуктивного моделювання залежності захворюваності населення Черкаської області на астму від концентрації техногенних забруднювачів у повітрі жилої зони. Дані представлені Черкаською обласною санітарно-епідеміологічною станцією.

Для синтезу моделей застосовувався багаторядний алгоритм МГУА. Результати моделювання порівнювались за критерієм регулярності [3] – визначалось значення середнього квадратичного відхилення результатів моделювання від дійсних значень на послідовності даних, які не використовувались у процесі синтезу моделей. Критерій регулярності розраховувався за виразом (2).

У таблиці подані значення критерію регулярності при використанні додаткових параметрів вхідного масиву даних, отриманих за коефіцієнтом кореляції Пірсона та рангових коефіцієнтів Спірмена та Кендала [6].

Результати досліджень дозволяють стверджувати, що застосування характеристик зв'язків між спостереженнями вхідного масиву даних дозволяють отримати суттєве (до 44 %) покращення якостей моделей при застосування кожного із додаткових параметрів масиву вхідних даних. Додаткові параметри на основі коефіцієнтів рангової кореляції дозволяють отримати більш якісні моделі. Порівняно із параметрами, отриманими на основі коефіцієнта Пірсона, використання рангових коефіцієнтів кореляції дозволяє покращити адекватність моделей на (8,79 – 10,01) %, точність моделі покращилась на (4,25–9,95) %.

Таблиця 1

Характеристики результатів моделювання

	Без застосування нового методу	За новим методом із використанням коефіцієнтів кореляції		
		Пірсона	Кендала	Спірмена
Критерій регулярності	15,48	10,19	8,64	8,83
Середня похибка, %.	28,73	18,70	17,48	15,84

Висновки. Таким чином знайшла підтвердження гіпотеза про ефективність застосування додаткових параметрів моделювання, порядок розрахунку яких не залежить від відповідності розподілу значень вхідних даних нормальному закону. Експериментально підтверджено ефективність нового методу підвищення інформативності масиву вхідних даних при зміні їх властивостей. Наступні дослідження доцільно проводити в напрямку створення методики оцінки інформативності вхідного масиву даних.

Бібліографічні посилання

1. **Голуб С.В.** Багаторівневе моделювання в технологіях моніторингу оточуючого середовища. / С.В. Голуб – Черкаси, 2007. – 220 с.
2. **Голуб С.В.** Застосування агрегатного підходу до моделювання структури інформаційних технологій соціо-екологічного моніторингу / С.В. Голуб // Вісник інженерної академії України. – 2007. –№ 3-4. – С. 93-97.
3. **Ивахненко А.Г.** Индуктивный метод самоорганизации моделей сложных систем. / А.Г. Ивахненко – Киев, 1981. 296 с.
4. **Голуб С.В.** Визначення структури інформаційних моделей в технологіях управління довкіллям / С.В. Голуб // Вісник інженерної академії України. – 2007. – № 2. – С. 35-42.
5. **Голуб С.В.** Визначення інформативності індуктивних моделей евристичних систем спостереження /С.В. Голуб // Вісник Східноукраїнського національного університету імені Володимира Даля. – 2007. – № 5 (111), – 1. – С. 265–269.
6. **Лапач С.Н.** Статистические методы в меедико-биологических исследованиях с использованием Excel. / С.Н. Лапач, А.В. Чубенко, П.Н. Бабич – 2-е из., переб. и доп.– К., 2001.– 408 с.

Надійшла до редколегії 05.10.10