

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
"КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"

На правах рукопису

КОВАЛЕНКО Ігор Іванович

УДК 681.323-523.8

**МЕТОДИ РОБАСТНОЇ ОБРОБКИ НЕОДНОРІДНИХ
ВИМІРЮВАЛЬНИХ ДАНИХ В СИСТЕМАХ УПРАВЛІННЯ**

05.13.04

Автоматизовані системи управління і системи
обробки інформації

АВТОРЕФЕРАТ ДИСЕРТАЦІЇ

на здобуття наукового ступеня доктора технічних наук

Київ - 1996



00752634 (R)

Роботу виконано на кафедрі матем.
 Національного технічного університету
 інститут”.

Науковий консультант:

академік НАН України, професор

Згуровський М.З.

Офіційні опоненти:

доктор технічних наук, професор

Барабаш Ю.Л.

доктор фізико-математичних наук

Андреев М.В.

доктор технічних наук

Данилов В.Я.

Провідна організація -

Харківський технічний університет
 радіоелектроніки.

Захист відбудеться “30” “вересня” 1996 р. о 15 годині
 на засіданні спеціалізованої Ради по присудженню наукового ступеня
 доктора технічних наук при Національному технічному університеті
 “Київський політехнічний інститут” Д 01.02.08

Адреса: 252056, м. Київ, пр. Перемоги, 37, ауд. 56 - 14 корп.

З дисертацією можна ознайомитися у бібліотечі НТУУ “КПІ”.

Автореферат розіслано “20” “серпня” 1996 р.

Учений секретар
 спеціалізованої Ради
 д.т.н., проф.

В.Д.Романенко

АНОТАЦІЯ

Мета роботи. Метою роботи є розробка аналітичного підходу, спрямованого на створення інформаційного забезпечення систем управління на основі робастної обробки неоднорідних вимірювальних даних, які характеризують складні фізичні системи (СФС) шляхом розробки моделей неоднорідних даних, методів їх класифікації та стискування, алгоритмів обробки для використання у автоматизованих системах обробки інформації, як складових частин систем управління.

Основні задачі дослідження. Для досягнення поставленої мети розв'язувались наступні задачі:

1. Аналіз відомих структур неоднорідних вимірювальних даних за їх геометричними ознаками та фізико-хімічними властивостями і подальша їх систематизація на ряд класів.
2. Розробка комплексу статистичних моделей, що описують визначені класи структур неоднорідних вимірювальних даних.
3. Теоретичне обґрунтування методів робастної статистики з метою визначення їх можливостей для вирішення задач класифікації неоднорідних вимірювальних даних. Проведення чисельного аналізу робастних оцінок на незсуєність з метою їх використання для оцінювання параметрів моделей неоднорідних даних.
4. Розробка методу класифікації та стискування неоднорідних даних, в якому визначено розв'язувальну функцію і комплекс розв'язувальних правил на основі основних показників робастності.
5. Розробка методик обробки неоднорідних вимірювальних даних з використанням виважених порядкових статистик, адаптивних та рекурентних процедур робастного оцінювання, перешкодостійкого згладження.
6. Розв'язання завдання синтезу структури автоматизованої системи обробки неоднорідних вимірювальних даних та інформаційного забезпечення для прийняття рішень в СФС.
7. Розробка алгоритмів і прикладних програм для реалізації методик обробки неоднорідних вимірювальних даних (аналіз даних радіаційного моніторингу водного середовища, обробка зображень, аналіз даних, одержаних при випробуваннях механічних пристроїв та ін.).

Автор захищає проблему дослідження якісних та прикладних задач інформаційного забезпечення управління СФС на основі робастної обробки неоднорідних вимірювальних даних та вносить на захист:

1. Принципи побудови моделей неоднорідних даних, зображених симетричними (несиметричними), двох- та багатокomпонентними сумішами скупчення розподілу ймовірностей.

ЛНБ ім. В. Стефаника
АН України

2. Теоретичне обґрунтування вибору чотирьох класів робастних оцінок на базі порядкових та виважених порядкових статистик, які широко використовуються на практиці.
3. Метод класифікації та стискування неоднорідних даних на базі основних показників робастності: функцій чутливості і порогових течок.
4. Метод формування комплексу розв'язувальних правил класифікації неоднорідних даних на основі перетворення початкових масивів даних у варіаційні ряди з подальшим їх розчленуванням за призначеними рівнями. "зрізання".
5. Методику робастної обробки неоднорідних даних з використанням виважених порядкових статистик, ("зрізаного" середнього, вибіркової медіани, вінзоризованого середнього та ін.).
6. Методику обробки неоднорідних даних адаптивними процедурами робастного оцінювання з використанням стабільних аналогів коефіцієнтів асиметрії і ексцесу.
7. Методику короткочасного прогнозування часових рядів на основі рекурентних процедур робастного оцінювання.
8. Методологічні основи синтезу структур автоматизованих систем обробки неоднорідних вимірювальних даних та інформаційного забезпечення для прийняття рішень в управлінні СФС. Математичне і програмне забезпечення робастної обробки даних при вирішенні деяких прикладних задач.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність проблеми. Складні фізичні системи (СФС) характеризуються великою кількістю різномірних підсистем, в основі яких лежать фізико-хімічні, біологічні, інформаційні процеси та взаємозв'язки між ними. Прикладом СФС може бути сучасний технологічний комплекс, що включає велику кількість взаємопов'язаних технологічних процесів, взаємодіючих із навколишнім середовищем. Ефективним засобом дослідження вказаних процесів і систем є системний аналіз, як комплекс математичних, методологічних і організаційних засобів, призначених для прийняття рішень в СФС при наявності інформації різної природи. На кафедрі математичних методів системного аналізу (ММСА) НТУУ "КПІ" школою М.З.Згуровського розробляється науковий напрям "системний аналіз у дослідженнях складних фізичних процесів і полів".

У попередніх роботах кафедри розроблено теорію системного аналізу стохастичних розподілень процесів (М.З. Згуровський, О.М.Новіков), розглянуто питання дослідження складних фізичних процесів і полів на прикладах еколого-економічних систем з позицій системного аналізу (М.З.Згуровський, О.М.Демченко, О.М.Новіков, І.І.Коваленко).

Одним із завдань вказаного наукового напрямку є розробка засобів формування інформаційного забезпечення управління СФС на основі обробки вимірювальних даних. Сучасний етап у розвитку методів обробки даних характеризується їх великою інформаційною ємністю, відносно високою швидкістю їх надходження на вхід автоматизованих систем обробки інформації (АСОІ).

Масиви даних, як правило, не систематизовані, мають велику надмірність і неоднорідність, що не дозволяє розглядати їх в якості емпіричних моделей об'єктів (систем) управління і будувати шукані залежності. Звідси можна зробити висновок, що інформаційне забезпечення, яке необхідне для прийняття рішень, і формується у АСОІ за допомогою обробки даних, повинно включати засоби класифікації та стискування початкової інформації з метою визначення найбільш інформативних параметрів (характеристик) об'єктів, що досліджуються, це і створює основу для формування керуючих дій.

Особливий інтерес у формуванні інформаційного забезпечення систем управління викликає обробка неоднорідних вимірювальних даних, які характеризуються термінами, як "аномалія", "пляма", "помилка" та інші, а їх фізико-хімічна інтерпретація дуже різноманітна.

Такі дані описуються моделями сумішей ймовірних розподілів, а для їх класифікації (розщеплення) використовуються методи, що базуються на байєсівській ідеології: метод максимальної правдоподібності, метод моментів та інші. Але ці методики мають ряд недоліків, які створюють труднощі при вирішенні задач класифікації. Наприклад, підхід, заснований на методі максимальної правдоподібності статистичного оцінювання параметрів суміші розподілів (крім необхідності "вгадати" загальний параметричний вид розподілу, що задає кожний із класів) веде до обмеження функції правдоподібності високої складності і трудоемкості процесу обчислювальної реалізації та повільної збіжності ітераційних алгоритмів.

Головними недоліками підходу, що базується на методі моментів статистичного оцінювання параметрів суміші розподілів, є громіздкість його обчислювальної реалізації (особливо у разі високих розмірностей розподілів, що аналізуються і великого числа класів) та відносно невисока якість статистичних властивостей оцінок. Крім того, треба відзначити, що ефективність зазначених методів істотно залежить від вибора їх "початкової позиції", коли кожному із таких алгоритмів потрібен етап розвідувального аналізу, який дозволяє сформулювати робочі гіпотези про число класів, тип ймовірних розподілів усередині кожного із класів, величини апріорних ймовірностей і т.п.

В цих умовах уваги заслуговують інші методичні підходи і зокрема підхід оцінювання параметрів сумішей розподілів та послідовної класифікації на основі робастної обробки неоднорідних вимірювальних даних, що і визначає актуальність досліджень, виконаних у даній роботі.

Методи дослідження. Для розв'язання поставлених задач в дисертації використані методи прикладної статистики (класифікація та зниження розмірності), робастної статистики, сучасної теорії управління та системного аналізу, методи статистичного моделювання.

Вірогідність поданих у дисертації результатів теоретично обґрунтована строгими математичними викладами і підтверджується погодженістю, теоретичних результатів з результатами обчислювальних експериментів.

Наукова новизна роботи визначається новизною методологічного підходу до розв'язання теоретичних та прикладних задач класифікації та стискування неоднорідних вимірювальних даних, що будується на їх робастній обробці.

Запропоновано новий підхід до вирішення задач класифікації та стискування неоднорідних даних за допомогою розщеплення сумішей ймовірних розподілів, який на відміну від вищеперелічених методів дозволяє працювати із повнорозмірними околицями параметричних моделей засобами робастної статистики. Механізм впливу процедур робастного оцінювання на неоднорідні ("забруднені") дані складається в їх підчищенні та підгонці до параметричних моделей, що безпосередньо веде до задач класифікації.

Розроблено комплекс моделей неоднорідних вимірювальних даних, поданих симетричними (несиметричними), двох- та багатокomпонентними сумішами щільностей розподілу ймовірностей, які характеризують різноманітні розподіли даних.

Виконано теоретичне обґрунтування методів робастної статистики, в результаті якого була доведена можливість їх використання для вирішення задач класифікації неоднорідних даних. Отримано нові результати дослідження робастних оцінок на їх незсуеність.

Вперше запропоновано розв'язувальні функції класифікації, що базуються на основних показниках робастності: функції чутливості та порогові точки.

Розроблено метод побудови рішачючих правил класифікації із використанням різних робастних оцінок, в основі якого лежать процедури перетворення варіаційних рядів з урахуванням критеріїв інформативності даних.

Практична цінність теоретичних результатів і проведених експериментальних досліджень полягає:

- в розробці математичного і програмного забезпечення обробки неоднорідних даних різними типами процедур робастного оцінювання: виваженими порядковими статистиками, адаптивними та рекурентними оцінками для вирішення ряду прикладних задач;
- у створенні інженерної методики перевірки вибірки значень часових рядів на однорідність (неоднорідність) із використанням непараметричних критеріїв перевірки гіпотез;

- у створенні методичних основ системного аналізу СФС та принципів синтезу структур автоматизованих систем обробки неоднорідних вимірювальних даних та інформаційного забезпечення для прийняття рішень в СФС.

Реалізація результатів роботи. Основні результати дисертації використані:

- при написанні учбового посібника для студентів вузів України;
- в учбовому процесі кафедри математичних методів системного аналізу НТУУ "КПІ";
- при створенні методики та пакету прикладних програм аналізу даних про поля радіаційного забруднення водного середовища, які були використані в КБ "Шторм" НТУУ "КПІ";
- при створенні програмного модуля для проведення аналізу вимірювальних даних, одержаних при виконанні доводочно-випробувальних робіт механічних вузлів спеціального призначення в Конструкторському бюро артилерії Мінмашпрома України;
- при створенні пакету прикладних програм для проведення аналізу вимірювальних даних, одержаних при конструюванні супінаторів стопи, який використовується у Науково-виробничому об'єднанні "ІРВУС-Ф" (м. Київ).

Акти впровадження із зазначенням техніко-економічного ефекту наведені у Додатках до дисертації.

Загальний економічний ефект від впровадження перелічених програмних розробок складає 266,0 млн.крб. на рік (за станом на 1995 р.).

Апробація роботи. Основні положення дисертаційної роботи доповідались і обговорювались на:

- регіональному науково-технічному семінарі "Автоматизированные системы управления технологическими комплексами на базе микропроцессоров, микро- и мини ЭВМ" (м. Новочеркаськ, 1983);
- II всесоюзній конференції "Перспективы и опыт внедрения статистических методов в АСУ ТП" (м. Москва, 1984);
- республіканській конференції "Робототехнические системы для программных технологических процессов" (м. Ворошиловград, 1985);
- всесоюзному семінарі "Программное обеспечение гибких автоматизированных систем" (м. Калінін, 1986);
- всесоюзній науково-технічній конференції "Актуальные проблемы моделирования и управления системами с распределенными параметрами" (м. Одеса, 1987);
- III польсько-радянській науково-технічній конференції "Комплексная автоматизация промышленности" (м. Врошлав, 1988);
- IV міжнародній конференції "Проблемы комплексной автоматизации" (м. Київ, 1990);
- республіканському семінарі "Ідентифікація і управління нестационарними розподільними процесами" (м. Київ, 1995, 1996);

- республіканському семінарі Інституту кібернетики Національної Академії Наук (м. Київ, 1996);
- міжнародній науково-технічній школі-семінарі "Альтернативные подходы к исследованию систем (техника, экология, экономика)" (м. Київ, 1996).

Публікації. Результати дисертаційної роботи опубліковані у 41 науковій роботі автора, в тому числі - 1 підручник для студентів вузів, 1 препринт, 39 статей, публікацій та звітів з НДР.

Структура і об'єм роботи. Робота складається із вступу, 7 глав, висновку та додатку з актами впровадження і містить _____ 282 _____ сторінок тексту, 55 _____ малюнків, 13 _____ таблиць, 16 _____ сторінок додатків.

ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми, сформувано мету роботи, зроблено її загальну характеристику.

В першій главі "Неоднорідні вимірювальні дані і проблеми їх аналізу" з позицій інформаційної теорії вимірювань показано, що під неоднорідними сукупностями даних розуміються такі, які формуються під впливом різних причин і умов, і їх закони розподілів мають складну структуру (наприклад, крива скупчення розподілу ймовірностей таких даних може бути багатомодальною) або вони можуть бути зображені у вигляді об'єднання деякого числа однорідних сукупностей із більш простою структурою законів розподілу.

Звідси було формалізовано критерії неоднорідності вимірювальних даних із застосуванням таких статистичних характеристик, як скупчення розподілу ймовірностей $F(x)$ і параметр положення μ .

Таким чином, деяка сукупність даних $X = \{x_1, x_2, \dots, x_n\} \Rightarrow X_{het}$ (het - скорочення від англійського слова "heterogeneous" - неоднорідний) є неоднорідною, якщо виконуються наступні умови:

$$\begin{aligned} F(x) &= \{f_1(x), f_2(x), \dots, f_n(x)\} \supset (f_{mod_1}(x), \dots, f_{mod_m}(x)) \\ |\mu(x) - \hat{\mu}(x)| &= \theta, \theta \neq 0, \end{aligned} \quad (1)$$

або

$$\begin{aligned} F(x) &= \{f_1(x), f_2(x), \dots, f_n(x)\} \supset f_{mod}(x) \\ |\mu(x) - \hat{\mu}(x)| &= 0 \end{aligned} \quad (2)$$

Тут: $f_i(x)$ - значення функції $F(x)$; $f_{\text{mod}}(x)$ - модальні значення цієї ж функції; μ і $\hat{\mu}$ - відповідно теоретичне та оцінне значення параметру положення (наприклад, математичне очікування і середнє арифметичне) функції $F(x)$; θ - параметр зсуву між значеннями μ і $\hat{\mu}$, який характеризує зсування оцінок.

Розглянуті критерії (1), (2) приводять до сумішей розподілів, що виражаються:

$$F(x) = \sum_{i=1}^k \pi_i f_i(x; \mu_i), \quad (3)$$

де $f_i(x; \mu)$, $F(x)$ - скупчення (у безперервному випадку), або полігона скупчень (у дискретному випадку) відповідно i -ої компоненти суміші та результуючого закону розподілу; π_i - апріорна ймовірність появи у випадковій вибірці спостережень із законом розподілу $f_i(x; \mu_i)$; k - число компонент суміші; μ_i - параметр положення (вибіркове середнє).

Суміші розподілів (3) можуть бути при цьому несиметричними і симетричними, двох- та багатокомпонентними. Вираз для несиметричної суміші має вигляд:

$$F(x) = \varepsilon_1 f_1(x; \mu_1, \sigma_1) + \varepsilon_2 f_2(x; \mu_2, \sigma_2, \theta_1) + \dots + \varepsilon_l f_l(x; \mu_l, \sigma_l, \theta_j) \quad (4)$$

$\varepsilon = \sum_{i=1}^k \varepsilon_i = 1$ - питома вага значень виборок, за якою побудований кожний

із розподілів f_i , θ - параметр, що визначає величину зсування параметрів положень μ відносно один одного, σ_i - середнє квадратне відхилення.

Симетричну функцію можна записати у такий спосіб:

$$F(x) = \varepsilon_1 f_1(x; \mu_1, \sigma_1) + \varepsilon_2 f_2(x; \mu_2, \sigma_2) + \dots + \varepsilon_l f_l(x; \mu_l) \quad (5)$$

Далі в главі дається коротке ознайомлення з основними положеннями теорії класифікації та її особливостями.

Особлива увага приділена методам класифікації за допомогою процедур розщеплення сумішей ймовірних розподілів.

В загальному виді задача розщеплення сумішей розподілів формується у такий спосіб. За вибіркою спостережень x_1, x_2, \dots, x_n , яка є сумішшю розподілів, необхідно побудувати статистичні оцінки для числа компонентів суміші n , їх питомої ваги (апріорних ймовірностей) p_1, p_2, \dots, p_n та, саме головне для кожного із компонентів f_i , суміші, що аналізуються. В деяких

випадках наявні апріорні відомості дають дослідникові точне значення числа компонентів суміші n , а інколи і апріорних ймовірностей p_1, p_2, \dots, p_n . Тоді задача розщеплення суміші зводиться лише до оцінювання функцій $f_i, i = \overline{1, n}$.

Задачі розщеплення сумішей розподілів вирішуються методами, розробленими в рамках параметричного підходу, який потребує точного опису функцій f_i конкуруючих генеральних сукупностей, за якими по черзі підраховується значення функції правдоподібності для кожного x_i в рамках кожної генеральної сукупності, що розглядається, (тобто підраховують значення $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$) і відносять x_i до того класу, функція правдоподібності якого максимальна. До числа основних вказаних методів відносяться процедури, що базуються на методі максимальної правдоподібності, методи моментів, методи адаптивного ймовірного навчання та інші. Відзначено ряд недоліків перелічених методів, а саме: необхідність пред'являти вимоги обмеження до функцій правдоподібності, невелика якість статичних властивостей отримуваних оцінок, висока складність обчислювальних процедур. Все це визначило необхідність розвитку інших підходів до рішення задач розщеплення сумішей ймовірних розподілів.

В другій главі "Структури неоднорідних вимірних даних та їх зображення статистичними моделями" присвячена розробці комплексу математичних моделей неоднорідних даних.

Перш за все було виділено три основних типи структур неоднорідних даних за їх геометричними ознаками:

- G_I - сегментовані структури, які зображені непересіченими областями різної форми;
- G_{II} - структури, які характеризуються наявністю локальних областей, вкладених в загальну площину;
- G_{III} - комбіновані структури, зображені сегментами з вкладеними в них локальними областями.

Кожну таку структуру в оцифрованому виді можна зобразити слідуючою матрицею:

$$G = \begin{bmatrix} g_{11} & g_{12} & \dots & g_{1n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ g_{m1} & g_{m2} & \dots & g_{mn} \end{bmatrix} = [g_{ij}], \quad (6)$$

де g_{ij} - числові значення, наприклад, динамічного діапазону градацій яскравості зображення; $i = \overline{1, m}$; $j = \overline{1, n}$; $m = n$; $N = m \cdot n$. Рядки і стовпці матриці (6) відповідають просторовим координатам.

Сегментовані структури даних, які складаються із I сегментів, описані наступним чином:

$$\begin{aligned} G &= \{G_1, G_2, \dots, G_I\}; \quad G_1 = \{g_1^{(1)}, g_2^{(1)}, \dots, g_n^{(1)}\}, \\ G_2 &= \{g_1^{(2)}, g_2^{(2)}, \dots, g_n^{(2)}\}, \dots, G_I = \{g_1^{(I)}, g_2^{(I)}, \dots, g_n^{(I)}\} \\ G_1 \cap (G_2, \dots, G_I) &= \emptyset; \quad G_2 \cap (G_1, \dots, G_I) = \emptyset, \\ G_I \cap (G_1, G_2, \dots, G_{I-1}) &= \emptyset. \end{aligned} \quad (7)$$

Модель такої структури описано несиметричною сумішню розподілів, кожна складова якої характеризується асимптотичним нормальним законом:

$$\begin{aligned} F_G(g) &= \varepsilon_1 f_{G_1}(g_1, \mu_1, \sigma_1) + \varepsilon_2 f_{G_2}(g_2, \mu_2, \sigma_2, \theta_1) + \dots + \varepsilon_I f_{G_I}(g_I, \mu_I, \sigma_I, \theta_p) \\ \mu_1 &\neq \mu_2 \neq \dots \neq \mu_I; \quad \sigma_1 \neq \sigma_2 \neq \dots \neq \sigma_I; \\ \mu_2 &= \mu_1 \pm \theta_1; \quad \mu_3 = \mu_2 \pm \theta_2, \dots, \mu_{I-1} \pm \theta_{p-1}; \end{aligned} \quad (8)$$

$\varepsilon = \sum_{i=1}^I \varepsilon_i = 1$ - рівень неоднорідності, тобто питома вага значень $g^{(i)}$, що складають площину кожного сегменту G_i .

Структури даних G_{II} , що характеризуються вкладеними локальними областями, описано наступним чином:

$$\begin{aligned} G &= \{G_0, G_1, G_2, \dots, G_I\}; \quad G_0 = \{g_1^{(0)}, g_2^{(0)}, \dots, g_n^{(0)}\}, \\ G_1 &= \{g_1^{(1)}, g_2^{(1)}, \dots, g_n^{(1)}\}, \quad G_2 = \{g_1^{(2)}, g_2^{(2)}, \dots, g_n^{(2)}\}, \dots, G_I = \{g_1^{(I)}, g_2^{(I)}, \dots, g_n^{(I)}\} \\ G_1 \cap (G_2, \dots, G_I) &= \emptyset; \quad G_2 \cap (G_1, \dots, G_I) = \emptyset, \dots, G_I \cap (G_1, G_2, \dots, G_{I-1}) = \emptyset. \end{aligned} \quad (9)$$

$S_{G_0} \ll S_{G_i}$; $\sum_{i=1}^I S_{G_i} < S_{G_0}$ відповідно площі локальних областей та загального фону зображення.

$$1 = \sum_{i=1}^I \sum_{j=1}^n \dots$$

Моделі таких структур даних в залежності від значень параметрів μ та σ описуються двох- та багатокомпонентними симетричними (несиметричними) сумішами ймовірних розподілів:

$$F_G(g) = \varepsilon_0 f_{G_0}(g_0, \mu_0, \sigma_0) + \varepsilon_1 f_{G_1}(g_1, \mu_1, \sigma_1, \theta_1) + \dots + \varepsilon_p f_{G_p}(g_p, \mu_p, \sigma_p, \theta_p) \quad (10)$$

$$\mu_0 \neq \mu_1 \neq \dots \neq \mu_p; \sigma_0 \neq \sigma_1 \neq \dots \neq \sigma_p; l = q + r;$$

$$\varepsilon = \varepsilon_0 + \varepsilon_1 + \dots + \varepsilon_p; \sum_{i=1}^p \varepsilon_i \leq 0,5.$$

Комбіновані структури даних описано таким чином:

$$\begin{aligned} G &= \{G_1, G_2, \dots, G_l\}; G_i = \{G_i^{(1)}, G_i^{(2)}, \dots, G_i^{(k)}\}, \\ G_2 &= \{G_2^{(2)}, G_2^{(2)}, \dots, G_2^{(2)}\}, \dots, G_l = \{G_l^{(l)}, G_l^{(l)}, \dots, G_l^{(l)}\}, \\ G_1^{(1)} &= \{g_1^{(1)}, g_2^{(1)}, \dots, g_p^{(1)}\}, G_1^{(2)} = \{g_1^{(2)}, g_2^{(2)}, \dots, g_k^{(2)}\}, \dots \\ \dots, G_k^{(1)} &= \{g_1^{(k)}, g_2^{(k)}, \dots, g_p^{(k)}\}, \dots, G_l^{(l)} = \{g_1^{(l)}, g_2^{(l)}, \dots, g_k^{(l)}\}, \\ G_1 \cap (G_2, \dots, G_l) &= \emptyset, G_2 \cap (G_1, \dots, G_l) = \emptyset, \dots, G_l \cap (G_1, G_2, \dots, G_l) = \emptyset. \\ G_1^{(1)} \cap (G_2^{(2)}, \dots, G_k^{(k)}) &= \emptyset, \dots, G_2^{(2)} \cap (G_1^{(1)}, \dots, G_k^{(k)}) = \emptyset, \dots, G_l^{(l)} \cap (G_2^{(2)}, \dots, G_k^{(k)}) \end{aligned} \quad (11)$$

Модель такої структури має наступний вигляд:

$$\begin{aligned} F_G(g) &= \varepsilon_1 F_{G_1}(g_1, \mu_1, \sigma_1) + \varepsilon_2 F_{G_2}(g_2, \mu_2, \sigma_2, \theta_2) + \dots + \varepsilon_l F_{G_l}(g_l, \mu_l, \sigma_l, \theta_l) \\ F_{G_1} &= \varepsilon_1^{(1)} f_{G_1^{(1)}}(g_1^{(1)}, \mu_1^{(1)}, \sigma_1^{(1)}) + (1 - \varepsilon_1^{(1)}) f_{G_1}(g_1, \mu_1^{(1)}, \sigma_1^{(1)}) \\ F_{G_l} &= \varepsilon_l^{(l)} f_{G_l^{(l)}}(g_l^{(l)}, \mu_l^{(l)}, \sigma_l^{(l)}) + (1 - \varepsilon_l^{(l)}) f_{G_l}(g_l, \mu_l^{(l)}, \sigma_l^{(l)}) \end{aligned} \quad (12)$$

Тут:

$$\mu_1 = \mu_1^{(0)} = \dots = \mu_1^{(l)}; \sigma_1 \neq \sigma_1^{(0)} \neq \dots \neq \sigma_1^{(l)};$$

$$\mu_l = \mu_l^{(0)} = \dots = \mu_l^{(l)}; \sigma_l \neq \sigma_l^{(0)} \neq \dots \neq \sigma_l^{(l)}$$

$$\varepsilon = \sum_{j=1}^l \sum_{i=1}^k \varepsilon_{ij} = 1$$

Розглянуті моделі дозволили намітити подальший хід розробок та досліджень по створенню методів класифікації неоднорідних даних.

Третя глава "Вибір і обґрунтування математичного апарату для аналізу (класифікації та стискування) неоднорідних вимірювальних даних. Дослідження робастних оцінок" присвячена теоретичному узагальненню методів робастної статистики з метою обґрунтування можливості їх використання для оцінювання параметрів сумішей ймовірних розподілів та реалізації схеми "від оцінювання до класифікації".

Як вже відмічалось, основним призначенням АСОІ є одержання інформації, призначеної для прийняття рішень про належність невідомого об'єкту (явища) до того, чи іншого класу.

Математичне забезпечення таких систем визначається алгоритмами побудови розв'язувальних правил прийняття рішень, математичними моделями об'єктів, що класифікуються, методами та алгоритмами обробки вимірювальної інформації. При цьому базою для створення перелічених процедур стали класичні результати теорії статистичних рішень (метод Байєса та інші), які з успіхом працюють в рамках строгих параметричних моделей.

Проте на практиці досить часто дані, що аналізуються не можуть бути описані строгими параметричними моделями внаслідок їх неоднорідної природи. Такі дані описуються моделями ймовірних розподілів, які лежать в деякій околиці параметричних моделей і зображаються сумішами розподілів, які в свою чергу характеризуються такими ознаками, як важкі "хвости", багатомодальність, явно виявлені асиметричність та ексцес. Такі обставини заважають використовувати класичні методи статистичного виводу для вирішення задач класифікації стискування даних.

Математичним апаратом, що працює із вказаними моделями, є робастна статистика, предметом вивчення якої є поведінка статистичних процедур не тільки на самих строго параметричних моделях, але і на менших чи більших околицях таких моделей. Механізм дії процедур робастного оцінювання на неоднорідні ("забруднені") дані полягає в їх підчистці та підгонці до параметричних моделей, що прямо веде до задач класифікації.

Таким чином, основна ідея, покладена в рішення задач класифікації та стискування неоднорідних даних, полягає у тому, що з використанням процедур робастного оцінювання проводиться виведення частки початкових даних, яка є причиною появ "хвостів" чи "горбів" у розподілах, та виділенні тієї стійкої частки даних, які характеризуються параметричною моделлю. Це дозволяє перейти до слідувочої ітерації аналізу: а саме, до тлумачення даних, що різко виділяються. В залежності від конкретної мети дослідження, такі дані можуть бути розтлумачені подвійно. У першому випадку їх можна інтерпретувати як помилки та вилучати із сукупності даних, що розглядаються. У другому випадку ці значення можуть бути

“хорошими” чи інформативними, які зможуть підвести до нової цікавої моделі чи явища.

Таким чином, використання процедур робастного оцінювання дозволяє вирішувати задачі класифікації та стискування даних комплексно.

В роботі визначено чотири класи робастних оцінок, які будуються на основі порядкових статистик:

- виважені порядкові статистики;
- адаптивні робастні процедури;
- рекурентні процедури стійкого оцінювання;
- оцінки на основі перешкодостійкості згладнення.

Виважені порядкові статистики

1. Урізане середнє рівня α ($0 \leq \alpha < 0,5$) для x_1, \dots, x_n визначається формулою:

$$\hat{\mu}(\alpha) = \frac{1}{n-2m} \sum_{i=m+1}^{n-m} x_{(i)}, \quad (13)$$

де m - найбільше число, яке не перевищує αn .

2. Середнє по Вінзору (α - вінзоризоване середнє) рівня α ($0 \leq \alpha < 0,5$) для x_1, \dots, x_n визначається:

$$\hat{\mu} = \frac{1}{n} \left[\sum_{i=m+2}^{n-m+1} x_{(i)} + m(x_{(m+1)} + x_{(n-m)}) \right] \quad (14)$$

3. Вибіркова медіана для вибірки x_1, \dots, x_n визначається:

- якщо n - непарне $\rightarrow \hat{\mu}_{med}(n) = x_{(\frac{n+1}{2})}$

- якщо n - парне $\rightarrow \hat{\mu}_{med}(n) = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right)$ (15)

Адаптивні робастні процедури

При адаптивному підході конкретний вид статистичної процедури обирається на основі допоміжної оцінки, наприклад, вибіркової ексцесу та асиметрії.

Одна із таких процедур, наприклад, будеється на визначенні стабільного значення асиметрії і має наступний вигляд:

$$\hat{\mu} = \begin{cases} \bar{x}(0,125), & \Gamma < 1,81 \\ \bar{x}(0,185), & 1,81 \leq \Gamma \leq 1,87 \\ \bar{x}(0,375), & \Gamma > 1,87, \end{cases} \quad (16)$$

де $\Gamma_1 = \frac{\bar{E}_{(0,2)} - \bar{L}_{0,2}}{\bar{E}_{0,5} - \bar{L}_{0,5}}$ стабільний аналог коефіцієнту асиметрії.

Тут: \bar{E} і \bar{L} - відповідно середнє α -n старших та молодших членів варіаційного ряду: $\bar{x}(0,125)$, $\bar{x}(0,185)$, $\bar{x}(0,375)$ - оцінки зрізаного середнього з рівнями зрізання відповідно $\alpha=0,125$, $\alpha=0,185$, $\alpha=0,375$. Описана процедура працює у такій спосіб. В залежності від підрахованої величини Γ обирається одна із зрізаних оцінок, яка в свою чергу і визначає оцінку $\hat{\mu}$.

Рекурентні процедури стійкого оцінювання.

Такі процедури базуються на перетворюванні варіаційних рядів з метою виділення стійкої частки даних з мінімальною дисперсією. Принципи побудови таких оцінок наступні. Вибираються дві опорні точки, які характеризують вибіркові кватилі:

$$h_1 = \begin{cases} x_{(1/4)}, & n \neq 4k \\ 0,5 \{x_{(n/4)} + x_{(1+n/4)}\}, & n = 4k \end{cases} \quad (17)$$

$$h_2 = \begin{cases} x_{(n+1 - [(n+3)/4])}, & n \neq 4k \\ 0,5 \{x_{(n+1-n/4)} + x_{(n-n/4)}\}, & n = 4k \end{cases}$$

і відкидаються вимірювання, які лежать поза інтервалами $[h_1 - \alpha(h_2 - h_1), h_2 + \alpha(h_2 - h_1)]$, $\alpha = 1,0; 1,5; 2,0$. Далі для залишеної частки виборки вибираються нові опорні точки h_3, h_4 і повторюється процедура відбраковки даних. Таку операцію повторюють доки із виборки не виділиться стійка центральна частина, яка не змінюється при подальших ітераціях. В якості процедури перешкодостійкого згладження в роботі використовується алгоритм ковзної медіани, що виконує наступні кроки:

- формується початкова вибірка значень x_1, x_2, \dots, x_n ;

- визначається інтервал ковзання $l = 3; 5; \dots, k$ - непарне число;
- початкова сукупність даних ранжується по l , наприклад, якщо $l=3$, то маємо:

$$\begin{aligned}x_1 &\leq x_2 \leq x_3, \dots, x_{l-1}, \dots, x_n, \\x_1, x_2 &\leq x_3 \leq x_4, \dots, x_{l-1}, \dots, x_n, \\&\dots \\x_1, x_2, x_{l-1}, \dots, x_{l-1}, \dots, x_{n-2} &\leq x_{n-1} \leq x_n.\end{aligned}$$

Залежна сукупність складається із значень x_i які дорівнюють медіані:

$$X = x_2, x_3, \dots, x_{n-1} \quad (18)$$

Розглянуті методи робастного оцінювання були використані для розробки розв'язувальних правил класифікації неоднорідних даних.

Четверта глава "Розробка розв'язувальних правил класифікації та стискування неоднорідних даних на основі стійких оцінок" присвячена розробці метода класифікації із використанням основних показників робастності: функції чутливості і порогових точок. Функція чутливості (мал.1) описує ефекти на різні оцінки від спостережень, що різко виділяються, (вибіркове середнє, різані, середні, медіана), та формалізує їх зсування. Із розглядання кривої цієї функції видно, що вона не обмежена для середнього арифметичного \bar{x} , тому одне значення, що різко виділяється може привести до скільки завгодно великого його зсування. В той же час, для медіани та різаних середніх вона обмежена.

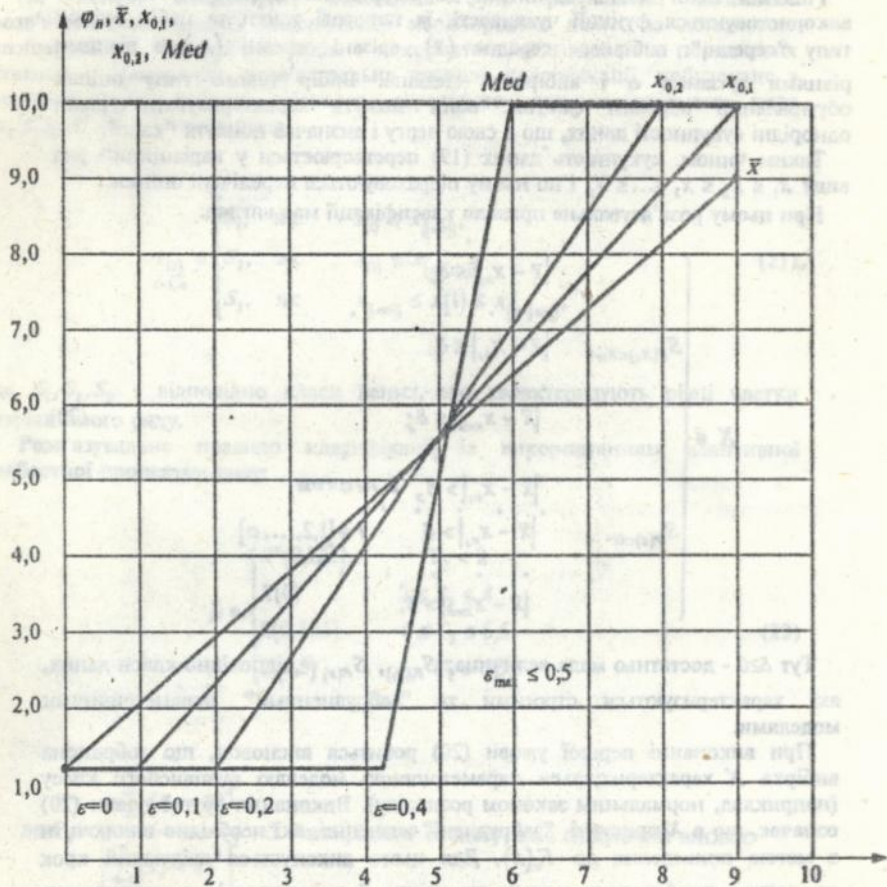
Порогові точки в нерозривно пов'язані з функціями чутливості та характеризують ту відстань від розподілу прийнятого у моделі, при дослідженні якого статистика стає зовсім ненадійною та неінформативною.

На основі цих показників розроблено розв'язувальні правила класифікації неоднорідних даних, для чого розглядалась сукупність даних, яка характеризується присутністю у ній деякого числа "забруднених" значень, тобто

$$X = (x_1, x_2, x'_1, \dots, x_l, x'_{l+1}, \dots, x_n), \quad (19)$$

де x' - "забруднені" значення, питома вага яких визначається максимальною величиною порогової точки $\epsilon_{max} \leq 0,5$. У складі (19) присутня підсукупність даних

$$X_0 = (x_1, x_2, \dots, x_l, \dots, x_n) \in X,$$



Мал. 1.

яка характеризується визначеною параметричною моделлю ймовірного розподілу $F_0(x)$, що належить деякому параметричному простору $R(F_0(x) \in R)$.

Рішення такої задачі зводиться до ітераційної процедури, для чого використовуються функції чутливості, їх порогові точки та набір оцінок типу "середні": вибіркоче середнє (\bar{x}), зрізані середні (x_α) з різними рівнями зрізання α і вибіркоче медіана. Вибір такого типу оцінок обумовлений відомим фактом: вони можуть характеризувати тільки однорідні сукупності даних, що в свою чергу і визначає поняття "клас".

Таким чином, сукупність даних (19) перетворюється у варіаційний ряд виду $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$ і по ньому підраховуються перелічені оцінки.

При цьому розв'язувальне правило класифікації має вигляд:

$$X \in \begin{cases} S_{F(x) \in R_0}, & \begin{cases} |\bar{x} - x_{\alpha_1}| \leq \delta, \\ \cdot \\ \cdot \\ \cdot \\ |\bar{x} - x_{\alpha_i}| \leq \delta \\ \cdot \\ \cdot \\ |\bar{x} - x_{med}| \leq \delta; \end{cases} \\ S_{F(x) \in A}, & \begin{cases} |\bar{x} - x_{\alpha_1}| > \delta, \text{ і такон} \\ |\bar{x} - x_{\alpha_i}| > \delta \quad i \in [1, 2, \dots, n] \\ \cdot \\ \cdot \\ |\bar{x} - x_{med}| > \delta, \end{cases} \end{cases} \quad (20)$$

Тут $\delta \geq 0$ - достатньо мала величина; $S_{F(x) \in R_0}$, $S_{F(x) \in A}$ - відповідно класи даних, які характеризуються строгими та "забрудненими" параметричними моделями.

При виконанні першої умови (20) робиться висновок, що зображена вибірка X характеризується параметричною моделлю визначеного класу (наприклад, нормальним законом розподілу). Виконання другої умови (20) означає, що в X присутні "забруднені" значення, які необхідно виключити з метою приведення до $F_0(x)$. Для цього виконується наступний крок процедури, який полягає у тому, що оцінка \bar{x} виводиться із спостереження, задається порогова точка $\varepsilon_1 = 0,1$ і підраховується зрізані оцінки $x_{\alpha_1} = 0,1$; $x_{\alpha_2}, \dots, x_{med}$ та проводиться аналогічна перевірка на виконання заданих умов (20).

Ітераційна процедура закінчується при досягненні $\epsilon_{\max} = 0,5$, при якому медіана залишає свою робастну якість. Необхідно відмітити, що медіана є найбільш стійкою з усіх розглянутих оцінок і витримує рівні "забруднення" до $\epsilon \leq 0,5$.

На основі розглянутих положень було розроблено комплекс розв'язувальних правил класифікації неоднорідних даних за допомогою розшарування варіаційних рядів з використанням анважених порядкових статистик. Наприклад, розв'язувальне правило класифікації, побудоване з використанням оцінок зрізаного середнього для варіаційного ряду $x_1 \leq x_2 \leq \dots \leq x_n$, має вигляд:

$$x_{(j)} \in \begin{cases} S_1, & \text{як } x_{(j)} \leq x_{(a)}, \\ S_2, & \text{як } x_{(j)} \leq x_{(a-[a/2])}, \\ S_3, & \text{як } x_{(a)} \leq x_{(j)} \leq x_{(a-[a/2])}, \end{cases} \quad (21)$$

де S_1, S_2, S_3 - відповідно класи даних, що характеризують різні частки варіаційного ряду.

Розв'язувальне правило класифікації із використанням адаптивної робастної процедури виду:

$$\hat{\mu} = \begin{cases} \bar{x}^*(0,125), & \Gamma_2 < 2 \\ \bar{x}(0) & 2 \leq \Gamma_2 \leq 4 \\ \bar{x}(0,125) & 4 \leq \Gamma_2 \leq 5,5 \\ \bar{x}(0,5) & \Gamma_2 > 5,5, \end{cases} \quad (22)$$

де $\Gamma_2 = \frac{n \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} - 3$ - коефіцієнт експесу, має наступний вигляд:

$$\begin{aligned}
 x_l \in \begin{cases} S_1, & \Gamma_2 < 2 \Rightarrow \begin{cases} x_{(l)} \leq x_{(l)} \leq x_{(a,n)} \\ x_{(\lfloor 3a, n \rfloor + 1)} \leq x_{(l)} < x_{(a)} \end{cases} \\ S_2, & \Gamma_2 < 2 \Rightarrow \left(x_{(\lfloor a, n \rfloor + 1)} \leq x_{(l)} \leq x_{(3a, n)} \right), \\ \alpha_1 = 0,25 \end{cases} \\
 x_l \in S_0, & 2 \leq \Gamma_2 \leq 4 \Rightarrow \left(x_{(l)} \leq x_{(l)} \leq x_{(a)} \right), \\ \alpha_2 = 0 \\
 x_{(l)} \in \begin{cases} S_1, & 4 < \Gamma_2 \leq 5,5 \Rightarrow \left(x_{(l)} \leq x_{(l)} \leq x_{(a, n)} \right) \\ S_2, & 4 < \Gamma_2 \leq 5,5 \Rightarrow \left(x_{(\lfloor a, n \rfloor + 1)} \leq x_{(l)} \leq x_{(3a, n)} \right) \\ S_3, & 4 < \Gamma_2 \leq 5,5 \Rightarrow \left(x_{(\lfloor 3a, n \rfloor + 1)} \leq x_{(l)} \leq x_{(a)} \right) \\ \alpha_3 = 0,25 \end{cases} \quad (23) \\
 x_l \in \begin{cases} S_1, & \Gamma_2 > 5,5 \Rightarrow \left(x_{(l)} \leq x_{(l)} \leq x_{(a, n)} \right) \\ S_2, & \Gamma_2 > 5,5 \Rightarrow \left(x_{(\lfloor a, n \rfloor + 1)} \leq x_{(l)} \leq x_{(a)} \right) \\ \alpha_4 = 0,5 \end{cases}
 \end{aligned}$$

Розв'язувальне правило класифікації побудоване на рекурентній процедурі (17) запишеться таким чином:

$$X = \begin{cases} \left[x_{(1)}, x_{(b-1)} \right], \left[x_{(b)}, x_{(b)} \right], \left[x_{(b+1)}, x_{(a)} \right], & \sigma_1^2 < \sigma_0^2, \\ \left[x_{(1)}, x_{(b-1)} \right], \left[x_{(b)}, x_{(b)} \right], \left[x_{(b+1)}, x_{(a)} \right], & \sigma_2^2 < \sigma_1^2, \\ \dots \\ \left[x_{(1)}, x_{(b_m-1)} \right], \left[x_{(b_m)}, x_{(b_m+1)} \right], \left[x_{(b_m+1)}, x_{(a)} \right], & \sigma_2^2 \rightarrow \min, \end{cases} \quad (24)$$

де σ_0^2 - дисперсія вихідного варіаційного ряду.

$$x_{(l)} \in \begin{cases} S_1, x_{(1)} \leq x_{(l)} \leq x_{(b_m-1)} \\ S_2, x_{(b_m)} \leq x_{(l)} \leq x_{(b_m+1)} \\ S_3, x_{(b_m+1)} \leq x_{(l)} \leq x_{(b)} \end{cases}$$

Аналогічно будуються розв'язувальні правила класифікації з використанням інших типів робастних оцінок.

В п'ятій главі "Аналіз даних локального радіаційного моніторингу гідросфери, що виконуються методами прямих вимірювань" розроблено методологічні основи синтезу автоматизованих систем обробки неоднорідних вимірювальних даних та інформаційного забезпечення СФС (мал.2). Запропоновано методику обробки даних локального радіаційного моніторингу гідросфери, в рамках якої вирішені наступні задачі:

- запропоновані технологічні вимоги до проведення прямих вимірювань;
- розроблено метод класифікації неоднорідних даних з використанням гістограм;
- розроблено комплекс статистичних моделей, які характеризують неоднорідну структуру полів радіаційного забруднення водного середовища, що складається з двох фаз: локальних турбулентних областей та ламінарних прошарків. Виходячи з цього положення, загальний вигляд моделі такий:

$$G = U(T_{\lambda}, V_{\lambda}, F_{\lambda}, T_{\Pi}, V_{\Pi}, F_{\Pi}), \quad (25)$$

Тут $T_{\lambda} = \sum_{m=1}^k t_m$, $T_{\Pi} = \sum_{m=1}^n t_m$ - відповідно часова тривалість локальних областей та ламінарних прошарків; k і n - число часових відліків t , що складають їх тривалість;

$V_{\lambda} = \sum_{m=1}^k v_x \cdot t_m$; $V_{\Pi} = \sum_{m=1}^n v_x t_m$ - відповідно просторовий протяг локальних областей і прошарків; v_x - середня швидкість зондування датчиків вимірювального пристрою ($v_x = const$).

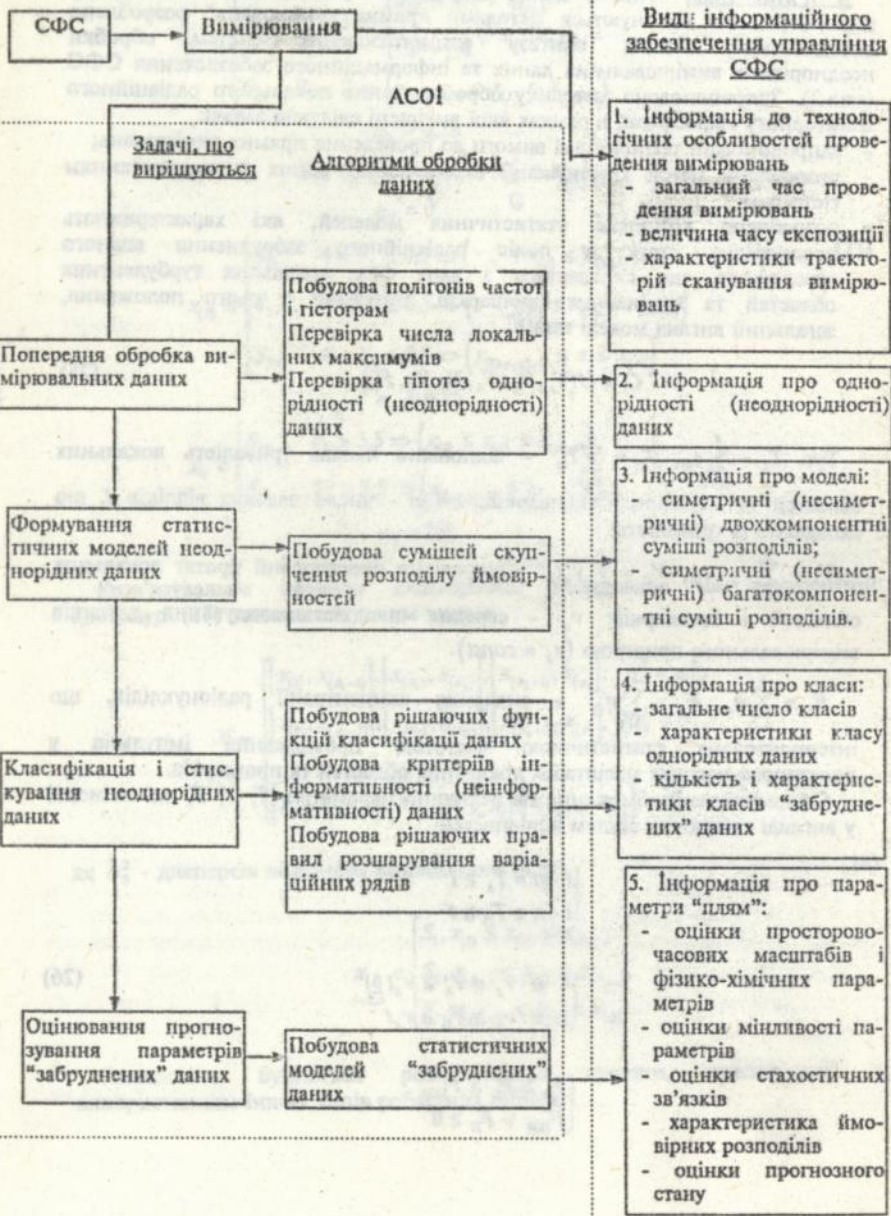
$F_{\lambda} = \sum_{m=1}^k q_m$; $F_{\Pi} = \sum_{m=1}^n q_m$ - значення концентрації радіонуклідів, що інтерпретована статистичною частотою проходження імпульсів у просторово-часових масштабах локальних областей та прошарків.

Сформульовані обмеження на величини параметрів T , V , F , які записані у вигляді наступних систем нерівностей

$$\begin{cases} t \cdot n > T_{\lambda} \geq t \\ t \cdot n \geq T_{\Pi} \geq t \end{cases}$$

$$\begin{cases} t \cdot n \cdot v_x > V_{\lambda} \geq v_x t \\ t \cdot n \cdot v_x \geq V_{\Pi} \geq v_x t \end{cases} \quad (26)$$

$$\begin{cases} F_{\Pi} \geq F_{\lambda} \geq F_{пор} \\ F_{пор} > F_{\Pi} \geq 0 \end{cases}$$



Мал.2

Тут: t - величина часу експозиції, n - число часових відліків t у вибірці, що досліджується; v_x - середня швидкість переміщення датчиків вимірального пристрою; $F_{ГДР}$ - порогова частота алгоритму знаходження;

$$F_B = \frac{1}{J(t)} - \text{верхня частотна межа вимірального датчика; } J(t) -$$

розв'язувальна здатність датчика вимірювача.

Виходячи із (26) розроблено:

- моделі просторово-часових масштабів неоднорідних структур даних;
- моделі мінливості просторово-часових масштабів локальних областей та прошарків;
- моделі стохастичних зв'язків між локальними областями.

В главі подаються результати досліджень за даними натурних експериментів, які були одержані в рамках виконання ряду НДР у 1981-83 рр., а також при обробці даних радіаційного фону Київського водосховища у 1986-89 рр.

Шоста глава "Робастні оцінки в задачах аналізу зображень, поданих методами дистанційного зондування об'єктів різної фізичної природи" присвячена розробці методу класифікації зображень для випадків, коли гістограми побудовані за значеннями діапазону градацій яскравостей, не мають помітних піків і розглядаються як одномодальні.

У цих випадках зображення можуть вважатися однорідними, в той час, як до їх складу можуть входити неоднорідні області. Метод побудований на використанні адаптивних робастних процедур, розглянутих в третій главі. За значеннями матриці (6) будується варіаційний ряд, підраховується згортання Γ з використанням виразу (16), згідно якого призначаються рівні зрізання ряду α , які в свою чергу дозволяють розчленити зображення G на однорідні сукупності. У формалізованому виді такий підхід виглядає:

$$G = \begin{cases} (g_{(1)}, g_{(\alpha_1, n)}) \cdot (g_{([a_1, n] \cdot 1)}, g_{(n - [a_1, n])}) \cdot (g_{(n - [a_1, n] \cdot 1)}, g_n), & \Gamma < 1,81 \\ (g_{(1)}, g_{(\alpha_2, n)}) \cdot (g_{([a_2, n] \cdot 1)}, g_{(n - [a_2, n])}) \cdot (g_{(n - [a_2, n] \cdot 1)}, g_n), & 1,81 \leq \Gamma < 1,81 \\ (g_{(1)}, g_{(\alpha_3, n)}) \cdot (g_{([a_3, n] \cdot 1)}, g_{(n - [a_3, n])}) \cdot (g_{(n - [a_3, n] \cdot 1)}, g_n), & \Gamma < 1,87 \end{cases} \quad (27)$$

$$\alpha_1 = 0,125; \quad \alpha_2 = 0,185; \quad \alpha_3 = 0,375$$

Спільне використання двох адаптивних процедур робастного оцінювання, одна з яких побудована на відслідкуванні величини асиметрії, а друга базується на урахуванні показників ексцесу, розширяє можливості методу в плані більш детального розшарування варіаційних рядів, що дозволяє виділити не тільки основні класи, але і деяку кількість підкласів.

Розглянемо дві процедури адаптивного оцінювання, одна з яких була зображена раніше виразом (16), а друга має наступний вигляд :

$$\hat{\mu} = \begin{cases} \bar{x}(0), & 2,0 \leq \Gamma_2 \leq 4,0 \\ \bar{x}(0,25), & 4,0 \leq \Gamma_2 \leq 5,5 \\ x(0,5), & \Gamma_2 > 5,5, \end{cases} \quad (28)$$

де $\Gamma_2 = \frac{\frac{1}{n} \sum_{i=1}^n (g_i - \bar{g})^4}{\left[\frac{1}{n} \sum_{i=1}^n (g_i - \bar{g})^2 \right]^2} - 3$ - вибірковий ексцес, $\bar{x}(0)$, $\bar{x}(0,25)$, $x(0,5)$ -

відповідно вибіркове середнє, зрізане середнє з рівнем зрізання $\alpha=0,25$, вибіркова медіана.

Результати класифікації даних із спільним використанням (16) і (28) зображений в таблиці 1.

В главі подані результати обробки зображень із використанням розробленого методу заданих космічних зйомок земної поверхні, та даними, отриманими при конструюванні супутників стопи.

ТАБЛИЦЯ 1

$\Gamma_2 \backslash \Gamma_1$	$\Gamma_1 < 1,81$ $\bar{x}(0,125)$	$1,81 \leq \Gamma_1 < 1,87$ $\bar{x}(0,185)$	$\Gamma_1 > 1,87$ $\bar{x}(0,375)$
$2 \leq \Gamma_2 \leq 4$ $\bar{x}(0)$	1 S_1 $S'_1, S''_1 \subset S_1$ $[g(1), g(n)] \in S_1$ $[g(1), g(0,125n)] \in S'_1$ $[g(0,875n+1), g(n)] \in S''_1$	4 S_1 $S'_1, S''_1 \subset S_1$ $[g(1), g(n)] \in S_1$ $[g(1), g(0,165n)] \in S'_1$ $[g(0,835n+1), g(n)] \in S''_1$	7 S_1 $S'_1, S''_1 \subset S_1$ $[g(1), g(n)] \in S_1$ $[g(1), g(0,375n)] \in S'_1$ $[g(0,625n+1), g(n)] \in S''_1$
$4 \leq \Gamma_2 \leq 5,5$ $\bar{x}(0,25)$	2 S_1, S_2, S_3 $S'_1 \in S_1; S'_2 \subset S_2$ $[g(1), g(0,25n)] \in S_1$ $[g(1), g(0,125n)] \in S'_1$ $[g(0,25n+1), g(0,75n)] \in S_2$ $[g(0,75n+1), g(n)] \in S_3$ $[g(0,875n+1), g(n)] \in S'_3$	5 S_1, S_2, S_3 $S'_1 \subset S_1; S'_2 \subset S_2$ $[g(1), g(0,25n)] \in S_1$ $[g(1), g(0,165n)] \in S'_1$ $[g(0,25n+1), g(0,75n)] \in S_2$ $[g(0,75n+1), g(n)] \in S_3$ $[g(0,835n+1), g(n)] \in S'_3$	8 S_1, S_2, S_3 $S'_1 \subset S_1; S'_2 \subset S_2$ $[g(1), g(0,375n)] \in S_1$ $[g(1), g(0,25n)] \in S'_1$ $[g(0,375n+1), g(0,625n)] \in S_2$ $[g(0,625n+1), g(n)] \in S_3$ $[g(0,75n+1), g(n)] \in S'_3$
$\Gamma_2 > 5,5$ $\bar{x}(0,5)$	3 S_1, S_2 $S'_1 \subset S_1; S'_2 \subset S_2$ $[g(1), g(0,5n)] \in S_1$ $[g(1), g(0,125n)] \in S'_1$ $[g(0,5n+1), g(n)] \in S_2$ $[g(0,875n+1), g(n)] \in S'_2$	6 S_1, S_2 $S'_1 \subset S_1; S'_2 \subset S_2$ $[g(1), g(0,5n)] \in S_1$ $[g(1), g(0,165n)] \in S'_1$ $[g(0,5n+1), g(n)] \in S_2$ $[g(0,835n+1), g(n)] \in S'_2$	9 S_1, S_2 $S'_1 \subset S_1; S'_2 \subset S_2$ $[g(1), g(0,5n)] \in S_1$ $[g(1), g(0,375n)] \in S'_1$ $[g(0,5n+1), g(n)] \in S_2$ $[g(0,625n+1), g(n)] \in S'_2$

У сьомій главі "Алгоритми оцінювання параметрів неблагоприємних впливів в задачах забезпечення живучості і надійності технічних систем" присвячена розробці методики аналізу параметрів середовища, в якому функціонують технічні системи (ТС). Вважається, що середовище функціонування ТС ділиться на внутрішнє та зовнішнє. Внутрішнє середовище ТС - це її апаратні і програмні компоненти, а зовнішнє середовище подається кількістю енергетичних, хімічних, механічних та інших явищ.

Особлива увага приділяється впливу неблагоприємних дій зовнішнього середовища на функціонування ТС, параметри чи характеристики яких треба прогнозувати.

Для отримання короткочасних значень широко використовуються методи експоненціального згладження часових рядів. Нехай d_t - часовий ряд значень деякого показника d , $0 < \alpha < 1$ - константа згладження, тоді експоненціально згладженим рядом для q_t , $t = \overline{1, N}$ буде ряд U_t , що одержано за рекурентною формулою

$$U_t = \alpha \cdot q_t + (1 - \alpha)U_{t-1} \quad (29)$$

Такий вираз за значенням показника на момент $t=p$ дозволяє побудувати прогноз на момент часу $p+l$:

$$U_{p+1} = \alpha \cdot q_p + (1 - \alpha)U_p \quad (30)$$

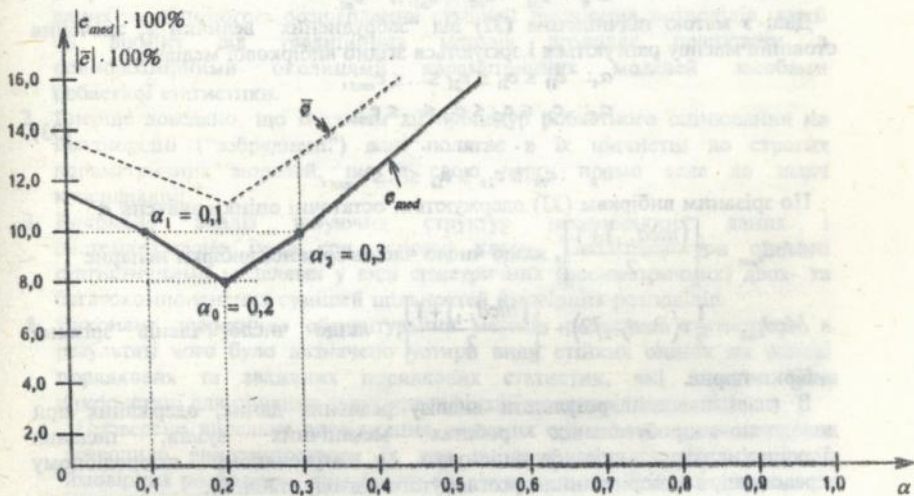
Важлива роль при цьому надається вибору оптимальних значень константи α , тому що саме вона визначає результати прогнозу. На сьогоднішній день не існує загальної методики вибору α , і вона у кожному конкретному випадку визначається по-різному.

В главі запропонована процедура адаптивного вибору α з використанням методу навчальної вибірки (НВ), яка відрізняється від існуючих підходів в одержанні залежностей виду

$$\begin{aligned} \sigma &= f(\alpha_1, \alpha_2, \dots, \alpha_k) \\ \hat{\sigma}(\alpha_j) &= |d_{p+1} - U_p| \rightarrow \min, \end{aligned} \quad (31)$$

де $\hat{\sigma}$ - оцінка помилки прогнозу, одержана за НВ.

Іншими словами, на етапі навчання будується крива залежності $\sigma=f(\alpha)$ (мал. 3), на якій вибирається ділянка з мінімальними значеннями $\hat{\sigma}$, яка в свою чергу також використовується для вибору оптимальних значень α , що дають хороші прогнозні результати.



Мал. 3

0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
10.0	8.0	10.0	12.0	14.0	16.0	18.0	20.0	22.0	24.0

Для реалізації критерія (30) задаємо сітку значень α ($\alpha=0; 0,1; 0,2; \dots; 0,9; 1,0$), та використовуємо вираз (31), підраховуємо значення прогнозів та їх помилок для кожного α впродовж усієї НВ. В результаті отримуємо масив наступного виду:

$$\begin{array}{cccccc}
 & \alpha_1 & \alpha_2 & \alpha_3 \dots & \alpha_k & \\
 U_1 & e_{11} & e_{12} & e_{13} \dots & e_{1k} & \\
 U_2 & e_{21} & e_{22} & e_{23} \dots & e_{2k} & \\
 U_3 & e_{31} & e_{32} & e_{33} \dots & e_{3k} & \\
 \dots & \dots & \dots & \dots & \dots & \\
 U_n & e_{n1} & e_{n2} & e_{n3} \dots & e_{nk} &
 \end{array} \quad (32)$$

Далі з метою підчищення (32) від "забруднених" величин e , значення стовпців масиву ранжуються і зрізуються згідно вибіркової медіани:

$$\begin{array}{l}
 \alpha_1: e_{11} \leq e_{31} \leq e_{21} \leq \dots \leq e_{med}, \\
 \alpha_2: e_{21} \leq e_{12} \leq e_{32} \leq \dots \leq e_{med}, \\
 \dots \\
 \alpha_k: e_{3k} \leq e_{2k} \leq e_{1k} \leq \dots \leq e_{med},
 \end{array} \quad (33)$$

По зрізанім вибіркам (33) одержуються остаточні оцінки значень e :

$$Med_{e_{med}} = c \left[\frac{(med_{j-1}) + 1}{2} \right], \text{ якщо число членів зрізаної вибірки непарне}$$

$$Med_{e_{med}} = \frac{1}{2} c (med_{j-2}/2) + c \left[\frac{(med_{j-1}) + 1}{2} \right], \text{ якщо число членів зрізаної}$$

вибірки парне.

В главі наведені результати аналізу реальних даних, одержаних при доводочно-випробувальних роботах механічних вузлів, поданих гідроциліндрами, які функціонують у агресивному газоподібному середовищі, з використанням розглянутого підходу (табл. 2).

Таблиця 2

q_t	492	466	476	484	477	451	504	464	456	469
U_t	503	504	505	504	504	504	504	501	502	501
$ e \cdot 100\%$	2,2	8,0	6,1	4,1	5,6	11	0,0	8,2	10	6,8

Тут q_t, U_t - відповідно дійсні та прогнозні значення часового ряду при величині $\alpha = \alpha_0 = 0,2$, забезпечующій мінімальну помилку прогнозу. Аналіз таблиці показує, що навіть при величині горизонту прогнозування $n_t = 10$ величина помилки практично не перевершує 10%, що указує на досить високу точність прогнозу.

ОСНОВНІ РЕЗУЛЬТАТИ РОБОТИ І ВИСНОВКИ

Робота присвячена розв'язанню актуальної проблеми побудови аналітичного підходу по створенню інформаційного забезпечення для керування складними фізичними системами (СФС) на основі робастної обробки неоднорідних вимірювальних даних. В якості інструментарію для формування інформаційного забезпечення систем управління СФС в роботі розроблено нові методи оцінювання, класифікації та стискування неоднорідних даних.

При розв'язанні даної проблеми одержано такі нові теоретичні та практичні результати:

1. Запропоновано новий підхід до рішення задач класифікації неоднорідних даних за допомогою розщеплення сумішей ймовірних розподілів, який на відміну від традиційних методів дозволяє працювати з повнорозмірними околицями параметричних моделей засобами робастної статистики.
2. Вперше доведено, що механізм дії процедур робастного оцінювання на неоднорідні ("забруднені") дані полягає в їх підчистці до строгих параметричних моделей, що в свою чергу прямо веде до задач класифікації.
3. Виконано аналіз існуючих структур неоднорідних даних і систематизовано їх в три основні класи. Такі структури описані статистичними моделями у виді симетричних (несиметричних) двох- та багатокомпонентних сумішей щільностей ймовірних розподілів.
4. Виконано теоретичне обґрунтування методів робастної статистики, в результаті чого було визначено чотири види стійких оцінок на основі порядкових та зважених порядкових статистик, які можуть бути використані для рішення задач класифікації неоднорідних даних.

Проведено чисельне дослідження вказаних оцінок на незсування, що дозволило використовувати їх для оцінювання параметрів сумішей ймовірних розподілів і тим самим реалізувати схему "від оцінювання до класифікації".

5. Розроблено новий метод класифікації неоднорідних даних, який використовує основні показники робастності: функції чугливості та порогові точки.

Запропоновано комплекс розв'язувальних правил класифікації з використанням робастних оцінок, в основі яких покладена процедура перетворення поданих масивів даних у варіаційні ряди з послідовним їх розчленуванням за призначеними рівнями "зрізання".

6. Запропоновано концепцію системного аналізу СФС та методологічні основи синтезу структур АСОІ і інформаційного забезпечення для прийняття рішень в управлінні СФС.
7. Розроблено методики робастної обробки неоднорідних даних з використанням виважених порядкових статистик, адаптивних процедур

робастного оцінювання, рекурентних стійких оцінок, процедур перешкодостійкого згладження.

8. На основі одержаних у роботі теоретичних результатів розроблено математичне і програмне забезпечення: обробки вимірювань полів радіаційного забруднення водного середовища, аналізу різноманітних зображень, короткочасного прогнозування часових рядів, отриманих при випробуваннях різних технічних пристроїв.

Публікації. Основні результати дисертації опубліковані у таких роботах:

1. Батиенко Л.Ю., Коваленко И.И. Методы и средства снижения размерности данных (подход, основанный на расщеплении смесей вероятностных распределений) //В сб. Перспективные средства информационного и программного обеспечения в сложных системах. Изд. ин-та кибернетики, - 1992.-С.31-39.
2. Згуровский М.Э., Демченко А.М., Новиков А.Н., Коваленко И.И. Системный анализ в исследовании сложных физических процессов и полей //Препр. АН Украины. Ин-т кибернетики им. В.М.Глушкова;- Киев, 1993.-37 с.
3. З.Игнатенко Б.В., Коваленко И.И., Краскевич В.Е., Мишаков В.А. Банк данных экологического состояния гидросферы //Автоматика. - 1988.- №4.-С.30-34.
4. Коваленко И.И. Анализ измерительных данных методами устойчивого оценивания //В учеб. пособии для студ. вузов В.А.Поджаренко, А.Д.Азаров, А.В.Власенко, И.И.Коваленко. Избыточные системы счисления, моделирование, обработка данных и системное проектирование в технике преобразования информации. - Киев.: Выща шк., 1990.-208 с.
5. Коваленко И.И. Сравнительный анализ адаптивных алгоритмов робастного оценивания //Адаптив. системы автомат. упр.: Респ. межвед. научн.-технн. сб. - 1988. - Вып.16. - С. 141-143.
6. Коваленко И.И. Устойчивое оценивание значений критериальных статистик //Вестн. Киев. политехн. ин-та. Техн. кибернетика. - 1989. - Вып.13. - С. 71-73.
7. Коваленко И.И. Автоматизированные системы обработки неоднородных измерительных данных //Тр. IV международн. конф. "Проблемы комплексной автоматизации". Секция 3. - Киев, 1990. - С. 98-93.
8. Коваленко И.И. Адаптивная классификация изображений //Адаптив. системы автомат. упр.: Респ. межвед. науч.-техн. сб. - 1992. - Вып.19. - С.109-112.
9. Коваленко И.И., Козлова Г.П. Анализ случайных процессов по непараметрическим критериям проверки гипотез// Вестн. Киев. политехн. ин-та. Техн. кибернетика. - 1986. - Вып.10. - С. 43-46.
10. Коваленко И.И., Гец А.В., Ануреева А.Ю., Тарасов Ю.Ф. Адаптивное прогнозирование временных рядов //В сб. научн. тр. "Новое в

- разработках и исследованиях электробытовых машин и приборов". Изд. ин-та ВНИЭКИЭМП, - 1988. - С. 96-100.
11. Коваленко И.И., Козлова Г.П. Помехоустойчивое сглаживание временных рядов //Адаптив. системы автомат. упр.: Респ. межвед. науч.-техн. сб. - 1989. - Вып.17. - С. 72-75.
 12. Коваленко И.И., Коц А.П. База моделей локального экологического мониторинга гидросферы //В сб. Энергетика и электрификация. - 1989. - №2. - С.39-41.
 13. Коваленко И.И., Рябов К.К. Некоторые вопросы информационного обеспечения робототехнических комплексов //Тр. Респ. науч.-техн. конф. "Робототехнические системы для промышленных технологических процессов". Ворошиловград, 1985. - С.42-43.
 14. Коваленко И.И., Коломыцев М.В., Пасько В.П. Алгоритм классификации реализаций пуассоновского процесса //Вестн. Киев. политехн. ин-та. Техн. кибернетика. - 1984. - Вып.8. - С. 27-30.
 15. Краскевич В.Е., Коваленко И.И., Коломыцев М.В. Адаптивные алгоритмы робастного оценивания характеристик локально-нестационарных процессов //Адаптив. системы автомат. упр.: Респ. межвед. науч.-техн. сб. - 1986. - Вып.14. - С.69-73.
 16. Краскевич В.Е., Коваленко И.И. Устойчивое оценивание характеристик локально-нестационарных процессов // Адаптив. системы автомат. упр.: Респ. межвед. науч.-техн. сб. - 1987. - Вып.15. - С.77-81
 17. Краскевич В.Е., Коваленко И.И., Макаровский А.П., Ивахненко Г.А. Анализ случайных пространственных изображений //Автоматика. - 1987. - №5. - С. 65-69.
 18. Краскевич В.Е., Коваленко И.И., Литвинов Е.Н., Пасько В.П. Структурно-параметрическая идентификация объектов с распределенными параметрами //Тр. III польско-советской науч.-техн. конф. "Комплексная автоматизация промышленности". Ворцлав, 1988, - ч.1 доп. С. 162-165.
 19. Краскевич В.Е., Коваленко И.И., Ивахненко Г.А. Некоторые структуры пространственных изображений и их анализ //Автоматика. - 1988. - №6. - С. 3-9.
 20. Компанец Л.Ф., Коваленко И.И., Коломыцев М.В. Особенности организации экспресс-режимов автоматизированной записи и воспроизведения импульсных потоков //Вестн. Киев. политехн. ин-та. Техн. кибернетика. - 1985. Вып.9. - С.73-75.
 21. Краскевич В.Е., Коваленко И.И., Коломыцев М.В., Алгоритмы устойчивого оценивания в математическом обеспечении СУГАП //Тр. всесоюзн. науч.-техн. семинара "Программное обеспечение гибких автоматизированных систем", Калинин, 1986. - Часть 3. - С. 75-77.
 22. Коваленко И.И. Модель случайного процесса на основе неоднородных (сложных) сред //деп. рук. УКР НИИНТИ, №322, Ук-Д83, 1983.

Коваленко И.И. Методы робастной обработки неоднородных измерительных данных в системах управления.

Диссертация на соискание ученой степени доктора технических наук по специальности 05.13.04 - "Автоматизированные системы управления и системы обработки информации", Национальный технический университет Украины "Киевский политехнический институт", Киев, 1996.

Защищается рукопись на основе 22 научных трудов, в которой рассматривается подход к проблеме информационного обеспечения систем управления на основе робастной обработки неоднородных измерительных данных, характеризующих сложные физические системы. Разработаны модели неоднородных данных, методы их классификации и сжатия посредством расщепления смесей вероятностных распределений.

Предложены методики обработки неоднородных данных с использованием взвешенных порядковых статистик, адаптивных и рекуррентных процедур робастного оценивания, помехоустойчивого сглаживания. Результаты теоретических исследований апробированы и подтверждены численным анализом данных натурных экспериментов.

Kovalenko I.I. "Robust methods for heterogeneous measurement data processing in control systems".

Doctor of Technical Sciences thesis to speciality - 05.13.04 "Automatic control systems and information processing systems", National Technical University of Ukraine, Kiev, 1996.

The manuscript is defended on the base of 22 scientific works in which has been developed an approach to the problem of information support for control systems based on robust processing of heterogeneous measurement data which characterize complex physical systems. The models of heterogeneous data are developed as well as the methods of their classification and compression through probability distribution mixtures splitting.

The methodics to heterogeneous data processing using suspended order statistics, adaptive and recurrence procedures of robust evaluating and noise-immunity smoothing are proposed. The results of theoretical investigations are verified and confirmed by digital simulation of natural experiment data.

Ключові слова: інформаційне забезпечення, суміші ймовірних розподілів, робастне оцінювання, класифікація, статистичні моделі, прогнозування, варіаційний ряд, порядкові статистики.

I. I. Kovalenko

438726

ДФОП.НТУУ. № 6.1996.Зам.173-100.

438726

AB 35.478