

Є. В. ГАВРИЛКО, О. А. КУРЧЕНКО, І. В. ТЕРЕЩЕНКО, А. І. ТЕРЕЩЕНКО

## МЕТОД БАГАТОВИМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ ЧАСОВИХ БАГАТОВИМІРНИХ КРИТИЧНИХ АТРИБУТІВ ЯКОСТІ ПРОЦЕСУ ВИРОБНИЦТВА З ФАКТОРИЗАЦІЄЮ ДАНИХ

**Об'єктом** дослідження є процес забезпечення якості продукції на етапі первинного проектування процесу виробництва. **Предметом** дослідження є інформаційні технології оцінки факторного впливу критичних параметрів процесу виробництва (CPPs, critical process parameters) на критичні атрибути якості продукту (CQAs, critical quality attributes). **Мета** – застосування запропонованого методу для визначення особливостей і характеристик залежності змінювання критичних атрибутів якості продукту від зміни критичних параметрів процесу виробництва. **Завдання**: визначення структури та ієрархії часових багатовимірних даних критичних параметрів процесу виробництва та критичних атрибутів якості продукту, а також визначення якісної та кількісної міри відносин між сформованими об'єктами зазначених параметрів. **Методи**. Послідовно використовуються: статистичні процедури розвідувального багатовимірного аналізу даних; перетворення однорідних матриць спостережень значень CPPs і CQAs в таблицю з факторизованими даними; побудова дерев регресії багатовимірних CPPs з багатовимірними відповідями CQAs. Методи використовують програмне забезпечення пакетів мови R. Отримано такі **результати**. Запропоновано метод розв'язання задачі забезпечення якості продукції на етапі первинного проектування процесу виробництва відповідно до актуальної процесно-аналітичної технології конструювання сучасних сертифікованих виробництв – "якість через дизайн" QbD (Quality-by-Design). Запропонований метод факторизації часових багатовимірних критичних параметрів процесу виробництва та критичних атрибутів якості продукту дозволяє збільшити число ступенів свободи для багатовимірного статистичного аналізу MSA (Multivariate Statistical Analysis) впливу факторів (класів) критичних параметрів процесу виробництва на багатовимірні критичні атрибути якості продукту. Наведені та проаналізовані результати використання методу на прикладі масивів даних з застосуванням пакетів об'єктно-орієнтованої мови програмування R. **Висновки**. Запропонований метод статистичного аналізу спільно з статистичними багатовимірним канонічним аналізом представляють актуальну інформаційну технологію детального оцінювання впливу об'єктів часових багатовимірних даних критичних параметрів процесу виробництва і окремих складових на критичні атрибути якості продукту. Метод орієнтований на практичне застосування з метою забезпечення якості продукції на етапі проектування (вдосконалення) виробництв.

**Ключові слова**: якість через дизайн; атрибути критичної якості; критичні параметри процесу; дизайн експерименту; багатовимірний статистичний аналіз.

### Вступ

Удосконалення будь-якого виробництва неминує породжує проблемну ситуацію, пов'язану з перманентною високою якістю продукту виробництва. Кон'юнктура ринку і посилення конкуренції об'єктивно ставлять задачу розробки та вдосконалення методів забезпечення високої якості продукції на етапі первинного проектування процесу виробництва з подальшим менеджментом всього виробничого циклу (PDCA, Plan-Do-Check-Act cycle [1]) відповідно до цільового профілю якості продукту TPQP (Target Product Quality Profile) [2].

Одним з актуальних шляхів вирішення цього завдання є методологія "якість через дизайн" QbD (Quality-by-Design) [2, 3]. Дана методологія зміщує акценти забезпечення якості продукту в напрямку інформаційних технологій статистичних досліджень і ретельного планування експерименту на етапі розробки продукту і проектування процесу виробництва (процесу).

Концепція QbD об'єднує категорії під загальним терміном "критичні атрибути якості": критичні матеріальні атрибути CMAs (Critical Material Attributes), критичні параметри процесу CPPs (Critical Process Parameters) і критичні атрибути якості продукту CQAs (Critical Quality Attributes) з TPQP. У свою чергу TPQP визначає кількісні параметри безпеки та ефективності продукту [2] і дозволяє

зіставити їх з вимогами стандартів, як показано на рис. 1.

Залежність критичних атрибутів якості продукту від критичних параметрів процесу виробництва і критичних матеріальних атрибутів трактується як функція якості:  $CQAs = F(CPP_i)$ . Ця функція якості структурована відповідно основним етапам виробничого/технологічного процесу QFD (Quality Function Deployment) [4], для яких визначені критичні атрибути якості і застосовні принципи QbD, як показано на рис. 1.

Важливим інструментом QbD є аналітична технологія PAT (Process Analytical Technology) [5, 6], яка активно використовує інформаційні технології MSA для забезпечення і підтримки критичних атрибутів якості в межах встановлених стандартами значень простору розробки (design space).

Результати MSA даних багатфакторного експерименту на етапі проектування дозволяють встановити допустимі діапазони змінності CPPs і CMAs виходячи зі ступеня впливу на CQAs.

Слід зауважити, що при жорстких вимогах стандартів (ISO 9001 [7], ISO 22000 [8]) до CMAs, частіше розглядають вплив CPPs на CQAs [5]. При цьому часові багатовимірні статистичні дані CPPs у загальному випадку надають груповий/об'єктний вплив на CQAs і тому потребують програмної факторизації комп'ютерного формату часових багатовимірних даних.

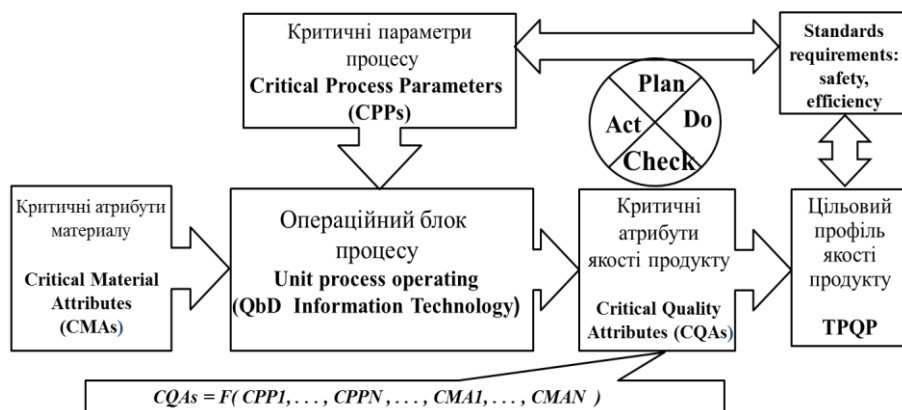


Рис. 1. Структура зв'язку категорій критичних атрибутів якості з TPQP згідно з концепцією QbD та удосконалення процесу забезпечення якості за циклом PDCA

Відповідно до вищевикладеного завдання забезпечення якості продукції на етапі первинного проектування процесу виробництва конкретизуємо як визначення CPPs і їх об'єктів, які суттєво впливають на CQAs. Запропоновано рішення цього завдання реалізувати послідовним застосуванням таких методів MSA:

- визначення структури та ієрархії часових багатовимірних даних: CPPs і CQAs;
- визначення якісної та кількісної міри відношення між сформованими об'єктами (факторами) CPPs і CQAs.

Разом з комп'ютерно-інтенсивними інформаційними технологіями (computer-intensive information technologies) статистичного канонічного аналізу [9], факторизовані багатовимірні дані критичних атрибутів якості дають додаткові принципові можливості для багатовимірного статистичного аналізу впливу CPPs на CQAs тобто дозволяють урахувувати додаткову змінну у вигляді факторних даних.

Відомий метод Г. Тагути (Genichi Taguchi) вирішення подібної задачі шляхом дробного факторного планування експерименту з суб'єктивними тестовими наборами параметрів проектування і факторів завод. Ці дані формують матрицю ортогональних розташувань [10]. За результатами впливу факторів завод на групи ортогональних параметрів проектування оцінюють їх робастність і аналізують можливі втрати якості за квадратичною функцією.

На відміну від такого рішення запропоновано застосувати MSA і комп'ютерно-інтенсивні об'єктно-орієнтовані інформаційні технології [11] для проведення саме статистичного повнофакторного експерименту з безпосередньою кластеризацією кожної з таблиць спостережень після проведення розвідувального аналізу даних. При цьому статистичні оцінки багатовимірного кластерного, а потім і факторного аналізу адекватно доповнюються оцінками компонентного впливу CPPs на CQAs.

Запропонований в статті метод MSA критичних атрибутів якості процесу виробництва з факторизацією даних доцільно розглядати як розвиток

концепції QbD забезпечення якості продукту на етапі проектних розробок. Метод спрямований на експериментальне дослідження ступеня впливу CPPs на CQAs, що дозволить застосовувати результати у процесі виробництва.

### Постановка проблеми

Статистика експерименту оперує даними об'єктно-орієнтованих  $n$  спостережень, сформованих в два масиви випадкових багатовимірних змінних (величин)  $X$  і  $Y$ ,  $X \rightarrow Y$ .

$X$  – близький до постійного або випадково варійований масив спостережень змінних процесу виробництва. Компоненти (складові)  $X$  – предиктори, екзогенні (пояснюючі) змінні – критичні параметри процесу виробництва – CPPs.

$X^j = \{x_1, \dots, x_m\}$  –  $m$ -мірний вектор  $j$ -го ( $j = 1, 2, \dots, n$ ) спостереження  $m$  критичних параметрів процесу або  $j$ -те значення випадкової багатовимірної величини  $X \supset \{X^j\}$ .

Випадкова величина  $X$  потенційно визначає властивості випадкової величини  $Y$ , де  $Y$  – масив з  $n$  спостережень випадкових багатовимірних відгуків процесу виробництва.

Компоненти (складові)  $Y$  – ендегенні, залежні змінні – атрибути критичного якості продукту – CQAs.

Категорія  $X \rightarrow Y$  описується функцією якості процесу виробництва –  $CQAs = F(CPP_i)$ ,  $CQAs \supset \{CQA_k\}$ .

Спостереження  $Y^j = \{y_1, \dots, y_k\}$  ( $k$  – число складових атрибутів критичного якості продукту) однозначно відповідають спостереженням  $X^j = \{x_1, \dots, x_m\}$   $m$  критичних параметрів процесу і складають масив  $Y \supset \{Y^j\}$ .

З формальної точки зору результати моніторингу (вибірки) випадкових багатовимірних величин  $X$  і  $Y$  є даними простору розробки QbD, проте не є

категоризованим інформаційним простором відношення  $X \rightarrow Y$ .

Для визначення і характеристики каузальних відносин  $X \rightarrow Y$  в інформаційному просторі змінних  $X$  і  $Y$  необхідно досліджувати структуру та ієрархію їх даних, а також якісну і кількісну міри відношення між ними.

Об'єктивною проблемою вирішення цього завдання є те, що емпіричні масиви часових даних критичних атрибутів якості  $X$  і  $Y$  мають комп'ютерний формат у вигляді часових взаємопов'язаних але не структурованих матриць, які не підготовлені для статистичного аналізу.

Важливою для практики проблемою є встановлення зв'язку між змінністю критичних атрибутів якості та дією зовнішніх і внутрішніх факторів, явний вплив яких прагнуть мінімізувати у процесі виробництва. З практичної точки зору це ускладнює формування факторних змінних у фреймі досліджуваних даних, без яких каузальний аналіз буде неповним. Між тим змінність параметрів має місце, що робить фактори об'єктом прихованого впливу і потребує дослідження методів аналізу цього впливу.

Запропонований метод пропонує варіант вирішення цих проблем послідовним застосуванням інструментів MSA.

#### Аналіз останніх досліджень і публікацій

Процесна методологія QFD [4, 12], метод досягнення TRQP за допомогою детального проектування QbD [2, 5, 13] і ризик орієнтований процесний підхід [4, 14, 3] – є домінантами сучасних систем менеджменту якості [7].

Високі вимоги до якості продукції і посилення конкуренції призводять до необхідності забезпечення якості продукції на стадії проектних розробок QbD [13, 5] при процесній декомпозиції функції якості (QFD) [4]. Адекватним інструментом QbD є аналітичні методи PAT [6, 15], в основу яких покладені інформаційні технології MSA. На етапі проектування методи PAT сфокусовані на дослідженні багатфакторних зв'язків між CPPs і CQAs, факторами середовища і їх впливом на якість кінцевого продукту [12, 16].

В роботі [17] наведено актуальний систематичний огляд великої кількості джерел з тематики застосування методів MSA. Відомі методи визначення номінальних значень CQAs, CPPs і допусків на них пропонують статистичні рішення вибору багатовимірних змінних (VS variable selection). Статистичні методи VS (по суті методи MSA) досліджують внесок екзогенних факторів в змінювання значень функції якості (багатовимірної ендегенної змінної).

Таким чином, методи MSA створюють адекватну аргументовану основу для перетворення функціональних вимог у параметри проектування та обґрунтування їх допусків (DP, design parameter) [18]. Адекватні CPPs і CMAs визначаються на основі

прийнятної зміни даних функції якості і чутливості цієї функції до змінювання значень CPPs і CMAs.

Дана проблематика також може розглядатися з позицій системної інженерії [19, 20, 21], коли моделі відіграють ключову роль на одному або декількох етапах процесу розробки і є частиною проектування систем або процесів на основі MBSE (model based systems engineering) [22]. Основні етапи дослідження та конструювання продукту/процесу виробництва: від загальних вимог – до системи [21] в рамках робастного проектування [23] неминуче приводять до архітектури (структура + логіка) проектування/експерименту [16, 21, 24]. Однією із загальних тенденцій цієї логіки є декомпозиція функції якості (QFD) [3, 4, 18] з використанням технології дизайну експерименту DoE (Design of Experiments) [19] подібно методології ретельного планування експерименту (QdD) [2, 3, 5] з інструментами PAT [6].

Важливо відзначити також одну з особливостей сучасного проектування, яка полягає у використанні технологій (наприклад, методів MSA) безпосередньо не пов'язаних з розробкою (NDI, non-development item) [24]. На рівні проектування і проведення експериментів цей напрямок розвивається в зв'язку з інтенсивним впровадженням інформаційних технологій об'єктно-орієнтованого MSA і високої довіри інженерів до розроблених і апробованих програмних пакетам.

Таким чином, огляд літератури дозволяє акцентувати такі особливості завдання дослідження:

- визначення вимог до методів технологічного наближення до заданих значень атрибутів CQAs, як певної функції якості від значень CPPs, проводиться на етапі первинного проектування процесу виробництва;

- гарантована і сертифікована якість продукту на ієрархічному рівні первинного проектування процесу виробництва досягається за залученням інформаційних технологій комп'ютерно-інтенсивного багатовимірного статистичного аналізу (MSA) даних критичних атрибутів якості процесів і продукту в багатовимірному просторі їх проектування.

#### Формулювання мети статті

Мета статті – навести результати застосування запропонованого методу визначення особливостей і характеристик залежності змінювання критичних атрибутів якості продукту CQAs ( $Y$ ) від зміни критичних параметрів процесу виробництва CPPs ( $X$ ).

Завдання дослідження:

1. Попередня обробка даних у вигляді статистичних процедур розвідувального аналізу (exploratory data analysis), кластеризації, факторизації та інтерпретації отриманих результатів для визначення структури та ієрархії часових багатовимірних даних критичних параметрів процесу виробництва CPPs ( $X$ ) та критичних атрибутів якості продукту CQAs ( $Y$ ).

2. Дослідження каузальних відносин  $X \rightarrow Y$  у інформаційному просторі дизайну (design space) змінних  $X$  і  $Y$  для визначення якісної та кількісної мір відношення між сформованими об'єктами зазначених параметрів за допомогою статистичного апарату багатовимірної ординації, а також використання процедури побудови дерев регресії з багатовимірним відгуком MRT (Multivariate Regression Trees).

Процедури 1 і 2 узагальнено формулюють завдання статистичного дослідження структури і ієрархії багатовимірних даних спостережень  $X$  і  $Y$ , а також якісної і кількісної мір відносин між ними.

На практиці спостереження екзогенних змінних  $X^j$  не є незалежними, а уявляють собою групи зв'язаних CPPs. Ці групи утворюють кластери з інтерпретованою залежністю змінних, для яких представляє профільний інтерес оцінювання впливу на багатовимірні відгуки процесу –  $Y^j$ .

Запропоновано провести кластеризацію часових випадкових багатовимірних даних CPPs ( $X^j$ ) і ототожнити їх з факторами впливу на багатовимірні відгуки процесу CQAs ( $Y^j$ ). Ці операції дозволяють

визначити потенційні закономірності і зв'язки між кластеризованими спостереженнями змінних  $X$  і масивом багатовимірних відгуків  $Y$ , а також розширюють можливості застосування MSA для дослідження залежності  $X \rightarrow Y$  як функції якості  $CQAs = F(CPPi)$ .

## Матеріали та методи

Для забезпечення стандартизованої якості продукції на етапі первинного проектування процесу виробництва завдання полягає у визначенні за даними  $n$  спостережень  $m$  критичних параметрів процесу виробництва –  $\{x_1, \dots, x_m\}$  і їх об'єктів (факторів), які суттєво впливають на  $k$  критичних атрибутів якості продукту –  $\{y_1, \dots, y_k\}$  (багатовимірний відгук процесу).

Методи кластерного аналізу об'єднують вибірки сукупності даних, яка спостерігається, в групи однорідних у певному сенсі об'єктів, які називають кластерами або класами, як показано на рис. 2.

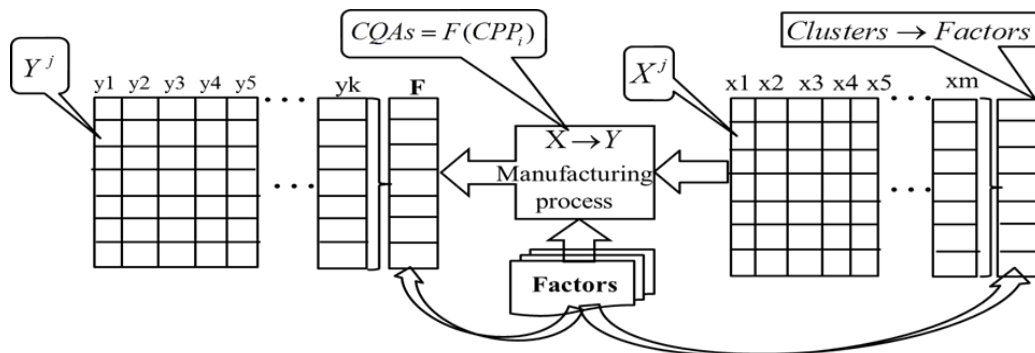


Рис. 2. Схема формування факторів часових багатовимірних даних  $X^j$  і  $Y^j$

Нехай кожне з спостережень  $X^j$  задається як точка –  $X_v^j$  в  $s$ -вимірному просторі ознак  $X_1^j, X_2^j, \dots, X_s^j$ . Сукупність цих точок можна трактувати як вибірку –  $q \times m$  ( $q=1, 2, \dots, n$  – число спостережень відповідних  $v$ -ій ознаці) з багатовимірної генеральної сукупності ознаки  $v$ .

Деяке число спостережень  $q$  ознаки  $v$  утворюють об'єкт (кластер), який є фактором однорідного впливу  $q$  спостережень на CQAs. Деякі спостереження  $Y^j$  групуються в об'єкти за цією ознакою шляхом проведення прямої ординації. За результатами такого аналізу будується діаграма – "ординаційний триплет", яка показує варіабельність критичних атрибутів якості продукту по відношенню до двох осей проектування CCA1-CCA2 (canonical correspondence analysis) і статистичні зв'язки з кожним критичним параметром процесу –  $x_m$ .

При дослідженні були використані алгоритми неієрархічного розподілу спостережень (partitioning algorithms) мови R, які здійснюють декомпозицію набору даних  $n$  спостережень на  $s$  кластерів із заздалегідь невідомими параметрами. При цьому

виконується пошук центроїдів тобто максимально віддалених один від одного центрів ущільнень точок з мінімальним розкидом всередині кожного кластера та з перевіркою якості кластеризації [11].

Взаємозв'язок між компонентами багатовимірних даних масивів і  $X$  і  $Y$  встановлювався в ході багатовимірного статистичного канонічного аналізу за допомогою пакетів прикладних програм мови R [11]. Зокрема був використаний статистичний апарат багатовимірної канонічної ординації: канонічний аналіз відповідностей CCA і неметричне багатовимірне шкалювання NMDS (nonmetric multidimensional scaling). Ці методи позиціонуються як розширення багатовимірного статистичного регресійного аналізу при моделюванні багатовимірного відгуку процесу виробництва (змінна  $Y^j$ ) на багатовимірне значення критичного параметра процесу (змінна  $X^j$ ).

Отже, послідовне виконання визначених процедур MSA даних CPPs і CQAs складає метод багатовимірного каузального аналізу, який дає можливість проведення повнофакторного експерименту для дослідження ступеня впливу CPPs на CQAs.



У даному випадку експеримент за запропонованим методом передбачає дослідження причинно-наслідкових або каузальних, залежностей багатовимірних змінних  $X^j$  і  $Y^j$ , для чого повинні виконуватися три умови [25]:

1. Події і факти, пов'язані зі змінною  $X^j$  повинні передувати подіям і фактам, пов'язаним із змінною  $Y^j$ .

2. Зміна даних  $X^j$  повинна супроводжуватися зміною даних  $Y^j$ .

3. Повинно бути виключено вплив третьої змінної, яка передбачає альтернативне, конкуруюче пояснення емпіричних спостережень і статистично підтверджений зв'язок між змінними  $X^j$  і  $Y^j$ .

Таким чином, на основі вищевикладеного запропонований метод аналізу каузальних відносин  $CPPs \rightarrow CQAs$ , який складається з наступних кроків:

1. Проведення розвідувального аналізу даних екзогенних змінних  $X^j$ : визначення корельованих змінних, оцінка тенденції кластеризації.

2. Проведення кластеризації пояснюючих змінних  $X^j$  і оцінка якості цієї кластеризації.

3. Трансформація однорідних таблиць даних спостережень змінних  $X^j$  і  $Y^j$  у таблиці з факторними змінними.

4. Проведення та оцінка непрямої ординації відгуків  $Y^j$  і прямої ординації відгуків  $Y^j$  за факторами екзогенних змінних масиву  $X$ .

4а. Порівняння результатів непрямої і прямої ординації багатовимірних відгуків  $Y^j$  (без урахування і з урахуванням факторів екзогенних змінних  $X^j$ ).

4б. Порівняння оцінок прямої ординації відгуків  $Y^j$  з оцінками канонічного аналізу методами CCA і NMDS.

4с. Побудова дерев регресії MRT (Multivariate Regression Trees) масиву екзогенних змінних  $X$  з багатовимірним відгуком  $Y$ .

5. Оцінка результатів багатовимірному статистичного аналізу зв'язку масиву екзогенних змінних  $X$  з масивом багатовимірних відгуків  $Y$ .

У загальному випадку даний алгоритм відповідає також методології SEMMA (Sample, Explore, Modify, Model, and Assess) [26] визначення фізичних закономірностей при статистичному аналізі. SEMMA орієнтована на структурування процесів (подібно QFD), а також застосування методів статистичного аналізу та візуалізації результатів на кожному з етапів процесу і дає розробнику свободу вибору концепції і реалізації досліджень.

### Результати досліджень

Наведені результати відповідають багатовимірному статистичному аналізу масивів часових даних предикторів  $X$  і відгуків  $Y$ , заданих у вигляді рівномірно розподілених випадкових величин простору розробки в межах допусків передбачених стандартом.

Розвідувальний аналіз даних використовувався для описової статистики властивостей вибірок. Цей аналіз включав: обробку та систематизацію емпіричних даних, наочне представлення багатовимірних даних у формі таблиць і графіків, а також визначення кількісних характеристик основних статистичних показників.

Оцінка ваги або "важливості" предикторів  $X$  (variable importance) і "схильність" їх до кластеризації з урахуванням міри мультиколінеарності оцінювалися за допомогою методу головних компонент PCA (principle components analysis) і візуалізації тенденцій утворення кластерів VAT (Visual Assessment of cluster Tendency), як показано відповідно на рис. 3, а, б.

Пошук оптимальної схеми об'єднання предикторів  $X^j$  у кластери проводився перебором різних комбінацій числа груп, метрик дистанцій і методів кластеризації за допомогою 30 індексів якості. За результатами обчислень було вибрано оптимальне число кластерів, як показано на рис. 3, б.

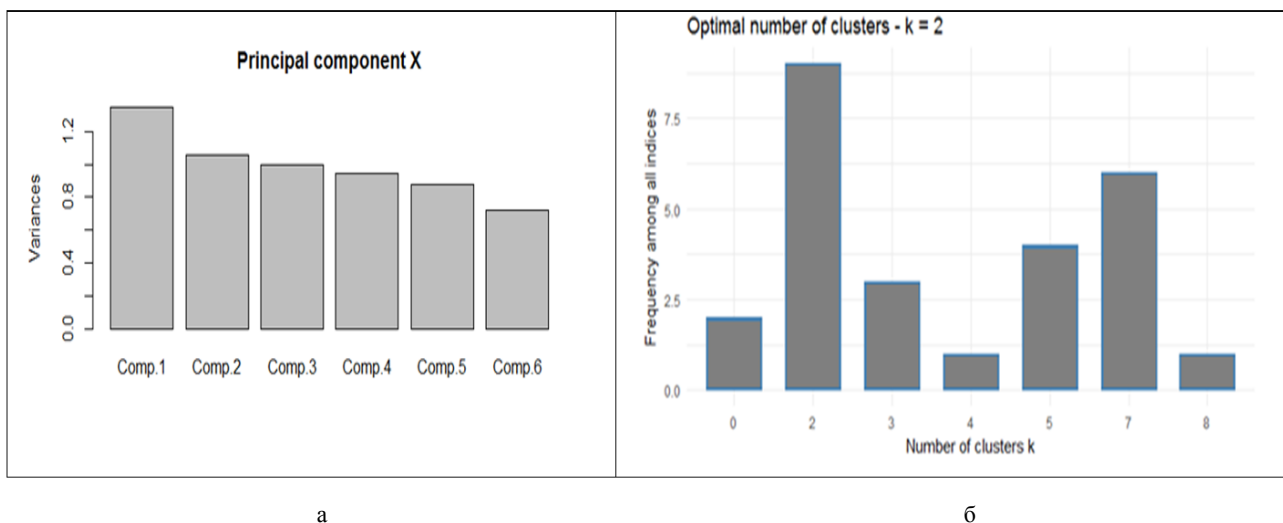


Рис. 3. Результат оцінки ваги компонент предикторів – а. Результат визначення оптимального числа кластерів за оцінками різних індексів – б

Операція трансформації однорідних масивів чисельних змінних  $X$  і  $Y$  в таблиці з факторними (кластерними) ознаками проводилася шляхом отождолення фактора з кластером групи спостережень  $X^j$  і його присвоєння відповідній групі багатовимірних відгуків  $Y^j$ . В результаті формувалися таблиці (data.frame) з факторизованими масивами даних  $X$  і  $Y$ , як показано на рис. 2.

На відміну від непрямої (unconstrained) ординації, пряма (constrained) ординація дозволяє зв'язати змінність багатовимірних значень відгуків  $Y^j$  з впливом предикторів  $X^j$ . На рис. 4, а, б показана ординація складових багатовимірних відгуків  $Y$  на площині головних компонент (непряма ординація, рис. 4, а) і розподіл цих значень по об'єктах факторів масиву  $X$  на площині головних компонент PCA (пряма ординація, рис. 4, б).

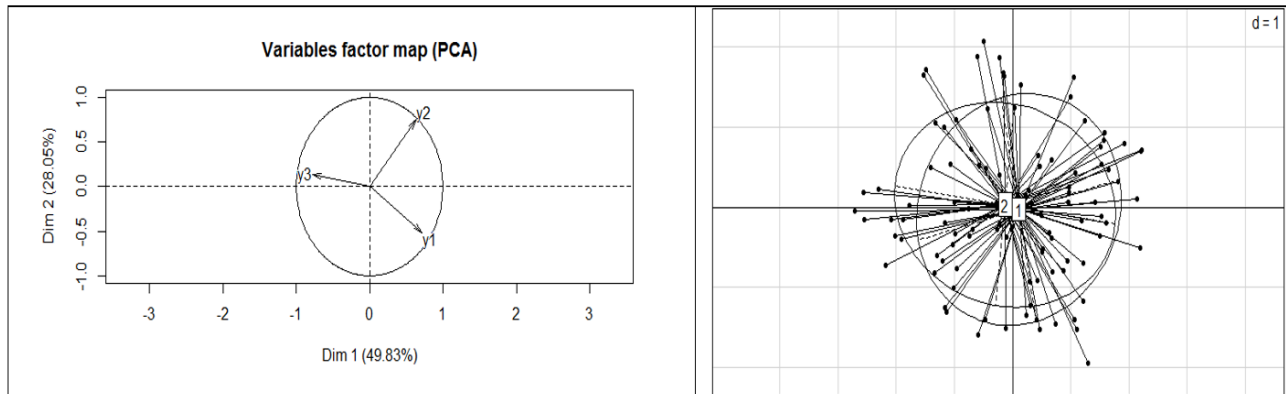


Рис. 4. Діаграми непрямої – а й прямої – б ординації масиву відгуків  $Y$  на площині головних компонент PCA

Непрямий ординаційний аналіз матриці відгуків  $Y$  дозволяє отримати відображення в ортогональній системі координат тільки деяких окремих структурних особливостей досліджуваного масиву даних у формі графічних проєкцій складових багатовимірних відгуків  $Y^j$  на площині головних компонент, як показано на рис. 4, а.

На діаграмі рис. 4, б показана угруповання спостережень  $Y^j$  за факторними ознаками на тлі пояснюють складових  $x_m$  (пунктирні лінії) зв'язаних з власними векторами масиву  $X$  (ординаційний триплот).

Відзначимо, що використання ординації PCA представляє собою метод згладжування випадкових флуктуацій даних і сприяє побудові більш стабільних (робастних) кластерних структур. Застосування прямої ординації дозволяє поєднати операції ієрархічного утворення об'єктів по розділяючим предикторам масиву  $X$  і проєціювання багатовимірних

даних масиву відгуків  $Y$  у просторі головних компонент.

Для оцінювання компонентного взаємозв'язку між двома масивами багатовимірних змінних  $X$  і  $Y$  використовувалися статистичні процедури канонічної ординації: канонічний аналіз відповідностей (CCA) і неметричне багатовимірне шкалювання (NMDS).

Наведемо результати NMDS, коли визначається найменш перекручена метрика дистанцій між двовимірними проєкціями значень  $Y^j$  на осі головних компонент. Дана метрика відстані визначається за коефіцієнтом кореляції Спірмена для кожного з варіантів метрики дистанцій у просторах відгуків  $Y^j$  і предикторів  $X^j$ . Результати NMDS: проєкції значень відгуків  $Y^j$  на осі ординації NMDS1-NMDS2 і зв'язок з компонентами предикторів  $\{x_1, \dots, x_m\}$  показані на рис. 5.

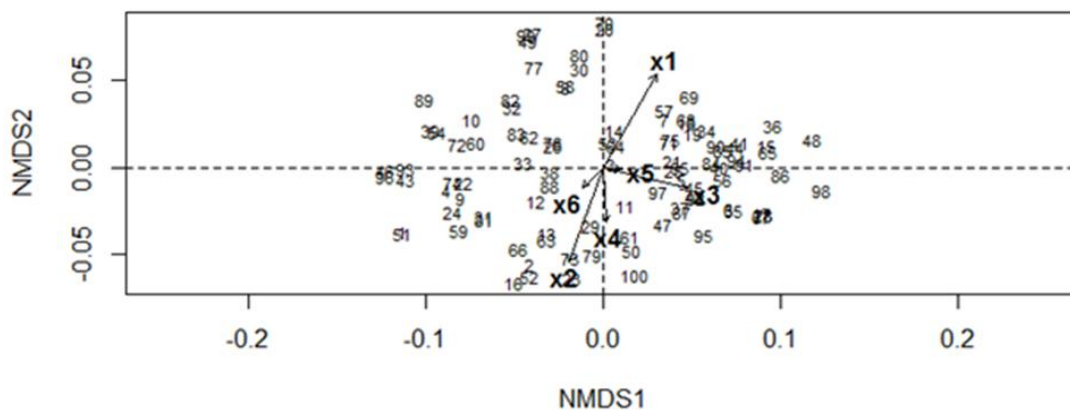


Рис. 5. Результати ординації значень відгуків  $Y^j$  на осі NMDS1-NMDS2 і зв'язок з компонентами предикторів  $\{x_1, \dots, x_m\}$

При побудові дерев регресії масиву екзогенних змінних  $X$  з багатовимірними відгуками  $Y$  (MRT) вирішується задача оцінювання впливу складових екзогенних змінних  $\{x_1, \dots, x_m\}$  на сукупну змінність складових багатовимірних відгуків  $\{y_1, \dots, y_k\}$ , як показано на рис. 6.

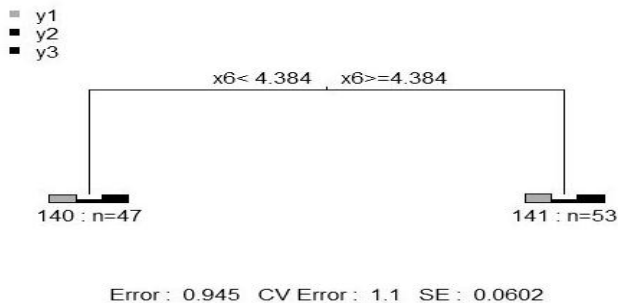


Рис. 6. MRT з багатовимірним відгуком і розділяючими компонентами масиву екзогенних змінних  $X$

MRT формуються в результаті рекурсивної операції поділу рядків масиву даних відгуків  $Y$  на однорідні підмножини з використанням масиву екзогенних змінних  $X$  як розділяючих параметрів.

Дерева MRT дають візуалізацію впливу складових екзогенних змінних  $\{x_1, \dots, x_m\}$  на формування масиву багатовимірних відгуків  $Y$  і показують, які складові масиву  $X$  ініціюють розбиття досліджуваної сукупності  $Y$  в вузлах регресійного дерева а також визначають склад сформованих об'єктів  $Y$ .

Результати свідчать про значні можливості MSA при аналізі часових багатовимірних даних для

Таблиця 1. Параметри очікуваного внеску компонент  $x_m$  у загальну варіацію даних масиву  $X$

Parameters	Comp.	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation		1.1597	1.0287	0.9966	0.9712	0.9378	0.8486
Proportion of Variance		0.2264	0.1781	0.1672	0.1588	0.1480	0.1212
Cumulative Proportion		0.2264	0.4046	0.5718	0.7306	0.8787	1.0000

В основу критерію результативності кластеризації покладено метрична гіпотеза компактності. Якщо вдається знайти таке розбиття спостережень на групи, що відстані між спостереженнями з однієї групи (intra-cluster distances) будуть менше деякого значення  $\varepsilon > 0$ , а між спостереженнями з різних груп (cross-cluster distance) – більше  $\varepsilon$ , то кластеризація успішна. За цим критерієм отримані результати оцінки якості кластеризації за тестом Загоруйко для кластерів масиву багатовимірних даних  $X$  (рис. 3б), як показано в табл. 2.

У табл. 2 по головній діагоналі розміщені середні внутрішньокластерні відстані, які менше, ніж міжкластерні відстані (недіагональні елементи таблиці), що говорить про допустимість розбиття простору значень масиву  $X$  на два кластери. Операція

практичного застосування забезпечення якості продукції на етапі первинного проектування процесу виробництва.

### Обговорення результатів дослідження

У статті представлені результати дослідження на етапі проектування процесу виробництва, в ході якого проводився багатовимірний статистичний аналіз причинно-наслідкових зв'язків між змінністю компонентів відгуку процесу і дією певних факторів, які ототожнюються з технологічними параметрами процесу. Дослідження містить елементи розвідувального, описового та аналітичного статистичного аналізу.

Оцінка "важливості" предикторів  $\{x_1, \dots, x_m\}$  (variable importance) і тенденція до кластеризації масиву  $X$  з урахуванням можливої мультиколінеарності компонент оцінювалися за допомогою методу головних компонент PCA і візуалізації тенденцій утворення кластерів (VAT). Метод PCA дає значення відносного очікуваного вкладу кожної компоненти (Comp.m) у загальну варіацію даних, як показують дані в табл. 1.

Відзначимо, що всі компоненти багатовимірних масиву  $X$  характеризуються практично рівним внеском (proportions of variance 16.6%) в очікуваний розкид і разом дають сумарний внесок 100%. Доступна в цьому тесті оцінка тенденції кластеризації характеризується статистикою Хопкінса (The Hopkins statistic, 0.5087), яка показує помірну схильність до утворення кластерів при випадковому розкиді даних.

трансформації однорідної таблиці спостережень ендogenous змінних  $Y$  в таблицю з факторними змінними була проведена після дослідження особливостей цього масиву.

Таблиця 2. Результати оцінки якості кластеризації за тестом Загоруйко

№	Cl. dist.	Cl. dist. 1	Cl. dist. 2
1		3.044665	3.623451
2		3.623451	3.009617

Результати проєкції в ортогональній системі координат головних компонент деяких внутрішніх особливостей зв'язку складових масиву багатовимірних відгуків  $Y$  (непряма ординація) показані в табл. 3.

**Таблиця 3.** Результати непрямой ординації масиву даних  $Y$  на осі головних компонент

Dim.	Dim. 1	Dim. 2	Dim. 3
$y_1$	43.56287	10.7237730	45.71336
$y_2$	8.30943	88.6417708	3.04880
$y_3$	48.12770	0.6344562	51.2378

З табл. 3 видно, що складові масиву відгуків  $Y = \{y_1, \dots, y_k\}$  за ступенем свого вкладу в компоненти розкладання Dim1, Dim2, Dim3 фактори не утворюють і мають слабку кореляцію. Непряма ординація обмежується тільки оцінкою розподілу складових  $\{y_1, \dots, y_k\}$  на площині головних компонент і інтерпретацією вкладу кожної складової. Пряма ординація дозволяє встановити відповідність змінювання багатовимірних значень відгуків  $Y$  з впливом екзогенних змінних  $\{x_1, \dots, x_m\}$ .

Пряма ординація масиву відгуків (рис. 4б) показана на тлі складових  $\{x_1, \dots, x_m\}$  (пунктирні лінії) і пов'язує внутрішні закономірності відгуків  $Y^j$  з факторизованим впливом критичних параметрів процесу  $Y^j$ . Попадання відгуків  $Y^j$  у обидва факторних об'єкти свідчить про слабку кореляцію складових  $\{y_1, \dots, y_k\}$  і переважання випадкових впливів компонент  $\{x_1, \dots, x_m\}$  на змінність складових відгуків. Цей факт обумовлює необхідність провести дослідження компонентного впливу пре дикторів  $\{x_1, \dots, x_m\}$  на змінність складових багатовимірної залежної випадкової величини  $Y^j$ .

Результат переважання впливу випадкової складової варіабельності компонент  $\{x_1, \dots, x_m\}$  на компоненти багатовимірних відгуків  $Y^j$  підтверджується неметричним багатовимірним шкалюванням (NMDS) проєкції значень масиву  $Y$  на осі ординації NMDS1-NMDS2.

Метод NMDS найбільш застосовується для масивів даних з істотним впливом випадкових або неврахованих факторів, як у даному випадку. Результати застосування методу NMDS зведені в табл. 4.

**Таблиця 4.** Результати NMDS масиву відгуків  $Y$  на осі NMDS1-NMDS2 і зв'язок з предикторами  $\{x_1, \dots, x_m\}$ 

NMDS	NMDS1	NMDS2	r2	Pr(>r)
$x_1$	0.48665	0.87360	0.0329	0.201
$x_2$	-0.33369	-0.94268	0.0292	0.239
$x_3$	0.97049	-0.24114	0.0213	0.327
$x_4$	0.06534	-0.99786	0.0087	0.665
$x_5$	0.99269	-0.12071	0.0010	0.960
$x_6$	-0.71850	-0.69552	0.0024	0.900

З табл. 4 випливає, що при низькій статистичній значущості впливу екзогенних змінних на відгуків  $Y^j$

має місце висока кореляція деяких складових  $\{x_1, \dots, x_m\}$  з осями ординації (наприклад,  $x_5$  з NMDS1 і  $x_4$  з NMDS2).

При інтерпретації (рис. 5) необхідно мати на увазі, що на результат неметричного шкалювання багатовимірних відгуків  $Y^j$  накладаються складові екзогенних змінних  $\{x_1, \dots, x_m\}$  у залежності від розрахованих коефіцієнтів кореляції кожної зі складових  $x_m$  з осями ординації NMDS1-NMDS2. З рис. 5 видно, що модель NMDS навіть при невеликій кількості компонент багатовимірних даних  $X$  і  $Y$  дає тільки умовне уявлення про вплив екзогенних змінних  $\{x_1, \dots, x_m\}$  на відгуків  $Y^j$ .

Деталізація впливу компонент  $\{x_1, \dots, x_m\}$  здійснюється шляхом проєкції на ординацію відгуків  $Y^j$  ізольованих згладжених компонент  $x_m$  за даними масиву  $X$ . Відзначимо, що при збільшенні числа компонент а також при слабкій тенденції значень масивів  $X$  і  $Y$  до кластеризації аналіз ускладнюється.

На заключному етапі аналізу було сформовано MRT масиву екзогенних змінних  $X$  з багатовимірним відгуком  $Y$ , як показано на рис. 6. Для даного MRT "листя" є кластери об'єктів  $Y$  з мінімальними відмінностями між даними в багатовимірному просторі у межах кожного кластера. Зауважимо, що для пояснення впливу змінних  $x_m$  на відгуки  $Y^j$  програмний пакет MRT має опцію вибору різного числа кластерів. Для даного MRT (рис. 6) розділюючою змінною є компонента  $x_6$ , яка утворює рівнозначні кластери з практично рівним числом спостережень (47 і 53) при відсутності явного домінування складових  $y_k$ .

Обговорення показує, що експериментальні дані комп'ютерного формату потребують на процедури обробки, аналізу та інтерпретації, які визначають потенційно можливі закономірності і залежності між окремими компонентами і об'єктами часових багатовимірних даних з метою практичної інтерпретації і застосування на етапі проектних розробок процесу виробництва.

## Висновки

У статті запропоновано та розкрито метод вирішення актуального завдання забезпечення якості продукції на етапі первинного проектування процесу виробництва відповідно до сучасної процесно-аналітичної методології – "якість через дизайн" QbD.

Наукова новизна запропонованого методу полягає в послідовному використанні процедур кластеризації і факторизації критичних атрибутів якості для аналізу характеру і міри впливу критичних параметрів виробництва на критичні атрибути якості товару. При цьому важливо, що кластеризація критичних параметрів виробництва ототожнюється з факторами зовнішнього і внутрішнього впливу на



процес виробництва, який виражається в змінності досліджуваних параметрів.

Ідея конструювання функції якості процесу виробництва на кожному з його етапів з метою отримання передбачуваних значень атрибутів критичної якості продукту при варіабельності значень критичних параметрів процесу в межах допусків реалізується за допомогою методів багатовимірного статистичного аналізу. Ці методи застосовуються в ході проведення статистичних експериментів з багатовимірними масивами критичних атрибутів якості для забезпечення якості продукції вже на етапі первинного проектування процесу виробництва.

Результати проведених досліджень є наслідком комбінування каузальних методів багатовимірного статистичного аналізу, які дозволяють отримувати кількісні показники з адекватною практичною інтерпретацією впливу критичних параметрів процесу на критичні атрибути якості продукту.

Перспектива подальших досліджень полягає в синтезі або застосуванні робастних методів забезпечення стандартних показників критичних атрибутів якості продукції, які стійкі до впливу різних факторів середовища і до варіабельності критичних параметрів процесу виробництва.

### Список літератури

- Deming W. E. Out of the crisis. Cambridge: Cambridge University Press, 1986. 507 p.
- Lionberger R. A., Lee S. L., Lee L. M., Raw A., Lawrence X. Yu. Quality by Design: Concepts for ANDAs. *American Association of Pharmaceutical Scientists journal*. Jun. 2008. No. 10 (2). P. 268–276. DOI: <https://doi.org/10.1208/s12248-008-9026-7>.
- ICH Q8. Pharmaceutical Development. URL: <http://www.ich.org/products/guidelines/quality/quality-single/article/pharmaceutical-development.html> (Last accessed: 12.08.2018).
- Eshan S. J. A Case Study on Quality Function Deployment (QFD). *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)*. Nov-Dec. 2012. Vol. 3. Issue 6. P. 27–35. DOI: <https://doi.org/10.9790/1684-0362735>.
- Zhang L., Mao S. Application of quality by design in the current drug development. *Asian journal of pharmaceutical sciences*. 12 (2017). No. 103. P. 1–8. DOI: <https://doi.org/10.1016/j.ajps.2016.07.006>.
- Guidance for Industry PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. URL: <http://www.fda.gov/cder/OPS/PAT.htm> (Last accessed: 12.08.2018).
- ISO 9001: 2015. Quality management systems. URL: <https://www.iso.org/standard/62085.html> (Last accessed: 02.08.2018).
- ISO 22000: 2005. Food safety management systems. URL: <https://www.iso.org/standard/35466.html> (Last accessed: 02.08.2018).
- Legendre P. Short course on Advanced spatial ecology. Research Center of Plant Ecology and Biodiversity Conservation. Institute of Botany, Chinese Academy of Sciences. 1-5 July 2006. P. 575–600. URL: [http://biol09.biol.umontreal.ca/PLcourses/Legendre\\_Canonical.pdf](http://biol09.biol.umontreal.ca/PLcourses/Legendre_Canonical.pdf) (Last accessed: 12.08.2018).
- Rao S., Samant P., Kadampatta A., Shenoy R. An Overview of Taguchi Method: Evolution, Concept and Interdisciplinary Applications. *International Journal of Scientific & Engineering Research*. October 2013. Vol. 4, Issue 10. P. 621–626. URL: <https://www.ijser.org/onlineResearchPaperViewer.aspx?An-Overview-of-Taguchi-Method-Evolution-Concept-and-Interdisciplinary-Applications.pdf> (Last accessed: 02.08.2018).
- Coghlan A. A Little Book of R for Multivariate Analysis, Release 0.1. 19 2017. 47p. URL: <http://a-little-book-of-r-for-time-series.readthedocs.org/> (Last accessed: 12.08.2018).
- Food and Drug Administration. Final Report on Pharmaceutical cGMPs for the 21st Century – A Risk Based Approach. URL: <https://www.fda.gov/downloads/drugs/developmentapprovalprocess/manufacturing/questionsandanswersoncurrentgoodmanufacturingpracticescgmppfordrugs/ucm176374.pdf> (Last accessed: 12.08.2018).
- Kannissery P., Tahir M. A., Charoo N. A., Ansari S. H., Ali J. Pharmaceutical product development: A quality by design approach. *International Journal of Pharmaceutical Investigation*. Jul-Sep 2016. No. 6 (3). P. 129–138. DOI: <https://doi.org/10.4103/2230-973X.187350>.
- ISO 31000 Risk management. URL: <https://www.iso.org/iso-31000-risk-management.html> (Last accessed: 02.08.2018).
- Guidance for Industry. PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. URL: <https://www.fda.gov/downloads/drugs/guidances/ucm070305.pdf> (Last accessed: 02.08.2018).
- Mohamed I. Progressive Modeling: The Process, the Principles, and the Applications. *Procedia Computer Science*. 2013. Vol. 16. P. 39–48. DOI: <https://doi.org/10.1016/j.procs.2013.01.005>.
- Araujo F. Variable selection methods in multivariate statistical process control: A systematic literature review. Department of Industrial Engineering: Federal University of Rio Grande do Sul, 90035-190 Porto Alegre, RS, Brazil. *Computers & Industrial Engineering*. January 2018. Vol. 115. P. 603–619. DOI: <https://doi.org/10.1016/j.cie.2017.12.006>.
- Göhler S. M., Husung S., Howard T. J. The Translation between Functional Requirements and Design Parameters for Robust Design. *Procedia College International pour la Recherche en Productique (CIRP)*. 2016. Vol. 43. P. 106–111. DOI: <https://doi.org/10.1016/j.procir.2016.02.028>.
- Gausemeier J., Gaukster T., Tschirner C. Systems Engineering Management Based on a Discipline-Spanning System Model. *Procedia Computer Science*. 2013. Vol. 16. P. 303–312. DOI: <https://doi.org/10.1016/j.procs.2013.01.032>.
- MacCalman A., Kwak H., McDonald M., Upton S. Capturing Experimental Design Insights in Support of the Model-based System Engineering Approach. *Procedia Computer Science*. 2015. Vol. 44. P. 315–324. DOI: <https://doi.org/10.1016/j.procs.2015.03.030>.
- Kaiser L., Bremer C., Dumitrescu R. Exhaustiveness of Systems Structures in Model-Based Systems Engineering for Mechatronic Systems. *Procedia Technology*. 2016. Vol. 26. P. 428–435. DOI: <https://doi.org/10.1016/j.protcy.2016.08.055>.
- MacCalman A., Lesinski G., Goerger S. Integrating External Simulations Within the Model-Based Systems Engineering Approach Using Statistical Metamodels. *Procedia Computer Science*. 2016. Vol. 95. P. 436–441. DOI: <https://doi.org/10.1016/j.procs.2016.09.309>.

23. Lemazurier L., Chapurlat V., Grossetête A. An MBSE Approach to Pass from Requirements to Functional Architecture. *International Federation of Accountants (IFAC): Papers OnLine*. July 2017. Vol. 50, Issue 1. P. 7260–7265. DOI: <https://doi.org/10.1016/j.ifacol.2017.08.1376>.
24. Clement S., Register A., Wise R. Key Enablers for Leveraging Non-Development Items in a System. *Procedia Computer Science*. 2015. Vol. 44. P. 164–173. DOI: <https://doi.org/10.1016/j.procs.2015.03.009>.
25. Cohen J., Cohen P., West S. G., Aiken L. S. Applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, New Jersey, US: Lawrence Erlbaum Associates Publishers. Third Edition. 2003. 703 p.
26. Azevedo A., Santos M. F. KDD, SEMMA and CRISP-DM: a parallel overview. IADIS European Conference on Data Mining: Amsterdam, the Netherlands, 22-27 July 2008. Proceedings of the IADIS European Conference on Data Mining, 2008. P. 182–185. URL: <https://pdfs.semanticscholar.org/7dfe/3bc6035da527deaa72007a27cef94047a7f9.pdf> (Last accessed: 12.08.2018).

## References

1. Deming, W. E. (1986), *Out of the crisis*, Cambridge University Press, Cambridge, 507 p.
2. Lionberger, R. A., Lee, S. L., Lee, L. M., Raw, A., Lawrence, X. Yu. (2008), "Quality by Design: Concepts for ANDAs", *American Association of Pharmaceutical Scientists journal*, No. 10 (2), P. 268–276. DOI: <https://doi.org/10.1208/s12248-008-9026-7>.
3. ICH Q8, "Pharmaceutical Development", available at: <http://www.ich.org/products/guidelines/quality/quality-single/article/pharmaceutical-development.html> (Last accessed: 02.08.2018).
4. Eshan, S. J. (2012), "A Case Study on Quality Function Deployment (QFD)", *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)*, Vol. 3, Issue 6, P. 27–35. DOI: <https://doi.org/10.9790/1684-0362735>.
5. Zhang, L., Mao, S. (2017), "Application of quality by design in the current drug development", *Asian journal of pharmaceutical sciences*, No. 103, P. 1–8. DOI: <https://doi.org/10.1016/j.ajps.2016.07.006>.
6. "Guidance for Industry PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance", available at : <http://www.fda.gov/cder/OPS/PAT.htm> (Last accessed: 02.08.2018).
7. ISO 9001: 2015, "Quality management systems", available at : <https://www.iso.org/standard/62085.html> (Last accessed: 02.08.2018).
8. ISO 22000: 2005, "Food safety management systems", available at : <https://www.iso.org/standard/35466.html> (Last accessed: 02.08.2018).
9. Legendre, P. (2006), "Short course on Advanced spatial ecology", Research Centre of Plant Ecology and Biodiversity Conservation. Institute of Botany, Chinese Academy of Sciences, P. 575–600, available at : [http://biol09.biol.umontreal.ca/PLcourses/Legendre\\_Canonical.pdf](http://biol09.biol.umontreal.ca/PLcourses/Legendre_Canonical.pdf) (Last accessed: 12.08.2018).
10. Rao, S., Samant, P., Kadampatta, A., Shenoy, R. (2013), "An Overview of Taguchi Method: Evolution, Concept and Interdisciplinary Applications", *International Journal of Scientific & Engineering Research*, Vol. 4, Issue 10, P. 621–626, available at : <https://www.ijser.org/onlineResearchPaperViewer.aspx?An-Overview-of-Taguchi-Method-Evolution-Concept-and-Interdisciplinary-Applications.pdf> (Last accessed: 12.08.2018).
11. Coghlan, A. (2017), *A Little Book of R for Multivariate Analysis*, Release 0.1, available at : <http://a-little-book-of-r-for-time-series.readthedocs.org/> (Last accessed: 12.08.2018).
12. "Food and Drug Administration. Final Report on Pharmaceutical cGMPs for the 21st Century – A Risk Based Approach", available at : <https://www.fda.gov/downloads/drugs/developmentapprovalprocess/manufacturing/questionsandanswersoncurrentgoodmanufacturingpracticescgmppfordrugs/ucm176374.pdf> (Last accessed: 12.08.2018).
13. Kannissery, P., Tahir, M. A., Charoo, N. A., Ansari, S. H., Ali, J. (2016), "Pharmaceutical product development: A quality by design approach", *International Journal of Pharmaceutical Investigation*, No. 6 (3), P. 129–138. DOI: <https://doi.org/10.4103/2230-973X.187350>.
14. ISO 31000, "Risk management", available at : <https://www.iso.org/iso-31000-risk-management.html> (Last accessed: 12.08.2018).
15. "Guidance for Industry. PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance", available at : <https://www.fda.gov/downloads/drugs/guidances/ucm070305.pdf> (Last accessed: 12.08.2018).
16. Mohamed, I. (2013), "Progressive Modeling: The Process, the Principles, and the Applications", *Procedia Computer Science*, Vol. 16, P. 39–48. DOI: <https://doi.org/10.1016/j.procs.2013.01.005>.
17. Araujo, F. (2018), "Variable selection methods in multivariate statistical process control: A systematic literature review", Department of Industrial Engineering: Federal University of Rio Grande do Sul, 90035-190 Porto Alegre, RS, Brazil, *Computers & Industrial Engineering*, Vol. 115, P. 603–619. DOI: <https://doi.org/10.1016/j.cie.2017.12.006>.
18. Göhler, S. M., Husung, S., Howard, T. J. (2016), "The Translation between Functional Requirements and Design Parameters for Robust Design", *Procedia College International pour la Recherche en Productique (CIRP)*, Vol. 43, P. 106–111. DOI: <https://doi.org/10.1016/j.procir.2016.02.028>.
19. Gausemeier, J., Gaukster, T., Tschirner, C. (2013), "Systems Engineering Management Based on a Discipline-Spanning System Model", *Procedia Computer Science*, Vol. 16, P. 303–312, DOI: <https://doi.org/10.1016/j.procs.2013.01.032>.
20. MacCalman, A., Kwak, H., McDonald, M., Upton, S. (2015), "Capturing Experimental Design Insights in Support of the Model-based System Engineering Approach", *Procedia Computer Science*, Vol. 44, P. 315–324. DOI: <https://doi.org/10.1016/j.procs.2015.03.030>.
21. Kaiser, L., Bremer, C., Dumitrescu, R. (2016), "Exhaustiveness of Systems Structures in Model-Based Systems Engineering for Mechatronic Systems", *Procedia Technology*, Vol. 26, P. 428–435. DOI: <https://doi.org/10.1016/j.protcy.2016.08.055>.
22. MacCalman, A., Lesinski, G., Goerger, S. (2016), "Integrating External Simulations Within the Model-Based Systems Engineering Approach Using Statistical Metamodels", *Procedia Computer Science*, Vol. 95, P. 436–441. DOI: <https://doi.org/10.1016/j.procs.2016.09.309>.
23. Lemazurier, L., Chapurlat, V., Grossetête, A. (2017), "An MBSE Approach to Pass from Requirements to Functional Architecture", *International Federation of Accountants (IFAC): Papers OnLine*, Vol. 50, Issue 1, P. 7260–7265. DOI: <https://doi.org/10.1016/j.ifacol.2017.08.1376>.

24. Clement, S., Register, A., Wise, R. (2015), "Key Enablers for Leveraging Non-Development Items in a System", *Procedia Computer Science*, Vol. 44, P. 164–173. DOI: <https://doi.org/10.1016/j.procs.2015.03.009>.
25. Cohen, J., Cohen, P., West, S. G., Aiken, L. S. (2003), *Applied multiple regression/correlation analysis for the behavioral sciences*, Mahwah, New Jersey, US: Lawrence Erlbaum Associates Publishers, Third Edition, 703 p.
26. Azevedo, A., Santos, M. F. (2008), "KDD, SEMMA and CRISP-DM: a parallel overview", IADIS European Conference on Data Mining : Amsterdam, the Netherlands, 22-27 July 2008, *Proceedings of the IADIS European Conference on Data Mining*, P. 182–185, available at : <https://pdfs.semanticscholar.org/7dfe/3bc6035da527deaa72007a27cef94047a7f9.pdf> (Last accessed: 12.08.2018).

Надійшла (Received) 20.08.2018

*Відомості про авторів / Сведения об авторах / About the Authors*

**Гаврилко Євген Володимирович** – доктор технічних наук, доцент, Державний університет телекомунікацій, професор кафедри комп'ютерної інженерії, м. Київ, Україна; e-mail: [dut.aspirantura@ukr.net](mailto:dut.aspirantura@ukr.net); ORCID ID: <https://orcid.org/0000-0002-6107-8918>.

**Гаврилко Евгений Владимирович** – доктор технических наук, доцент, Государственный университет телекоммуникаций, профессор кафедры компьютерной инженерии, г. Киев, Украина.

**Navrylko Yevhen** – Doctor of Sciences (Engineering), Associate Professor, State University of Telecommunications, Professor at the Department of Computer Sciences, Kyiv, Ukraine.

**Курченко Олег Анастасійович** – кандидат технічних наук, доцент, Державний університет телекомунікацій, завідувач кафедри управління інформаційною та кібернетичною безпекою, м. Київ, Україна; e-mail: [kuro1@ukr.net](mailto:kuro1@ukr.net); ORCID ID: <https://orcid.org/0000-0002-3507-2392>.

**Курченко Олег Анастасиевич** – кандидат технических наук, доцент, Государственный университет телекоммуникаций, заведующий кафедрой управления информационной и кибернетической безопасности, г. Киев, Украина.

**Kurchenko Oleg** – PhD (Engineering Sciences), Associate Professor, State University of Telecommunications, Head at the Department of Management of Information and Cyber Security, Kyiv, Ukraine.

**Терещенко Ігор Володимирович** – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри інфокомунікаційної інженерії, м. Харків, Україна; e-mail: [iter@ukr.net](mailto:iter@ukr.net); ORCID ID: <https://orcid.org/0000-0002-6197-1914>.

**Терещенко Игорь Владимирович** – кандидат технических наук, доцент, Харьковский национальный университет радиоэлектроники, доцент кафедры инфокоммуникационной инженерии, г. Харьков, Украина.

**Tereshchenko Igor** – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Assistant Professor at the Department of Infocommunication Engineering, Kharkiv, Ukraine

**Терещенко Антон Ігорович** – Державний університет телекомунікацій, аспірант кафедри управління інформаційною та кібернетичною безпекою, м. Київ, Україна; e-mail: [vti62@ukr.net](mailto:vti62@ukr.net); ORCID ID: <https://orcid.org/0000-0002-3507-2392>.

**Терещенко Антон Игоревич** – Государственный университет телекоммуникаций, аспирант кафедры управления информационной и кибернетической безопасности, г. Киев, Украина.

**Tereshchenko Anton** – State University of Telecommunications, Post-graduate Student at the Department of Management of Information and Cyber Security, Kyiv, Ukraine.

## МЕТОД МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ВРЕМЕННЫХ МНОГОМЕРНЫХ КРИТИЧЕСКИХ АТРИБУТОВ КАЧЕСТВА ПРОЦЕССА ПРОИЗВОДСТВА С ФАКТОРИЗАЦИЕЙ ДАННЫХ

**Объектом** исследования является процесс обеспечения качества продукции на этапе первоначального проектирования процесса производства. **Предметом** исследования являются информационные технологии оценки факторного влияния критических параметров процесса производства (CPPs, critical process parameters) на критические атрибуты качества продукта (CQAs, critical quality attributes). **Цель** – применение предложенного метода для определения особенностей и характеристик зависимости изменения критических атрибутов качества продукта от изменения критических параметров процесса производства. **Задача**: определение структуры и иерархии временных многомерных данных критических параметров процесса производства и критических атрибутов качества продукта, а также качественной и количественной мер отношений между сформированными объектами указанных параметров. **Методы**. Последовательно используются: статистические процедуры разведывательного многомерного анализа данных; преобразование однородных наблюдаемых матриц значений CPPs и CQAs в таблицу с факторизованными данными; построение деревьев регрессии многомерных CPPs с многомерными ответами CQAs. Методы реализуют программное обеспечение пакетов языка R. Получены следующие **результаты**. Предложен метод решения задачи обеспечения качества продукции на этапе первоначального проектирования процесса производства в соответствии с актуальной процессно-аналитической технологии конструирования современных сертифицированных производств – "качество через дизайн" QbD (Quality-by-Design). Предложенный метод факторизации временных многомерных критических параметров процесса производства и критических атрибутов качества продукта позволяет увеличить число степеней свободы для многомерного статистического анализа MSA (Multivariate Statistical Analysis) влияния факторов (классов) критических параметров процесса производства на многомерные критические атрибуты качества продукта. Приведены и проанализированы результаты использования метода на примере массивов данных с применением пакетов объектно-ориентированного языка программирования R. **Выводы**. Предложенный метод статистического анализа совместно со статистическим многомерным каноническим анализом представляют актуальную

информационную технологию детального оценивания влияния объектов временных многомерных данных критических параметров процесса производства и отдельных составляющих на критические атрибуты качества продукта. Метод ориентирован на практическое применение с целью обеспечения качества продукции на этапе проектирования (совершенствования) производств.

**Ключевые слова:** качество через дизайн; атрибуты критического качества; критические параметры процесса; дизайн эксперимента; многомерный статистический анализ.

## THE METHOD OF MULTIVARIATE STATISTICAL ANALYSIS OF TIME MULTIVARIATE CRITICAL ATTRIBUTES OF MANUFACTURING QUALITY WITH DATA FACTORIZATION

The **object** of the study is the process of product quality assurance at the stage of the initial design of the manufacturing process. The **subject** matter is informational technologies for assessing the factor influence of critical parameters of the process of manufacturing (CPPs) on the critical quality attributes of a product (CQAs). The **goal** of the study is to use the method of multivariate statistical analysis for assessing the character and features of the influence of time multivariate critical process parameters on time multivariate critical quality attributes at the stage of designing the manufacturing process. The **task** of the study is to determine the structure and hierarchy of time multivariate data of CPPs and CQAs and to determine qualitatively and quantitatively the relationship among the formed objects of the specified parameters. The following **methods** were consistently used – statistical procedures of the exploratory analysis of multivariate data; transforming the homogeneous observed values matrices of CPPs and product CQAs into the data table with factorized data; deriving the regression trees of multivariate CPPs with multivariate responses (CQAs). The methods implement the software packages of the R language. The following **results** were obtained – the method to solve the problem of product quality assurance at the stage of designing the initial manufacturing process in accordance with the process-analytical technology for designing modern certified manufacturing standards such as QbD (Quality-by-Design) is suggested. The method uses the information technologies of multivariate statistical analysis (MSA) to evaluate the influence of time multivariate critical process parameters (CPPs) on the time product critical quality attributes (CQAs). Preparatory transformation of clusters of critical process (manufacture process) parameters into factors of product critical quality attributes was carried out. Factorized time multivariate CPPs enable using the methods of multivariate statistical analysis for assessing the impact of CPP factors on the time multivariate CQAs. **Conclusions.** This method of statistical analysis along with statistical multivariate canonical analysis present the up-to-date information technology for the detailed assessment of the influence of time multivariate CPP objects and some CPP components on CQAs. The method is oriented to practical application to assure the quality of products at the stage of designing (improving) manufacturing processes.

**Keywords:** quality-by-design; critical quality attributes; critical process parameters; the design of experiments; multivariate statistical analysis.

---