УДК 519.237.5

**С.Г. РАДЧЕНКО**

# АНАЛИЗ ЭКПЕРИМЕНТАЛЬНЫХ ДАННЫХ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ МНОГОФАКТОРНЫХ СТАТИСТИЧЕСКИХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

***Abstract:*** *The methods for obtaining of multifactorial statistical models of real systems are considered, which allow to interpret the reasonal, structural and quantitative relations between separate factors and criteria of the system quality. The general and the special properties of statistical models necessary for the analysis of the data with the use of the models are formulated. The classes of experimental designs for obtaining such models are given. Examples of creation of multifactorial statistical models and performing on their basis of the analysis of data about systems are presented.*

***Key words:*** *statistical models, plans of experiments, analysis of data.*

***Анотація:*** *Розглядаються методи отримання багатофакторних статистичних моделей реальних систем, які дозволяють інтерпретувати причинні, структурні та кількісні зв'язки між окремими факторами та критеріями якості системи. Сформульовано загальні та спеціальні властивості статистичних моделей, необхідні для аналізу даних. Перелічено класи планів експериментів для отримання таких моделей. Наведено приклади створення багатофакторних статистичних моделей і виконання на їх основі аналізу даних про системи.*

***Ключові слова:*** *статистичні моделі, плани експериментів, аналіз даних.*

***Аннотация:*** *Рассматриваются методы получения многофакторных статистических моделей реальных систем, которые позволяют интерпретировать причинные, структурные и количественные связи между отдельными факторами и критериями качества системы. Сформулированы общие и специальные свойства статистических моделей, необходимые для анализа данных. Перечислены классы планов экспериментов для получения таких моделей. Приведены примеры создания многофакторных статистических моделей и выполнения на их основе анализа исходных данных о системах.*

***Ключевые слова:*** *статистические модели, планы экспериментов, анализ данных.*

## 1. Введение

Математическое моделирование, применяемое для описания сложных систем, будучи универсальным средством разработки новых и совершенствования известных систем (объектов, процессов), может привести к принципиально новым прикладным результатам в производстве и эксплуатации изделий.

Математическое моделирование сложных систем, как правило, осуществляется на основе многофакторных статистических моделей (уравнений регрессии), полученных путем аппроксимации данных экспериментов, статистических испытаний, экспертных оценок, трудоемких вычислений на ЭВМ. Под данными будем понимать исходные условия (факторы) проведения статистического исследования в виде плана эксперимента и полученные результаты (критерии качества), упорядоченные в таблице по столбцам и строкам.

Применение упрощенных форм описания систем позволяет решать многие технические и технологические задачи, которые невозможно решить традиционными методами в необходимые сроки с доступными затратами физических ресурсов. Поэтому разработка методов получения статистических моделей, линейных относительно параметров и, в общем случае, нелинейных относительно факторов, является актуальной проблемой.

Статистические модели, получаемые формализованными методами, должны обладать также свойствами, которые позволяют проводить смысловой анализ причинных, структурных и количественных связей между комплексом начальных условий и критериями качества моделируемых систем, объектов, процессов.

Некоторые специалисты считают, что модели, полученные с использованием метода наименьших квадратов, являются формальными и не несут необходимую семантическую и эвристическую информацию о моделируемой системе (объекте, процессе). Это имеет место во многих решаемых задачах, если факторы мультиколлинеарны (взаимно сопряжены) между собой.

**2. Анализ исследований и публикаций**

Практика использования статистических многофакторных моделей при решении задач в различных предметных областях часто приводит к тому, что задача становится некорректно поставленной. Основные причины некорректности: исходная мультиколлинеарность факторов; не известная исследователю структура многофакторной статистической модели; ограниченные возможности в использовании физических ресурсов, необходимых для проведения экспериментов, подтверждающих правильность выбранной модели.

Демиденко Е.З., д.ф.-м.н., считает, что «мультиколлинеарность – одно из основных препятствий эффективного применения аппарата регрессионного анализа» [1, с. 186].

Петрович М.Л., к.т.н., анализируя получение статистических моделей в условиях мультиколлинеарности, приходит к выводу, что полученная модель может не иметь прикладного смысла, т.е. может быть не пригодна к использованию [2, с. 46–47, 189, 192, табл. П9].

Решение проблемы мультиколлинеарности видится авторами учебника «Эконометрика» так: «Однозначного ответа на этот вопрос нет, и среди эконометристов есть разные мнения на этот счет. Существует даже такая школа, представители которой считают, что не нужно ничего делать, поскольку “так устроен мир”... Мы здесь не ставим цель дать достаточно полное описание методов борьбы с мультиколлинеарностью» [3, с. 94].

Мультиколлинеарность факторов является основной причиной того, что решаемая задача становится некорректно поставленной и требует специальных методов ее решения. Мультиколлинеарность факторов приводит к следующим отрицательным последствиям.

1. Информационная матрица $X^T X$ ($X$ – расширенная матрица эффектов, главных и взаимодействий) системы нормальных уравнений плохо обусловлена, что является причиной падения точности оценивания и неустойчивости определяемых коэффициентов регрессии.

2. Оценки коэффициентов уравнения регрессии статистически взаимосвязаны; установить истинное влияние эффекта на зависимую переменную трудно или невозможно.

3. Оценки коэффициентов уравнения регрессии становятся чувствительными к изменениям совокупности исходных данных и к вычислительному процессу получения этих коэффициентов.

**3. Постановка задачи**

Моделирование сложных систем целесообразно осуществлять на основе системно-структурного подхода, а полученные многофакторные модели должны позволять интерпретировать в предметной области как качественную – причинную, структурную, так и количественную информацию. При соблюдении названных условий можно будет реализовывать смысловой анализ данных, описывать механизмы происходящих явлений и использовать полученные знания при создании наукоемких изделий, высоких технологий, интеллектуальных средств измерений, композиционных материалов.

Причинные связи между факторами и критериями качества сложных систем могут быть установлены, если коэффициенты модели статистически значимы и определены независимо относительно друг друга. Последнее условие будет выполнено, если соответствующие эффекты модели при коэффициентах будут попарно ортогональны друг другу.

Под структурой системы (процесса) будем понимать совокупность существенных связей между отдельными частями целого, сохраняющую свою неизменность и единство в процессе функционирования и, следовательно, неизбежных внутренних и внешних изменений. В математической модели целое – моделируемый критерий качества $y_t$ $(1 \le t \le m$, $m$ – общее число критериев качества моделируемой системы). Частями целого являются эффекты – главные и взаимодействия. Главные эффекты отражают собственное влияние факторов: знак коэффициента – направление, а величина – силу влияния. Взаимодействия показывают взаимосвязанное влияние факторов и подчеркивают важнейшее системное свойство – эмергентность (emergence – эмерджентность, «внезапное появление»): несводимость свойств системы (целого) к сумме свойств ее отдельных подсистем (частей), внутреннюю целостность системы. Структура системы правильно отображена в математической модели, если ее коэффициенты статистически независимы относительно друг друга.

Рассмотрим математическую модель для двух факторов $X_1$, $X_2$, каждый из которых изменяется в полном факторном эксперименте $3^2//9$ на трех уровнях. Структура такой модели задается выражением

$$(1+x_1+z_1)(1+x_2+z_2) \to N_{\Pi},$$

где 1 – элемент структуры модели, соответствующий фиктивной переменной $x_0$; $x_1$, $x_2$ – линейные контрасты факторов $X_1$, $X_2$; $z_1$, $z_2$ – квадратичные контрасты факторов $X_1$, $X_2$; $N_{\Pi}$ – общее число структурных элементов, равное числу опытов полного факторного эксперимента.

Общий вид данной математической модели:

$$\hat{y} = b_0 + (b_1x_1 + b_2z_1 + b_3x_2 + b_4z_2) + (b_5x_1x_2 + b_6x_1z_2 + b_7z_1x_2 + b_8z_1z_2).$$

Все эффекты модели разделены на две группы: главные эффекты – в первых скобках, эффекты взаимодействий – во вторых скобках. В первой группе эффектов влияние факторов на моделируемый критерий качества $y$ независимо друг от друга аддитивно. Существенное значение здесь имеет сложение эффектов. Во второй группе представлено совместное влияние факторов, где важно перемножение их эффектов, т.е. мультипликативное влияние. Взаимодействие факторов исследователю трудно предугадать и прогнозировать интуитивно.

Отметим, что при использовании ортогональных контрастов в полном факторном эксперименте все эффекты ортогональны друг другу и, следовательно, все коэффициенты модели статистически независимы относительно друг друга. Модель со статистически независимыми коэффициентами и нормированными эффектами позволяет проводить смысловой анализ данных.

При этом причинные связи характеризуют силу влияния факторов на моделируемый критерий качества, а структурные связи – влияние главных эффектов и взаимодействий факторов.

Выполнение требований адекватности полученной модели, а также статистической значимости коэффициентов модели и их статистической независимости относительно друг друга для анализа причинных и структурных связей является достаточным и для анализа количественных связей.

Модели, используемые для анализа данных, должны иметь определенные общие и специальные свойства.

Под общими свойствами статистических моделей будем понимать соответствие их критериям адекватности, информативности, устойчивости. Анализ указанных критериев позволяет сделать вывод о возможности использования полученных моделей.

Специальные свойства статистических полиномиальных моделей позволяют формализованно устанавливать причинные и структурные связи между факторами и критериями качества, если исследователю они априори не известны (именно такая ситуация и является типичной при исследовании сложных систем). К специальным свойствам такого вида моделей относят следующие:

1) статистическая независимость коэффициентов модели (или достаточно слабая зависимость для отдельных коэффициентов);

2) возможность выделения необходимой («истинной») структуры многофакторной модели по исходному плану эксперимента и определенному алгоритму (например, RASTA3);

3) статистическая эффективность коэффициентов модели.

Модели со специальными свойствами характеризуются как модели семантически информационные: их можно использовать для анализа механизмов происходящих явлений в изучаемых системах, объектах, процессах.

Для придания многофакторным статистическим моделям специальных свойств необходимо применять специальные методы планирования экспериментов.

**4. Концепция решения**

Известно, что модель, построенная по результатам проведения любого полного факторного эксперимента, характеризуется необходимыми специальными свойствами: эффекты модели ортогональны друг другу, а коэффициенты в статистическом смысле независимы относительно друг друга, план эксперимента соответствует критериям *D*-, *A*-, *E*-, *G*-оптимальности, и, следовательно, все коэффициенты модели определяются с минимальными погрешностями. План такого эксперимента будет равномерным, т.е. уровни любого фактора будут встречаться в плане одинаковое для данного фактора число раз. Коэффициенты полученной модели максимально устойчивы: число обусловленности $cond = 1$. Математическая модель адекватна в точках аппроксимации поверхности отклика [4, с. 102–103]. Будем считать такую модель «истинной» и «наилучшей». Ее и следует использовать для смыслового анализа данных.

Однако при сравнительно большом числе факторов $(k \geq 5)$ использование полного факторного эксперимента в прикладных исследованиях нереально, поскольку требует проведения значительного числа опытов. В подобных ситуациях необходимо выбирать дробный факторный эксперимент, наиболее близкий по своим свойствам к схеме полного факторного эксперимента.

Необходимые свойства математических моделей можно получить при использовании следующих классов планов экспериментов.

1. Формирование плана эксперимента на основе условия пропорциональности частот уровней факторов [4, с. 114–118]. Такие планы относятся к классу многофакторных регулярных планов экспериментов. Возможно также последовательное и специальное планирование эксперимента на основе многофакторных регулярных планов [5, с. 322–328].

Многофакторные регулярные планы позволяют сформировать структуру модели практически произвольной сложности. Их следует использовать, если известна информация о сложности влияния факторов на моделируемый критерий качества. Сложность характеризуется степенью полинома по каждому фактору, необходимого для адекватного описания моделируемого критерия качества.

2. Планирование эксперимента на основе $ЛП_{\tau}$ равномерно распределенных последовательностей [4, с. 121–127, 129, 132] позволяет выделить необходимую структуру математической модели при значительном числе факторов ($k = 20...50$ при $N = 64...128$) и попутно решить (если необходимо) задачу многокритериальной оптимизации.

3. При отсутствии необходимого по числу уровней факторов и опытов многофакторного регулярного плана необходимо синтезировать последовательный квази-*D*-оптимальный план эксперимента [4, с. 119–120, 128–131; 5, с. 326–328].

Рекомендуется также применять специальные алгоритмы выделения требуемой структуры многофакторной модели [4, с. 111–113], реализованные в программном обеспечении для обработки исходных данных [6].

Для придания коэффициентам многофакторных моделей свойств статистической независимости необходимо использовать систему ортогональных контрастов, или систему ортогональных полиномов Чебышева [7].

Известные методы планирования эксперимента разработаны лишь для многомерного факторного пространства на кубе, сфере и симплексе. Но если факторное пространство задачи не соответствует указанным формам, определяемые коэффициенты становятся статистически зависимыми, или мультиколлинеарными, а сама задача некорректно поставленной. При линейных нестандартных ограничениях факторного пространства предлагается использовать алгоритм RASTA4 [8] и алгоритм RASTA4K – при криволинейных ограничениях [9]. Указанные алгоритмы реализуют топологическое отображение (гомеоморфизм) хорошо обусловленного факторного пространства – прообраза – в плохо обусловленное – образ. Достигается получение устойчивых структур моделей и устойчивых коэффициентов из исходных некорректных условий при сравнительно высокой коррелированности факторов: $0,6 < \left| r_{ij}\left(X_{io}, X_{jo}\right) \right| < 1$ ($r_{ij}$ – парный коэффициент корреляции; $X_{io}$, $X_{jo}$ – натуральные значения факторов в факторном пространстве; $1 \le i < j \le k$; $k$ – общее число факторов).

Задача устойчивого определения структуры и коэффициентов моделей при исходных некорректных условиях может быть решена также путем использования модификации алгоритма RASTA – фрагмента PRELE [10], где реализовано *планирование эксперимента с фиктивными факторами*. Необходимость привлечения фиктивных факторов возникает тогда, когда реальные исходные условия, описываемые группой факторов, невозможно представить известным стандартным способом, вытекающим из теории планирования эксперимента и определяемым известными оптимальными планами многофакторного эксперимента. Типичной причиной таких исходных условий является мультиколлинеарность факторов.

В большинстве решаемых научных и прикладных задач используется ортогональная система координат факторного пространства, где уровни варьирования факторов определяются пересечением отрезков прямых линий.

Устойчивого выделения структуры и оценки коэффициентов в некоторых случаях при исходной мультиколлинеарности факторов можно также достичь, если использовать *сложные функции* или *оптимальную систему координат* факторного пространства вместо ортогональной системы [11].

Выполнение комплекса вычислений и проверок по многим критериям качества полученных моделей, проведение вычислительного эксперимента, визуализация результатов и их документирование требуют специального программного обеспечения, представленного в [6].

**5. Применение экспериментально-статистического подхода**

**Пример 1.** Рассмотрим построение многофакторной статистической модели и выполнение смыслового анализа данных на основе этой модели в случае решения задачи повышения точности измерения цифровых весов с диапазоном взвешивания 0–100 кгс.

Предварительное исследование весов показало, что изменения температуры окружающей среды (воздуха) в диапазоне 0…60°С и питающего напряжения автономного источника в диапазоне 12,3…11,7 В при расчетном номинальном значении 12 В сравнительно мало влияют на показания емкостного датчика и, следовательно, на результаты взвешивания [12, с. 82–88]. Если не учитывать влияние изменений температуры и питающего напряжения, среднее абсолютных величин относительных погрешностей аппроксимации $\left|\bar{e}_{u\text{отн}}\right|$ составляет 0,16 %, а среднеквадратичная погрешность остатка (в единицах измерения выходной величины взвешивания) равна 53,92. Но по техническим требованиям необходима была более высокая точность взвешивания.

Применение традиционного подхода предполагает в этом случае стабилизацию указанных внешних и внутренних условий с необходимой точностью и поддержку их в процессе функционирования весов. Однако выполнить это не представлялось возможным ввиду того, что весы должны эксплуатироваться не в стационарных (лабораторных) условиях, а на борту объекта.

Для получения многофакторной математической модели $\hat{y}(D)$, где $y(D)$ – показание датчика, выраженное в милливольтах, с учетом системного влияния факторов были приняты следующие условия.

$X_1$ – гистерезис; уровни: 0 (нагрузка); 1 (разгрузка); фактор качественный;

$X_2$ – температура окружающей среды; уровни: 0; 22; 60°С;

$X_3$ – напряжение питания; уровни: 11,7; 12,0; 12,3 В;

$X_4$ – измеряемый вес; уровни: 0; 20; 40; 60; 80; 100 кгс.

Поскольку количество уровней варьирования факторов невелико и объем испытаний сравнительно небольшой, было решено провести полный факторный эксперимент, т.е. $2 \times 3^2 \times 6 // 108$ (исходные данные испытаний были представлены проф. Новицким П.В.). Каждый опыт проводился только один раз, что в общем случае нельзя признать хорошим решением. Желательно повторение каждого опыта два раза.

Натуральные значения уровней варьирования факторов были преобразованы в ортогональные контрасты, т.е. в систему ортогональных полиномов Чебышева.

С использованием системы ортогональных контрастов структура полного факторного эксперимента имеет следующий вид:

$$(1+x_1)(1+x_2+z_2)(1+x_3+z_3)(1+x_4+z_4+v_4+w_4+g_4) \to N_{108},$$

где $x_1,...,x_4$, $z_2,...,z_4, v_4, w_4, g_4$ – соответственно линейные, квадратичные, третьей, четвертой и пятой степени контрасты факторов $X_1,...,X_4$; $N_{108}$ – число структурных элементов для схемы полного факторного эксперимента.

Все эффекты (главные и взаимодействия) нормированы:

$$\sum_{u=1}^{108} \left[x_{iu}^{(p)}\right]^2 = 108,$$

где $x_{iu}^{(p)}$ — значение *p*-го ортогонального контраста $i$-го фактора для $u$-й строки матрицы планирования, $1 \le u \le 108$, $1 \le i \le 4$; $1 \le p \le s_i - 1$; $s_i$ – число уровней фактора $X_i$, $2 \le s_i \le 6$.

Предварительный расчет математической модели показал, что в качестве оценки дисперсии воспроизводимости может быть выбрана (приближенно) величина 20,1. Число степеней свободы (условно) принято $f_{\text{восп}} = 108$. Дисперсия была использована для определения стандартной ошибки коэффициентов уравнения регрессии.

Использование ПС ПРИАМ позволило выделить структуру математической модели следующего вида:

$$\hat{y} = b_0 + b_1 x_4 + b_2 x_3 + b_3 z_2 + b_4 x_2 + b_5 z_4 + b_6 z_3 + b_7 x_2 z_3 + b_8 z_2 x_4 + b_9 x_1 x_2 + b_{10} z_2 x_3$$

и получить математическую модель

$$\begin{aligned}\hat{y}(D) = 28968{,}9 - 3715{,}13 x_4 + 45{,}2083 x_3 - 37{,}5229 z_2 + 23{,}1658 x_2 - 19{,}0708 z_4 - \\ - 19{,}6574 z_3 - 9{,}0094 x_2 z_3 - 9{,}27434 z_2 x_4 + 1{,}43465 x_1 x_2 + 1{,}65431 z_2 x_3,\end{aligned} \quad (1)$$

где $x_1 = 2(X_1 - 0{,}5)$; $x_3 = 3{,}33333(X_3 - 12)$;

$$x_2 = 0{,}0306122(X_2 - 27{,}333);$$
$$z_2 = 1{,}96006(x_2^2 - 0{,}237337x_2) - 0{,}575594;$$
$$z_3 = 1{,}5(x_3^2 - 0{,}666667);$$
$$x_4 = 0{,}02(X_4 - 50);$$
$$z_4 = 1{,}875(x_4^2 - 0{,}466667).$$

Результаты статистического анализа [4, с. 169–172] полученной многофакторной математической модели (1) (табл. 1) таковы: модель адекватна ($F_{97;108}^{\text{эксп}} = 1{,}05 < F_{0{,}05;97;108}^{\text{крит}} = 1{,}38$); доля рассеивания, объясняемая моделью, весьма высока, так как модель высокоточная, изменчивость функции отклика велика, а ее случайная изменчивость сравнительно мала; коэффициент множественной корреляции $R$ весьма близок к 1 и устойчив (будучи скорректированным с учетом степеней свободы, практически не меняется), его статистическая значимость велика, т.е. модель информативна (высокая информативность модели подтверждается также значением критерия Бокса и Веца); коэффициенты модели максимально устойчивы: число обусловленности $cond = 1$.

Таблица 1. Результаты статистического анализа полученной математической модели $\hat{y}(D)$

| Параметры статистического анализа | | Значения параметров для модели |
|---|---|---|
| Проверка гипотезы об адекватности математической модели | Дисперсия воспроизводимости (задана пользователем) | 20,1 |
| | Дисперсия адекватности | 21,1084 |
| | Уровень значимости $\alpha$ для адекватности | 0,05 |
| | Число степеней свободы для адекватности $f_{\text{ад}}$ | 97 |
| | Экспериментальное значение $F$ -критерия | 1,05017 |
| | Критическое значение $F$ -критерия для адекватности | 1,3844 |
| | Критическое значение $F$ -критерия (при отсутствии повторных опытов) | 1,02681 |
| | Среднеквадратичная погрешность аппроксимации | 4,59439 |
| | Среднеквадратичная погрешность аппроксимации, скорректированная с учетом степеней свободы | 4,80072 |
| | МОДЕЛЬ АДЕКВАТНА | |
| Анализ математической модели на информативность | Коэффициент множественной корреляции $R$ | 0,999999 |
| | Введено коэффициентов (эффектов) | 11 |
| | Число степеней свободы для коэффициентов модели $f_{k'}$ | 10 |
| | Число степеней свободы для остаточной суммы квадратов $f_R$ | 97 |
| | Коэффициент множественной корреляции, скорректированный с учетом степеней свободы | 0,999998 |
| | Уровень значимости $\alpha$ | 0,01 |
| | Экспериментальное значение $F$ -критерия для $R$ | $3{,}29697 \times 10^6$ |
| | Критическое значение $F$ -критерия для $R$ | 2,50915 |
| | МОДЕЛЬ ИНФОРМАТИВНА | |
| | Критерий Бокса и Веца | >49 |
| | Информативность модели | Очень высокая |
| | Число обусловленности cond | 1 |
| Среднее абсолютных величин относительных погрешностей аппроксимации $\lvert \bar{e}_{u\text{отн}} \rvert$, % | | 0,012 |
| Доля рассеивания, объясняемая моделью | | 0,999997 |

Полученная модель (1) семантична в информационном смысле, так как все ее коэффициенты ортонормированы: они статистически независимы и могут сравниваться по абсолютной величине друг с другом. Знак коэффициента показывает направление влияния, а его абсолютная величина – силу влияния. Построенная модель наиболее удобна для интерпретации в данной предметной области.

*Анализ влияния факторов на результаты измерений.* Учитывая семантические свойства математической модели (1) и доли участия каждого из ее эффектов в объясняемой моделью общей доле рассеивания, можно провести содержательный информационный анализ формирования результата измерения в исследуемых цифровых весах.

Превалирующая доля участия в результатах моделирования, равная 0,999557, создается линейным главным эффектом $x_4$ (с коэффициентом $b_1 = -3715{,}13$), т.е. измеряемым весом (фактор $X_4$).

Нелинейность $z_4$ (с коэффициентом $b_5 = -19{,}07$) сравнительно мала (доля участия равна $3{,}16 \cdot 10^{-5}$), и ее учет в модели повышает точность измерения. Линейный эффект $x_4$ (измеряемый вес) сравнительно слабо (доля участия $3{,}19 \cdot 10^{-6}$) взаимодействует с квадратичным эффектом $z_2$ температуры окружающей среды: взаимодействие $z_2 x_4 (b_8 = -9{,}27)$. Следовательно, математическая модель должна включать и эффект влияния температуры окружающей среды, фактор которого $X_2$ является неуправляемым:

$$\hat{y}_1 = 28968{,}90 - 3715{,}13 x_4 - 19{,}07 z_4 - 9{,}27 z_2 x_4 ,$$

где $\hat{y}_1$ – математическая модель, учитывающая только измеряемый вес и его взаимодействие с температурой окружающей среды.

Напряжение питания (фактор $X_3$) изменяет результаты взвешивания в виде линейного $x_3 (b_2 = 45{,}21)$ и квадратичного эффектов $z_3 (b_6 = -19{,}66)$. Их суммарная доля участия составляет $2{,}41 \cdot 10^{-4}$.

Температура окружающей среды влияет в виде квадратичного $z_2$ ($b_3$ = -37,52) и линейного $x_2 (b_4 = 23{,}17)$ эффектов с суммарной долей участия $1{,}60 \cdot 10^{-4}$.

Температура окружающей среды и напряжение питания (факторы $X_2$ и $X_3$) образуют парное взаимодействие $x_2 z_3 (b_7 = -9{,}01)$ с долей участия $3{,}63 \cdot 10^{-6}$.

Доказательность статистической значимости двух последних эффектов $x_1 x_2$ и $z_2 x_3$ не может быть обоснована, так как они существенно меньше эффектов $x_2 z_3$ и $z_2 x_4$, а уточненного значения дисперсии воспроизводимости по результатам повторных опытов в представленных исходных данных, к сожалению, не было.

*Таким образом, анализ моделирования в математической модели (1) с учетом разнообразных систематических погрешностей, нелинейностей, взаимодействий и неуправляемых факторов показал, что возможно повышение точности исследуемого средства* *измерения (табл. 1) по критерию «среднее абсолютных величин относительных погрешностей аппроксимации»* $\left|\bar{e}_{\text{иотн}}\right|$ *до 0,012 %, т. е. в 13,3 раза, а по критерию «среднеквадратичная погрешность аппроксимации» – до 4,80, т. е. в 11,2 раза*. Для этого необходимо показания датчиков, фиксирующих фактические значения температуры окружающей среды, напряжения питания, вводить во встроенный в прибор микропроцессор, использующий модель (1) для коррекции полученного результата измерения.

**Пример 2.** Рассмотрим получение многофакторной математической модели $\hat{y}(L)$ с использованием дробного регулярного плана эксперимента для оптимизации конструкции и технологии изготовления спиральных монолитных твердосплавных сверл (СМТС) диаметром 1 мм по критерию стойкости [13].

Системный анализ позволил выделить не менее 139 факторов, от которых зависят состояние печатных плат, условия и режимы сверления, качество СМТС. По результатам экспертного анализа для моделирования было выбрано 20 факторов (в том числе блоковый фактор), характеризующих конструкцию сверла и технологию образования качественного отверстия для последующей его металлизации. Среди них качественные и количественные, простые и сложные, дискретные и непрерывные факторы и их уровни:

$X_1$ – тип смазочно-охлаждающей технологической среды (СОТС) для вышлифовки стружечных канавок СМТС; уровни: сунграйнл – 600Х; МР-10;

$X_2$ – стадийность спекания заготовок СМТС; уровни: одностадийный; двухстадийный;

$X_3$ – тип связки алмазного круга для окончательного шлифования рабочей части СМТС; уровни: В2–01; В2–08;

$X_4$ – вид пооперационной очистки заготовок СМТС; уровни: бензин; ультразвуковая обработка + Лабомид;

$X_5$ – зернистость алмазного круга для вышлифовки стружечных канавок СМТС; уровни:50/40; 40/28; 28/20 мкм/мкм;

$X_6$ – тип связки алмазного круга для вышлифовки стружечных канавок СМТС; уровни: В2-08; МС6П; Винтер (ФРГ);

$X_7$ – вид упрочняющей обработки; уровни: исходные (неупрочненные) СМТС; ультразвуковая обработка (УЗО) (выдержка 1,0 мин); УЗО (выдержка 1,5 мин).

$X_8$ – зернистость алмазного круга для окончательного шлифования рабочей части СМТС; уровни: 40/28; 28/20; 20/14;

$X_9$ – марка твердого сплава; уровни: ВК6-М; ВК6-ОМ; ОМС;

$X_{10}$ – величина конусности (обратной) по рабочей части СМТС; уровни: 0,005; 0,015; 0,025 мм/10 мм;

$X_{11}$ – диаметр по спинке СМТС; уровни: 0,75; 0,85; 0,95 мм;

$X_{12}$ – тип связки алмазного круга для доводки СМТС; уровни: D15CC50, DTC50, D3C50 (HAWERA, ФРГ); БСТК; МС6П; Б11-Л;

$X_{13}$ – величина уголка СМТС; уровни: 0; 0,05; 0,10; 0,15 мм;

$X_{14}$ – величина конусности (прямой) по сердцевине СМТС; уровни: 0,05; 0,10; 0,15; 0,20 мм/10мм;

$X_{15}$ – ширина ленточки по спирали СМТС; уровни: 0,04; 0,10; 0,16; 0,22 мм;

$X_{16}$ – ширина пера СМТС; уровни: 0,3; 0,4; 0,5; 0,6 мм;

$X_{17}$ – толщина сердцевины СМТС; уровни: 0,13; 0,19; 0,25; 0,31 мм;

$X_{18}$ – основной задний угол СМТС; уровни: 12; 14; 15; 17 град;

$X_{19}$ – износостойкое покрытие; уровни: iC; TiN; без покрытия; $Al_2O_3$;

$X_{20}$ – разбиение плана эксперимента на ортогональные блоки, уровни: 0; 1; 2; 3; 4; 5; 6; 7.

Главным критерием качества работы СМТС служила максимальная длина сверления $L$ (в погонных метрах) после всех переточек сверла. Эта величина коррелирует со всеми 9 показателями, характеризующими интенсивность, надежность и эффективность процесса сверления, а также со стоимостью получения одного отверстия в печатных платах, и является обобщенным технико-экономическим критерием.

Из приведенного можно заключить: сверление отверстий в печатных платах представляет собой сложный технологический процесс, описываемый многими факторами и критериями качества. Для его изучения целесообразно применить экспериментально-статистический подход.

С учетом количества факторов и числа их уровней был выбран дробный регулярный многофакторный план эксперимента $2^4{\times}3^7{\times}4^8{\times}8^1//64$. (Отметим, что полный факторный эксперимент без блокового фактора включает 2293235712 опытов). Выбранный план близок к *Q*-оптимальному [14, с. 65], при котором достигается минимум средней дисперсии оценки поверхности отклика и функция эффективности составляет $\varphi = 1$ [14, с. 170–171], для нашего плана $\varphi = 0{,}962$. План эксперимента был разбит на 8 ортогональных блоков, по 8 опытов в каждом блоке.

Среди возможных дробных планов экспериментов данный план выделяется наилучшими статистическими характеристиками. Однако эффекты взаимодействий неизбежно коррелированы с главными эффектами и друг с другом.

В результате с использованием ПС ПРИАМ [6] получена математическая модель следующего вида:

$$\begin{aligned}\hat{y}(L) &= 14{,}859 - 2{,}466x_{17} - 2{,}253x_{16} - 1{,}951x_{15} - 0{,}560x_2z_7 + 1{,}717x_{20} - \\ &- 1{,}453z_7 - 2{,}090x_8 + 1{,}269x_2 - 1{,}231x_5 + 1{,}121x_4 + 1{,}098x_9 + 1{,}493x_1z_{14} + \\ &+ 0{,}822z_{20} - 1{,}008x_{19} - 0{,}865z_9 - 1{,}804x_1x_2x_5x_{14} + 1{,}091x_1 + 0{,}720v_{20} + \\ &+ 0{,}758z_{12} + 0{,}720x_6 + 0{,}698x_{12} + 0{,}515x_3 - 0{,}497x_2x_{12},\end{aligned} \tag{2}$$

где $x_i, z_i, v_i$ – соответственно линейные, квадратичные, кубические ортогональные контрасты *i*-го фактора $(1 \le i \le 20)$. Для вычисления значений $\hat{y}(L)$ необходимо также использовать нормировочные коэффициенты $k_{xi}$, $k_{zi}$, $k_{vi}$.

Результаты статистического анализа (табл. 2) многофакторной математической модели (2) таковы: дисперсии в незначительной степени неоднородны; проверка коэффициентов математической модели, проводимая по двум критериям устойчивости (в табл. 2 приведены два значения числа обусловленности cond), показала, что модель устойчива, но в незначительной степени неадекватна. (Неадекватность модели – следствие рассеивания полезной информации по значительному числу (многие тысячи) взаимодействий факторов. Математическая модель объясняет 96 % рассеивания результатов. Если учесть, что фактор разбиения эксперимента на ортогональные блоки объясняет 11,3 % систематических неоднородностей в его результатах, то адекватность и предсказывающие свойства полученной модели следует в системном смысле считать хорошими). Коэффициент множественной корреляции $R = 0{,}98$ достаточно близок к 1 и статистически значим, т.е. информативность модели высокая.

Таким образом, модель (2) можно использовать для изучения процесса сверления печатных плат, прогноза стойкости, поиска оптимальной конструкции сверла и технологии его изготовления. Свойства модели подтверждают правильность принятых исходных теоретических положений и методологии ее получения.

*Анализ данных на основе использования полученной математической модели показывает*, что наибольшее системное влияние (в виде взаимодействий) на моделируемый критерий качества оказывают факторы $X_1$, $X_2$, $X_5$, $X_7$, $X_{12}$, $X_{14}$.

Наиболее сильные главные эффекты и взаимодействия факторов математической модели (2) объясняют следующие доли рассеивания максимальной длины сверления *L* после всех переточек сверла:

1) толщина сердцевины СМТС, $X_{17}$ – 14,7 %;

2) ширина пера СМТС, $X_{16}$ – 12,3 %;

Таблица 2. Результаты статистического анализа полученной математической модели $\hat{y}(L)$

| Параметры статистического анализа | | Значения параметров для модели |
|---|---|---|
| Проверка гипотезы об однородности дисперсий | Дисперсия воспроизводимости результатов экспериментов | 0,590 |
| | Среднеквадратичная погрешность воспроизводимости | 0,768 |
| | Уровень значимости $\alpha$ | 0,01 |
| | Число степеней свободы для дисперсии воспроизводимости | 192 |
| | Экспериментальное значение $G$ -критерия | 0,114 |
| | Критическое значение $G$ -критерия | 0,106 |
| Проверка гипотезы об адекватности математической модели | Дисперсия адекватности | 2,724 |
| | Уровень значимости $\alpha$ | 0,05 |
| | Число степеней свободы для дисперсии адекватности | 40 |
| | Экспериментальное значение $F$ -критерия | 4,618 |
| | Критическое значение $F$ -критерия для адекватности | 1,458 |

Продолжение табл. 2

| | | |
|---|---|---|
| Анализ математической модели на информативность | Коэффициент множественной корреляции $R$ | 0,979 |
| | Уровень значимости $\alpha$ | 0,01 |
| | Число степеней свободы для коэффициентов модели $f_{k'}$ | 23 |
| | Число степеней свободы для остаточной суммы квадратов $f_R$ | 232 |
| | Экспериментальное значение $F$ -критерия для $R$ | 40,4 |
| | Критическое значение $F$ -критерия | 1,89 |
| | Критерий Бокса и Веца | 4 |
| | Информативность модели | Высокая |
| Число обусловленности $cond1$ | | 5,40 |
| $cond2$ | | 4,16 |
| Среднее абсолютных величин относительных погрешностей аппроксимации $\left\|\bar{e}_{u\text{ОТН}}\right\|$, % | | 9,24 |
| Доля рассеивания, объясняемая моделью | | 0,959 |

3) блоковый фактор – совокупность неуправляемых и неконтролируемых факторов как систематических неоднородностей, $X_{20}$ – 11,3 %;

4) ширина ленточки по спирали СМТС, $X_{15}$ – 9,2 %;

5) эффект взаимодействия факторов: тип СОТС для вышлифовки стружечных канавок и величина конусности (прямой) по сердцевине СМТС, $X_1X_{14}$ – 9,0 %;

6) вид упрочняющей обработки, $X_7$ – 5,1 %;

7) зернистость алмазного круга для окончательного шлифования рабочей части СМТС, $X_8$ – 4,7 %;

8) марка твердого сплава, $X_9$ – 4,7 %;

9) стадийность спекания заготовок СМТС, $X_2$ – 3,9 %;

10) зернистость алмазного круга для вышлифовки стружечных канавок СМТС, $X_5$ – 3,7 %.

Другие эффекты в математической модели имеют доли рассеивания менее 3,11 %.

Проведенное исследование охватывало 19 факторов, описывающих конструкцию, технологию изготовления, упрочнения, эксплуатации и восстановления СМТС. Обобщенный критерий качества работы СМТС – максимальная длина сверления $L$ – учитывал в статистическом смысле 9 показателей. Это проявление системного подхода в данном исследовании, результаты которого позволили изменить технологию сверления так, что ресурс работы сверла повысился в 5-6 раз и производительность сверления печатных плат была доведена до 100…150 ходов/мин. Стало возможным гарантировать 100% – нанесение сплошного плотного мелкокристаллического слоя химически осажденной меди без разрывов и царапин на последующей операции металлизации стенок отверстий благодаря высокому качеству просверленных отверстий. Работоспособность СМТС удалось повысить до $L = 32$ погонных метров.

Приведенные результаты исследования были внедрены на производственном объединении им. С.П. Королева (г. Киев), и в лаборатории твердосплавного инструмента налажен выпуск СМТС по усовершенствованной технологии. При работе на станках с ЧПУ группой сверл рекомендуется уменьшить общую длину просверленных отверстий до 20...22 м.

Разработанные методы, алгоритмы и программное обеспечение позволяют получить многофакторные статистические модели с семантическими свойствами. Такие модели по сравнению с моделями, полученными по другим подходам, дают возможность анализировать механизмы изучаемых явлений.

## 6. Выводы

1. Применение разработанных методов и алгоритмов устойчивого оценивания структур и коэффициентов статистических моделей при моделировании в условиях исходной мультиколлинеарности факторов следует считать целесообразным.
2. Предложенные классы многофакторных планов экспериментов обеспечивают получение специальных свойств статистических моделей и использование их для анализа влияния причинных, структурных и количественных связей факторов (в виде их главных эффектов и взаимодействий) на моделируемый критерий качества.
3. Использование моделей с функционирующими системами позволяет реализовать информационный и физический принципы в единой надсистеме – прибор, технологический процесс и математическая модель. Сочетание информационного и физического принципов создает эффект взаимодействия (эмергентность) и способствует улучшению критериев качества эксплуатируемых и производимых изделий.

## СПИСОК ЛИТЕРАТУРЫ


1. Демиденко Е.З. Линейная и нелинейная регрессии. – М.: Финансы и статистика, 1981. – 302 с.
2. Петрович М.Л. Регрессионный анализ и его математическое обеспечение на ЕС ЭВМ. – М.: Финансы и статистика, 1982. – 199 с.
3. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник. – М.: Дело, 2001. – 400 с.
4. Радченко С.Г. Математическое моделирование технологических процессов в машиностроении.– К.: ЗАО “Укрспецмонтажпроект”, 1998. – 274 с.
5. Радченко С.Г. Последовательное и специальное планирование эксперимента на основе многофакторных регулярных планов // Вестник Национального технического университета Украины “Киевский политехнический институт”. Машиностроение. – 2000. – Вып. 39.– С. 322–328.
6. Лапач С.Н., Радченко С.Г., Бабич П.Н. Планирование, регрессия и анализ моделей PRIAM (ПРИАМ). SCMC–90; 325, 660, 668 // Каталог. Программные продукты Украины. – К.: СП "Текнор", 1993.– С. 24–27.
7. Радченко С.Г., Добрянский С.С. Методические рекомендации по самостоятельному изучению дисциплин “Основы научных исследований и технического творчества”, “Оптимизация и моделирование технологических процессов и объектов в машиностроении”. – К.: НТУУ «КПИ», 1987. – 68 с.
8. Радченко С.Г. Топологический метод устойчивого оценивания коэффициентов многофакторного уравнения регрессии в условиях мультиколлинеарности факторов // Математичні машини і системи. – 2001. – № 1, 2. – С. 114–121.
9. Радченко С.Г. Алгоритм устойчивого оценивания коэффициентов статистических моделей (алгоритм RASTA4K) // УСиМ. – 2002. – № 1. – С. 27–36.
10. Радченко С.Г. Алгоритм устойчивого оценивания коэффициентов статистических моделей с использованием фиктивных факторов // Математичне моделювання. – 2001. – № 1(6). – С. 56–58.
11. Радченко С.Г. Планирование многофакторного эксперимента с использованием сложных функций и оптимальных координат факторного пространства // Математичне моделювання. – 2001. – № 2(7). – С. 15–20.
12. Радченко С.Г., Бабич П.Н. Информационная коррекция переменных систематических погрешностей средств измерений и измерительных информационных систем // Радиоэлектроника и информатика. – 1999. – № 3. – С. 82–88.
13. Радченко С.Г. Моделирование и оптимизация конструкции и технологии изготовления спиральных монолитных твердосплавных сверл // Надежность режущего инструмента и оптимизация технологических систем. – Т. 1. – Краматорск: ДГМА, 1997. – С. 50–59.
14. Бродский В.З. Введение в факторное планирование эксперимента. – М.: Наука, 1976. – 224 с.