



УДК 519.254;691.3

А.М. Данилов – доктор технических наук, профессор**В.И. Логанина** – доктор технических наук, профессор**В.А. Смирнов** – кандидат технических наук**Пензенский государственный университет архитектуры и строительства (ПГУАС)**

АНАЛИЗ ПОКАЗАТЕЛЕЙ ЛАКОКРАСОЧНЫХ ПОКРЫТИЙ МЕТОДОМ ГЛАВНЫХ КОМПОНЕНТ

АННОТАЦИЯ

Дается сравнительный анализ методов понижения размерности задач строительного материаловедения (метод наименьших квадратов и главных компонент); приводится оценка на примере управления качеством лакокрасочных покрытий.

КЛЮЧЕВЫЕ СЛОВА: Метод главных компонент, понижение размерности, оценка качества

A.M. Danilov – doctor of technical sciences, professor**V.I. Loganina** – doctor of technical sciences, professor**V.A. Smirnov** – candidate of technical sciences**Penza State University of Architecture and Construction (PSUAC)**

ANALYSIS OF LACQUER COATING CHARACTERISTICS BY MEANS OF PRINCIPAL COMPONENT ANALYSIS

ABSTRACT

The comparison analysis of dimension's reduction methods (least squares and principal component analysis) in materials science is performed. The principal component analysis is illustrated by the example of quality control of lacquer coating.

KEYWORDS: Principal component analysis, reduction of the dimension, quality estimation.

Обычно в задачах строительного материаловедения предполагается, что эмпирические данные представляются детерминированным процессом с наложением флуктуаций. Внутренние механизмы этого процесса в большинстве случаев неизвестны; анализу подлежат только результаты измерений.

Возможность содержательного анализа достигается после надлежащего представления данных – в форме, наиболее удобной для осмысления. Фактически, здесь налицо один из видов *моделирования* – замена объекта его абстрактным описанием (работа с ним часто возможна в терминах, отличных от терминов предметной области). Так, *математическое моделирование* предполагает замену объекта описанием, работа с которым возможна в терминах *математического анализа*.

В отличие от математического, моделирование в *теории эксперимента* минует стадию исследования модели средствами анализа (возможно, включая лишь решение оптимизационных задач); полученная *экспериментально-статистическая модель* (ЭС-модель)

с формально найденными параметрами сразу допускает анализ в терминах прикладной области. Подобное моделирование нередко называют *информационным*; полученные ЭС-модели имеют *прогностическую* ценность, однако могут выполнять и *смыслообразующую роль* (количественные оценки имеют второстепенную роль по сравнению с качественными представлениями) [1]. В последнем случае на основе ЭС-модели формулируются гипотезы, касающиеся закономерностей функционирования системы; в дальнейшем это может привести к получению математической модели. Смыслообразующая роль предполагает эвристический анализ, результативность которого существенно возрастает тогда, когда его исходным пунктом является *графическое, визуальное представление* первичных данных. Найденная ЭС-модель обычно используется для *визуализации*; однако можно считать, что ее построение (переход от эмпирической информации к набору параметров) преследует и другую цель – *понижение размерности исходной задачи, редукцию*,



свертку эмпирической информации. Выводы, полученные на основе анализа одних лишь численных значений параметров, нередко носят спекулятивный характер [2].

Построение ЭС-модели является лишь одним из *параметрических методов* статистики. Для полного восстановления исходного распределения в теории статистических оценок Фишером определено понятие *достаточных статистик*. Как подмножество методов информационного моделирования, параметрические методы позволяют выделить достаточные статистики, которые с заданной достоверностью позволяют сделать максимальное число выводов (в терминах предметной области), либо заданное число выводов с максимальной достоверностью. Предполагается выполнение для исходных данных тех предпосылок, на которых основан вычислительный аппарат поиска параметров модели.

В математической теории эксперимента базовым является предположение о нормальном распределении эмпирической информации в каждой точке факторного пространства. Это, наряду с принципом максимального правдоподобия, приводит к методу наименьших квадратов (МНК) – вычислительному аппарату построения ЭС-модели [3]. При других распределениях МНК неприменим.

Возможные подходы к понижению размерности задач не исчерпываются наиболее распространенными в практике строительного материаловедения методами регрессионного анализа. Так, эффективным является *метод главных компонент* (МГК), предложенный К. Пирсоном (часто называется методом *собственного ортогонального разложения* или *дискретным преобразованием Карунена – Лоэва* [4]).

МГК состоит в отыскании многомерного *эллипсоида рассеяния* эмпирических данных в факторном пространстве, который определяется расположением и длинами полуосей – *главными направлениями* и стандартными отклонениями в пространстве главных направлений.

Вычислительный аппарат МГК допускает компактное представление [1, 5, 6]. Здесь для выборки $\{x_{iu}\}$,

$i = \overline{1, k}$, $u = \overline{1, N}$, значений первичных признаков (здесь k – число признаков, N – число измерений) последовательно выполняются следующие процедуры.

1. *Центрирование признаков (частных критериев)*:

$$\xi_{ui} = x_{ui} - \bar{x}_i, \quad i = \overline{1, k}, \quad u = \overline{1, N}, \quad (1)$$

$$\text{где } \bar{x}_i = \frac{1}{N} \sum_{u=1}^N x_{ui} \text{ – выборочное среднее } i\text{-го}$$

признака.

2. *Определение матрицы ковариаций* *:

$$C = (c_{ij}) = \Xi^T \Xi, \quad (2)$$

где $\Xi = (\xi_{ui})$ – матрица центрированных признаков.

3. *Определение собственных значений* l_i и *собственных векторов* матрицы ковариаций (всегда имеет k действительных неотрицательных собственных значений, включая кратные).

4. *Сортировка собственных векторов* в порядке *убывания собственных значений*. Единичные собственные векторы, определяющие главные направления, составляют строки матрицы k -го порядка L . Линейный однородный оператор, определяемый полученной матрицей, производит преобразование исходных центрированных данных в *некоррелированные* и с *убывающими дисперсиями* [4] (переход от исходного факторного пространства в пространство главных компонент).

В МГК предположение о нормальном распределении эмпирической информации не используется (в отличие от МНК применим для произвольных данных). Как и в непараметрических методах, в МГК возможно предположение об *автоинформативности данных* – «...за данными нет ничего, кроме них самих» [1]. Это сближает МГК с методами описательной статистики.

Сходство МНК и МГК проявляется в поиске по возможности «более простого» многообразия (в МГК – всегда линейного), выбор которого минимизирует заранее заданную норму. В МГК минимизации подлежит сумма квадратов евклидовых расстояний от точки до многообразия (в МНК используется другая норма). Как правило, расстояние в пространстве признаков не имеет никакого содержательного смысла, кроме меры «различия» объектов. Предобработка данных (например, нормировка на стандартное отклонение) для обоснованного выбора метрики может существенно изменить вид эллипсоида рассеяния. В некоторых случаях уже на этапе предварительного анализа данных можно выбрать линейное или нелинейное преобразование, сводящее нелинейную задачу (поиск *главных многообразий*) к рассматриваемой линейной (МГК) [1].

Отметим: если каждой точке исходных данных сопоставить единичную массу, то матрица ковариаций совпадет с тензором инерции системы частиц (механическая интерпретация МГК); задача поиска главных компонент перейдет в задачу приведения тензора инерции к главным осям [5].

* Часто эту матрицу называют ковариационной, однако подобная терминология входит в противоречие с рядом нормативных документов.



Понижение размерности (разделение исходных данных на содержательную часть и шумы [6]) в рамках МГК достигается *отбрасыванием направлений, соответствующих малым собственным значениям*. По-видимому, общих правил выбора числа значимых главных компонент не существует. Это число определяется не только величинами собственных значений матрицы ковариаций, но и задачами исследования (визуализация на плоскости или в пространстве), интуицией исследователя и т.п.

Существует эвристический метод оценки необходимого числа главных компонент, а именно – *правило сломанной трости*. Он состоит в сравнении упорядоченных (по убыванию) k собственных значений матрицы ковариаций с длинами l_i «обломков трости» единичной длины, сломанной в $(k-1)$ -й точке (координаты изломов распределены равномерно на отрезке $[0; 1]$) [7]. Очередное i -е главное направление считается значимым, если

$$\frac{\lambda_j}{\text{tr}C} > l_j, \quad j = \overline{1, i}, \quad (3)$$

где $\text{tr}C$ – матрицы ковариаций.

Правило сломанной трости является вероятностным. Оно дает математическое ожидание числа значимых главных компонент (не является целым числом). Кроме этого, во многих практических задачах величины собственных значений матрицы ковариаций отличаются на порядки, и правило (3) избыточно.

В качестве иллюстрации приведем результаты практического применения МГК к оценке показателей качества лакокрасочных покрытий строительных изделий и конструкций. В соответствии с действующей нормативной документацией качество покрытий определяется совокупностью критериев, к числу которых относятся шероховатость и время высыхания.

Было выполнено экспериментальное исследование влияния рецептурных факторов на указанные показатели для покрытий, изготовленных на основе краски ПФ-115. Наряду с шероховатостью и временем высыхания, регистрировалось значение показателя (розлив), характеризующего реологические свойства лакокрасочного состава. Результаты приведены в таблице.

Предварительная обработка данных (соотношения (1) и (2)) выполнена с использованием табличного процессора Microsoft Excel. Нахождение собственных значений матрицы ковариаций произведено в системе численной математики Mathcad. Найденная по таблице матрица ковариаций имеет вид:

$$C = (x_{ui})^T (x_{ui}) = \begin{pmatrix} 16,225 & -7,355 & 135,3 \\ -7,355 & 4,369 & -82,34 \\ 135,3 & -82,34 & 1624,4 \end{pmatrix};$$

ее собственные значения l_i и собственные векторы $\mathbf{v}_i$:

$$l_1 = 0,148, \quad \mathbf{v}_1 = (0,098; 0,994; 0,042);$$

$$l_2 = 4,964; \quad \mathbf{v}_2 = (0,992; -0,094; -0,088);$$

$$l_3 = 1640; \quad \mathbf{v}_3 = (0,083; -0,05; 0,995).$$

Матрица перехода к главным компонентам:

$$L = \begin{pmatrix} 0,083 & -0,05 & 0,995 \\ 0,992 & -0,094 & -0,088 \\ 0,098 & 0,994 & 0,042 \end{pmatrix}.$$

Вид множества экспериментальных точек со стороны наименее значимого главного направления показан на рис. 1**;

со стороны наиболее значимого (главной оси эллипсоида рассеяния) – на рис. 2.

Так как $\lambda_3 \gg \lambda_1$ и $\lambda_3 \gg \lambda_2$, то применение правила сломанной трости избыточно: значимая главная компонента единственна и соответствует главному направлению, определяемому вектором .

Главные компоненты связаны с исходными показателями (розливом q_1 , шероховатостью q_2 и временем высыхания q_3) линейно:

$$Pc_1 = 0,083q_1 - 0,05q_2 + 0,995q_3;$$

$$Pc_2 = 0,992q_1 - 0,094q_2 - 0,088q_3;$$

$$Pc_3 = 0,098q_1 + 0,994q_2 + 0,042q_3.$$

** Переход к главным направлениям – суть введение смешанных координат (линейных комбинаций исходных переменных, где коэффициентами выступают главные компоненты). Поэтому оси на рисунках 1 и 2 лишены обозначений.

Таблица

Наименование показателей	Составы (объекты)				
	1	2	3	4	5
Розлив, балл.	10	10	10	7	8
Шероховатость, мкм	8,5	8	8,2	9,6	9,8
Время высыхания, мин.	40	45	43	20	7

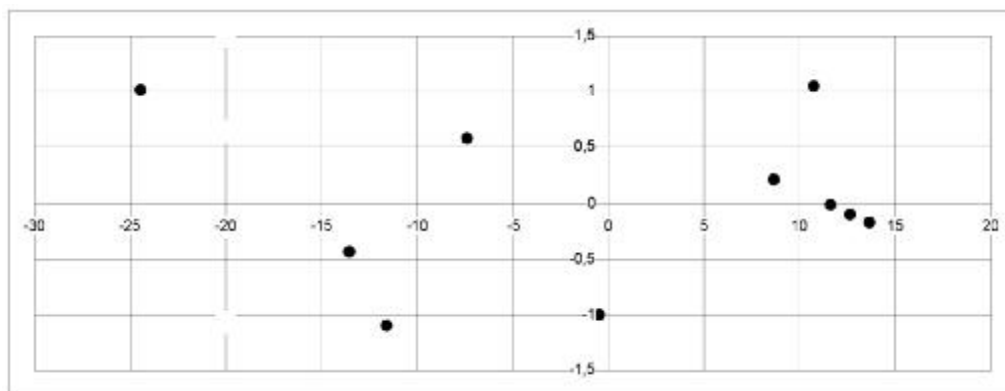


Рис. 1. Проекция на плоскость двух первых главных направлений
(по горизонтали – первое главное направление)

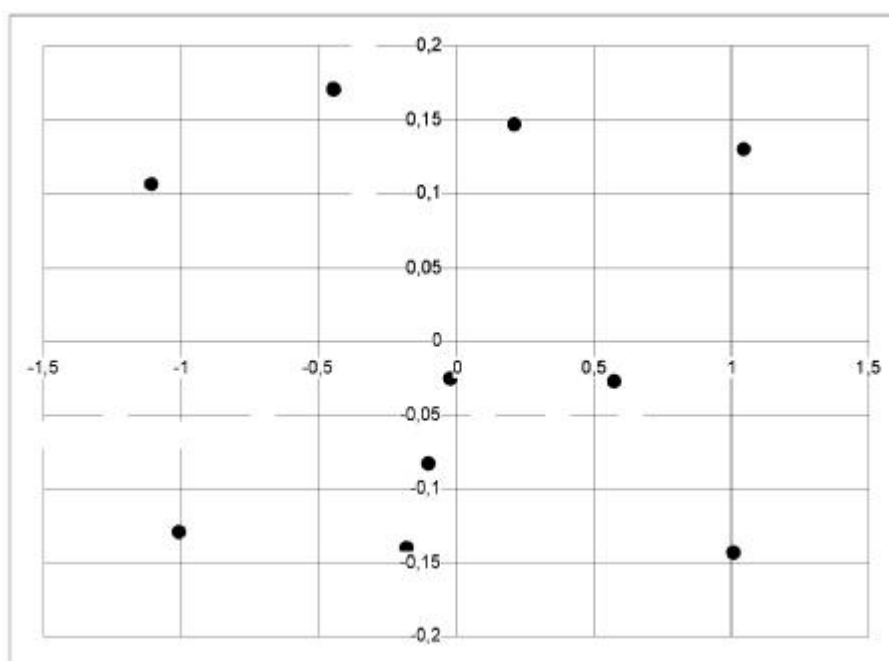


Рис. 2. Остатки, возникающие при отбрасывании главного направления
(по горизонтали – второе главное направление)

Вектор первого главного направления образует весьма малый угол с осью третьей исходной переменной. В рассматриваемой задаче доминирующим является третий показатель (время высыхания).

Литература

1. Зиновьев А.Ю. Визуализация многомерных данных. – Красноярск: Изд-во КГТУ, 2000. – 180 с.
2. Налимов В.В., Голикова Т.И. Логические основания планирования эксперимента. – М.: Наука, 1981. – 149 с.
3. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. – М.: ИНФРА-М, 1998. – 528 с.
4. Солодовников А.Ю., Платонов А.К. Исследование метода Карунена-Лоэва // URL: http://www.keldysh.ru/papers/2006/source/prep2006_19.doc
5. Метод главных компонент // URL: <http://ru.wikipedia.org>
6. Эсбенсен К. Анализ многомерных данных. Избранные главы / Пер. с англ. С.В. Кучерявского; под ред. О.Е. Родионовой. – Казань: Изд-во КазГАСУ, 2008. – 158 с.
7. Cangelosi R., Goriely A. Component retention in principal component analysis with application to cDNA microarray data // *Biology Direct*, 2007, 2:2.
8. Логанина В.И., Орентлихер Л.В. Управление качеством лакокрасочных покрытий строительных изделий и конструкций. – М.: АСВ, 2007.