УДК 563.08:519.237.7

Швейкин А.Л. *(УкрНИИгаз, г. Харьков)*

ВОЗМОЖНОСТЬ ВЫЯВЛЕНИЯ СКРЫТЫХ ВЗАИМОДЕЙСТВИЙ ВНУТРИ ИЗМЕРИТЕЛЬНОЙ СИСТЕМЫ ПРИ ИСПОЛЬЗОВАНИИ ФАКТОРНОГО АНАЛИЗА

Швейкин О.Л. Можливість виявлення прихованих взаємодій всередині вимірювальної системи при використанням факторного аналізу. Наявність прихованих зв'язків між окремими вузлами вимірювальної системи може призводити до небажаним взаємним впливам окремих компонентів всередині її і, як наслідок, до виникнення додаткового-котельної випадкової складової похибки вимірювань проведених такою системою. Для мінімізації кількості помилок при проведенні випробувань та виключення додаткових складових похибки вимірювань, ще на етапі проведення попередніх випробувань, необхідно виявити можливий взаємний вплив вузлів вимірювальної системи. Для проведення такого виявлення запропоновано використовувати один з методів математичної статистики - факторний аналіз. Перевага застосування даного методу полягає в тому, що вихідні дані, які використовуються для його проведення, отримані експериментальним шляхом, та кількість гіпотез які висуваються на початковому етапі, може довільно змінюватися. У статті наведено приклад обробки експериментальних даних для визначення впливу функціональних і вимірювальних вузлів реальної вимірювальної системи на поведінку корисного сигналу, який відображає наявності процесу конденсації компонентів досліджуваного газу.

Ключові слова: РІВНОВАЖНІ ТЕРМОДИНАМІЧНІ УМОВИ, ПРОЦЕС КОНДЕНСАЦІЇ, ФАКТОРНИЙ АНАЛІЗ, КОРЕЛЬОВАНІ ОЗНАКИ

Швейкин А.Л. Возможность выявления скрытых взаимодействий внутри измерительной системы при использовании факторного анализа. Наличие скрытых связей между отдельными узлами измерительной системы может приводить к нежелательным взаимным влияниям отдельных компонентов внутри ее и, как следствие, к возникновению дополнительной случайной составляющей погрешности измерений проведенных такой системой. Для минимизации количества ошибок при проведении испытаний и исключения дополнительных составляющих погрешности измерений, еще на этапе проведения предварительных испытаний, необходимо выявить возможное взаимное влияние узлов измерительной системы. Для проведения такого выявления предложено использовать один из методов математической статистики – факторный анализ. Преимущество применения данного метода состоит в том, что исходные данные, которые используются для его проведения, получены экспериментальным путем, и количество гипотез которые выдвигаются на начальном этапе, может произвольно изменяться. В статье приведен пример обработки экспериментальных данных для определения влияния функциональных и измерительных узлов реальной измерительной системы на поведение полезного сигнала, который отражает наличия процесса конденсации компонентов исследуемого газа.

Ключевые слова: РАВНОВЕСНЫЕ ТЕРМОДИНАМИЧЕСКИЕ УСЛОВИЯ, ПРОЦЕСС КОНДЕНСАЦИИ, ФАКТОРНЫЙ АНАЛИЗ, КОРРЕЛИРОВАННЫЕ ПРИЗНАКИ

Shveikin O.L. Possibility exposure of the hidden co-operations into measuring system at the use factor analysis. This work is devoted the problem of exposure of presence of cross-couplings of components into the measuring system. The presence of such connections cans at-lead to the origin of additional random error making term of measurings of conducted by such system. For minimization of amount of errors during testing and exception of additional constituents of error of measurings, yet on the stage of lead through of preliminary tests, it is necessary to expose the possible cross-coupling of knots of the measuring system. For the lead through of such exposure it is suggested to use one of methods of mathematical statistics is a factor analysis. Advantage of application of this method consists of that basic data which are used for his lead through are got experimental a way, and amount of hypotheses which are pulled out on the initial stage, can arbitrarily change. The example of processing of experimental data is resulted for determination of influence of functional and measurings knots of the real measuring system on the conduct of useful signal which reflects the presences of process of condensation of components of the probed gas.

Keywords: EQUILIBRIUMS THERMODYNAMICS TERMS, PROCESS OF CONDENSATION, FACTOR ANALYSIS, CORRELATED SIGNS

Довольно часто при разработке измерительных систем, в состав которых входит достаточное количество исполнительных узлов и измерительных каналов, на этапе проектирования не всегда удается предусмотреть их взаимное влияние. От возможного наличия таких влияний во многом зависят результаты испытаний разработанных систем и определение их метрологических характеристик. Выявление нежелательных взаимодействий исполнительных узлов и измерительных каналов является нетривиальной задачей, для которой в настоящее время не разработано функциональных методик. Практика применения статистических методов в разных областях науки и народного хозяйства позволяет предположить, что их применение позволит решить вышеуказанные проблемы при разработке и построении измерительных систем.

Цель исследований данной работы заключается в том, чтобы на основе факторного анализа [1] определить индивидуальные различия данных проведенных измерений отдельными измерительными каналами во время функционирования исполнительных узлов измерительной системы и вскрыть закономерности, объясняющие эти различия.

Основное предположение факторного анализа можно сформулировать следующим образом: составляющие данных, которые получены по различным измерительных каналам измерительной системы конденсационного типа, несмотря на свою разнородность и изменчивость признаков, могут быть описаны относительно небольшим числом функциональных единиц или факторов. Факторный анализ определяет эти факторы на основе корреляции, существующей между отдельными признаками.

Пусть исходные данные $X_k(t_i)$ и $Y_k(t_i)$ записаны в виде матрицы, где индекс $i = 1, 2, ..., n$ относится к переменной, а индекс $k = 1, 2, ..., m + r$ – к числу составляющих элементов измерительных каналов. Коэффициент корреляции K_{ik} между переменными i и k вычисляется по известной формуле [2]. Если все переменные $X_k(t_i)$ и $Y_k(t_i)$ пронормировать, то получим матрицу $\boldsymbol{Z}$ со стандартизированными элементами z_{ik} , а для корреляционной матрицы $\boldsymbol{R}$ имеет место соотношение:

$$R = \frac{1}{n-1} ZZ' . \tag{1}$$

Целью любого метода факторного анализа является представление величины z_{ik}, т.е. элемента матрицы Z, в виде линейной комбинации нескольких гипотетических переменных, или факторов. Основную модель факторного анализа можно выразить следующей формулой:

$$z_{ij} = a_{i1}F_{1k} + a_{i2}F_{2k} + \ldots a_{il}F_{lk} . \tag{2}$$

Здесь $a_{i1} \ldots a_{il}$ – неизвестные постоянные коэффициенты; $F_{1k} \ldots F_{lk}$ – значения факторов у k-го объекта. Используя матричную форму записи, для всех z_{ik} имеем:

$$Z = A \cdot F , \tag{3}$$

Матрица стандартизированных переменных Z порядка $(m+r) \times n$ является матрицей исходных данных. Матрица A порядка $l \times n$ представляет факторное отображение, где элементы матрицы являются факторными нагрузками. Матрица F порядка $(m+r) \times l$ представляет значения всех факторов у всех объектов. Таким образом, матрица A отражает связи переменных с факторами, а матрица F описывает отдельные объекты.

Подставив (3) в (1), получим:

$$R = \frac{1}{n-1} ZZ' = \frac{1}{n-1} AF(AF)' = \frac{1}{n-1} AFF'A' = A\frac{1}{n-1} FF'A' . \tag{4}$$

По аналогии с формулой (1) выражение $FF'/(n-1) = C$ является матрицей коэффициентов корреляции между факторами, а соотношение (4) примет вид:

$$R = ACA' . \tag{5}$$

Введем в равенство (5) условие некоррелированности факторов, т.е. представим матрицу C в виде единичной матрицы, тогда в результате получим

$$R = ACA' = AA' . \tag{6}$$

Формулы (5) и (6) описывают фундаментальную теорему факторного анализа. Эта теорема утверждает, что корреляционная матрица может быть воспроизведена с помощью факторного отображения и корреляций между факторами [1].

Задачей факторного анализа является определение матрицы факторного отображения A. При ортогональных факторах факторные нагрузки принимают значения между –1 и +1. Если факторы не ортогональны, то элементы могут принимать большие значения. Каждый фактор характеризуется столбцом, каждая переменная – строкой матрицы A. Если факторная нагрузка значительно больше или меньше нуля, то принята упрощенная форма записи в виде крестика в соответствующем месте факторного отображения (рис. 1).

Фактор называется генеральным, если все его нагрузки значительно отличаются от нуля (он имеет нагрузку во всех переменных) – столбик *D*. Фактор называется общим, если хотя бы две его нагрузки значительно отличаются от нуля – столбики *D, B, C*.

Такие факторы могут взаимно перекрываться, т.е. одни и те же переменные могут давать нагрузки на несколько факторов. Генеральный фактор является частным случаем общих факторов.

В противоположность перечисленным факторам у индивидуальных факторов значительно отличается от нуля только одна нагрузка – столбики $U_1 \ldots U_8$. В этом случае наблюдаются только характерные факторы, которые представляют одну переменную.

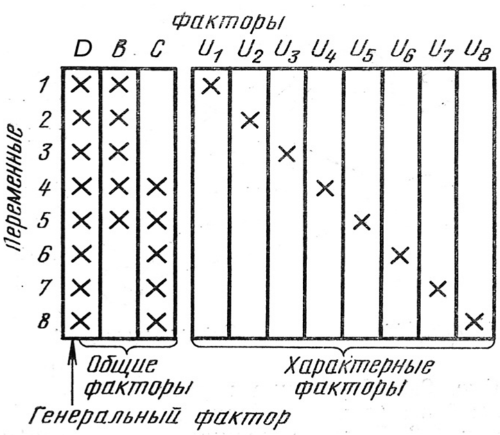

Рис. 1. Схематическое изображение матрицы факторного отображения (А)

По аналогии с факторами можно провести классификацию переменных по числу достаточно высоких нагрузок. Число высоких нагрузок переменной на общие факторы называется ее сложностью. Например, первая переменная на рис. 1 имеет сложность два, четвертая переменная – три. Таким образом, решающее значение в факторном отображении играют общие факторы *D, B, C*.

Система уравнений, соответствующая (6), имеет однозначное решение с вводом дополнительных условий, а именно: сумма квадратов нагрузок первого фактора должна составлять максимум от полной дисперсии; сумма квадратов нагрузок второго фактора должна составлять максимум оставшейся дисперсии и т.д., т.е. максимизирует функцию

$$S_1 = \sum_{i=1}^{n} a_{i1}^2 = \max \qquad (7)$$

при $n(n-1)/2$ независимых друг от друга условиях $R_{ik} = a_{i1} \cdot a_{k1}$ $(i < k)$.

Для максимизации функции, связанной некоторым числом дополнительных условий, пользуются методом множителей Лагранжа. В результате приходят к системе n однородных уравнений с n неизвестными a_{i1} [1]. Необходимым и достаточным условием существования нетривиального решения этой системы является равенство нулю детерминанта матрицы коэффициентов этих уравнений.

$$\begin{vmatrix} (1-\lambda) & R_{12} & \ldots & R_{1n} \\ R_{21} & (1-\lambda) & \ldots & R_{2n} \\ \vdots & \vdots & & \vdots \\ R_{n1} & R_{n2} & \ldots & (1-\lambda) \end{vmatrix} = 0 . \qquad (8)$$

Детерминант (определитель), записанный в виде (8), называется характеристическим, а в развернутом виде – характеристическим уравнением. Все n корней этого уравнения действительны, т.е. $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$ являются возможными, иногда совпадающими решениями. Найденное значение корня λ_1 соответствует вектору решения $(\alpha_{11}, \ldots, \alpha_{n1})$, причем $\sum_{i=1}^{n} \alpha_{i1}^2$ является максимумом в отношении оставшейся дисперсии.

Система равенств (8) составляет так называемую проблему собственных значений действительной симметрической матрицы. В общем, она записывается в следующем виде:

$$R\alpha_l = \lambda_l \alpha_l \quad \text{или} \quad (R - \lambda_l \mathrm{I})\alpha_l = 0 , \qquad (9)$$

где λ_l – собственные значения, соответствующие собственным векторам α_l матрицы R; I – единичная матрица.

Тот факт, что максимизация функции (7) приводит к классической проблеме собственных значений, облегчает численное решение системы уравнений (6), так как проблема собственных значений достаточно разработана.

Известно [3], что факторы F_{lk} пропорциональны собственным векторам матрицы R. Путем нормирования несложно получить искомые значения α_{il} матрицы A по компонентам собственных векторов матрицы R [1]:

$$\alpha_{il} = \alpha_{il} \cdot \sqrt{\lambda_l} \Big/ \sqrt{\alpha_{1l}^2 + \alpha_{2l}^2 + \ldots \alpha_{nl}^2} \,. \qquad (10)$$

Для приведенного примера, после проведения предварительных испытаний был проведен факторный анализ с целью подтверждения влияния процесса конденсации на изменение полезного сигнала либо поиск дополнительного фактора который может влиять на значение полезного сигнала.

Информация, которая касается исследуемого явления, приведена в форме табличных данных (как количественных, так и качественных). В этой таблице строки соответствуют множеству исследуемых объектов, а столбцы – множеству характеристик признаков, которые описывают данные объекты. Для данного случая таблица содержит укрупненные по среднему значению параметры работы измерительной системы на протяжении одного цикла измерения, а именно данные проведенных измерений по шести измерительным каналам:

1) Значение перепада давлений (сигнал описывающий процесс конденсации) ΔР, КПа.
2) Температура измерительного газопровода, Ткп, ºС.
3) Температура исследуемого газа в измерительном газопроводе, Тг, ºС.
4) Температура окружающей среды, Тос, ºС.
5) Величина электрического тока который потребляется системой охлаждения, Іб, мА.
6) Значение температуры измеренной компенсационным температурным каналом, Тк, ºС.

Созданная таблица представляет собой лишь набор чисел и только путем проведения специального анализа можно выявить скрытые закономерности, которые скрыты за набором чисел. Связь факторного анализа с изучением индивидуальных особенностей поведения измерительной системы возникла потому, что использование этого метода возможно лишь при наличии изменяющихся величин. Факторные модели используются в тех случаях, когда удается уловить индивидуальные расхождения между параметрами, а именно вариацию явления, которое изучается [3].

Соответственно, на первом этапе необходимо определение индивидуальных различий. Обычно целью исследований является выявление закономерностей, которые объясняют эти различия. В сравнении с другими методами математической статистики, при помощи факторного анализа гипотезы не проверяются, а формулируются.

Факторный анализ является методикой, которая в определенном смысле сама является источником возникновения гипотез. Рассчитанные статистические данные исходных значений приведены в табл. 1.

Статистические данные **Табл. 1**

Variable	Descriptive Statistics (Книга1.sta)				
	Valid N	Mean	Std.Dev.	Minimum	Maximum
dP	67	112,0853	11,25087	98,00000	140,8571
Ткп	67	16,6136	7,77748	5,50000	27,7429
Тг	67	20,1521	0,84257	18,87500	22,2625
Тос	67	19,1507	1,09633	16,32500	20,4500
Іб	67	97,8632	69,57254	0,02387	229,4118
Тк	67	0,6731	0,17726	0,40000	1,1500

Процедуру расчетов начинаем с определения коэффициентов корреляции внутри совокупности переменных, которая изучается. Для данного случая, когда количество выборок N составляет значения $N \geq 50$ [1] используем коэффициент корреляции смешанного момента Пірсона-Браве, который для данного случая имеет вид:

$$r = \frac{MXY - MX \cdot MY}{\sqrt{MX^2 - (MX)^2} \cdot \sqrt{MY^2 - (MY)^2}},$$

где X і Y – результати наблюдений двух переменных;

$MX = \dfrac{\sum X}{N}$; $MY = \dfrac{\sum Y}{N}$; $MX^2 = \dfrac{\sum X^2}{N}$; $MY^2 = \dfrac{\sum Y^2}{N}$; $MXY = \dfrac{\sum XY}{N}$ (N – число наблюдений).

Для поиска факторных нагрузок и общностей используют аппарат матриц, элементами которых являются коэффициенты корреляции. Матрица коэффициентов корреляции полученных, как правило, экспериментальным путем, называют матрицей корреляции или корреляционной матрицей. Элементами этой матрицы являются коэффициенты корреляции между всеми переменными данной совокупности. То есть имеется n=6 переменных, для которых число коэффициентов корреляции, полученных опытным путем, составит $n(n\text{-}1)/2$ = 15. Эти коэффициенты заполнят половину матрицы, которая находится по одну сторону от ее главной диагонали. По другую сторону находятся те же коэффициенты, так как $r_{12} = r_{21}$; $r_{13} = r_{31}$ и т.д. Поэтому корреляционная матрица симметрична (табл. 2).

Корреляционная матрица Табл. 2

Variable	dP	Ткп	Тг	Тос	Іб	Тк
	Correlations (Книга1.sta) Marked correlations are significant at p < N=67 (Casewise deletion of missing data					
dP	1,00	-0,06	-0,26	-0,67	-0,11	0,20
Ткп	-0,06	1,00	-0,67	0,53	-0,87	-0,63
Тг	-0,26	-0,67	1,00	0,05	0,62	0,52
Тос	-0,67	0,53	0,05	1,00	-0,38	-0,51
Іб	-0,11	-0,87	0,62	-0,38	1,00	0,60
Тк	0,20	-0,63	0,52	-0,51	0,60	1,00

Для вычисления собственных значений и собственных векторов матрицы R применена ЭВМ со специализированными пакетами программ. Вычисление статистических параметров, коэффициентов корреляции и факторов временных рядов расхода проведено с использованием пакетов программ MathCad и Statsoft Statistica [4].

Анализ полученных факторных загрузок начали с выделения наиболее весомых факторов с нагрузкой большей, чем 0,7.

Приведенное в табл. 3 факторное отображение показывает, что для измерительной системы существует пять общих факторов: F_1 – генеральный фактор охлаждения измерительного газопровода; F_2 – общий фактор конденсации компонентов газа; F_3 – фактор эффективности работы устройства охлаждения; F_4 – фактор температуры газа; F_5 – фактор температуры корпуса измерителя.

Факторное отображение ряда переменных Табл. 3

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	Factor Loadings (Unrotated) (Книга1.sta) Extraction: Principal components (Marked loadings are > ,700000)				
dP	-0,157562	-0,906537	-0,203401	0,293735	0,157422
Ткп	0,940650	-0,080647	-0,209641	-0,149487	0,111174
Тг	-0,695864	0,557395	-0,307773	0,307305	-0,073078
Тос	0,601188	0,739138	-0,193350	0,077746	0,183726
Іб	-0,883352	0,233444	0,307597	-0,051946	0,252103
Тк	-0,823487	-0,100486	-0,411590	-0,374155	0,022323
Expl.Var	3,213743	1,749922	0,481452	0,351798	0,140291
Prp.Totl	0,535624	0,291654	0,080242	0,058633	0,023382

Фактор охлаждения со значением $F_1 \approx 0{,}94$, определяет генеральный фактор изменения температуры измерительного газопровода F_1 (Ткп) (табл. 3). Он характеризуется значениями F_1(Тг)=-0,69, F_1(Іб)=-0,88 і F_1(Тк)=-0,82. При этом они составляют отрицательные значения то есть противоположны F_1(Ткп) и объясняется тем, что с увеличением тока потребления устройством охлаждения увеличивается температура корпуса как следствие увеличения мощности рассеивания тепла "горячим" спаем устройства соответственно и увеличивается температура газа движущегося па каналам в корпусе измерителя. Однако фактор изменения разности давлений F_1(ΔP)==-0,15, то есть практически не зависит от температуры измерительного газопровода, а его незначительная ортогональность обусловлена обратной зависимостью изменения в процессе конденсации компонентов исследуемого газа. Значение F_1(Тос)=0,6 подтверждает прямую зависимость минимально достигаемой температуры измерительного газопровода от температуры окружающей среды. Фактор конденсации компонентов газа со значением $F_2 \approx$-0,91 определяет фактор F_2(ΔP) и характеризуется значениями F_2(Тг)=0,56, F_2(Іб)=0,23, F_2(Тк)=-0,1, F_2(Тос)=0,73 и F_2(Ткп)=-0,08 является вторым общим фактором и определяет изменение разности давлений в процессе конденсации компонентов исследуемого газа. При этом факторы F_2(Ткп) и F_2(Тк) являются незначительно зависимыми, что обусловлено слабой прямой зависимостью плотности исследуемого газа от температуры корпуса и измерительного газопровода в процессе цикла измерения при отсутствии конденсации. Ортогональность факторов F_2(Тг), F_2(Іб) і F_2(Тос) вероятно обусловлена тем фактом, что процесс конденсации начинается и проходит при температурах, которые ниже температуры корпуса. Таким образом, фактор $F_2(\Delta P)$ в значительной мере определен процессом конденсации компонентов исследуемого газа в средине измерительного газопровода, что приводит к возникновению разности давлений.

Графическое представление общности факторов дает графическая интерпретация. Тут факторами являются нормированные координатные оси на которые “натянуто” факторное пространство [3]. Применив общую формулу для определения длинны вектора в l-мерном пространстве[2], получим: $d_i = \sqrt{\sum_{k=1}^{l} a_{ik}^2} = \sqrt{a_{i1}^2 + a_{i2}^2 + \ldots + a_{il}^2}$.

Таким образом, длина вектора d_i равна корню квадратному из общности. Наибольшая длина такого вектора равна единице и указывает на то, какая часть дисперсии каждой переменной является общей с факторами, что является графическим отображением общности. Угол φ между двумя векторами в пространстве общих факторов является мерой корреляции обоих рядов переменных (коэффициенты корреляции между двумя переменными равны скалярному произведению векторов), то есть справедливо равенство

$$R_{ij} = \cos\varphi_{ij} = \sum_{k=1}^{l} a_{ik} a_{jk} \Big/ \left(d_i d_j\right).$$

Графическое отображение общности двух факторов F_1 и F_2 приведена на рис. 2. Векторы сгруппированы по первому F_1 (охлаждение измерительного газопровода) и второму F_2 (конденсация компонентов газа) факторам. Угол между факторами практически равен $\varphi_{1,2} \approx \pi/2$, коэффициент корреляции $R_{1,2} = \cos\varphi_{1,2} \approx 0$, что отражает практически отсутствие зависимости F_1 и F_2.

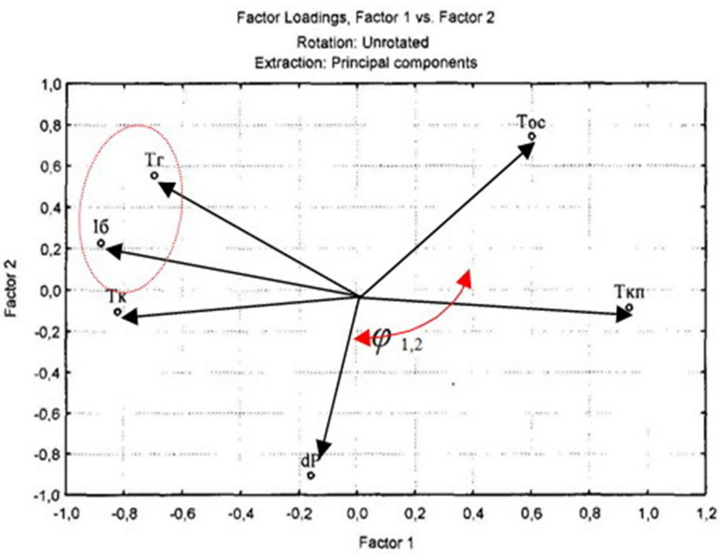

Рис. 2. Общность факторов F_1 и F_2

Таким образом, показан пример возможности определения индивидуальных различий данных проведенных измерений отдельными измерительными каналами во время функционирования исполнительных узлов измерительной системы и вскрытия закономерностей, объясняющих эти различия, благодаря применению факторного анализа. Установлено отсутствие влияния узлов измерительной системы на поведение полезного сигнала, значения которого зависят в основном от наличия сконденсированных компонентов исследуемого газа.

Литература

1. Иберла К. Факторный анализ: пер. с нем. В.М. Ивановой / К. Иберла. – М.: Статистика, 1980. – 398 с.

2. Игуменцева Н.В. Статистический анализ результатов экспериментов и наблюдений / Н.В. Игуменцева, В.И. Пахомов. – Харьков: СМИТ, 2005. – 236 с.

3. Ахмедов Т. Использование непараметрических критериев статистики Уилкоксона и Миллера в медико-биологических исследованиях. Основы факторного анализа: методические рекомендации / Т. Ахмедов. – Харьков, 1987. – С. 11-29.

4. Дьяконов В. Mathcad 8/2000. Специальный справочник / В. Дьяконов. – СПб.: Питер, 2000. – 592 с.