

**ИСПОЛЬЗОВАНИЕ КОМПОНЕНТНОГО АНАЛИЗА ПРИ ОБРАБОТКЕ
ПОЛЕТНОЙ ИНФОРМАЦИИ**

к. т. н. *В.Д. Шаров*
(ООО «Волга-Днепр-Москва»)

Рассматривается возможность практического применения метода главных компонент для анализа больших массивов полетной информации с целью выявления наиболее значимых факторов, снижающих качество выполнения полета.

Полетная информация (ПИ) является важнейшим источником объективных данных о режимах полета, пространственном положении ВС, действиях экипажа и состоянии контролируемых систем. Использование ПИ в предприятиях ГА РФ в соответствии с документом [1] вносит значительный вклад в повышение безопасности полетов. Однако можно утверждать, что информативность больших массивов ПИ востребована далеко не полностью. Одним из методов «компьютерной статистики» [2], позволяющих повысить качество обработки данных может быть компонентный анализ или метод «главных компонент», являющийся частным «упрощенным» вариантом факторного анализа.

Как известно [3], в основу факторного анализа положены результаты изучения случайных векторов большой размерности, наблюдаемых на практике. Установлено, что распределение случайного вектора сосредоточено вблизи некоторого подпространства значительно меньшей размерности. Это позволяет с достаточной для практики точностью представить исходный случайный вектор как результат линейного преобразования случайного вектора меньшей размерности.

Рассматривая параметрическую ПИ, совокупность отклонений от рекомендуемых (номинальных) значений параметров полета (пилотирования) можно представить как исходный многомерный вектор, который в определенном смысле характеризует качество выполнения полетов. Применение компонентного анализа существенно уменьшает размерность данного вектора, выделяет скрытые (латентные) факторы, определяющие значительную часть отклонений. Это позволяет оптимизировать процесс разработки корректирующих действий, направленных на устранение причин отклонений.

Задача формулируется следующим образом: используя материалы объективного контроля n полетов заменить набор m отклонений параметров полета (исходных признаков Z) меньшим числом $k < m$ стандартизованных ортогональных факторов или компонент, представляющих собой наиболее существенные латентные факторы отклонений.

Модель компонентного анализа в матричном виде представляется как:

$$Z=WF, \tag{1}$$

где $Z=(Z_1, Z_2, \dots, Z_m)$ - случайный стандартизованный вектор исходных признаков;
 $F=(F_1, F_2, \dots, F_m)$ - вектор факторов; W – матрица факторных нагрузок.

Матрица W вычисляется на основе собственных чисел и собственных векторов корреляционной матрицы R исходных признаков из соотношения:

$$R=WW^T.$$

При этом достигается ортогональность столбцов матрицы W , что при решении системы уравнений (1) относительно F в свою очередь обеспечивает ортогональность найденных компонент.

Геометрически решение задачи можно интерпретировать следующим образом. Полученные факторы F представляют собой новую систему координат в гиперпространстве порядка m . Факторные нагрузки являются проекциями вектор-признаков Z на оси-факторы F . При этом первая компонента F_1 ориентирована вдоль направления наибольшей вытянутости эллипсоида рассеивания отклонений (имеет максимальную дисперсию). Вторая компонента F_2 имеет максимальную дисперсию из

всех оставшихся компонент и задает направление наибольшей вытянутости в гиперплоскости, перпендикулярной к первой компоненте и т. д. Для реальных данных относительно небольшое число первых компонент, как правило, содержит основную часть суммарной дисперсии исходных признаков (в нашем случае – наблюдаемых отклонений параметров полета). Это подтверждается и опытом применения факторного анализа при исследовании деятельности службы АТБ, описанном в [4], когда из 30 используемых для анализа параметров были выделены 5 укрупненных факторов, охватывающих практически всю деятельность службы.

Важной деталью является то, что, вообще говоря, не обязательно «привязываться» к структуре F , обеспечивающей максимум дисперсии на F_1 . Рекомендуется выбирать [3] такое расположение осей-факторов, при котором достигается наибольшее количество проекций признаков, близких к 0 или 1 (так называемая «простая» структура матричных нагрузок). Возможность «вращения главных компонент» на практике позволит выделить те латентные факторы, которые связаны с наиболее важными с точки зрения безопасности полетов отклонениями в летной эксплуатации.

Совершенно очевидно, что практическое применение компонентного и факторного анализа, как и других статистических методов, возможно только с использованием компьютерных программ. Так, только для решения матричного уравнения (1) в нашем сравнительно простом случае пришлось бы решать $n \times m = 120$ линейных уравнений. Наиболее подходящим инструментом для практической реализации метода представляется программа *Statistica* (*StatSoft Inc., USA*) с использованием руководства [2].

Для проверки пригодности программы в качестве исходных данных использована ПИ о $n=12$ заходах на посадку однопосадочных ВС на один тот же аэродром с фиксацией $m=10$ следующих параметров: отклонений от рекомендуемых значений скоростей выпуска закрылков на δ° ($dV-\delta$), отклонений скорости захода на посадку (dV_{ref}), крена в процессе захода ($K-зп$), крена, тангажа и вертикальной перегрузки в момент приземления ($K-п$; $T-п$; N_y). При этом ни одно из отклонений не выходило за допустимые пределы и не было зафиксировано в экспресс-анализе ПИ, все посадки по параметру N_y выполнены на «отлично», т. е. имеем массив ПИ нормальной эксплуатации.

Основным первичным документом для анализа является корреляционная матрица (табл. 1).

Таблица 1

Корреляционная матрица стандартизованных отклонений

	dV-н	dV-5	dV-10	dV-20	dV-30	K-зп	Ny-	K-п	T-п	dVref
dV-н	1,00	0,65	0,44	0,02	0,56	0,26	-0,12	-0,12	-0,16	-0,31
dV-5	0,65	1,00	0,50	-0,39	0,45	-0,03	0,39	0,03	0,05	0,08
dV-10	0,44	0,50	1,00	0,25	0,29	-0,38	-0,10	0,27	0,43	-0,22
dV-20	0,02	-0,39	0,25	1,00	0,25	0,10	-0,72	0,46	0,44	-0,18
dV-30	0,56	0,45	0,29	0,25	1,00	0,14	0,00	-0,02	-0,27	-0,21
K-зп	0,26	-0,03	-0,38	0,10	0,14	1,00	0,11	0,20	-0,15	-0,36
Ny-	-0,12	0,39	-0,10	-0,72	0,00	0,11	1,00	0,11	-0,15	-0,13
K-п	-0,12	0,03	0,27	0,46	-0,02	0,20	0,11	1,00	0,82	-0,32
T-п	-0,16	0,05	0,43	0,44	-0,27	-0,15	-0,15	0,82	1,00	-0,04
dVref	-0,31	0,08	-0,22	-0,18	-0,21	-0,36	-0,13	-0,32	-0,04	1,00

Задавая число факторов $k=4$ и минимальное собственное значение матрицы R , равное 1,000, получаем таблицу собственных значений факторов (табл. 2).

Таблица 2

Кумулятивные суммы собственных значений

Factor	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	2,630809	26,30809	2,630809	26,30809
2	2,488920	24,88920	5,119729	51,19729
3	1,687755	16,87755	6,807485	68,07485
4	1,582824	15,82824	8,390308	83,90308

Из таблицы следует, что четыре первых главных компоненты объясняют почти 84% всей дисперсии исходных признаков (последний столбец таблицы).

Вращение системы факторов дает возможность добиться лучшего представления исходных признаков в выделенных факторах (табл. 3). До вращения только два исходных признака были представлены в выявленных факторах с нагрузкой более 0,7, а после вращения *Varimax*-методом за счет перераспределения нагрузок их количество увеличилось до 8.

Таблица 3

Факторные нагрузки до и после вращения

Factor Loadings (Unrotated) (Исх-Ст)
Extraction: Principal components
(Marked loadings are > ,700000)
(до вращения)

	Factor-1	Factor-2	Factor-3	Factor-4
dV-н	-0,5589	-0,6634	0,2595	0,1513
dV-5	-0,4035	-0,7464	-0,4339	0,0141
dV-10	-0,7600	-0,1428	-0,3894	0,3049
dV-20	-0,5908	0,6020	0,4293	0,2413
dV-30	-0,4917	-0,5437	0,3370	0,1653
К-зп	-0,0960	-0,1212	0,6094	-0,6290
Ny	0,2214	-0,4645	-0,4525	-0,6690
К-п	-0,6276	0,4671	-0,2327	-0,5178
Т-п	-0,5700	0,5993	-0,4736	-0,1386
dVref	0,4584	0,0738	-0,3578	0,5008
Expl.Var	2,6308	2,4889	1,6878	1,5828
Prp.Totl	0,2631	0,2489	0,1688	0,1583

Factor Loadings (Varimax normalized) (Исх-Ст)
Extraction: Principal components
(Marked loadings are > ,700000)
(после вращения)

	Factor-1	Factor-2	Factor-3	Factor-4
dV-н	-0,1335	0,8795	0,0479	0,2215
dV-5	0,1114	0,7787	-0,5142	-0,1587
dV-10	0,5251	0,6802	0,0981	-0,3077
dV-20	0,3895	0,0539	0,8743	0,1870
dV-30	-0,1620	0,7576	0,1431	0,2399
К-зп	-0,1058	-0,0125	-0,0337	0,8823
Ny	0,0604	-0,0226	-0,9393	0,1751
К-п	0,9097	-0,0550	0,0179	0,3216
Т-п	0,9396	-0,0853	0,1419	-0,1317
dVref	-0,2386	-0,2256	-0,0271	-0,7090
Expl.Var	2,2661	2,4813	1,9657	1,6771
Prp.Totl	0,2266	0,2481	0,1966	0,1677

Программа позволяет применять и другие способы вращения, строить графики собственных значений матрицы *R*, 2-х и 3-х мерные графики факторных нагрузок, а также получать ряд других характеристик в виде таблиц.

Ограниченный объем исходных данных и их характер (полеты выполнялись разными экипажами) не позволяет в полной мере использовать метод, тем ни менее некоторые выводы, главным образом о возможностях метода, сделать можно.

Около 84% всех отклонений объясняются четырьмя скрытыми факторами.

Фактор 1, объясняющий более 22% суммы всех отклонений, имеет нагрузки, близкие к 1, для отклонений по крену и тангажу в момент приземления. В данном случае можно предположить, что эта ситуация характерна для рассматриваемого аэродрома. Возможно, это связано с его особенностями, которые необходимо учитывать при подготовке к полету.

Фактор 2, на который приходится около 25% суммы отклонений, «нагружен» отклонениями по скорости выпуска закрылков. В данном случае аналитик должен задуматься, не связаны ли это с недостатками схемы захода данного аэродрома или

технологии работы службы УВД (например, «высоко подводят»), и как можно скомпенсировать эти недостатки.

При анализе данных, относящихся к конкретному экипажу за определенный период, достоинством метода является то, что он позволяет установить, какие отклонения главным образом снижают общее качество полетов этого экипажа и принять соответствующие меры.

Разумеется, для практического использования компонентного или факторного анализа необходимо определиться по номенклатуре учитываемых в исходных данных отклонений, возможно установление определенных приоритетов или даже весовых коэффициентов таким образом, чтобы выявленные латентные факторы объективно отражали снижение качества с точки зрения безопасности полетов.

Необходимо отметить, что данный метод не заменяет, а дополняет анализ ПИ, проводимый в соответствии с [1] в летной службе. При надлежащей организации автоматизированной обработки и продуманной системе учета отклонений метод предоставит в распоряжение эксперта-аналитика дополнительные, научно обоснованные данные о качестве всех полетов каждого КВС за длительный период в компактном виде, а также позволит прогнозировать тенденции. Это поможет реализовать на практике проактивный подход к предотвращению авиационных происшествий, базирующийся в значительной степени на анализе массивов различных данных нормальной эксплуатации.

Работа выполнена при финансовой поддержке РФФИ, грант № 06-08-01518.

ЛИТЕРАТУРА

1. Руководство по организации сбора, обработки и использования полетной информации в авиапредприятиях ГА РФ. – М.: Воздушный транспорт, 2001.- 80 с.
2. Боровиков В. STATISTICA. Искусство анализа данных на компьютере. Для профессионалов. 2-е изд. (+CD). – СПб.: Питер, 2003. – 688 с.: ил.
3. Пугачев В.С. Теория вероятностей и математическая статистика. – М.: Наука, 1979. – 495 с.
4. Сакач Р.В., Зубков Б.В., Давиденко М.Ф., и др. Безопасность полетов. – М.: Транспорт, 1989. – 239 с.