УДК 629.76/.78:519.25

Л. В. КРИВОБОКОВ, Д. В. ДУНАЕВ, А. В. ДЕМЧЕНКО

ОПРЕДЕЛЕНИЕ АДЕКВАТНОСТИ УСЛОВИЙ ОТРАБОТКИ СИСТЕМ РАКЕТ КОСМИЧЕСКОГО НАЗНАЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ СТАТИСТИЧЕСКОГО КРИТЕРИЯ ПОДОБИЯ

Государственное предприятие «Конструкторское бюро «Южное», ул. Криворожская 3, 49008, Днепр, Украина, e-mail: dimor9diit@gmail.com

У статті розглянуто алгоритм розрахунку критерію статистичної подібності умов відпрацювання систем ракет космічного призначення (РКП) у вигляді співвідношення між об'ємами еліпсоїдів розсіювання на порівнюваних етапах випробувань (в наземних і натурних умовах). При розрахунку критерію використовується метод головних компонент, який має обмеження при підготовці вихідних даних. Це, наприклад, спотворення матриці коваріацій через великі непомічені «викиди» в даних або через невдале нормування, вибіркова матриця може бути негативно визначеною. Для подолання обмежень при підготовці вихідних даних у методі головних компонент пропонуються нормування, масштабування і використання «комплектних зразків». При нормуванні потрібен певний досвід його застосування для врахування «викидів» у даних, щоб не отримати негативну коваріаційну матрицю завдяки нелінійній залежності даних. Масштабування застосовується при різній розмірності вихідних компонент, але з огляду на те, що вектора не інваріантні відносно зміни масштабу, це може призвести до порушення взаємозв'язків між вихідними даними і, як наслідок, виникає проблема перетворення коваріаційної матриці в кореляційну. При застосуванні «комплектних зразків» з матриці вихідних даних виділяється максимально можливий її фрагмент (виключаються стовпці з пропущеними значеннями параметрів або пропущене значення ознаки заміняється його середнім арифметичним значенням за наявними реалізаціями). В результаті застосування нормування зменшується вплив «викидів» даних, при застосуванні масштабування виключається негативна вибіркова матриця, а при «комплектних зразках» виключаються обидва обмеження при підготовці вихідних даних в методі головних компонент.

Використання критерію статистичної подібності умов відпрацювання систем РКП, який характеризує «близькість» наземних і льотних випробувань, дозволяє оптимізувати процес відпрацювання РКП за рахунок зменшення тривалості і вартості випробувань.

В статье рассмотрен алгоритм расчета критерия статистического подобия условий отработки систем ракет космического назначения (РКН) в виде соотношения между объемами эллипсоидов рассеяния на сравниваемых этапах испытаний (в наземных и натурных условиях). При расчете критерия используется метод главных компонент, который имеет ограничения при подготовке исходных данных. Это, например, искажение матрицы ковариаций из-за больших незамеченных «выбросов» в данных или из-за неудачной нормировки, выборочная матрица может быть отрицательно определенной. Для преодоления ограничений при подготовке исходных данных в методе главных компонент предлагаются нормировка, масштабирование и использование «комплектных образцов». При нормировке требуется определенный опыт её применения для учета «выбросов» в данных, чтобы не получить отрицательно определенную ковариационную матрицу ввиду нелинейной зависимости данных. Масштабирование применяется при различной размерности исходных компонент, но ввиду того, что вектора не инвариантны относительно изменения масштаба, возможно искажение взаимосвязей между исходными данными и, как следствие, проблема преобразования ковариационной матрицы в корреляционную. При применении «комплектных образцов» из матрицы исходных данных выделяется максимально возможный ее фрагмент (исключаются столбцы с пропущенными значениями параметров либо пропущенное значение признака заменяется его средним арифметическим значением по имеющимся реализациям). В результате применения нормировки уменьшается влияние «выбросов» данных, при применении масштабирования исключается отрицательно определенная выборочная матрица, а при «комплектных образцах» обходятся оба ограничения по подготовке исходных данных в методе главных компонент.

Использование критерия статистического подобия условий отработки систем РКН, который характеризует «близость» наземных и летных испытаний, позволяет оптимизировать процесс отработки РКН за счет уменьшения длительности и стоимости испытаний.

This paper considers an algorithm for calculating a statistical similarity criterion for space launch vehicle (SLV) system development conditions in the form of the ratio of the dispersion ellipsoid volumes at the test stages under comparison (ground tests and full-scale tests). The criterion is calculated using the principal components method, which has some limitations in input data preparation. For example, the covariance matrix may be distorted due to large unnoticed outliers in the data or due to an inappropriate normalization, or the sample matrix may be negatively definite. To overcome these limitations in input data preparation, normalization, scaling, and the use of "complete samples" are proposed. Normalization calls for some experience in its use to account for outliers in the data so as not to get a negatively definite covariance matrix due to a nonlinear relationship between the data. Scaling is used when the input components have different dimensions, but since vectors are not invariant under scale change, the relationships between the input data may be distorted, as a consequence of which the problem of transformation of the covariance matrix into a correlation one may arise. When using “complete samples”, a maximum possible fragment of the input data matrix is set off (by deleting the columns with missing parameter values or replacing a missing value of an attribute with its arithmetical mean value over the available realizations). Normalization reduces the effect of data outliers, scaling rules out a negatively definite sample matrix, and “complete samples” allow one to avoid both limitations involving source data preparation in the principal components method.

The statistical similarity criterion for SLV system development conditions, which characterizes the “proximity” of ground and flight tests, allows one to optimize the SLV development process by reducing test duration and cost.

Ключевые слова: *статистический критерий, эллипсоид рассеяния, метод главных компонент, ракета космического назначения, ковариационная матрица.*

Введение. Повышение надежности ракет космического назначения (РКН) современных систем при снижении общих затрат на экспериментальную отработку – одно из главных направлений развития современного ракетостроения. Одним из перспективных методов решения такой задачи является увеличение эффективности испытаний за счет применения различных методов пересчета результатов испытаний с использованием теории статистического подобия. В данной статье на основе зависимостей, приведенных в [1], предложен алгоритм расчета критерия адекватности (подобия) условий отработки, а также проанализированы особенности подготовки исходных данных при таких расчетах.

Метод решения задачи. В статье [1] был предложен и обоснован критерий статистического подобия условий отработки систем РКН в виде соотношения между объемами эллипсоидов рассеяния на сравниваемых этапах испытаний. Окончательное выражение для коэффициента подобия T_π имеет вид

$$P(T_\pi) = \varPhi\left[\frac{\lambda_{\text{Н}\max}\cos\theta - \lambda_{\text{Л}\max}}{\sqrt{2\left(\dfrac{\lambda^2_{\text{Н}\max}}{n_\text{Н}}\cos\theta + \dfrac{\lambda^2_{\text{Л}\max}}{n_\text{Л}}\right)}}\right], \tag{1}$$

где $P(T_\pi)$ – точечное значение вероятности подобия условий отработки сравниваемых этапов $(T_\pi \geq 1)$; $\varPhi$ – функция Лапласа; $\lambda_{\text{Н}\max}$, $\lambda_{\text{Л}\max}$ – максимальное значение собственного числа ковариационных матриц, полученных по результатам испытаний в наземных и натурных (летных) условиях соответственно; θ – угол, учитывающий расположение эллипсоидов рассеяния друг относительно друга при проецировании на плоскость; $n_\text{Н}$, $n_\text{Л}$ – объем выборки, по которой определялись значения $\lambda_{\text{Н}\max}$, $\lambda_{\text{Л}\max}$ соответственно.

Рассмотрим подробно алгоритм расчета критерия (1).

Исходными данными для расчета являются результаты испытаний системы РКН в наземных и натурных условиях в виде выборок по некоторому набору параметров (X и X' соответственно)

$$X=\begin{vmatrix} x_{11} & \ldots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{k1} & \ldots & x_{km} \end{vmatrix} \text{ и } X'=\begin{vmatrix} x'_{11} & \ldots & x'_{1n} \\ \vdots & \ddots & \vdots \\ x'_{k1} & \ldots & x'_{kn} \end{vmatrix}, \qquad (2)$$

где m, n – число испытаний для X и X' соответственно; k – число сравниваемых параметров.

Вычисляем выборочную ковариационную матрицу A и A' (для этапов наземных и натурных испытаний соответственно)

$$A=\begin{vmatrix} \operatorname{cov}(x_1,x_1) & \ldots & \operatorname{cov}(x_1,x_k) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(x_k,x_1) & \ldots & \operatorname{cov}(x_k,x_k) \end{vmatrix} \text{ и } A'=\begin{vmatrix} \operatorname{cov}(x'_1,x'_1) & \ldots & \operatorname{cov}(x'_1,x'_k) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(x'_k,x'_1) & \ldots & \operatorname{cov}(x'_k,x'_k) \end{vmatrix}, \qquad (3)$$

где $\operatorname{cov}(x_1,x_j)=\dfrac{\sum_{j=1}^{m}\left(x_{ij}-\bar{x}_i\right)\left(x_{kj}-\bar{x}_k\right)}{m-1}$, $\operatorname{cov}(x'_i,x'_p)=\dfrac{\sum_{p=1}^{n}\left(x'_{ip}-\bar{x}'_i\right)\left(x'_{kp}-\bar{x}'_k\right)}{n-1}$,

$\bar{x}_i=\dfrac{\sum_{i=1}^{m}x_i}{m}$, $\bar{x}'_i=\dfrac{\sum_{i=1}^{n}x'_i}{n}$, $\bar{x}_k=\dfrac{\sum_{i=1}^{m}x_k}{m}$, $\bar{x}'_k=\dfrac{\sum_{i=1}^{n}x'_k}{n}$; m, n – объем выборки, по которой определялись X и X' соответственно.

Вводим матрицу

$$|A-\lambda I|=\begin{vmatrix} \sigma_1^2-\lambda & \operatorname{cov}(x_1,x_2) & \cdots & \operatorname{cov}(x_1,x_k) \\ \operatorname{cov}(x_2,x_1) & \sigma_2^2-\lambda & & \operatorname{cov}(x_2,x_k) \\ \vdots & & \ddots & \vdots \\ \operatorname{cov}(x_k,x_1) & \operatorname{cov}(x_k,x_2) & \cdots & \sigma_k^2-\lambda \end{vmatrix} \qquad (4)$$

и точно такую же матрицу

$$|A'-\lambda' I|,$$

где λ и λ' – собственные числа матриц A_1 и A'_1 соответственно; I – единичная матрица.

Находим собственные значения λ как корни характеристического уравнения

$$\det|A-\lambda I|=0 \qquad (5)$$

и соответственно

$$\det|A'-\lambda' I|=0. \qquad (6)$$

Для вычисления (5) и (6) используем итерационные методы, например [2].

Вычисляем выборочную корреляционную матрицу для этапов наземной (R) и натурной (R') отработки соответственно

$$R = \begin{vmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & & r_{2k} \\ \vdots & & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{vmatrix} \text{ и } R' = \begin{vmatrix} 1 & r'_{12} & \cdots & r'_{1k} \\ r'_{21} & 1 & & r'_{2k} \\ \vdots & & \ddots & \vdots \\ r'_{k1} & r'_{k2} & \cdots & 1 \end{vmatrix}, \tag{7}$$

где $r_{ij} = \dfrac{\operatorname{cov}(x_i, x_j)}{\sqrt{\sigma_i \sigma_j}}$ – коэффициент корреляции случайных величин x_i, x_j.

Вводим матрицы

$$|R - \lambda_k I| = \begin{vmatrix} 1-\lambda_1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1-\lambda_2 & & r_{2k} \\ \vdots & & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1-\lambda_k \end{vmatrix}, \tag{8}$$

$$|R' - \lambda'_k I| = \begin{vmatrix} 1-\lambda'_1 & r'_{12} & \cdots & r'_{1k} \\ r'_{21} & 1-\lambda'_2 & & r'_{2k} \\ \vdots & & \ddots & \vdots \\ r'_{k1} & r'_{k2} & \cdots & 1-\lambda'_k \end{vmatrix}, \tag{9}$$

где λ_k и λ'_k – собственные числа матриц R и R' соответственно.

Находим собственные значения λ_k и λ'_k как корни характеристического уравнения

$$\det|R - \lambda_k I| = 0 \text{ и } \det|R' - \lambda'_k I| = 0. \tag{10}$$

Из вариационного ряда $\lambda_1 > \lambda_2 ... > \lambda_S$ выбирается максимальное значение $\lambda_i = \lambda_{\max}$ и, соответственно, находятся значения $\lambda'_{\max}$, $\lambda_{k\max}$, $\lambda'_{k\max}$ для матриц A, A', R и R'.

Вычисляем значение $\cos\theta$ по формуле

$$\cos\theta = \cos\left(\arccos\frac{\lambda_{k\max}}{k - \lambda_{k\max}} - \arccos\frac{\lambda'_{k\max}}{k - \lambda'_{k\max}}\right), \tag{11}$$

где k – количество параметров системы РКН, входящих в исходную матрицу (2).

Вычисляется значение критерия подобия по формуле (1)

$$P(T_\pi) = \Phi\left[\frac{\lambda_{\max}\cos\theta - \lambda'_{\max}}{\sqrt{2\left(\dfrac{\lambda^2_{\max}}{n_{\text{Н}}}\cos\theta + \dfrac{(\lambda'_{\max})^2}{n_{\text{Л}}}\right)}}\right],$$

где $\lambda_{\max} = \lambda_{\text{Н}\max}$, $\lambda'_{\max} = \lambda_{\text{Л}\max}$.

При расчетах по вышеприведенному алгоритму необходимо учитывать ряд особенностей подготовки исходных данных, связанных с применением метода главных компонент. Как следует из общей теории статистического оценивания, полученная в результате эксперимента матрица (2) и (3) является выборочной, поэтому, в отличие от теоретической, может быть отрицательно определенной (теоретически такая матрица всегда является положительно определенной [3] и собственные числа могут иметь только положительные значения). Последнее связано с тем, что метод главных компонент строго применим к исходным данным, связанным линейной зависимостью. Для преодоления этого недостатка рекомендуется использовать, например, нелинейные методы отображения данных в пространство малой размерности [4]. Еще одним недостатком метода главных компонент является искажение матрицы ковариаций из-за больших незамеченных «выбросов» в данных или из-за неудачной нормировки. От этого недостатка можно избавиться путем перехода к различного рода устойчивым оценкам, например взвешенным [5], либо путем предварительного удаления «выбросов» при анализе исходной статистики. Возможно также оценивание не по всей выборке, а только по какой-то ее части.

Применение метода главных компонент наиболее эффективно, когда все компоненты имеют общую физическую природу и, соответственно, измерены в одних и тех же единицах [4]. При различной размерности исходных компонент результат будет зависеть от выбора масштаба измерения, что может привести к преобразованию ковариационной матрицы в корреляционную и исказить действительную физическую картину взаимосвязи между исходными параметрами. Это связано с тем, что собственные числа и вектора не инвариантны относительно изменения масштаба [3]. В таком случае рекомендуется переходить к безразмерным величинам, например, с помощью соотношений

$$x_i^* = \frac{x_i - x_з}{x_з} \text{ или } x_i^* = \frac{x_i}{x_з}, \tag{12}$$

где $x_з$ – заданное значение параметра x; x_i – i-е значение параметра x.

Практический анализ имеющейся статистики испытаний различных систем РКН показывает, что в исходной матрице данных (см. (2), (3)) очень часто отсутствует часть измерений (по причине неправильных показаний измерительных приборов, либо грубой ошибки при подготовке данных, либо ошибок в телеметрических данных и т. д.). Введем понятие «комплектный образец» [4] – система (сборочная единица) РКН, у которого замерены все значения всех параметров. Когда выборка не содержит достаточного числа комплектных экземпляров, то рекомендуется выделять максимально возможный фрагмент матрицы исходных данных, в котором все столбцы будут комплектными, т. е. из исходной матрицы исключаются столбцы с пропущенными значениями параметров. Однако для рассматриваемых систем РКН в условиях выборок малых и средних объемов и высокой стоимости измерений (особенно при летных испытаниях системы) естественно попытаться использовать всю имеющуюся информацию. Один из наиболее распространенных и простых способов обработки данных с пропусками состоит в замене пропущенного значения признака x_i его средним арифметическим значением, которое оценивается по имеющимся реализациям [4]. В связи с этим рекомендуется для дальнейших расчетов использовать исходную матрицу, в которой пропущенные значения параметров заменяют их средними значениями. Полученные при этом оценки ковариационной и корреляционной матрицы, в частности, дисперсии будут смещены в сторону уменьшения. Такое смещение дисперсии можно устранить оцениванием их только по измеренным значениям соответствующих параметров, в то же время смещение недиагональных элементов матрицы ковариаций нельзя учесть без дополнительных предположений о виде распределения пропусков в матрице данных. Если предположить, что вероятность возникновения пропуска значения параметра x_{ij} не зависит от его измерения на других экземплярах этой же системы, то несмещенная оценка элемента ковариационной матрицы $\hat{S}_{ij}$ будет равна [5]

$$\hat{S}_{ij} = \delta_{ij}\,\frac{\hat{\sigma}_{ij}(n-1)}{n_i - 1} + (1-\delta_{ij})\frac{\hat{\sigma}_{ij}(n-1)}{(n_i-1)(n_j-1)}, \tag{13}$$

$$\delta_{ij} = \begin{cases} 1, \text{åñëè } i = j, \\ 0, \text{åñëè } i \neq j, \end{cases}$$

где $\hat{\sigma}_{ij}$ – оценка элемента ковариационной матрицы, полученная после подстановки средних значений вместо пропущенных записей параметров; n – объем выборки, по которой определялось значение $\hat{\sigma}_{ij}$; $n_{i(j)}$ – число экземпляров системы, у которых замерены все значения параметров.

Однако на практике независимость возникновения пропусков редко имеет место, поэтому более надежным является оценивание вектора средних значений и матрицы ковариаций только по имеющимся измерениям [4]. В качестве оценки среднего значения и дисперсии используются среднеквадратическое и средний квадрат отклонения соответственно, оцениваемые по имеющимся измерениям параметра, а недиагональные элементы ковариационной матрицы оцениваются по всем экземплярам, у которых измерена соответствующая пара признаков, т. е.

$$\bar{x} = \frac{1}{n_i}\sum_{i=1}^{n_i} x_i\ , \ S_{ij} = \frac{1}{n_{ij}-1}\sum_{i=1,\, j=1}^{n_{ij}} (x_i - \bar{x}_i)(x_j - \bar{x}_j), \tag{14}$$

где n_{ij} – число экземпляров системы, у которых измерена соответствующая пара x_i и x_j.

Согласно [4] оценка (14) несмещена и состоятельна, если $n_{ij} \to \infty$ при $n \to \infty$, что вполне соответствует действительности.

Выводы. Предложен алгоритм расчета коэффициента статистического подобия условий отработки систем РКН, базирующийся на соотношении объемов эллипсоидов рассеяния на сравниваемых этапах отработки.

Рассмотрены особенности подготовки исходных данных для расчета коэффициента подобия, связанные с использованием метода главных компонент.

Предложены способы преодоления ограничений при подготовке исходных данных в методе главных компонент: нормировка (выражения (2) – (11)), уменьшение размерности пространства исходных данных (масштабирование) и использование «комплектных образцов». Каждый из способов имеет свои особенности в применении:

– при нормировке исходных данных возможно получение ковариационной матрицы, отрицательно определенной ввиду нелинейных зависимостей данных. Тогда необходимо использовать нелинейные методы отображения данных в пространство малой размерности;

– при масштабировании возникает проблема преобразования ковариационной матрицы в корреляционную и, как следствие, искажение взаимосвязей между исходными данными. Для исключения этого недостатка рекомендуется использовать безразмерные величины, например (12);

– при применении «комплектных образцов» возникает проблема пропусков в матрице исходных данных. Для исключения этого недостатка предлагается определять закон распределения пропусков.

1. *Кривобоков Л. В., Дунаев Д. В., Демченко А. В.* Установление степени адекватности условий отработки изделий ракетной техники, как сложных систем, с применением теории статистического подобия. Техническая механика. 2017. №3. С. 64–71.
2. Итерационные методы решения систем линейных алгебраических уравнений. Метод Якоби. URL: http://bigor.bmstu.ru/?cnt/?doc=Parallel/ch030203.mod – 8.06.18.
3. *Кендалл М., Стьюарт А.* Многомерный статистический анализ и временные ряды. Том 3. Пер. с англ. Под ред. А. Н. Колмогорова. М.: Наука, 1976. 736 с.
4. *Айвазян С. А., Енюков Е.С., Мешалкин Л. Д.* Прикладная статистика. Основы моделирования и первичная обработка данных. М.: Финансы и статистика. 1983. 471 с.
5. *Курочкина А. И.* Оптимальные свойства главных компонент λ -взвешенной ковариационной матрицы. В кн. *Рубашкин Т. В., Айвазан С. А., Эников И. С.* Алгоритмическое и программное обеспечение прикладного статистического анализа. М.: Наука, 1980. 420 с.

Получено 13.09.2018,
в окончательном варианте 14.12.2018