

Переменные звезды **21**, № 2, 153–159, 1979.

Variable Stars **21**, No 2, 153–159, 1979.

Применение метода главных компонент к задаче классификации звезд типа RR Лиры

М. Б. Богданов

Для анализа группы звезд типа RR Lyr использован метод главных компонент, основанный на отыскании линейных комбинаций параметров объектов, учитывающих их суммарную дисперсию. Для каждой звезды рассматривалось семь параметров: период, амплитуда изменения блеска, амплитуда изменения показателя цвета, асимметрия кривой блеска, индекс металличности, индекс ускорения силы тяжести на поверхности звезды и индекс эффективной температуры. Построенные диаграммы в плоскостях главных компонент указывают на возможность существования среди звезд типа RRab по крайней мере трех групп объектов.

The Use of the Method of Principal Components to Problem of the Classification of RR Lyrae Type Stars

by M. B. Bogdanov

The method of principal components was used for analysis of RR Lyrae type stars ensemble. This method is based on search of linear combinations of objects parameters and this linear combination takes into account summary dispersion of parameters. For every star seven parameters were considered: period, amplitude of light variation, amplitude of color index variation, asymmetry of light curve, metal abundance index, surface gravity index and index of effective temperature. Diagrams drawing on the plane of principal components shows, that among RRab type stars there are perhaps three groups of objects.

В настоящее время считается общепринятым мнение о том, что ансамбль звезд типа RR Lyr галактического поля является смешанным. Многими исследователями предпринимались попытки разделения этого ансамбля на группы по тем или иным признакам. Подробное описание этих работ приведено в обзоре Цесевича (1970). В качестве параметров, по которым предпринимались попытки проведения классификации, использовались величина периода, степень асимметрии кривой блеска, спектральные и фотометрические особенности, кинематические признаки, а также характер и степень изменения периода. Тем не менее, единой классификации звезд типа RR Lyr до настоящего времени не создано. Это можно объяснить как малочисленностью объектов, для которых измерены все интересующие параметры, так и трудностью рассмотрения многих параметров объектов одновременно. С этой точки зрения представляет интерес использование для исследования ансамбля звезд

типа RR Lyg методов многомерной статистики, позволяющих включать в рассмотрение одновременно достаточно большое число параметров, которое ограничено только мощностью ЭВМ и в настоящее время может достигать нескольких десятков. В данной работе предпринята попытка применения к задаче классификации звезд типа RR Lyg метода главных компонент.

Метод главных компонент является одним из методов факторного анализа, основная задача которого заключается в компактном описании исследуемой многопараметрической совокупности объектов. Практически это осуществляется переходом к новым переменным — главным компонентам, являющимся линейными комбинациями исходных параметров. Во многих случаях основные свойства исследуемого ансамбля удается описать с помощью лишь двух-трех из этих главных компонент, чем достигается эффективное сжатие исходной информации. Благодаря этой особенности метод главных компонент часто используется для решения задач классификации и распознавания образов.

Достаточно подробное описание метода главных компонент приведено в монографии Хармана (1972), а примеры его практического применения даны в книгах Налимова (1971) и Андруковича (1973). Наибольшее распространение метод главных компонент получил в таких науках как психология, социология и экономика, хотя в последнее время он начинает довольно интенсивно применяться и в естественных науках. К сожалению, в астрономии он не получил достаточного распространения; единственная попытка применения метода главных компонент — при исследовании пульсаров, имела место лишь совсем недавно (Клякотко, 1977).

Предположим, что мы рассматриваем совокупность из N объектов, каждый из которых характеризуется M одноименными параметрами. Тогда исходную совокупность объектов можно представить в виде матрицы X , размером $N \times M$, каждый элемент которой — X_{ij} представляет собой значение i -го параметра у j -го объекта. В предположении линейной модели взаимосвязи между рассматриваемыми параметрами характеризуются корреляционной матрицей R , элементами которой r_{ij} являются обычные парные коэффициенты корреляции i -го параметра с j -тым, подсчитанные по всем N объектам. Очевидно, что R — симметричная матрица с единицами на главной диагонали.

Сущность метода главных компонент заключается в поиске некоррелированных нормированных линейных комбинаций параметров

$$F_i = \sum_{j=1}^M a_{ij} x_j,$$

обладающих особым свойством: первая из них имеет наибольшую дисперсию, вторая — наибольшую дисперсию из всех линейных комбинаций, некоррелированных с первой и т. д. Можно показать (Харман, 1972), что дисперсии этих главных компонент — λ_i являются собственными числами корреляционной матрицы R , т.е. корнями уравнения

$$|R - \lambda I| = 0,$$

где I — единичная матрица, а матрица нагрузок компонент a_{ij} образуется из собственных векторов матрицы R — U_i , удовлетворяющих условиям:

$$(R - \lambda_i I) U_i = 0 \quad \text{и} \quad U_i^* U_i = 1,$$

так что эти собственные вектора являются строками матрицы нагрузок. Таким образом, введя в рассмотрение вектор — столбец параметров — X можно записать:

$$F_i = U_i^* X.$$

Следует отметить, что в данном случае предполагается, что параметры объектов — X_{ij} приведены к стандартной форме (с нулевым средним и единичной дисперсией) и главные компоненты ищутся по корреляционной матрице R . Часто, однако, система главных компонент строится для ковариационной матрицы параметров. Можно показать, что использование ковариационной матрицы оправдано лишь в тех случаях, когда все параметры измеряются в одних и тех же единицах. Поскольку метод связан с суммарной дисперсией параметров, то из-за произвольности выбора единиц измерения, найденные главные компоненты могут не иметь содержательного смысла. Использование стандартизованных переменных и корреляционной матрицы устраняет эти неприятности. Однако в этом случае происходит уравнивание величин, несущих разную информацию. Тем не менее, на первых этапах исследования легче предположить одинаковую значимость параметров, чем ставить результаты анализа в зависимость от произвольности единиц измерения.

Сущность главных компонент можно пояснить на двумерном примере. Пусть дана совокупность объектов, каждый из которых характеризуется двумя параметрами — X_1 и X_2 . Тогда на плоскости этих параметров точки, соответствующие объектам, будут образовывать, при нормальном законе распределения, корреляционный эллипс. В этом случае переход к главным компонентам будет эквивалентен повороту осей координат так, что ось, соответствующая компоненте F_1 , будет направлена вдоль большой оси, а F_2 — вдоль малой оси корреляционного эллипса.

Таким образом, математически задача нахождения главных компонент сводится к вычислению корреляционной матрицы параметров и нахождению ее собственных чисел и собственных векторов. Стандартные программы, реализующие решения этих задач, имеются в математическом обеспечении большинства ЭВМ, и поэтому процедура вычисления главных компонент не представляет особого труда.

Описанный метод был использован для анализа совокупности 74 звезд типа RR Lyr, приведенных в списке Джонса (1973), из которого были исключены звезды типа δ Sct, карликовые цефеиды и звезды, для которых не были известны все рассматриваемые параметры. Всего рассматривалось семь параметров:

$$\begin{aligned} X_1 &= P && \text{— период изменения блеска;} \\ X_2 &= \Delta V && \text{— амплитуда изменения блеска;} \\ X_3 &= \Delta(B-V) && \text{— амплитуда изменения показателя цвета;} \end{aligned}$$

$X_4 = \epsilon$ — асимметрия кривой блеска;
 $X_5 = (K-b)_2$ — индекс металличности;
 $X_6 = (u-b)_3$ — индекс ускорения силы тяжести на поверхности звезды;
 $X_7 = \beta_1$ — индекс эффективной температуры.

Данные по первым четырем параметрам были взяты из третьего дополнения ОКПЗ.

После перевода параметров в стандартную форму для всей совокупности звезд типа RR Lyr была вычислена корреляционная матрица, верхний треугольник которой имеет вид:

1.00	0.29	0.20	-0.59	-0.32	0.10	-0.66
	1.00	0.81	-0.71	0.06	0.13	-0.23
		1.00	-0.68	0.29	0.03	-0.26
			1.00	-0.12	0.03	0.60
				1.00	-0.56	-0.19
					1.00	0.00
						1.00

При использовании метода главных компонент вопрос о значимости коэффициентов корреляции не ставится. Тем не менее, нельзя не отметить, что некоторые из них достаточно велики, хотя и указывают либо на очевидные, либо на уже известные закономерности между параметрами.

Для этой корреляционной матрицы были вычислены главные компоненты, причем вся процедура заняла 5 минут на ЭВМ "Минск-32". Матрица нагрузок для трех главных компонент, имеющих наибольшие дисперсии, имеет вид:

F_1	0.37	0.46	0.45	-0.53	0.08	0.02	-0.39	43.8%
F_2	-0.36	0.03	0.19	-0.02	0.70	-0.59	0.05	24.5%
F_3	0.44	-0.45	-0.41	-0.03	0.10	-0.36	-0.55	17.5%

В последнем столбце приведена доля суммарной дисперсии параметров, учитываемой главной компонентой.

Таким образом, в совокупности они учитывают 85.8% всей дисперсии, что говорит о существенном сжатии информации, получаемом при рассмотрении этих главных компонент.

Абсолютная величина коэффициентов матрицы нагрузок говорит о значении того или иного параметра для данной главной компоненты. Так, например, для F_1 существенны все параметры, за исключением индекса металличности X_5 и индекса ускорения силы тяжести X_6 , причем наибольшее значение имеет асимметрия кривой блеска X_4 , а для компоненты F_2 , наоборот, параметры X_5 и X_6 являются наиболее существенными.

Для всей совокупности исследуемых звезд были построены двумерные диаграммы на плоскостях, соответствующих различным параметрам главных компонент. Наиболее четкое разделение объектов наблюдается на диаграмме $F_1 F_3$, которая приведена на рис. 1а. В хорошо выделяющуюся группу из 12 объектов, имеющих большие отрицательные значения F_1 , попадают все имеющиеся в исследуемой совокупности звезды типа RRc. Этот результат, безусловно, не является неожиданным, так как обособленность группы звезд RRc по отношению к другим звездам типа RR Lyr в

настоящее время не вызывает сомнений. Тем не менее, он служит хорошим доказательством пригодности метода главных компонент для решения задач классификации объектов.

В дальнейшем все звезды типа RRc были исключены из рассмотрения и оставшийся ансамбль из 62 звезд RRab был вновь подвергнут процедуре анализ... Верхний треугольник корреляционной матрицы в этом случае выглядит следующим образом:

1.00	-0.37	-0.52	0.05	-0.65	0.21	-0.39
	1.00	0.68	-0.28	-0.09	0.26	0.51
		1.00	-0.36	0.23	0.11	0.44
			1.00	0.13	-0.17	0.02
				1.00	-0.59	-0.08
					1.00	0.00
						1.00

а матрица нагрузок для трех главных компонент имеет вид:

F'_1	-0.47	0.49	0.54	-0.20	0.21	0.00	0.40	37.3%
F'_2	-0.33	-0.28	-0.09	0.28	0.62	-0.59	-0.10	28.2%
F'_3	0.00	0.10	-0.16	0.76	-0.21	0.01	0.58	14.5%

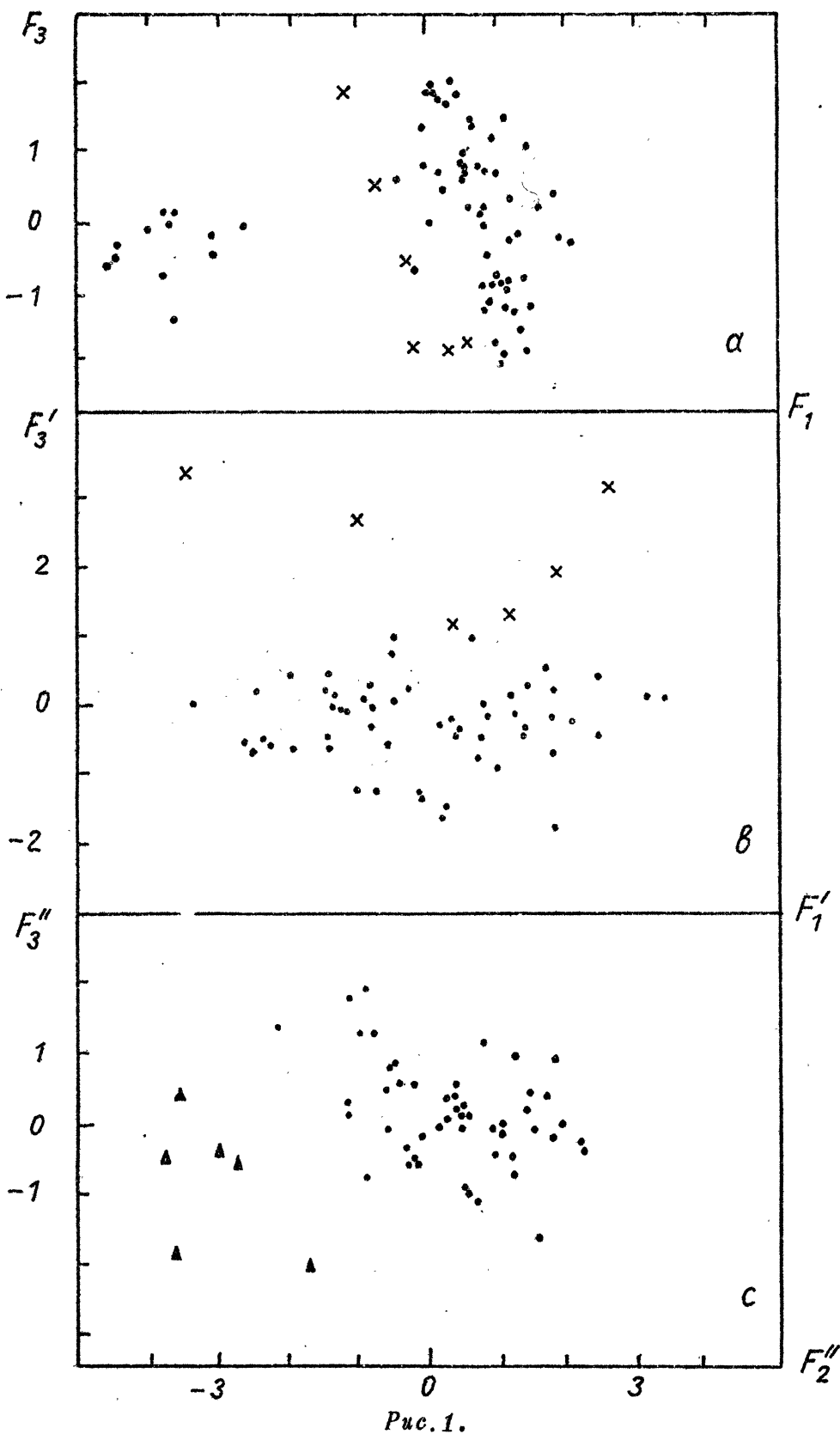
В совокупности эти главные компоненты учитывают 80% общей дисперсии.

Наиболее заметное разделение объектов, хотя и не такое четкое, как в первом случае, наблюдается на плоскости $F'_1 F'_3$ (рис. 1в). Выделенная группа из шести звезд отмечена на этом рисунке крестиками. В порядке возрастания компоненты F'_1 этими звездами являются SV Eri, RT Gru, RW Tra, RY Com, RV Car и V494 Sco. Для всех этих объектов характерны большие значения ϵ и, за исключением RW Tra, малые $(K-b)_2$ (в случае RV Car, кривая блеска которой испытывает сильные изменения, для ϵ было взято среднее значение). Интересно отметить, что на диаграмме $F'_1 F'_3$ эти звезды, также отмеченные крестиками, занимают промежуточное положение между основной группой объектов и звездами RRc.

Большая асимметрия кривой, как известно, свойственна звездам типа RRb по классификации Бейли. Однако для выделенных шести объектов это условие не является достаточным. В исследуемой совокупности имеются еще три звезды (DX Del, V445 Oph и BH Peg) с $\epsilon \geq 0.20$, которые, тем не менее, лежат глубоко в основной группе, как на диаграмме $F'_1 F'_3$, так и $F'_1 F'_2$.

После исключения этих шести звезд был снова проведен поиск главных компонент для оставшихся 56 объектов. Верхний треугольник корреляционной матрицы в этом случае имеет вид:

1.00	-0.33	-0.51	0.14	-0.66	0.23	-0.53
	1.00	0.65	-0.38	-0.09	0.25	0.54
		1.00	-0.39	0.26	0.09	0.49
			1.00	0.24	-0.25	-0.38
				1.00	-0.59	-0.10
					1.00	0.08
						1.00



а матрица нагрузок для трех главных компонент, учитывающих в совокупности 78.7% общей дисперсии, выглядит следующим образом:

F_1''	-0.39	0.49	0.52	-0.35	0.10	0.08	0.45	39.2%
F_2''	0.42	0.16	-0.07	-0.28	-0.64	0.54	0.10	29.9%
F_3''	0.08	-0.31	-0.25	-0.64	-0.13	-0.55	0.33	9.6%

Наиболее заметное разделение объектов наблюдается на плоскости $F_2'' F_3''$ (рис. 1с). Как и в предыдущем случае, выделенная группа насчитывает шесть объектов, которые отмечены треугольниками на рис. 1с. В порядке возрастания компоненты F_2'' этими звездами являются SW And, V445 Oph, AV Peg, SW Cru, DX Del и BV Pup. Для всех этих объектов характерны малые периоды ($P \leq 0.480$), сравнительно большая асимметрия кривой блеска ($\epsilon \geq 0.16$) и большие значения индекса металличности $((K-b)_2 \geq 0.22)$. Следует отметить, что на предыдущих диаграммах эти шесть звезд никак не выделяются из основной группы звезд RRab.

Процедуру нахождения главных компонент можно было бы продолжить и для оставшихся объектов, однако их немногочисленность не позволяет надеяться на получение каких-либо уверенных результатов.

Таким образом, проведенный анализ указывает на возможность существования среди звезд RRab по крайней мере трех групп объектов. К сожалению, число объектов в двух из этих групп невелико. Кроме того, полученные диаграммы имеют заметный разброс точек, который может быть вызван как случайными ошибками определения параметров звезд, так и, возможно, недостатками линейной модели, положенной в основу анализа. Поэтому полученные результаты нуждаются в дальнейшем подтверждении. Тем не менее, можно надеяться, что применение метода главных компонент к большему числу объектов и расширение набора используемых параметров будет способствовать решению задачи классификации звезд типа RR Lyr.

Литература:

- Андронович П.Ф., 1973, "Применение метода главных компонент в практических исследованиях", изд-во МГУ, М.
 Джонс, 1973 — Jones D.H.P., *ApJ Suppl.* **25**, 487.
 Клякотко М.А., 1977, Письма в АЖ **3**, 129.
 Налимов В.В., 1971, "Теория эксперимента", изд-во "Наука", М.
 Харман Г., 1972, "Современный факторный анализ", изд-во "Статистика", М.
 Цесевич В.П., 1970, в сб. "Пульсирующие звезды", изд-во "Наука", М., стр. 177.

Саратовский государственный
университет им. Н.Г. Чернышевского

Поступила в редакцию
25 апреля 1978 г.