THE ASTROPHYSICAL JOURNAL, 446:108–114, 1995 June 10

AN OBJECTIVE MULTICOLOR METHOD FOR THE CHARACTERIZATION OF LOW-RESOLUTION X-RAY SPECTRA

A. COLLURA,[1] G. MICELA,[2] S. SCIORTINO,[2] F. R. HARNDEN, JR.,[3] AND R. ROSNER[4]

Received 1993 October 14; accepted 1994 December 15

ABSTRACT

We describe a new technique for the characterization of low-resolution X-ray spectra, based on generalized colors defined by means of a principal component analysis. Application of the method to a sample of optically and radio-selected *Einstein Observatory* X-ray sources yields groups that are dominated either by stars or by extragalactic sources. The method thereby provides an immediate interpretation in terms of differences in the physics of the emission processes dominating the objects studied. The principal advantages of the method are that it provides an objective means of classifying source spectra and that it provides an objective means of obtaining the minimal set of spectral classification parameters. Though the method does not significantly improve results obtained with previously employed techniques when applied to *Einstein* data, its objective operation should prove powerful in aiding source classification in the more recent medium spectral resolution X-ray surveys based on the *ROSAT* and *ASCA* mission.

Subject headings: X-rays: galaxies — X-rays: stars

1. INTRODUCTION

The advantages of an objective, model-independent method for the efficient characterization of X-ray spectra can easily be seen by considering the large amount of data presently available from the low-resolution detectors employed in X-ray astronomy. Until now, most studies of X-ray spectra have employed either parameter estimation (via model fitting to observational data (see Schmitt et al. 1990) or construction of color indices (see Maccacaro et al. 1988; Kim, Fabbiano, & Trinchieri 1992). The former technique provides useful constraints on free model parameters (e.g., source temperature and line-of-sight absorption), but such an analysis requires a priori assumptions about the underlying X-ray emission mechanisms and can only be conducted on relatively strong sources, particularly when the number of free parameters is large. In contrast, while use of color indices requires no a priori assumptions about the objects under study, can be applied to relatively weak sources, and is easily implemented, color indices employed to date have not made optimal use of the data. Instead, the definition of such indices has often been based on limited empirical tests with small numbers of spectra, and has at times been dictated by considerations of uniformity with other studies.

For these reasons, an objective, model-independent method of characterizing objects through optimal use of available spectral information would find application in a variety of X-ray astronomy problems. Such applications include the selection of appropriate objects for further study in other wavelength regimes solely on the basis of X-ray-selected source classes, the study of extended sources (e.g., the X-ray background) through the identification of spatially correlated spectral features, and an extension of the method to permit joint analysis of X-ray data together with observations at other wavelengths.

[1] Istituto per le Applicazioni Interdisciplinari della Fisica, CNR, Via Archirafi 36, Palermo, Italy.
[2] Istituto ed Osservatorio Astronomico, Piazza del Parlamento 1, I-90134 Palermo, Italy.
[3] Smithsonian Astrophysical Observatory, Cambridge, MA 02138.
[4] Department of Astronomy & Astrophysics, University of Chicago, Chicago, IL 60637.

Our paper is organized as follows: in the next section, we describe a method for the characterization of X-ray spectra in terms of low-resolution "colors" that are "optimal" in the sense of providing the minimal set of spectral classification parameters. Section 3 then defines the composition of our test sample and describes the application of this method to a specific case of interest in X-ray astronomy, namely, one in which source "colors" are defined by low-resolution energy spectra. The final section is devoted to a summary and discussion of the results of applying our analysis to existing *Einstein Observatory* data. However, the advantages of this method become particularly evident when applied to data sets consisting of very large numbers of unclassified sources, viz., the *ROSAT* and *ASCA* data sets.

2. AN OBJECTIVE MULTIPLE-COLOR METHOD

The main features of the classification method described here—principal component analysis (PCA)—are (1) its model-independent classification scheme, (2) its objective use of the data, and (3) its ability to determine the minimal set of relevant classification parameters.

PCA is described in many textbooks on multivariate statistical data analysis (e.g., Murtagh & Heck 1987 and references therein). In the context of astronomical classification, it has been used extensively in the optical, as illustrated by the recent study of QSO optical and UV spectra by Francis et al. (1992). Briefly and qualitatively, PCA identifies the directions of maximum data variance (i.e., greatest "data spread") in the original N-dimensional data space defined by the N distinct observables for a data set consisting of M distinct objects (thus, there are M data points in this N-dimensional space); equivalently, PCA finds the directions of maximum variance in the M-dimensional object space (in which the ith data vector provides the observed values for the i observable for all M objects). This result is obtained by diagonalizing the correlation matrix of the original data and leads to a new N-dimensional orthogonal sets of vectors, ordered by the variance along each of these N "directions." Under many circumstances, a sharply reduced subset of n such vectors ($n \ll N$) turns out to describe most of the observed variance in the data; this is often the practically

most useful outcome of PCA, namely, the sharp reduction in the number of observationally determined parameters which describe the data set.

In the context of imaging X-ray astronomy, typical focal plane detectors provide moderate spectral resolution by sorting detected photons from a given source by their energy into pulse-height (PH) channels characterizing the amplitude of the electrical signal produced by each detected photon. The energy resolution of the detector is determined by the number of relatively independent PH channels within the energy range of the detector; however, in common practice the actual number of PH channels is substantially in excess of the number of truly independent energy channels, so that the PH channel values for a given source are not entirely independent measurements of the source's energy spectrum. Thus, each detected source is described by a set of N numbers, which is simply the N measured values ("colors") of the N pulse height (PH) channels of the detector (an imaging proportional counter in the case of the *Einstein Observatory*). For a set of M sources, we then have a set of M N-dimensional vectors, characterizing the N "colors" of our M sources. In most practical cases, this N-dimensional description introduces (substantial) redundancy because—as noted just above—the spectral resolution of the instrument is usually much coarser than the energy passband of the individual channels. PCA then provides an "objective" means of identifying the number, $n < N$, of "colors" that completely characterize the energy resolution of the detector; these are referred to as the "principal components."

An essential aspect of PCA is that the "optimal" set of colors, as defined above, is entirely determined by a combination of the instrumental characteristics *and* the nature of the sources making up the data set. Thus, one should expect that PCA will lead to different choices of "optimal" color sets for analyses based on (for example) stars alone, extragalactic sources alone, or stars and extragalactic sources. Hence, PCA does not intrinsically provide physical interpretations for the principal components; rather, it provides an objective means of describing the main differences in the nature of the spectra comprising the data set. In many practical cases, however, these differences can in turn be attributed to physically meaningful differences in the way that the radiation is produced by the sources themselves or in the way that intervening matter between us and the sources modify the source spectra.

As already noted, the eigenvalues $\{\lambda_i\}$ $(i = 1, \ldots, N)$ of the original data correlation matrix indicate the amount of "spread" accounted for by each principal component. The decision of how many principal components to retain is then usually based on the fraction of the total data variance which is taken into account by using $n < N$ principal components, as measured by, for example, the ratio $f = \sum_{i=1}^{n} \lambda_i / \sum_{i=1}^{N} \lambda_i$; this method allows one to determine how much of the "data spread" (i.e., the percentage of the total variance) is still unexplained. Alternatively, one can use the principal-component projected data (i.e., inverse-transformed back to the original N-dimensional space) to fit the original data and determine from the quality of the resulting fit (e.g., with a χ^2 test; see Francis et al. 1992) whether a larger number of components need be retained. Such a technique indicates at the same time peculiar spectra, i.e., those requiring an anomalously high number of components. We note in passing that this last task cannot be carried out systematically by looking at simple color-color scatter plots produced by standard techniques, if more than three colors are being used. In the present case, we have used the simpler of these two above-mentioned alternatives, namely, the ratio test.

Once the optimal "colors" have been found, an objective criterion for identifying distinct groups of objects with common spectral characteristics must be established. Among the several suitable techniques, we have chosen to use an ascendant hierarchical cluster analysis method known as the *complete linkage* method (see Murtagh & Heck 1987 and references therein). This technique constructs a hierarchy of groups, starting at the bottom (i.e., with single elements) and forming successive groups up to a single group (the top of the hierarchy) that contains the entire sample. The (Euclidean) distance between pairs of sample elements in the principal component space is the criterion used in forming the groups: the two closest elements form the first group. Successive clustering considers the previously formed groups as single elements and again designates the two closest elements as a new group. In this agglomerative method, the distance between a point and a group is taken as the greatest of the distances between that point and the elements of the group. Similarly, the distance between two groups is defined as the largest of the distances between elements of the two groups.

After "clustering" our sample, we determined the centroid of each group and reassigned the elements to groups on the basis of minimum distances from the centroids. This procedure was iterated until stable groups were obtained. This departure from the classical method makes the method almost independent from the different arbitrary possible choices of distance among clusters. Although this procedure is essentially automatic in its application, there is an element of arbitrariness in determining the number of groups to use in this classification. Having too few groups does not permit sufficiently fine classification, but retaining too many groups reduces the number of elements in individual groups and thereby renders their existence as distinct groups uncertain.

3. APPLICATION OF THE METHOD

In order to test the effectiveness of our method, we selected a sample of X-ray sources detected with the *Einstein Observatory* Imaging Proportional Counter (IPC) (Gorenstein, Harnden, & Fabricant 1981) and whose identifications were known from observations at other wavelengths. In this section, we describe the instrument and the sample we studied, followed by our results.

3.1. *IPC Spectra*

The *Einstein* IPC, a proportional counter with imaging capability and moderate energy resolution, provided information about the energy of individual photons selected from regions of interest within the X-ray telescope's field of view. The spectral resolution of the IPC is a function of photon energy, and it describes the statistical uncertainty with which photons of a given energy are assigned to detected pulse-height channels. The spectral resolution of the instrument at 1 keV is $\Delta\lambda/\lambda = 1$, much broader than the width of the individual channels. For most analysis purposes, the IPC's PH signals were binned in 16 channels. Since the gain of the instrument was not constant in time, the *Einstein* standard data analysis system (Rev1; see Harnden et al. 1984) compensated for these variations by reassigning the PH channels to pulse-height independent (PI) bins whose boundaries remained fixed in energy. In our analysis, we have employed PI bins 1 through 10, which

TABLE 1

TEST SAMPLE COMPOSITION

Parameter	All	Extragalactic	Woolley	BSC
Identified with:				
Detections.....	418	277	33	108
Sources	329	224	25	94
Number of observations				
1.....	275	174	21	80
2.....	36	26	1	9
3.....	10	6	2	2
4.....	3	1	1	1
5.....	2	2	0	0
6.....	2	2	0	0
7.....	1	1	0	0

span the energy range 0.04–3.5 keV; hence, each IPC source spectrum is a vector of 10 numbers.

3.2. *A Test Sample*

Our test sample included those sources detected in the broad energy band (0.16–3.5 keV) using the Local detection algorithm, as implemented in IPC Rev1 processing. We discarded any detections near the window-support structure or detector edges to minimize possible effects of electric field distortion (see Micela et al. 1988). In addition, we corrected our data[5] for a recently discovered Rev1 processing error (i.e., the "picorr bug"; see Rhoe & Harnden 1993; Prestwich et al. 1992), retaining only those detections whose X-ray spectra contained more than 50 (background subtracted) net counts and which could be identified with high Galactic latitude ($|b| > 20°$) entries in one of the following catalogs:[6] The Fourth Yale Bright Star Catalogue and its supplements (Hoffleit & Jaschek 1982; Hoffleit, Saladyga, & Wlasuk 1984, hereafter BSC), the Woolley catalog (Woolley et al. 1970), the Zwicky catalog (Zwicky et al. 1961–68), or the Burbidge catalog (Burbidge, Crowne, & Smith 1977). We refer to stars present in both the BSC and Woolley catalogs as BSC stars, and for an identification, we required the X-ray/optical position separation to be less than 2′.

Table 1 characterizes the resulting sample, which contains 418 X-ray detections, corresponding to 329 individual sources, and has recognized biases introduced by the adopted selection criteria. For example, OB and pre–main sequence stars are poorly represented in the BSC and are completely missing from Woolley. Since these stars are preferentially located near the Galactic plane, our restriction to high Galactic latitudes virtually eliminates them from the test sample. We also note that BL Lac objects are not represented in the Zwicky and Burbidge catalogs and are therefore largely absent from our sample. Also, most active galactic nuclei (AGNs) listed in the Burbidge catalog were originally identified from their radio position and are therefore radio loud.

[5] We note, however, that correcting for the picorr bug did not yield a statistical improvement in our results, probably because most of the sources we studied were located in the center of the field, where the PI bins were least affected by the gain-map roll error responsible for this bug.

[6] The BSC is a star catalog complete to $m_v = 6.3$ and containing objects to $m_v = 7.1$ (Bahcall, Casertano, & Ratmatunga 1987). The Woolley catalog contains all known stars (mainly late-type main-sequence stars) within 25 pc of the Sun. The Burbidge catalog is a sample of radio and optically selected QSOs; and the Zwicky catalog contains ~30,000 galaxies and ~10,000 clusters of galaxies, is complete for the northern hemisphere, and extendes a few degrees into the southern hemisphere.

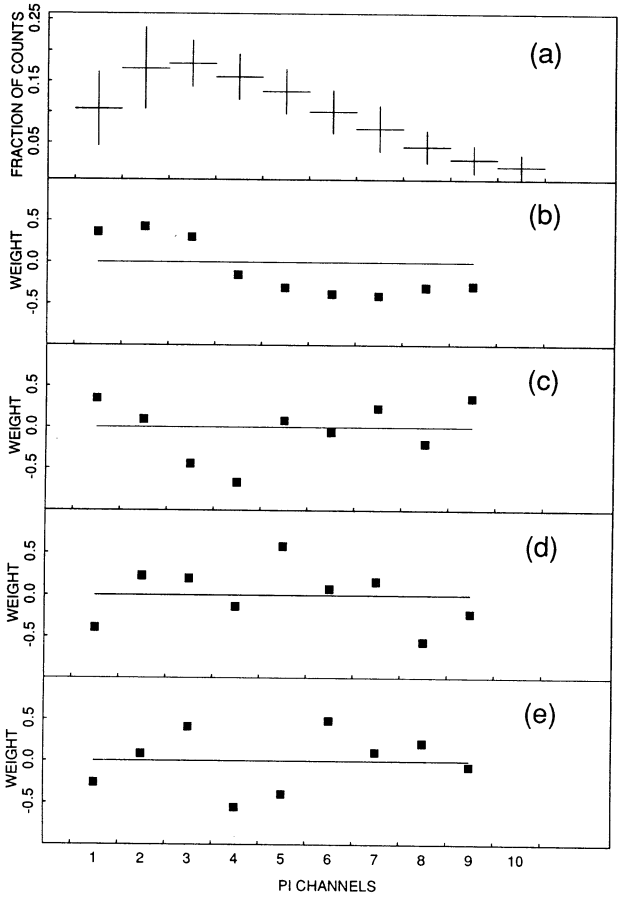

FIG. 1.—Mean spectrum (*a*) and weights of PI bins that determine the first four components for stars only. Panels (*b*), (*c*), (*d*), and (*e*) refer to the first, second, third, and fourth component, respectively.

The X-ray selection criteria also introduce a bias. Since the sensitivity of the observations employed spans of only about an order of magnitude, the cut at 50 X-ray counts renders the whole sample essentially X-ray flux limited. Our sample is therefore representative,[7] for each class, of objects belonging to the high-luminosity tails of the luminosity functions.

3.3. *Analysis of the Test Sample*

In Figures 1, 2, and 3 we show the mean PI spectrum (*a*) and the first four principal components (panels *b*, *c*, *d*, and *e*, respectively) for two different subsamples of the test sample, namely, stars only (Fig. 1) and extragalactic sources only (Fig. 2), and for the entire test sample (Fig. 3). The first component can be interpreted as being related to spectral hardness (since the PI bin weights that determine this first component are such that soft spectra tend to have a positive projection on this component, while hard spectra have a negative one); the second component emphasizes counts falling in intermediate PI bins (viz., bins 4–5) and discriminates soft, absorbed spectra from hard, unabsorbed spectra; interpretation of the third and fourth components remains unclear. In order to get an insight into how theoretical spectra are described by the first two

[7] For instance, although the BSC is populated with both A stars and giants, A stars virtually disappear with the application of the X-ray selection criteria, owing to the low level of A star X-ray emission (Schmitt et al. 1985).

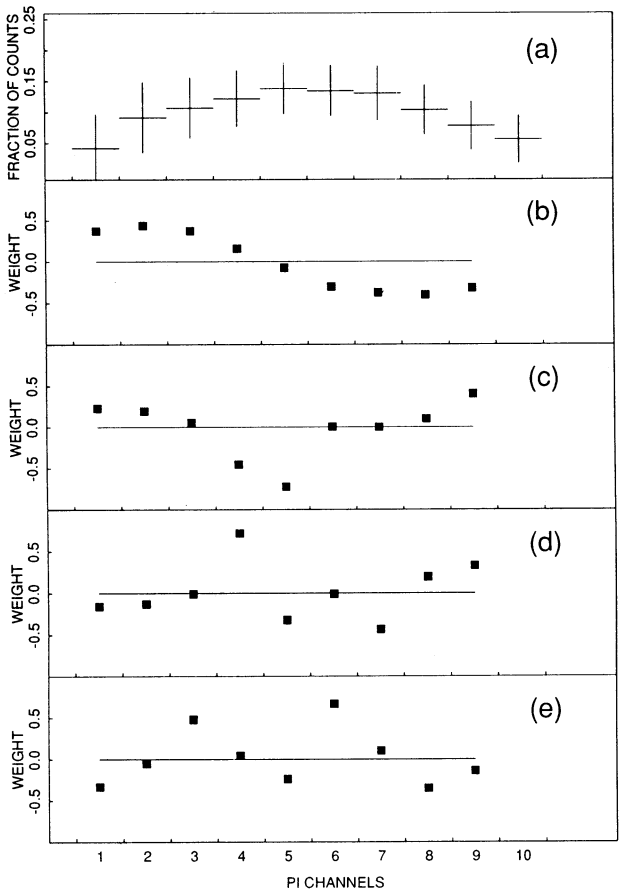

FIG. 2.—Same as Fig. 1, for extragalactic sources only

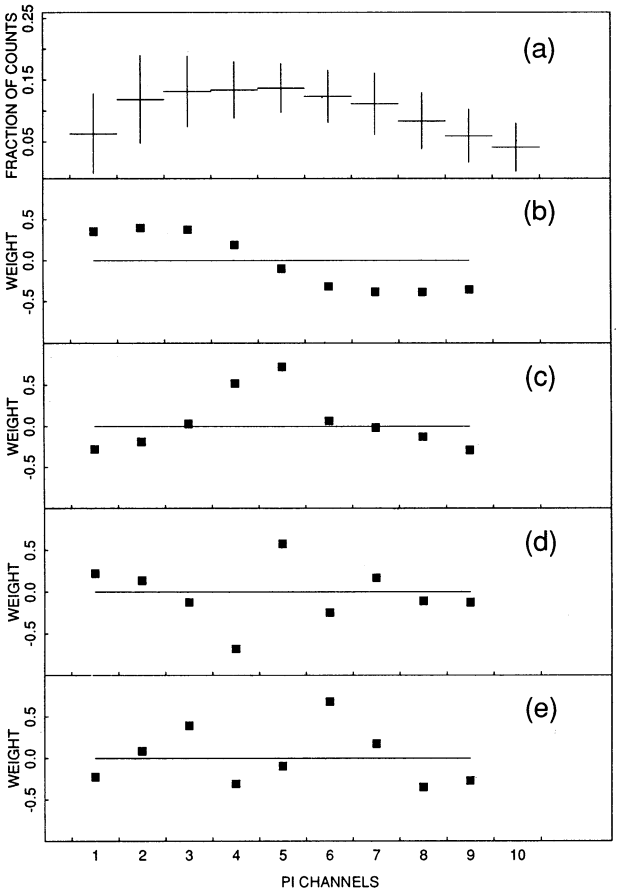

FIG. 3.—Same as Fig. 1, for all sources in the test sample

components, we have generated a number of noiseless isothermal Raymond-Smith (R-S) and single-slope power-law (PL) spectra with different parameters and projected them onto the plane generated by the first two components defined by the entire test sample. Figure 4*a* refers to R-S spectra. Each curve, identified by a particular symbol and type of line, refers to a value of N_H. In particular, N_H (cm^{-2}) ranges from 10^{19} (*squares*) to 10^{22} (*plus signs*); circles and triangles are at 10^{20} and 10^{21}, respectively. Points on the same curve differ by temperature value. Temperature increases leftward, taking at each data point the following values (keV): 0.01, 0.1, 0.3, 0.5, 0.7, 1.0, 1.3, 2.0. Though only isothermal spectra have been considered, the location of multitemperature spectra can easily be extrapolated: their position would be intermediate among the data points corresponding to the various temperatures and exactly determined by the emission measure distribution. Therefore, it is possible that temperature stratifications produce spectra filling the region underneath the bell-shaped curves of Figure 4*a*. Figure 4*b* refers to PL spectra. As in the case of Figure 4*a*, different symbols and lines refer to different N_H values; in particular, squares, circles, and triangles identify N_H values of 10^{20}, 10^{21}, and 10^{22} cm^{-2}. The energy index α increases rightward, taking at each data point the values 0.5, 1.0, 1.5, 2.0, respectively. It is interesting to note that different spectral models can, for certain spectral parameters, occupy the same region. This is a consequence of the coarse spectral resolution of the IPC and limits our ability in separating different spectra. Therefore, the effectiveness in separating classes of objects in a sample may depend strongly on the actual composition of the sample.

Application of our method to the test sample described above defined a set of principal components that provided good separation of stellar sources from extragalactic ones. The percentages of variance "explained" by each axis, together with the percentages of cumulative variance, were obtained from the PCA procedure and are given in Table 2. The first component alone accounts for ~48% of the variance, while inclusion of the next three components encompasses a further ~13%, 10%, and 8%, respectively, of the total variance. We have explored the effects of retaining both two and four components and report results for both below; the principal con-

TABLE 2

PRINCIPAL COMPONENT ANALYSIS

Component	Variance (%)	Cumulative Variance (%)
1	0.484	0.484
2	0.133	0.617
3	0.100	0.717
4	0.082	0.799
5	0.059	0.858
6	0.052	0.910
7	0.043	0.953
8	0.043	0.996
9	0.004	1.000

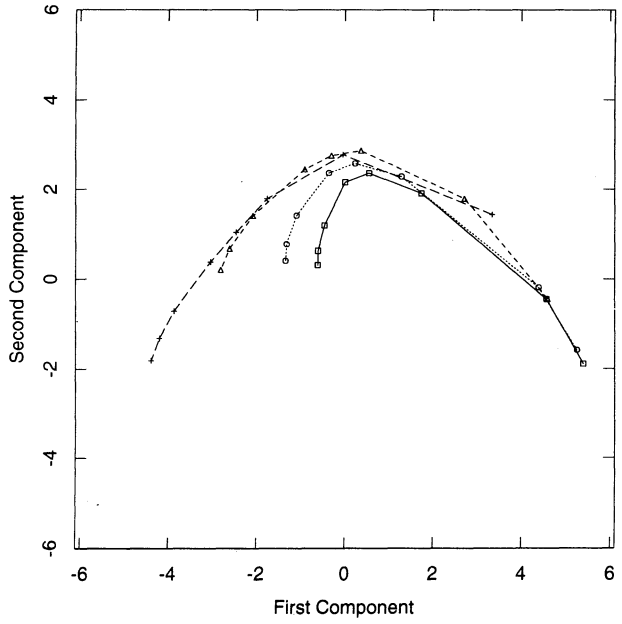

FIG. 4a

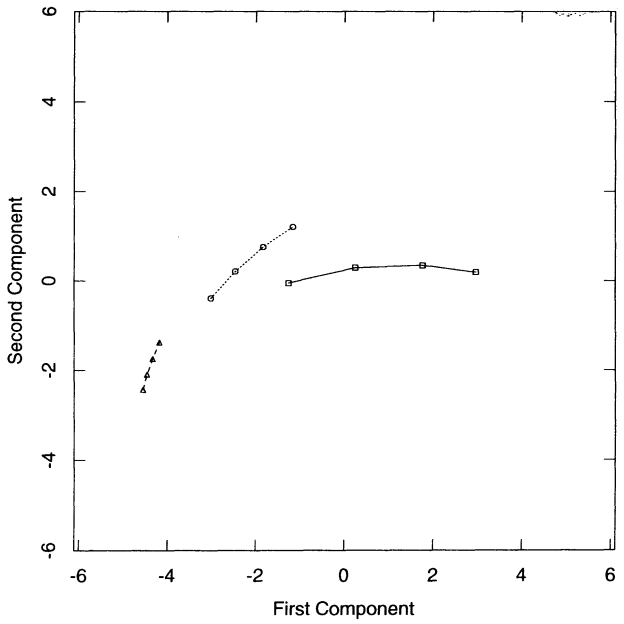

FIG. 4b

FIG. 4.—Projection on the first two component axes of noiseless spectra as seen with the IPC. (*a*) Raymond-Smith spectra: points on the same curve have the same N_H value. Values of log (N_H) (cm^{-2}) are 19 (*squares*), 20 (*circles*), 21 (*triangles*), 22 (*plus signs*). Temperature covers the values 0.01, 0.1, 0.3, 0.5, 0.7, 1.0, 1.3, 2.0 (keV), increasing leftward. (*b*) Power-law spectra: points on the same curve have the same N_H value. Values of log (N_H) (cm^{-2}) are 20 (*squares*), 21 (*circles*), 22 (*triangles*). The energy index covers the values 2.0, 1.5, 1.0, 0.5, increasing leftward.

clusion is that, based on our clustering analysis, only minor improvement in separation between stellar and nonstellar sources is gained by using four components. Thus, our analysis indicates that an enormous economy can be gained by using only two principal components (as opposed to, for example, the full set of nine independent "colors").

In order to test the effectiveness of the method, we explored its ability to classify sample sources into the broad categories of stellar versus extragalactic. Based upon content, we labeled each group obtained from the cluster analysis as extragalactic (E), stellar (S), or mixed (M), characterizing all groups with more than three sources and designating as "pure" (i.e., E or S) those groups whose fractional content in one pure category was at least twice that in the other. Our hierarchical cluster analysis results for the test sample are given in Tables 3A and 3B, which contains results for the two-component analysis. Parts A and B of the table refer to analyses conducted using 10 and 30 groups, respectively.

In the analysis with two components and 10 groups, 84% of stars fell into stellar groups, together with the 7% of extragalactic objects, while 88% of extragalactic objects fell into extragalactic groups, together with the 12% of stars. The two-component analysis with 30 groups yielded the following results: 83% of stars and 84% of extragalactic sources fell into stellar and extragalactic groups, respectively, with respective contaminations of 8% and 7%. Four-component, 30-group classification produced 88%-pure stellar groups, with 6% extragalactic-object contamination, and 87%-pure extragalactic groups, with 7% stellar contamination. Our attempts to improve on this last result went unrewarded. We note that replacing PC analysis with classical color indices like those defined by Kim et al. (1992) and maintaining cluster analysis, 78%-pure stellar groups and 82%-pure extragalactic groups, with 9% and 8% contamination, respectively, are obtained. Such a result is not significantly worse that that obtained with PC analysis and 2 components, thus demonstrating that Kim et al. (1992) color indices are a good choice for IPC data. Figure 5 shows a scatter plot of the projection of test sample spectra on the second principal component versus the projection on the first component. The stars (*squares* and *triangles*) are seen to occupy a well-defined region, largely separated from that occupied by the QSOs and galaxies, with a relatively sharp boundary between the two running obliquely through

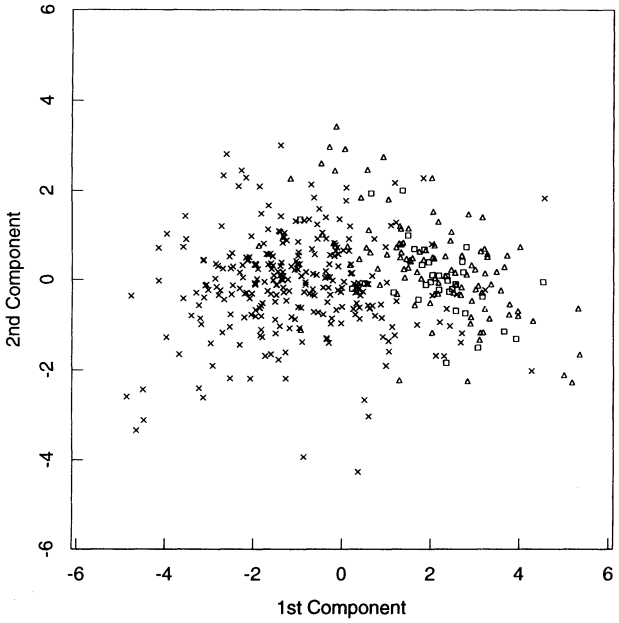

FIG. 5.—Projection on the first two component axes of the spectra of X-ray-detected objects contained in the optical sample. Open squares, triangles, and crosses indicate Woolley stars, BSC stars, and extragalactic sources, respectively.

TABLE 3A

RESULTS FOR TWO COMPONENTS, 10 GROUPS

Cluster Number	Type of Identification	Number of Extragalactic Detections	Number of Woolley Stars	Number of BSC Stars
1	S	6	6	19
2	E	51	0	1
3	S	7	6	32
4	E	19	0	0
5	E	5	0	1
6	E	71	1	4
7	M	15	0	6
8	E	47	3	6
9	E	50	0	1
10	S	6	17	38

TABLE 3B

RESULTS FOR TWO COMPONENTS, 30 GROUPS

Cluster Number	Type of Identification	Number of Extragalactic Detections	Number of Woolley Stars	Number of BSC Stars
1	S	3	11	14
2	E	6	0	0
3	E	22	1	1
4	E	27	0	1
5	S	6	4	7
6	E	37	0	0
7	E	49	0	1
8	S	0	2	7
9	S	1	0	3
10	S	1	3	18
11	E	5	0	0
12	S	0	1	9
13	M	28	0	0
14	E	8	0	1
15	S	2	1	3
16	S	2	1	2
17	E	4	0	0
18	S	4	5	23
19	E	27	3	3
20	M	19	0	8
21	M	2	0	0
22	E	6	0	0
23	S	1	0	5
24	E	11	0	0
25	M	2	0	0
26	M	1	0	0
27	M	1	0	0
28	M	1	0	1
29	M	1	0	0
30	M	0	1	1

the plane. This figure clearly demonstrates that these first two components can be used to separate stars from extragalactic objects. Incidentally, we note that a small number of stars occupies a region which does not correspond to any of the regions defined by simulated R-S spectra of Figure 6*a*, nor to any possible temperature stratification at a given N_H. Instead, a combination of different temperatures *and* absorptions, or of R-S *and* PL spectra could account for those points. It is interesting that some of these cases are high signal-to-noise ratio spectra.

In order to gauge the effects of source variability and multiple observations of a given object, we repeated the above analysis with all multiply observed sources excluded. This result, for 329 sources having only single detections, is in good agreement with those obtained using the entire test sample.

In order to evaluate how statistical noise affects the position of points in the PCA space, we have produced a set of simulated spectra using spectra of Figures 4*a* and 4*b* as input. For each data point of Figure 4 we produced 100 realizations with a distribution in counts closely reflecting that of sources in the test sample. Resulting scatter plots are shown in Figures 6*a* (R-S) and 6*b* (PL). We note that the intrinsic limit of the IPC in separating certain classes of spectra, which appeared already from Figure 4, can become more severe because of noise. Again we stress that the effectiveness of the method depends on the composition of the sample. As a matter of fact, we obtain fairly good results with our test sample, despite the shortcomings just pointed out and shown in Figures 4 and 6.

4. SUMMARY AND DISCUSSION

In this paper, we have shown that X-ray spectra can be characterized using an objective multiple-color scheme that requires no model assumptions, data fitting, or parameter estimation. When applied to *Einstein* IPC point-source PI spectra, defined by the numbers of counts in PI bins 1 through 10 and corresponding to the energy range from 0.04 to 3.5 keV, the individual spectra can be represented as points in a 10-dimensional space defined by these 10 selected PI channels. Our principal component analysis technique then applies hierarchical clustering to construct generalized "colors" with the property of maximally accounting for the variance in the IPC spectral data.

For the *Einstein* IPC data, we demonstrated that two such colors "explained" nearly two-thirds of the variance and that the use of four colors only marginally increased the ability to distinguish between source categories. This is a result of both the limited spectral resolution of the IPC and the intrinsic nature of the two classes of objects—stars and extragalactic sources—used on our study. The first two "colors" have a straightforward physical interpretation in terms of the IPC energy response (see Fig. 1): the first "color" much resembles the hardness ratio often used in previous analyses (see Cordova et al. 1990; Maccacaro et al. 1988), with large positive values for sources with soft spectra and large negative values for those with hard spectra. This "color" is therefore sensitive to the temperature distribution of the emitting plasma for the thermal sources. The second color is sensitive to the spectral flux in the middle of the bandpass and attains large positive values for absorbed spectra.

A cluster analysis of the distribution of sources in this multicolor space permitted the identification of maximally distinct source classes such that stellar sources were largely distinguished from extragalactic sources, with little mixing between the two types. This indicates that the IPC has sufficient spectral resolution to distinguish spectral features characteristic of most stars and extragalactic sources. A comparison of colors defined by PC analysis with color indices defined by Kim et al. (1992) has shown that the two methods are equally powerful when applied to *Einstein* data. Such a result confirms that Kim et al. (1992) colors, based on empirical definition, are a good choice for the interpretation of *Einstein* data. Our work therefore confirms validity of analyses based on these colors. We, however, wish to stress that empirical definition of colors is easily feasible only for a low energy resolution instrument like the *Einstein* IPC. It might become much more difficult in the presence of instruments with higher energy resolution and a higher number of energy channels, able to resolve detailed

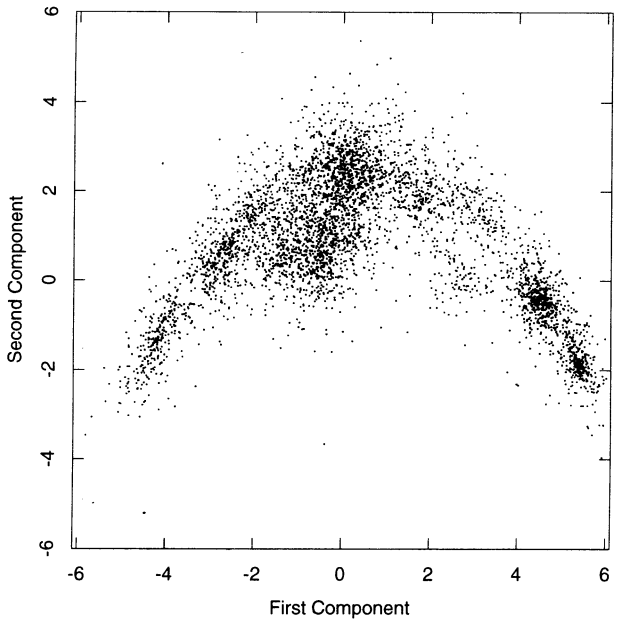

FIG. 6a

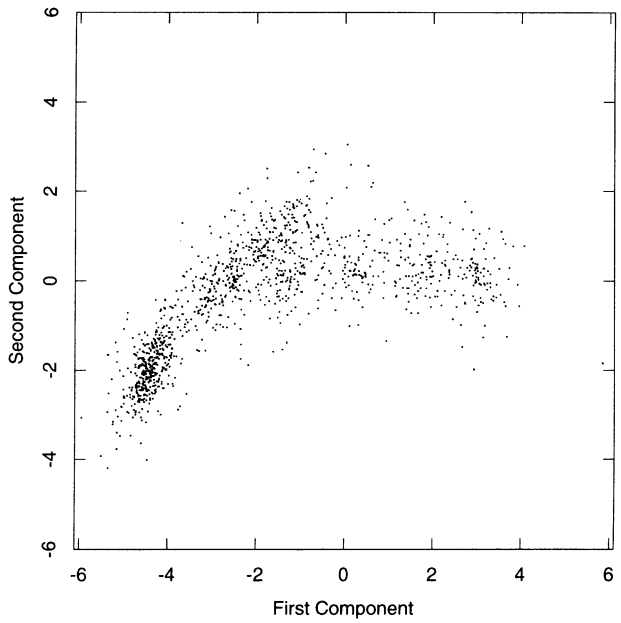

FIG. 6b

FIG. 6.—Projection on the first two component axes of noisy spectra. Input spectra are data points of Fig. 4. For each data point we plot 100 realizations with a count distribution resembling that of the test sample. Panel (*a*) refers to Raymond-Smith spectra, panel (*b*) to power-law spectra.

spectral features. In such a situation, the appropriate choice of colors depends strongly on both the instrument and the specific data set under study and may require an accurate weighting of individual channels which hardly can be achieved empirically.

The success of this technique with *Einstein* data suggests that it may prove useful for automated source classification in larger (and possibly higher energy resolution) data sets, such as those provided by *ROSAT* and *ASCA*.

The authors wish to remember the late G. S. Vaiana, who originally stimulated their interest in a nonparametric characterization of X-ray spectra. We thank the referee, K. Ebisawa, for his very constructive criticism.

REFERENCES

Bahcall, J. N., Casertano, S., & Ratmatunga, K. U. 1987, ApJ, 320, 515
Burbidge, G. R., Crowne, A. H., & Smith, H. E. 1977, ApJS, 33, 113
Cordova, F. A., Kartje, J., Mason, K. O., & Mittaz, J. P. D. 1990, in Imaging X-Ray Astronomy, ed. M. Elvis (Cambridge: Cambridge Univ. Press), 273
Francis, P. J., Hewett, P. C., Foltz, C. B., & Chaffee, F. H. 1992, ApJ, 398, 476
Gorenstein, P., Harnden, F. R., Jr., & Fabricant, D. G. 1981, IEEE Trans. Nucl. Sci., NS-28
Harnden, F. R., Jr., Fabricant, D. G., Harris, D. E., & Schwarz, J. 1984, SAO Special Rep. 393
Hoffleit, D., & Jaschek, C. 1982, The Bright Star Catalogue (4th ed.; New Haven: Yale Univ. Press)
Hoffleit, D., Saladyga, M., & Wlasuk, P. 1984, Supplement to the Bright Star Catalogue (New Haven: Yale Univ. Press)
Kim, D. W., Fabbiano, G., & Trinchieri, G. 1992, ApJS, 80, 645
Maccacaro, T., Gioia, I. M., Wolter, A., Zamorani, G., & Stocke, J. T. 1988, ApJ, 326, 680
Micela, G., Sciortino, S., Vaiana, G. S., Schmitt, J. H. M. M., Stern, R. A., Harnden, F. R., Jr., & Rosner, R. 1988, ApJ, 325, 798
Murtagh, F., & Heck, A. 1987, Multivariate Data Analysis (Boston: Reidel)
Prestwich, A., McDowell, J., Garcia, M., & Harnden, F. R., Jr. 1992, HEAO Newsletter, 1(7), 4
Rhode, K., & Harnden, F. R., Jr. 1993, Harvard-Smithsonian CfA internal memorandum
Schmitt, J. H. H. M., Collura, A., Sciortino, S., Vaiana, G. S., Harnden, F. R., Jr., & Rosner, R. 1990, ApJ, 365, 704
Schmitt, J. H. H. M., Golub, L., Harnden, F. R., Jr., Maxson, C. W., Rosner, R., & Vaiana, G. S. 1985, ApJ, 290, 307
Woolley, E., Epps, E. A., Penston, M. J., & Pocock, S. B. 1970, R. Obs. Ann., 3
Zwicky, F., et al. 1961–1968, Morphological Catalog of Galaxies (Pasadena: California Inst. of Tech.)