The Astrophysical Journal, 659:585–598, 2007 April 10

AN X-RAY SPECTRAL CLASSIFICATION ALGORITHM WITH APPLICATION TO YOUNG STELLAR CLUSTERS

S. M. Hojnacki,[1] J. H. Kastner,[1] G. Micela,[2] E. D. Feigelson,[3] and S. M. LaLonde[4]

Received 2006 November 22; accepted 2006 December 30

ABSTRACT

A large volume of low signal-to-noise, multidimensional data is available from the CCD imaging spectrometers aboard the *Chandra X-Ray Observatory* and the *X-Ray Multimirror Mission* (*XMM-Newton*). To make progress analyzing this data, it is essential to develop methods to sort, classify, and characterize the vast library of X-ray spectra in a nonparametric fashion (complementary to current parametric model fits). We have developed a spectral classification algorithm that handles large volumes of data and operates independently of the requirement of spectral model fits. We use proven multivariate statistical techniques including principal component analysis and an ensemble classifier consisting of agglomerative hierarchical clustering and *K*-means clustering applied for the first time for spectral classification. The algorithm positions the sources in a multidimensional spectral sequence and then groups the ordered sources into clusters based on their spectra. These clusters appear more distinct for sources with harder observed spectra. The apparent diversity of source spectra is reduced to a three-dimensional locus in principal component space, with spectral outliers falling outside this locus. The algorithm was applied to a sample of 444 strong sources selected from the 1616 X-ray emitting sources detected in deep *Chandra* imaging spectroscopy of the Orion Nebula Cluster. Classes form sequences in $N_{\rm H}$, A_V, and accretion activity indicators, demonstrating that the algorithm efficiently sorts the X-ray sources into a physically meaningful sequence. The algorithm also isolates important classes of very deeply embedded, active young stellar objects, and yields trends between X-ray spectral parameters and stellar parameters for the lowest mass, pre–main-sequence stars.

Subject headings: methods: statistical — open clusters and associations: general — stars: early-type — stars: pre–main-sequence — X-rays: stars

Online material: color figure, machine-readable table

1. INTRODUCTION

Chandra X-Ray Observatory (Weisskopf et al. 2002) studies often involve X-ray source populations with hundreds or thousands of members. Significant *Chandra* surveys to date include deep Galactic (Brandt & Hasinger 2005) and Galactic center (Muno et al. 2003, 2006) surveys, the *Chandra* Multiwavelength Project (ChaMP; Kim et al. 2004), and nearby galaxies (Kilgard et al. 2001). In addition, the *XMM-Newton* European Photon Imaging Camera (EPIC) catalog[5] already contains over 100,000 sources.

A substantial fraction of observing time has been devoted to the study of rich young stellar clusters. These observations have built on results from the earlier *Einstein Observatory*, *Röntgensatellit* (*ROSAT*), and *Advanced Satellite for Cosmology and Astrophysics* (*ASCA*) missions, demonstrating that strong X-ray emission is a defining characteristic of pre–main-sequence (PMS) stars (Feigelson & Montmerle 1999). Automated spectral classification and analysis is needed to sort and classify the vast database of X-ray spectra that have been collected by the CCD imaging spectrometers aboard *Chandra* and *XMM-Newton*, to characterize X-ray spectra in a nonparametric fashion (complementary to current parametric model fits), and to discriminate spectroscopically between various known classes of emission (stellar coronal emission, accretion, winds, shock fronts, and extragalactic contaminants).

A typical *Chandra* charge-coupled device (CCD) observation of a young stellar cluster results in detection of X-ray emission from tens to hundreds of very young stars. The foremost example of such an observation is the *Chandra* Orion Ultradeep Project (COUP; Getman et al. 2005b), an ~838 ks exposure of the Orion Nebula Cluster (ONC) obtained over a nearly continuous period of 13.2 days in 2003 January (Fig. 1). The COUP data set represents the most sensitive and comprehensive description of X-ray emission from a PMS star cluster (Getman et al. 2005a; Feigelson et al. 2007). *Chandra* has resolved more than 1600 X-ray emitting sources in the COUP image of the ONC (Fig. 1), ranging from X-ray sources associated with the massive Trapezium stars (Stelzer et al. 2005), to a Herbig-Haro object at the tip of a collimated outflow (Grosso et al. 2006), to the low-mass stars embedded within externally illuminated structures that are presumably planet-forming circumstellar disks (Kastner et al. 2005).

A new set of problems has been uncovered by the X-ray imaging observations of young stellar clusters. Among the challenges and puzzles are the following:

1. Only weak trends have been found when attempting to correlate model parameters derived from spectral fitting of individual sources with stellar properties (e.g., X-ray luminosity and X-ray temperature with stellar photospheric temperature; Preibisch et al. 2005).

2. The relationship between PMS accretion and X-ray emission is not simple. X-ray luminosities of accretors are systematically lower than those of nonaccretors (Preibisch et al. 2005), and

[1] Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623.

[2] INAF, Osservatorio Astronomico di Palermo G. S. Vaiana, 90134 Palermo, Italy.

[3] Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, PA 16802.

[4] Center for Quality and Applied Statistics, Rochester Institute of Technology, Rochester, NY 14623.

[5] The Second *XMM-Newton* Serendipitous Source Prerelease Catalogue, *XMM-Newton* Survey Science Centre, 2006 (http://xmmssc-www.star.le.ac.uk).

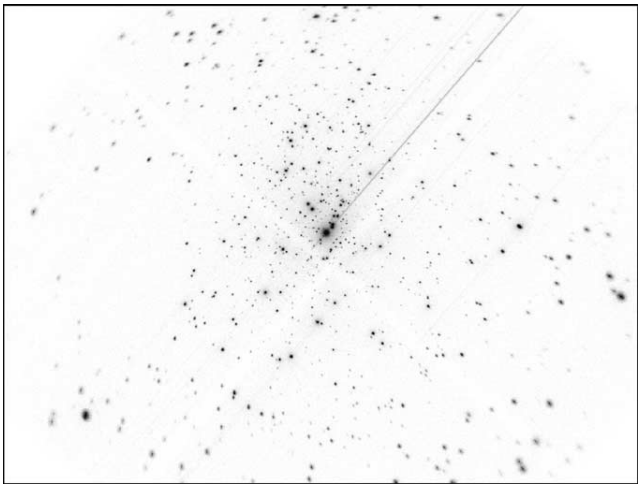

FIG. 1.—*Chandra* image of the ONC from the COUP observation (Getman et al. 2005b). More than 1600 X-ray emitting sources have been detected in this image. [*See the electronic edition of the Journal for a color version of this figure.*]

short-lived optical variations suggestive of accretional events do not seem to be accompanied by X-ray variations (Stassun et al. 2006). But accretors show soft excesses and spectral line signatures suggestive of emission in an accretion column rather than a coronal magnetic loop (Kastner et al. 2002; Telleschi et al. 2007), and strongly enhanced hard X-ray emission can accompany sustained accretion bursts (Kastner et al. 2004, 2006).

3. A very wide range of temporal behavior has been observed, ranging from rotational modulation to short, episodic X-ray bursts, to explosive X-ray flares with decays of many hours interfilled by hard quasi-constant emission attributed to smaller flares (Flaccomio et al. 2005; Wolk et al. 2005; Favata et al. 2005).

4. Approximately 18% of the 1616 detected X-ray sources in and around the ONC have no visible or infrared (IR) counterparts; many of these are background (extragalactic) sources (Getman et al. 2005a), but some are very deeply embedded protostars (Grosso et al. 2005).

The identification of X-ray spectral classes would greatly facilitate progress on these and other problems. For example, an automated spectral classification tool may aid in finding additional PMS accretors within this large volume of data for further study. To take advantage of the vast collection of high spatial resolution CCD spectral data now being acquired by *Chandra* requires the use of objective, model-independent methods for spectral grouping of X-ray sources in young star clusters, as opposed to fitting physical models to individual source spectra. The latter approach is essential for understanding the physics underlying individual, bright X-ray sources; however, it is an inefficient method for analyzing and classifying sources detected in the large data sets generated by observations of rich stellar clusters. Furthermore, once grouped spectrally, sources within the same spectral group may be sufficiently similar as to be treated identically (and perhaps analyzed en masse) for the purpose of further spectral analysis employing physical models.

The use of statistical clustering techniques and pattern recognition on spectral data is not new to astronomy. The literature in optical spectroscopy is diverse, including recent efforts of mainly supervised approaches using axis-parallel decision trees (Ball et al. 2006), mixture models (Bazell & Miller 2005), mixture models combined with neural networks (Qin et al. 2003), and independent component analysis (Lu et al. 2006). In addition, many studies, too numerous to list here, utilize neural networks or Bayesian classifiers for object classification.

McGlynn et al. (2004) developed an X-ray source classifier called ClassX that works with *ROSAT* data. The classifier is a supervised method that requires labeled data from multiple databases. It categorizes sources according to their physical basis into one of six categories: stars, galaxies, white dwarfs, X-ray binaries, active galactic nuclei, and clusters of galaxies. ClassX has also been used with Sloan Digital Sky Survey (SDSS) data after the classifier was suitably trained on SDSS objects whose nature was precisely known (Suchkov et al. 2005).

Examination of the X-ray spectral properties of a group of sources via analysis of X-ray hardness ratios (HRs) is common practice in X-ray astronomy. A wide variety of instrument-specific definitions of HRs exist; multiple definitions can exist even for a given telescope-detector combination. More importantly, the band definitions underlying HRs are generally developed without regard to optimizing the selection of spectral bins or to applying nonuniform weighting to spectral bins within these (somewhat arbitrarily) defined bands. A multivariate statistical technique, principal component analysis (PCA; Jolliffe 1986; Jackson 1991), may effectively provide a means to automatically define optimal spectral band definitions from the data itself, rather than imposing bands or colors on the data.

We are exploring combinations of PCA and unsupervised classification schemes to sort and group X-ray sources based solely on their spectral attributes. Our technique (Hojnacki 2005) employs PCA to reduce dimensionality of spectral bands for subsequent classification. PCA has been applied to optical spectra for galaxy classification (e.g., Connolly et al. 1995; Bromley et al. 1998; Folkes et al. 1999) and optical and UV spectra for quasar classification (e.g., Boroson & Green 1992; Yip et al. 2004; Suzuki 2006). In the only previous model-independent X-ray spectral classification study that employed PCA, Collura et al. (1995) successfully demonstrated a method to group X-ray sources detected with the *Einstein Observatory* Imaging Proportional Counter (IPC). Their method distinguishes between stellar sources and extragalactic sources, despite the severely limited spectral and spatial resolution and energy range of the *Einstein* IPC relative to contemporary X-ray CCD detectors.

We describe a novel X-ray source classification method developed for application to CCD spectra and originally applied to *Chandra* archival data (Hojnacki & Kastner 2004). In contrast to previous methods (e.g., ClassX), our method is an unsupervised technique that uses only raw X-ray CCD spectral data to find natural spectral groupings. Our method provides the ability to sort and classify X-ray sources in rich fields without a priori information as to their nature, thereby guiding and/or fine-tuning downstream automated fitting of physical source models. In addition, we present a study of the X-ray, optical, and IR properties of the X-ray spectral classes identified in the COUP subsample, the results of which will aid in deriving the nature of populations of X-ray sources in other PMS clusters. The method has been applied to young stellar objects with the goal of a deeper understanding of X-ray emission from PMS stars, but is not limited to this particular application.

Section 2 outlines the preparation of the COUP data. Section 3 describes the source classification algorithm. Additional detailed information for these two sections is contained in Hojnacki (2005). Results of running the algorithm on the COUP data set are presented in § 4, followed by an analysis and discussion of the results in § 5. We present a summary in § 6.

2. DATA PREPARATION

Data from the COUP (Getman et al. 2005b) observations (ObsIDs 4395, 3744, 4373, 4374, 4396, and 3498) of the ONC were used as the input data set for development of the X-ray source classification algorithm. The 838 ks total exposure consists of six consecutive exposures. A gap of 29 ks exists between exposures due to removal of the Advanced CCD Imaging Spectrometer (ACIS) from the *Chandra* boresight during five passages through the Van Allen belts during this period.

Getman et al. (2005b) describe in detail the fundamental data reduction performed on the 1616 source COUP data set. This data reduction included extracting valid events, detecting sources, flagging potential source detection problems and/or artifacts, and deriving source X-ray properties from fits of one- or two-component thermal plasma models with intervening absorption.

A subset of the COUP sources detected and analyzed by Getman et al. (2005b) was selected for use in developing and testing the algorithm described here. Specifically, sources in the COUP data set that were flagged as a double source, a pileup source, and/or in a source extraction region crossing a bright source readout trail were eliminated from the sample. Sources with less than 300 counts were also eliminated, to restrict the sample to those with higher signal-to-noise CCD spectra. This resulted in a sample of 444 sources for which high-quality ACIS spectra could be drawn from the COUP data set. Photon counts due to local background were subtracted from the spectrum for each source. Figure 2 shows the brightness distribution of the subset of 444 COUP sources.

With the expectation that most sources of interest to our research (i.e., young stars) will exhibit emission-line spectra characteristic of ionized plasma, the 0.4–8.2 keV spectra were rebinned into a number of non-overlapping spectral bands (see Appendix). Band locations and widths were defined based on hardware constraints, being careful not to bisect key emission features. Finally, the source spectra were normalized to the maximum number of counts obtained for the brightest source within the sample.

3. X-RAY SOURCE CLASSIFICATION ALGORITHM

Once the spectra were binned into X-ray bands and normalized, the algorithm was run without the use of models or a priori information. The source classification algorithm consists of the following steps: PCA, agglomerative hierarchical clustering (Johnson & Wichern 2002), and *K*-means clustering (Johnson & Wichern 2002).

3.1. *Spectral Dimension Reduction*

A nonlinear monotone transformation (Kruskal 1964) was performed on the input data to reduce nonlinearities and thereby prepare the data for the subsequent dimension reduction. PCA was then used to reduce the dimensionality in the data set by identifying X-ray spectral bands that are most discriminatory based on their contribution to the variance over the entire data set. This new, smaller set of uncorrelated variables is called the *principal components* (PCs), which together explain all or nearly all of the total variance in the data set. Each principal component is described by a linear combination of the original input variables, called an eigenvector. The variance accounted for by a component is called an eigenvalue. Clusters that are evident in PC space can indicate distinct subgroups within the data set.

Several stopping rules were applied to determine the number of principal components to retain for the clustering analysis (Hojnacki 2005). We applied Levene's statistical significance test for homogeneity of the eigenvalues (Levene 1960), Horn's stopping rule (Horn 1965), Jolliffe's broken stick stopping rule (Jolliffe 1986), the average root rule (Jackson 1991), and the scree test (Cattell 1966). These tests indicated that out of the 42 principal components, two to four principal components are sufficient to represent the variability in the data set. Hence, we chose four principal components to retain for the ensuing clustering steps. The variance explained by these four components is 62.3%, 30.7%, 2.2%, and 1.9%, representing 97.1% of the total variance in the data set. Plots of the eigenvectors that correspond to these first four components are shown in Figure 3. The *x*-axis of the eigenvector plots corresponds to the 42 X-ray spectral bands (see Appendix). The *y*-axis corresponds to the eigenvector coefficient (arbitrary units). Note that the sign of the eigenvectors is arbitrary.

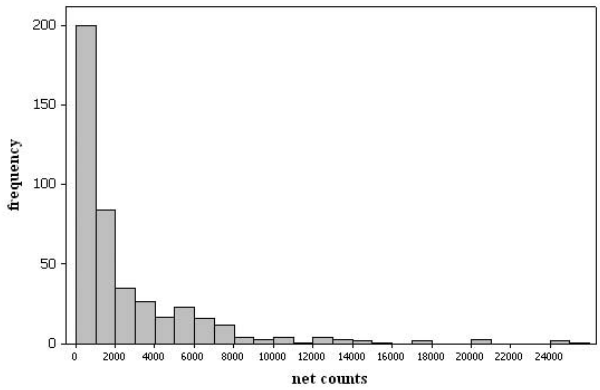

FIG. 2.—Distribution of net counts for the 444 source sample.

As can be seen in Figure 3, the first eigenvector peaks sharply at ~1 keV and becomes negative for energies >1.5 keV. Hence, large positive values of PC1 could be interpreted as indicative of sources with the softest X-ray spectra, while large negative values of PC1 would correspond to the hardest sources. The peak in the second eigenvector is broader and is shifted to a slightly higher energy than that of the first eigenvector (~1.5 keV); the value of this eigenvector also remains nearly zero (less negative than PC1) for energies >2.5 keV. Hence, large values of PC2 are indicative of sources with intermediate spectral hardness. The sources with the hardest observed spectra will have negative values of PC1.

The plots of the third and fourth eigenvectors indicate that large values of PCs 3 and 4 could be indicators of sources that have both soft and hard X-ray components. The hard X-ray component indicated by larger values of PC4 is broader and of higher energy than that of PC3.

3.2. *Source Clustering*

Two unsupervised classification methods—agglomerative hierarchical clustering and *K*-means clustering—were applied following PCA in an attempt to determine a "natural" partitioning of the data set into a number of relatively homogeneous groups (or "classes") of X-ray sources. An agglomerative hierarchical clustering method based on Euclidean distance and complete linkage (Johnson & Wichern 2002) was used on the first four principal components generated from the X-ray source spectra. The Euclidean distance values were calculated for each intermediate amalgamation step between successive mergers of the groups. To determine the final number of groups, the pattern of changes in the distance value was examined to find a large jump in the metric between amalgamations. The number of groups in the final clustering must be logical for the data set. For example, if the partitioning results in a large number of groups relative to the number of sources, then many groups will likely be small and indistinct. If too few groups are chosen, then the groups may be very large and inhomogeneous (Collura et al. 1995). The resulting dendrogram indicates where

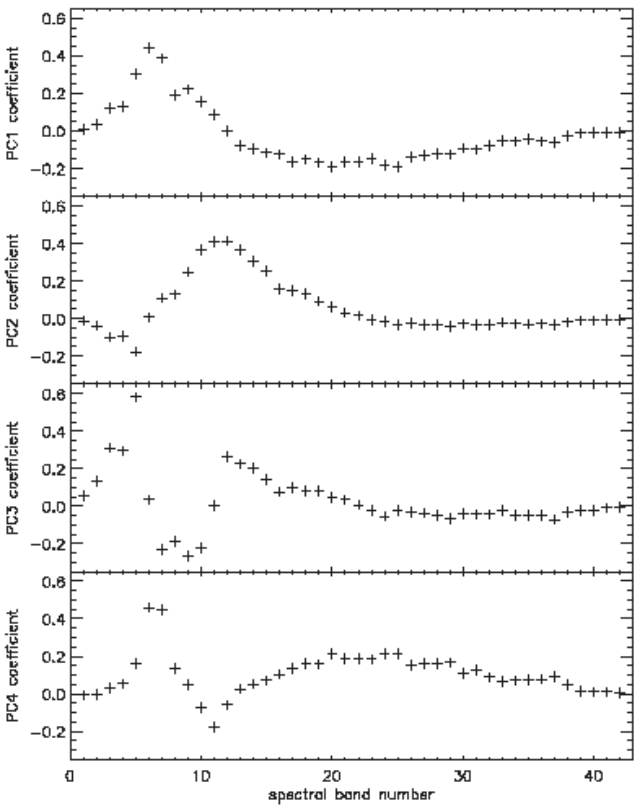

FIG. 3.—Eigenvector plots for the four principal components. The *x*-axis corresponds to the 42 X-ray spectral bands. The *y*-axis is the eigenvector coefficient.

and how the source sample was broken into groups. For our 444 source COUP subsample, this procedure yielded 17 groups in the final partition.

The class assignments resulting from the hierarchical clustering were used as the initial partition for the *K*-means clustering. Each source was examined and assigned to the group with the nearest centroid (multidimensional mean) in PCA space. In some cases, this resulted in a source being reassigned to a new group, after which the centroids were recalculated for the group receiving the reassigned source and for the group losing the source. This process was repeated until no further reassignments took place.

4. RESULTS

A score plot of the first two principal components is shown in Figure 4. Since these two principal components explain most of the dispersion in the data, the score plot of PC2 versus PC1—which is somewhat analogous to a color-color diagram for these spectral data—reveals the most information about the subsequent clustering that was performed on the results of the PCA (§ 3.1). The plot shows that nearly all of the PMS stars in the input sample lie in a single spectral sequence, with the notable exception of classes 1 and 17. However, the higher numbered classes do not appear to be physically distinct in this 2D score plot. The separation in the classes is more readily seen when this PC2 versus PC1 plot is rotated about the PC2 axis (see Fig. 5). There is curvature in the third dimension (PC3) that can be seen in the sequence of plots in Figure 5. We analyze this structure further in §§ 5.1 and 5.3.

Class sizes range from 1 source to 88 sources. Visual inspection of source spectra as grouped by the algorithm described in § 3 makes apparent that there are strong spectral similarities within a given class, while fundamental spectral differences can be seen between classes. The average spectrum for each of the 17 classes was calculated to illustrate the overall shapes of the X-ray SEDs of each class (Fig. 6). The only source in class 17 is readily distinguished from the other 443 sources by its position near the center of the horseshoe in the score plot (Fig. 4). This source, COUP 948, has a double-peaked SED (see Fig. 7) that is unique among the sample of 444 sources considered here (see § 5.3.3).

It is apparent from Figure 6 that the progression of classes moving clockwise around the horseshoe-shaped curve in the score plot

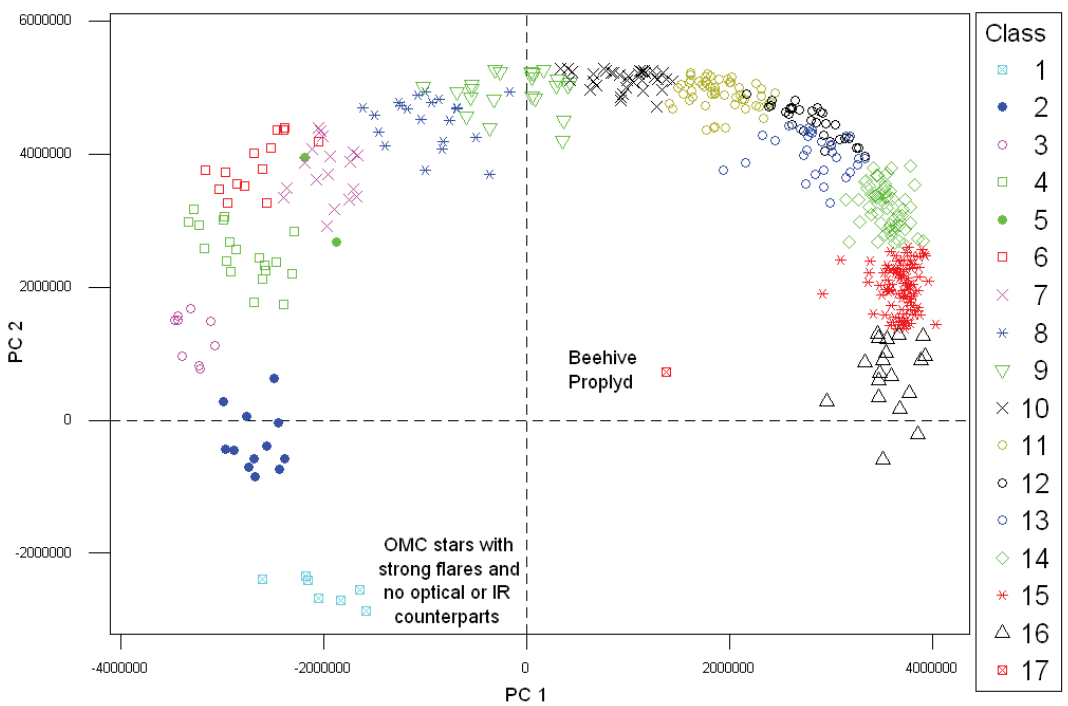

FIG. 4.—Plot of the first two principal components with the source classes shown. The class numbers increase clockwise around the horseshoe-shaped curve. Nearly all of the X-ray sources in the input sample lie in a single spectral sequence. Class 17 consists of one outlier.

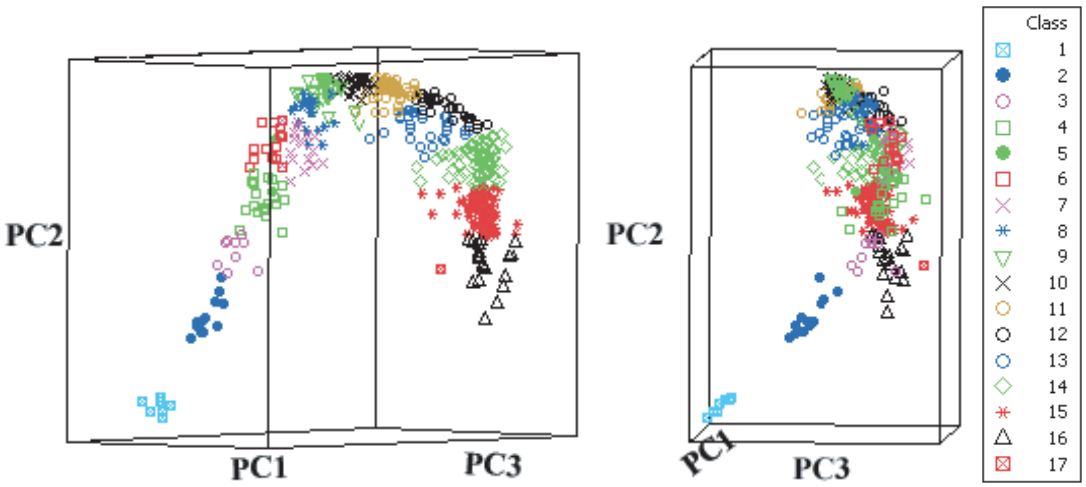

FIG. 5.—Sequence of two plots showing rotation of PC2 vs. PC1 plot about the PC2 axis. The curvature and depth in the third dimension (PC3) can be seen in these plots.

(Fig. 4) forms a sequence of decreasing spectral hardness. The lowest numbered classes contain sources with the hardest observed spectra. These classes are also more easily separated visually in this plot of only the first two principal components.

Closer examination of Figures 4 and 6 makes apparent that the three principal regions of the horseshoe in the PC2 versus PC1 score plot correspond to three fundamentally different groups of X-ray SEDs that are present in the 444 source COUP sample studied here. Classes 1–6 form a sequence of very hard spectra with decreasingly prominent Fe line emission in the 6.4–6.7 keV region as one moves "up" (increasing PC2) the left side of the horseshoe. These Fe line complexes are especially strong and broad in the average spectrum of class 1 and, to a lesser extent, class 2, both of which are readily separated from the "mass" of hard sources on the left side of the score plot. Classes 7–11 display SEDs with strong peaks at ~1.5 keV and prominent high-energy tails. The softest X-ray spectral classes, 12–16, have SEDs that peak at ~1.0 keV. The strength of this peak relative to the (weak) high-energy tails in the average spectra of these classes increases as one moves "down" the right side of the horseshoe in Figure 4.

4.1. *Relationship of X-Ray Spectral Classes to COUP X-Ray Hardness Ratios*

A plot of X-ray hardness ratios (HRs) for the 444 source subset is shown in Figure 8. Five subbands were defined for COUP: 0.5–1.7 keV (s_1), 0.5–2.0 keV (s_2), 1.7–2.8 keV (m), 2.0–8.0 keV (h_1), and 2.8–8.0 keV (h_2) (Getman et al. 2005b). The three COUP hardness ratios are then defined as HR1 $= (h_1 - s_2)/(h_1 + s_2)$, HR2 $= (m - s_1)/(m + s_1)$, and HR3 $= (h_2 - m)/(h_2 + m)$. HR1 represents a broadband hardness ratio, while HR2 is used to measure the softer part of the X-ray spectrum and HR3 the harder part of the X-ray spectrum (Getman et al. 2005b).

The class sequence identified via the algorithm described in § 3 appears more or less in the same sequence in the HR3 versus HR2 diagram of Figure 8, with the sources having the softest observed spectrum (highest numbered classes) found at the lower left of the Hertzsprung-Russell (H-R) diagram, and the sources having the hardest observed spectrum found at the upper right. However, in contrast to the PC2 versus PC1 score plot, no distinct source clusters are apparent in the HR3 versus HR2 plot, and source classes that are tightly clustered in the score plot are more spread out in the HR3 versus HR2 plot. Note in particular that class 1 sources, which are all found along the upper range of HR3, display a wide range of HR2, and that the distributions of class 2 and 3 sources merge together in the H-R plot.

5. ANALYSIS

5.1. *Simulated versus Observed Data*

To aid in the interpretation of the PC2 versus PC1 score plot (Fig. 4), simulations of X-ray spectral data were performed using XSPEC version 12.2. The model input to the simulations consisted of thermal plasma emission plus intervening absorption. The thermal plasma temperature kT was varied from 0.26 to 8.6 keV in steps of 0.86 keV, and the hydrogen column density, $\log N_{\rm H}(\rm cm^{-2})$, was varied from 20.5 to 23.5 in steps of 0.5, resulting in 84 unique models. The simulated spectra were run through the same spectral classification algorithm applied to the 444 source COUP subsample. The resulting 2D principal component plot showing PC2 versus PC1 is shown in Figure 9, where the simulations are superimposed on the distribution of COUP sources. Only 49 of the 84 simulations are shown here for clarity. The simulated sources are represented as solid diamonds and are labeled with their $\log N_{\rm H}$ values; simulated sources of equal kT are connected by lines.

The distribution of simulated X-ray sources in PC-space reproduces the fundamental shape of the horseshoe-like distribution formed by the COUP subsample. Moving clockwise around the horseshoe, the simulations form a sequence of monotonically decreasing $\log N_{\rm H}$, from 23.5 in the extreme lower left of the score plot to 20.5 in the lower right. This indicates that the position of a COUP source on the horseshoe is determined largely by the degree of its obscuration by intervening cloud or circumstellar material.

The importance of $\log N_{\rm H}$ to the membership of the source classes is particularly apparent along the left side of the horseshoe in this plot. In particular, the only simulations that can reproduce the position of the distinct grouping that forms class 1 are those for which $\log N_{\rm H} = 23.5$. It is also evident, however, that the positions of class 1 sources cannot be reproduced by low-temperature plasma emission ($kT \leq 0.5$ keV). At the other extreme, the higher numbered classes (14–16) appear to form a sequence of decreasing plasma temperature, indicating that temperature is more important than intervening absorption in determining the positions of these sources in the PC2 versus PC1 plot. Note that X-ray sources from the COUP subsample do not fall along the temperature lines for $kT = 0.26$ keV in the PC2 versus PC1 plot. This could be interpreted as a lack of such soft X-ray emission from sources in the

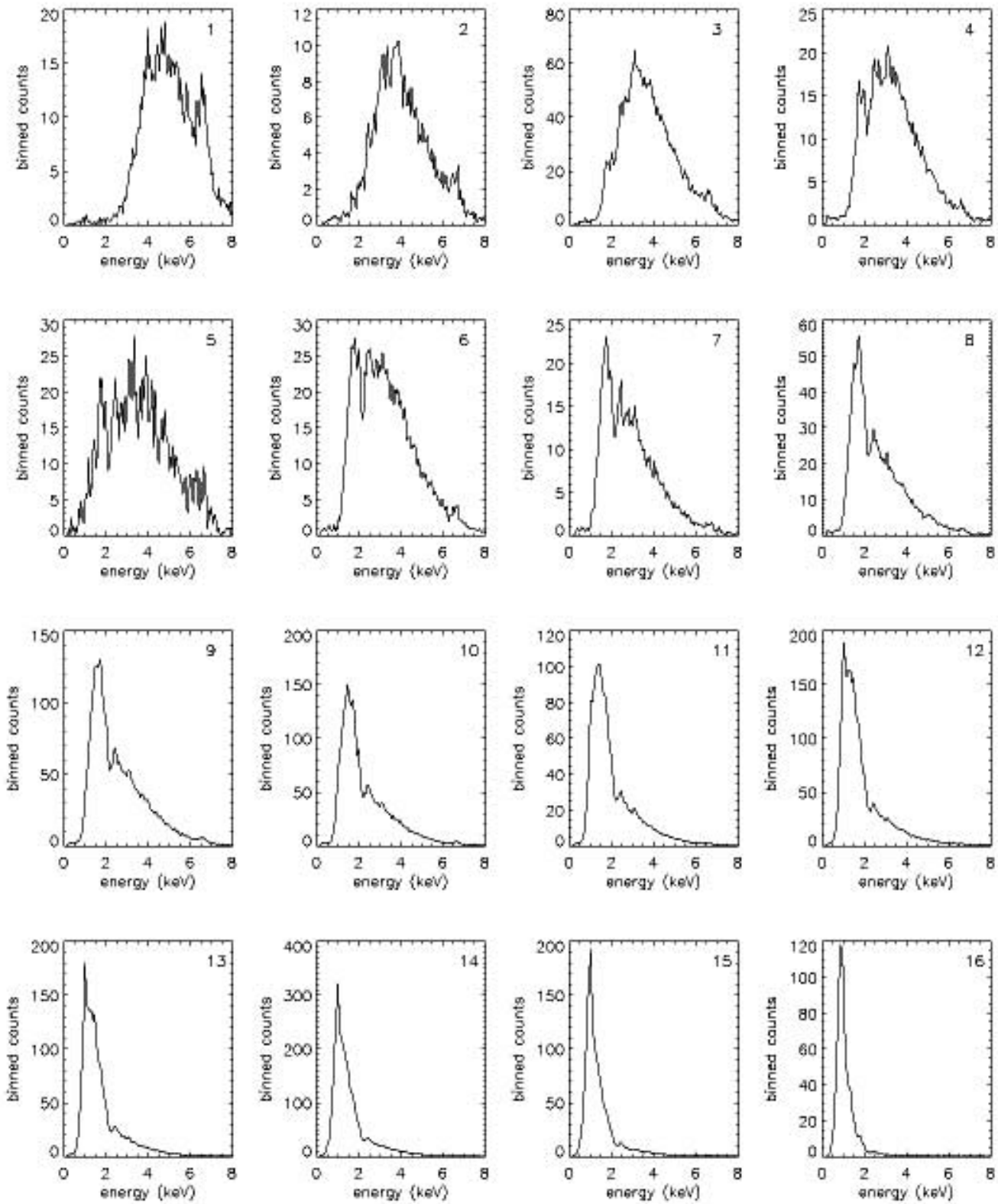

FIG. 6.—Average observed X-ray spectra for 16 of the 17 classes. The class number is shown in the upper right corner of each plot. The trend of decreasing spectral hardness with increasing class number can be seen in these plots. (See Fig. 7 for class 17.)

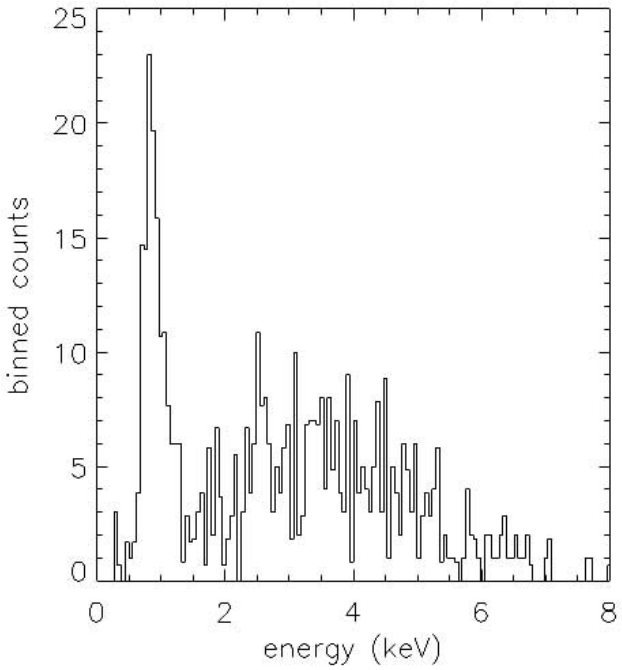

FIG. 7.— Observed X-ray spectrum of COUP 948, the only source in X-ray spectral class 17.

ONC. At the same time, few sources lie along the temperature lines for $kT = 8$ keV. Indeed, simulations with kT greater than 0.26 keV but less than 4 keV bound the majority of the data set.

Moving clockwise in the PC2 versus PC1 plot (Fig. 9), the spread in the COUP subsample in classes 4, 6, 7, and 8 is mimicked by the model distribution: the model lines "fan out" such that the lines for $kT = 0.86$ and 4.3 keV appear to describe the (relatively wide) spread in the observed distribution at the upper left of the horseshoe. At this point in the horseshoe, it appears the sources are characterized by very uniform log $N_{\rm H}$ ($\sim$22.5).

5.2. *Relationship of X-Ray Spectral Classes to Fundamental Stellar Parameters*

5.2.1. *Correlation of X-Ray and Optical Spectral Types*

The H-R diagram for the 444 COUP sources studied here is shown in Figure 10. The sources are color-coded with their X-ray spectral classes as determined by the X-ray source classification algorithm. The $L_{\rm bol}$, $T_{\rm eff}$, and spectral type data were obtained from Getman et al. (2005b). Figure 10 demonstrates that the vast majority of the sources found in the higher numbered classes are optically identified late-type (K and M) stars. For example, in classes 11–16, spectral-type data exist for 223 out of 282 sources, and $\sim$96% of these 223 sources are K- and M-type stars.

Furthermore, the six classes in Figures 10*b* and 10*c* appear to occupy slightly different regions in the H-R diagram, showing a trend of decreasing $T_{\rm eff}$ with increasing X-ray spectral class number. This is most evident in class 16, where the sources are clumped in the lower right part of the main sequence. Hence, these diagrams show a trend of increasing spectral softness with decreasing effective photospheric temperature for X-ray sources in the COUP subsample. A similar trend is hinted at in Figure 12 of Preibisch et al. (2005).

5.2.2. *X-Ray Sources Associated with A and B Stars*

There are six X-ray sources in our COUP data subset that correspond to intermediate-mass (A or B) stars. These sources are labeled on the H-R diagram (Fig. 10) with their COUP source numbers. All six of these sources are found in X-ray spectral classes 14, 15, and 16, which are associated with the softest observed X-ray spectra. Five of these sources, numbered 100, 113, 869, 1360, and 1415, have been included in a study by Stelzer et al. of the X-ray properties of O, B, and A stars (Stelzer et al. 2005). For four of these five sources (100, 113, 1360, and 1415), Stelzer et al. conclude that the X-rays are being emitted by low-mass companions to the A-type and B-type stars found in optical spectroscopy. These conclusions were based primarily on the X-ray luminosities and variability properties of these sources. For these four sources,

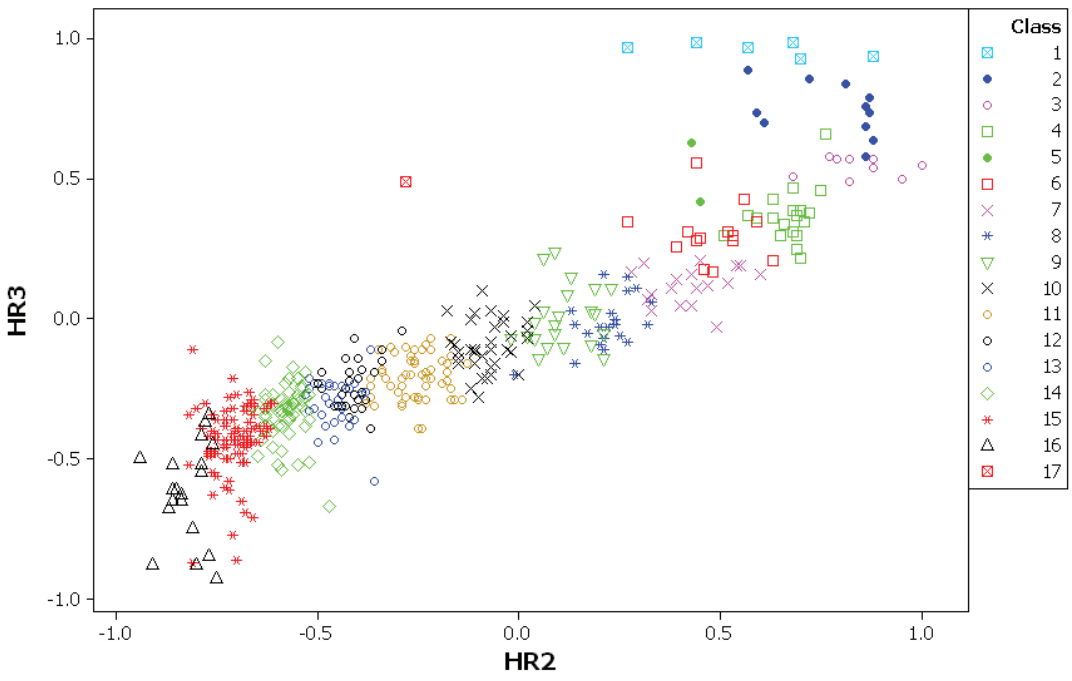

FIG. 8.—Hardness ratio diagram for the COUP 444 subset, excluding COUP 647 (no HR2 data available).

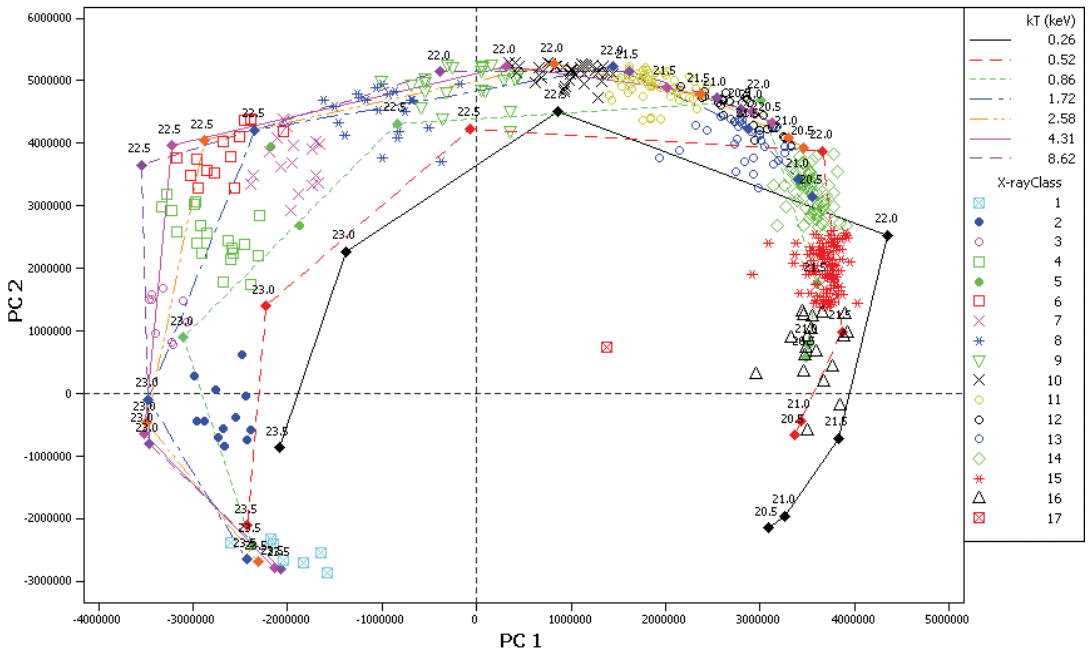

FIG. 9.—Simulated spectra with varying kT and $N_{\rm H}$ are overlaid on the plot of PC2 vs. PC1 for the 444 COUP sources. Lines connect simulations calculated for a given value of kT; the points on the lines represent the locations in PC space of the simulated spectra. Each point is labeled with the value of log $N_{\rm H}$ used in the simulation. The basic shape of the horseshoe is also produced by the simulated data.

our spectral analysis therefore supports the contention that the X-rays originate from late-type companions. In the case of COUP 869, Stelzer et al. concluded (based on apparent lack of variability) that the X-ray emission originates with the small-scale wind shocks of the massive B-type star itself. However, our X-ray spectral classification algorithm places COUP 869 in class 16, which is dominated by soft-spectra M-type stars, suggesting that its X-ray emission may also originate with a late-type companion. The other OB stars in the COUP observation are not part of our subsample, due to heavy pileup. Analysis of algorithm results on these and other OB stars will be described in a future paper.

5.2.3. *X-Ray Properties versus Optical/Near-IR Properties*

Clustering results were compared with known optical and near-infrared properties of the population of the ONC to assess the algorithm's ability to identify groups of sources that share common attributes. Table 1 lists the properties of the 444 COUP sources in our sample. Data are from Getman et al. (2005b). Table 2 lists the final number of X-ray sources contained in each of the 17 groups, as identified by hierarchical clustering and subsequently refined via K-means clustering. Properties of the 17 X-ray classes are also shown in Table 2: mean X-ray-determined hydrogen column density ($N_{\rm H}$), mean effective photospheric temperature ($T_{\rm eff}$), mean visual extinction (A_V), and mean K-band excess [$\Delta(I-K)$]. The final two columns illustrate the fraction of sources with equivalent width of the λ8542 Ca II infrared triplet lines [EW(Ca II)] of less than -1 Å, where a negative equivalent width represents an emission line. K-band excess has been attributed to existence of a hot circumstellar disk (Getman et al. 2005b). EW(Ca II) has been shown to be correlated with mass accretion rate (Muzerolle et al. 1998; Getman et al. 2005b) where EW(Ca II) < -1 is indicative of strong accretion (Flaccomio et al. 2003). The errors on the mean are also given in Table 2. These results were compiled from data available for the X-ray-emitting ONC population (Getman et al. 2005b). Data for the six A-type and B-type stars were not included in the mean calculations of optically derived properties (i.e., $T_{\rm eff}$) because we believe companion stars are emitting the X-rays associated with these six A-type and B-type stars.

We see that log $N_{\rm H}$ decreases monotonically from class 1 to class 16 (Fig. 11). The large log $N_{\rm H}$ characteristic of classes 1–8 is reflected in very small fractions of optical and near-infrared (ONIR) counterparts. The mean visual extinction correspondingly decreases monotonically for classes 11–16 (Fig. 12). For classes 12–16, which have relatively large fractions of ONIR counterparts and softer X-ray emission, both the mean near-infrared excess and the fraction of sources with EW(Ca II) < -1 are observed to decrease monotonically (Fig. 12), indicating a generally decreasing accretion rate. The mean $T_{\rm eff}$ decreases as the observed X-ray spectra become softer (Fig. 12). This same result is reflected in the H-R diagrams in Figure 10.

The mean stellar masses decrease with increasing spectral softness for classes 10–16. However, these masses are determined by comparing measurements of $T_{\rm eff}$ and luminosity with evolution models for PMS stars. The mass derived in this manner depends primarily on $T_{\rm eff}$, which is highly model-dependent and therefore potentially uncertain.

5.3. *Noteworthy X-Ray Sources Isolated by the Classification Algorithm*

5.3.1. *Class 1: Deeply Embedded Protostars*

Sources in class 1 are easily identified as a distinct X-ray spectral group by our classification technique (Figs. 4 and 5; Table 2). All of the class 1 sources lack ONIR counterparts and—although the classification algorithm is based solely on spectral information with no input from the temporal domain—all have high-amplitude, fast-rise X-ray flares (Getman et al. 2005a). Therefore, the objects in our class 1 are most likely very young protostars deeply embedded in the Orion molecular core.

In addition, all sources in class 1 have an X-ray median energy (MedE) > 4.6 keV, and the mean MedE for the class is 4.94 keV. Getman et al. (2007, their Fig. 8) find that all IC 1396N sources with MedE > 3 keV are protostars. They conclude that when MedE

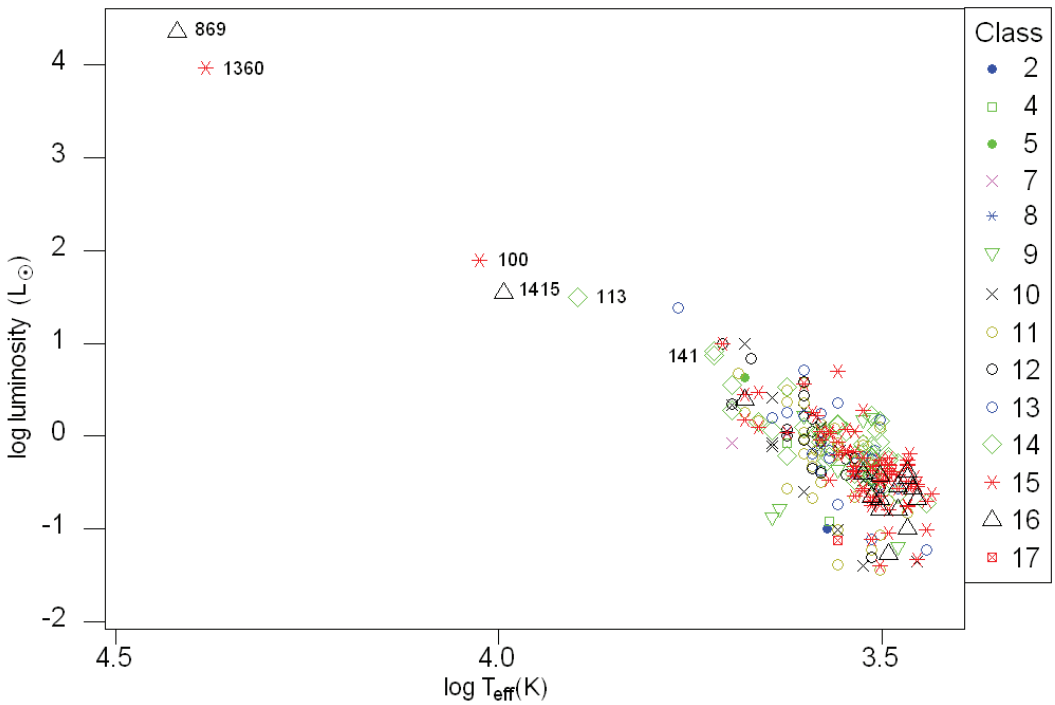

FIG. 10*a*

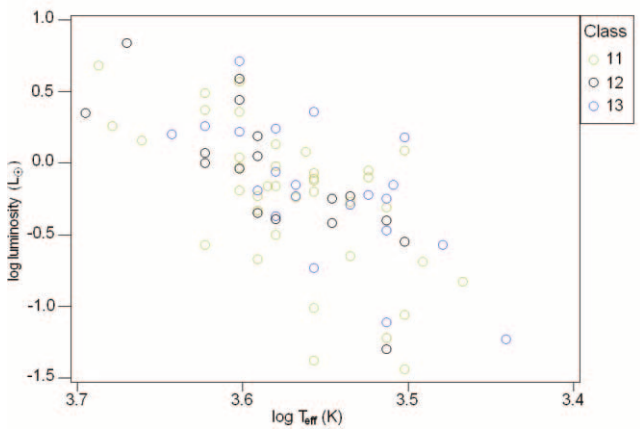

FIG. 10*b*

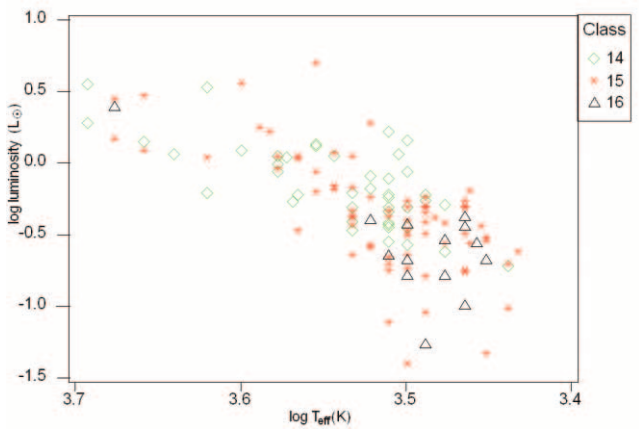

FIG. 10*c*

FIG. 10.—(*a*) H-R diagram of COUP 444 data set, color-coded by X-ray spectral class. The A-type and B-type stars are labeled with their corresponding COUP source number. (*b*) The *x*-axis scale of the H-R diagram for soft X-ray spectral classes 11–13 was restricted to focus on the main group of sources, which are of spectral types K and M. (*c*) The *x*-axis scale of the H-R diagram for the softest X-ray spectral classes, 14–16, was restricted to focus on the main group of sources, which are of spectral types K and M.

TABLE 1

PROPERTIES OF THE 444 COUP SOURCES

Sequence Number	X-Ray Class	HR2	HR3	log N_H (cm $^{-2}$)	kT_1 (keV)	kT_2 (keV)	kT_1/kT_2	log T_{eff} (K)	EW(Ca II) (Å)	A_V	$\Delta(I-K)$
6.....	14	−0.61	−0.46	21.12	0.78	2.32	0.34	3.502	0.0	0.46	0.12
8.....	9	0.06	0.22	22.05	15.00	...	...	...	...	...	...
11.....	12	−0.39	−0.18	21.69	0.52	4.13	0.13	3.602	−14.6	0.42	1.37
17.....	13	−0.53	−0.27	21.12	2.70	...	...	3.623	1.9	1.58	0.20
20.....	15	−0.81	−0.87	21.69	0.62	...	...	3.467	0.0	1.57	−0.35

NOTE.—Table 1 is published in its entirety in the electronic edition of the *Astrophysical Journal*. A portion is shown here for guidance regarding its form and content.

TABLE 2
ONIR Properties of the 17 X-Ray Classes

X-Ray Class	N	$\langle \log N_{\rm H} \rangle$ Value ($\rm cm^{-2}$)	N	$\langle \log T_{\rm eff} \rangle$ Value (K)	N	A_V Value (mag)	N	$\langle \Delta(I-K) \rangle$ Value (mag)	N	EW(Ca II) < -1 Fraction	N
1.....	7	23.40 ± 0.06	7	...	0	...	0	...	0	...	0
2.....	12	22.96 ± 0.03	12	3.57:	1	...	0	...	0	0	1
3.....	9	22.79 ± 0.02	9	...	0	...	0	...	0	...	0
4.....	19	22.66 ± 0.01	19	3.59:	3	1.34:	3	1.66:	3	0.25	4
5.....	2	22.52 ± 0.05	2	3.68:	1	3.67:	1	2.61:	1	1.00	1
6.....	14	22.48 ± 0.02	14	...	0	...	0	...	0	...	0
7.....	18	22.46 ± 0.02	18	3.70:	1	3.52:	1	0.98:	1	0.00	2
8.....	21	22.30 ± 0.02	21	3.55:	3	1.52:	3	0.30:	2	0.40	5
9.....	22	22.18 ± 0.01	22	3.56 ± 0.02	7	1.77 ± 0.99	7	1.10:	3	0.20	10
10.....	37	22.03 ± 0.02	37	3.58 ± 0.01	21	2.60 ± 0.45	20	1.31 ± 0.18	18	0.55	22
11.....	54	21.90 ± 0.02	54	3.57 ± 0.01	38	2.69 ± 0.31	38	0.91 ± 0.12	30	0.33	36
12.....	30	21.66 ± 0.03	30	3.59 ± 0.01	20	1.57 ± 0.29	19	0.80 ± 0.14	16	0.41	17
13.....	30	21.61 ± 0.03	30	3.56 ± 0.01	22	1.44 ± 0.27	22	0.62 ± 0.11	18	0.38	21
14.....	61	21.32 ± 0.03	61	3.55 ± 0.01	45	1.16 ± 0.16	44	0.49 ± 0.08	38	0.17	36
15.....	88	20.79 ± 0.05	86	3.52 ± 0.01	75	0.65 ± 0.11	72	0.25 ± 0.07	62	0.11	62
16.....	19	20.28 ± 0.11	19	3.50 ± 0.01	14	0.32 ± 0.14	16	0.11 ± 0.05	12	0.00	13
17.....	1	20.88:	1	3.56:	1	0.34	1	...	0	1.00	1

Notes.—N = number of sources. The six A-type and B-type stars in the COUP 444 source sample have not been included in mean calculations based on optically derived properties.

exceeds ~3 keV, the X-ray absorption is dominated by infall from the protostellar envelope rather than by intervening, ambient molecular cloud material. A similar conclusion would appear to apply to class 1 sources.

An analogous sample of deeply embedded, flaring protostars was selected for study by Tsujimoto et al. (2005). The strong Fe line emission at 6.4–6.7 keV seen in the class 1 average spectrum attests to their high X-ray emission temperatures. Furthermore, the breadth of the 6.4–6.7 keV Fe line complex in these sources—which is even evident in their class average spectrum (Fig. 6)—strongly suggests that the line complex includes a fluorescent, "cold disk" component (Tsujimoto et al. 2005). Five of the sources in our class 1 were not included in the sample of protostars selected by Tsujimoto et al. (2005) because these five sources failed their third criterion: in spectral model fitting, the addition of a 6.4 keV Gaussian component to the thermal model could not be justified statistically. The sixth source in our class 1 was not included in their sample because there were <100 photons in the 6.0–9.0 keV band. These sources are worthy of additional study as potential examples of fluorescent Fe line emission from circumstellar disks.

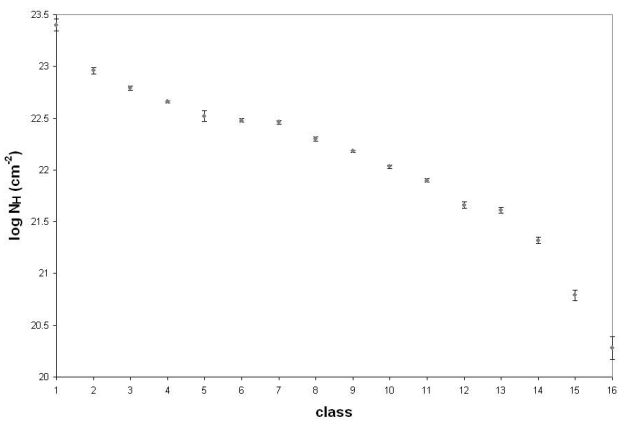

Fig. 11.—Mean log hydrogen column density plotted for each class.

5.3.2. *Evolved PMS Stars on the Far Side of the OMC?*

Sources in class 2, which are isolated in Figure 9, show very large $N_{\rm H}$. Only one source in this class has an optical counterpart.

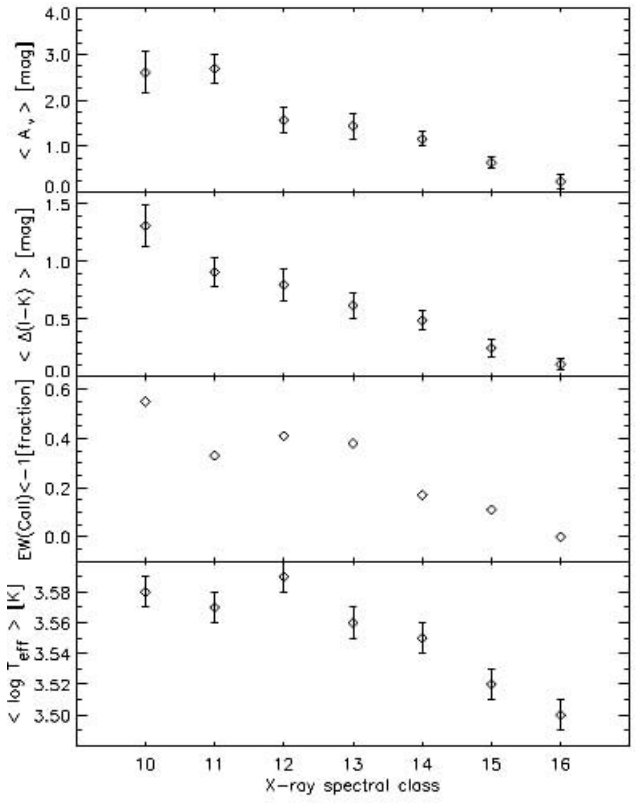

Fig. 12.—Mean visual extinction, mean near-IR K-band excess, the fraction of sources with EW(Ca II) < -1, and mean log effective photospheric temperature plotted by class, for classes with corresponding data.

In Figure 9, class 2 is bounded by simulations with $kT = 0.52$ and 0.86, the same range characterizing classes 15 and 16. Therefore, class 2 may consist of an *intrinsically* soft population ($kT \sim 0.5$ keV), comparable to the X-ray sources in classes 15 and 16. The principal difference between sources in these classes appears to be $N_{\rm H}$, suggesting that class 2 objects are similar to objects in classes 15 and 16—that is, fairly evolved K- and M-type PMS stars—but that class 2 stars are located on the far side of the Orion molecular cloud. The only source in class 2 with an optical counterpart is COUP 1139, an M-type star, supporting this interpretation.

A similar interpretation may hold for sources in classes 3–7, which also show large $N_{\rm H}$. When the horseshoe-shaped curve is rotated and viewed in the third dimension (see final panel of Fig. 5), there is some overlap between sources in these lower numbered classes and sources in the higher numbered classes 14 and 15. Indeed, there are four sources in class 4 that have optical counterparts: two K-type stars and two M-type stars. These four stars fall outside of the $kT = 0.52$–0.86 range in Figure 9.

Classes 2–7 lie at the upper right of the HR3 versus HR2 hardness ratio diagram (see Fig. 8). However, this diagram may be misleading for these heavily absorbed X-ray sources, because the HR2 subband is a measure of softer X-rays and cannot supply any information for sources with large $N_{\rm H}$. Indeed, the possible relationship between classes 2 and 15–16 cannot be inferred from this "standard" hardness ratio diagram. The classification algorithm applied here therefore may represent a unique means to identify evolved low-mass PMS stars "hidden behind" molecular cloud cores.

5.3.3. *Class 17: The Beehive Proplyd*

COUP 948 is isolated by the source classification algorithm. This extraordinary X-ray source is associated with an unusual optical jet source (the Beehive Proplyd; Bally et al. 2005). The X-ray spectrum of COUP 948 (Fig. 7) has distinct hard and soft components with the soft component peaking at around 0.85 keV and the hard component at about 3.5 keV. This unusual, double-peaked X-ray spectral energy distribution is indicative of a combination of strong shocks in the jet collimation region and a variable, magnetically active X-ray source (Kastner et al. 2005). Several PMS stars in the Taurus region display a similar phenomenon (Güdel et al. 2007).

Although the X-ray source associated with the Beehive Proplyd is isolated in both the hardness ratio diagram (Fig. 8) and the principal component plot (Fig. 4), its unusual nature is much more apparent using our method, since its position in the horseshoe plot is significantly offset from the locus of sources in all three dimensions. This holds promise for the discovery of additional examples of such hybrid X-ray spectra lurking among the thousands of PMS stars detected thus far by *Chandra* and *XMM-Newton*.

6. SUMMARY

We have developed an algorithm and corresponding input variable definition that has the advantage of being nonparametric and independent of any spectral fitting procedures. The algorithm performs without a priori information concerning the nature of the sources; it objectively groups X-ray sources based solely on observables. We have demonstrated that this method finds natural groupings of X-ray sources when applied to spectra extracted from observations of rich stellar fields. The apparent diversity of source spectra is reduced to a three-dimensional locus in principal component space, with sources exhibiting unusual X-ray spectra falling outside this locus.

The plot of the first two principal components (Fig. 4) contains a horseshoe-shaped curve. In general, the spectral hardness of the classes decreases, moving clockwise around the horseshoe. Certain noteworthy outliers are readily identified by our spectral classification algorithm in this plot. These results demonstrate that the algorithm can be used to place the sources in order of decreasing spectral hardness and can be used to identify sources with unusual spectra.

Trends between X-ray spectral parameters and stellar parameters have been found for very low-mass, soft-spectra, young sources. Specifically, the X-ray spectral classes form sequences in $N_{\rm H}$, A_V, and accretion indicators, demonstrating that the algorithm effectively sorts young stars into a physically meaningful sequence. Furthermore, these trends demonstrate that correlations exist between X-ray and ONIR properties of PMS stars in the ONC. These trends and correlations are of significance for the study of star formation and X-ray emission in young stellar clusters. In addition, our algorithm can be used to efficiently identify very young X-ray sources that lack optical and near-infrared counterparts, display strong Fe Kα line emission, and display large-amplitude, fast-rise impulsive flares versus more low-level activity.

The principal components, and therefore the derived classification, are dependent on the properties of the input sample. The final classes presented here are therefore suitable for young stars and perhaps stellar sources in general, but not for sources with very different spectra from young stars. However, the method itself is completely general and future plans include applying it to entirely different input samples, including those from different wavelength regimes.

The lack of distinct boundaries between the higher numbered X-ray spectral classes in the PC2 versus PC1 score plot—i.e., those at the "soft" end of the horseshoe distribution—suggests that some of these classes could be merged. However, more spread in these higher numbered classes can be seen in the rotated principal component plot of Figure 5, suggesting that these higher PCs are responsible for some of the statistical differentiation in these classes. Additional analysis is being performed to interpret PC3 and PC4 in this context.

We are also analyzing the results of applying the X-ray source classification algorithm to objects in other star formation regions, by projection of these additional sources onto the principal component space of the COUP subsample and then determining their class membership probabilities for the existing X-ray spectral classes. Ongoing work also includes a feasibility study of inclusion of measures of X-ray source variability as additional inputs to the algorithm.

We thank Ettore Flaccomio, Michael Richmond, and Carl Salvaggio for their valuable comments. This research is supported by NASA under AISRP award NNG04GQ07G and *Chandra* award AR5-6004X issued by the *Chandra X-Ray Observatory* Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of NASA under contract NAS8-03060.

APPENDIX

X-RAY SPECTRAL BAND SELECTION

The spectral resolution of the ACIS CCDs at the nominal operating temperature of *Chandra* (-120°C) and knowledge of the charge transfer inefficiency (CTI) correction were used in determining the width of the spectral bands. The full width at half-maximum (FWHM) of the front-illuminated detectors increases with increasing energy, so the spectral bands increase in width accordingly. Finally, absorption features apparent in the quantum efficiency curves of the ACIS-I CCDs were also considered when selecting ranges for the spectral bands. The edges of these features were avoided when defining the bands, creating the gaps from 526 to 544 eV and 1841 to 1851 eV. These hardware constraints resulted in the division of the X-ray spectrum between 0.4 and 8.2 keV into 42 X-ray spectral bands; these are given in Table 3. The spectral band edges were chosen to the extent possible, given these hardware constraints, to avoid bisecting strong lines of high-ionization species.

TABLE 3
X-Ray Spectral Bands

Band Number	Low–High (eV)	Ion	E[a] (eV)
1	425–525	N VII	500.345
2	545–631	O VII	561.117
		O VII	568.735
		O VII	574.000
3	632–721	O VIII	653.640
		O VII	665.676
		Fe XVIII	703.601
4	722–815	Fe XVII	725.290
		Fe XVII	727.204
		Fe XVII	738.948
		Fe XVIII	767.347
		Fe XIX	769.681
		Fe XVIII	771.548
		O VIII	774.682
		Fe XVIII	781.320
		Fe XVIII	783.592
		Fe XVIII	793.571
		Fe XVII	812.499
5	816–912	O VIII	817.050
		Fe XIX	822.306
		Fe XVII	825.866
		O VIII	836.621
		Fe XVIII	853.141
		Fe XX	869.107
		Fe XVIII	869.778
		Fe XVIII	869.778
6	913–1013	Ne IX	914.961
		Fe XIX	917.262
		Fe XIX	918.690
		Ne IX	922.106
		Fe XXII	972.209
		Fe XX	985.970
		Fe XXI	1009.407
		Fe XVII	1010.888
7	1014–1115	Fe XXIII	1019.616
		Ne X	1021.801
		Fe XVII	1022.728
		Fe XXII	1053.488
		Fe XXIII	1056.540
		Ne IX	1074.112
		Fe XVIII	1075.697
		Fe XVIII	1094.787
		Fe XXIV	1109.480
8	1116–1220	Fe XXIV	1124.268
		Fe XXIII	1125.288
		Fe XXIII	1129.183
		Fe XIX	1146.408
		Fe XVII	1151.305
		Fe XXIV	1162.858
		Fe XXIV	1167.676
		Ne X	1211.012

TABLE 3—*Continued*

Band Number	Low–High (eV)	Ion	E[a] (eV)
9	1221–1335	Ne X	1277.251
		Mg XI	1331.281
10	1336–1445	Mg XI	1343.397
		Mg XI	1352.334
		Fe XXII	1381.566
		Fe XXII	1406.643
11	1446–1558	Mg XII	1472.281
		Fe XXIV	1491.048
		Fe XXIII	1493.203
		Fe XXIV	1496.627
		Fe XXIV	1506.080
		Fe XXIV	1551.690
12	1559–1673	Al XII	1575.147
		Mg XI	1579.561
		Al XII	1598.499
13	1674–1840	Al XIII	1728.884
		Fe XXIV	1729.607
		Mg XII	1744.941
		Si XIII	1839.696
14	1852–1974	Si XIII	1854.278
		Si XIII	1865.156
15	1975–2100	Si XIV	2005.427
16	2101–2400	Si XIV	2376.759
17	2401–2537	S XV	2430.332
		S XV	2448.086
		S XV	2460.717
18	2538–2676	S XVI	2621.470
19	2677–3045	...	...
20	3046–3276	Ar XVII	3106.101
		Ar XVII	3124.888
		Ar XVII	3139.922
21	3277–3436	Ar XVII	3320.716
22	3437–3737	Ar XVII	3684.860
23	3738–3909	Ca XIX	3877.284
24	3910–4085	...	...
25	4086–4266	Ca XX	4104.453
26	4267–4452	...	...
27	4453–4643	...	...
28	4644–4838	...	...
29	4839–5038	...	...
30	5039–5243	...	...
31	5244–5454	...	...
32	5455–5670	...	...
33	5671–5891	...	...
34	5892–6118	...	...
35	6119–6351	...	...
36	6352–6590	Neutral Fe Kα	6400
37	6591–6834	Fe XXV	6662.845
38	6835–7086	Fe XXVI	6962.130
39	7087–7344	...	...
40	7345–7609	...	...
41	7610–7881	...	...
42	7882–8156	...	...

[a] Huenemoerder et al. (2003).

REFERENCES

Ball, N. M., Brunner, R. J., Myers, A. D., & Tcheng, D. 2006, ApJ, 650, 497
Bally, J., Licht, D., Smith, N., & Walawender, J. 2005, AJ, 129, 355
Bazell, D., & Miller, D. J. 2005, BAAS, 37, 1213
Boroson, T. A., & Green, R. F. 1992, ApJS, 80, 109
Brandt, W. N., & Hasinger, G. 2005, ARA&A, 43, 827
Bromley, B. C., Press, W. H., Lin, H., & Kirshner, R. P. 1998, ApJ, 505, 25
Cattell, R. B. 1966, Multivariate Behavioral Research, 1, 245
Collura, A., Micela, G., Sciortino, S., Harnden, F. R., Jr., & Rosner, R. 1995, ApJ, 446, 108
Connolly, A. J., Szalay, A. S., Bershady, M. A., Kinney, A. L., & Calzetti, D. 1995, AJ, 110, 1071
Favata, F., Flaccomio, E., Reale, F., Micela, G., Sciortino, S., Shang, H., Stassun, K. G., & Feigelson, E. D. 2005, ApJS, 160, 469
Feigelson, E. D., & Montmerle, T. 1999, ARA&A, 37, 363
Feigelson, E. D., Townsley, L., Güdel, M., & Stassun, K. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keil (Tucson: Univ. Arizona Press), 313
Flaccomio, E., Damiani, F., Micela, G., Sciortino, S., Harnden, F. R., Murray, S. S., & Wolk, S. J. 2003, ApJ, 582, 398

Flaccomio, E., Micela, G., Sciortino, S., Feigelson, E. D., Herbst, W., Favata, F., Harnden, F. R., Jr., & Vrtilek, S. D. 2005, ApJS, 160, 450
Folkes, S., et al. 1999, MNRAS, 308, 459
Getman, K. V., Feigelson, E. D., Garmire, G., Broos, P., & Wang, J. 2007, ApJ, 654, 316
Getman, K. V., Feigelson, E. D., Grosso, N., McCaughrean, M. J., Micela, G., Broos, P., Garmire, G., & Townsley, L. 2005a, ApJS, 160, 353
Getman, K. V., et al. 2005b, ApJS, 160, 319
Grosso, N., Feigelson, E. D., Getman, K. V., Kastner, J. H., Bally, J., & McCaughrean, M. J. 2006, A&A, 448, L29
Grosso, N., et al. 2005, ApJS, 160, 530
Güdel, M., Telleschi, A., Audard, M., Skinner, S. L., Briggs, K. R., Palla, F., & Dougados, C. 2007, A&A, in press (astro-ph/0609182)
Hojnacki, S. M. 2005, Ph.D. thesis, Rochester Institute of Technology
Hojnacki, S. M., & Kastner, J. H. 2004, Proc. SPIE, 5493, 474
Horn, J. L. 1965, Psychometrika, 30, 179
Huenemoerder, D. P., Canizares, C. R., Drake, J. J, & Sanz-Forcada, J. 2003, ApJ, 595, 1131
Jackson, J. E. 1991, A User's Guide To Principal Components (New York: Wiley)
Johnson, R. A., & Wichern, D. W. 2002, Applied Multivariate Statistical Analysis (5th ed.; Upper Saddle River: Prentice Hall)
Jolliffe, I. T. 1986, Principal Component Analysis (New York: Springer)
Kastner, J. H., Franz, G., Grosso, N., Bally, J., McCaughrean, M. J., Getman, K., Feigelson, E. D., & Schulz, N. 2005, ApJS, 160, 511
Kastner, J. H., Huenemoerder, D. P., Schulz, N. S., Canizares, C. R., & Weintraub, D. A. 2002, ApJ, 567, 434
Kastner, J. H., et al. 2004, Nature, 430, 429
———. 2006, ApJ, 648, L43
Kilgard, R. E., Kaaret, P. E., Krauss, M. I., McDowell, J. C., Prestwich, A. H., Raley, M. T., & Zezas, A. 2001, BAAS, 33, 1349
Kim, D.-W., et al. 2004, ApJS, 150, 19
Kruskal, J. B. 1964, Psychometrika, 29, 1
Levene, H. 1960, Contributions to Probability and Statistics (Stanford: Stanford Univ. Press)
Lu, H., Zhou, H., Wang, J., Wang, T., Dong, X., Zhuang, Z., & Li, C. 2006, AJ, 131, 790
McGlynn, T. A., et al. 2004, ApJ, 616, 1284
Muno, M. P., Bauer, F. E., Bandyopadhyay, R. M., & Wang, Q. D. 2006, ApJS, 165, 173
Muno, M. P., et al. 2003, ApJ, 589, 225
Muzerolle, J., Hartmann, L., & Calvet, N. 1998, AJ, 116, 455
Preibisch, T., et al. 2005, ApJS, 160, 401
Qin, D.-M., Guo, P., Hu, Z.-Y., & Zhao, Y.-H. 2003, Chinese J. Astron. Astrophys., 3, 277
Stassun, K. G., van den Berg, M., Feigelson, E., & Flaccomio, E. 2006, ApJ, 649, 914
Stelzer, B., Flaccomio, E., Montmerle, T., Micela, G., Sciortino, S., Favata, F., Preibisch, T., & Feigelson, E. D. 2005, ApJS, 160, 557
Suchkov, A. A., Hanisch, R. J., & Margon, B. 2005, AJ, 130, 2439
Suzuki, N. 2006, ApJS, 163, 110
Telleschi, A., Güdel, M., Briggs, K. R., Audard, M., & Scelsi, L. 2007, A&A, in press (astro-ph/0611024)
Tsujimoto, M., Feigelson, E. D., Grosso, N., Micela, G., Tsuboi, Y., Favata, F., Shang, H., & Kastner, J. H. 2005, ApJS, 160, 503
Weisskopf, M. C., Brinkman, B., Canizares, C., Garmire, G., Murray, S., & Van Speybroeck, L. P. 2002, PASP, 114, 1
Wolk, S. J., Harnden, F. R., Jr., Flaccomio, E., Micela, G., Favata, F., Shang, H., & Feigelson, E. D. 2005, ApJS, 160, 423
Yip, C. W., et al. 2004, AJ, 128, 2603