Available online at www.sciencedirect.com
ScienceDirect
Statistical Methodology 5 (2008) 350–360
Statistical Methodology
Official Journal of the International Indian Statistical Association
www.elsevier.com/locate/stamet

An unsupervised, ensemble clustering algorithm: A new approach for classification of X-ray sources

S.M. Hojnacki[a,*], G. Micela[b], S.M. LaLonde[d], E.D. Feigelson[c], J.H. Kastner[a]

[a] *Center for Imaging Science, Rochester Institute of Technology, 74 Lomb Memorial Drive, Rochester, NY 14623, United States*
[b] *INAF, Osservatorio Astronomico di Palermo G. S. Vaiana, Piazza del Parlamento 1, 90134 Palermo, Italy*
[c] *Department of Astronomy and Astrophysics, Pennsylvania State University, 525 Davey Laboratory, University Park, PA 16802, United States*
[d] *Center for Quality and Applied Statistics, Rochester Institute of Technology, 98 Lomb Memorial Drive, Rochester, NY 14623, United States*

Received 19 April 2007; received in revised form 20 February 2008; accepted 20 February 2008

Abstract

A large volume of CCD X-ray spectra is being generated by the Chandra X-ray Observatory (*Chandra*) and XMM-Newton. Automated spectral analysis and classification methods can aid in sorting, characterizing, and classifying this large volume of CCD X-ray spectra in a non-parametric fashion, complementary to current parametric model fits. We have developed an algorithm that uses multivariate statistical techniques, including an ensemble clustering method, applied for the first time for X-ray spectral classification. The algorithm uses spectral data to group similar discrete sources of X-ray emission by placing the X-ray sources in a three-dimensional spectral sequence and then grouping the ordered sources into clusters based on their spectra. This new method can handle large quantities of data and operate independently of the requirement of spectral source models and a priori knowledge concerning the nature of the sources (i.e., young stars, interacting binaries, active galactic nuclei). We apply the method to *Chandra* imaging spectroscopy of the young stellar clusters in the Orion Nebula Cluster and the NGC 1333 star formation region.

Keywords: Statistical methods; Unsupervised clustering; Nonparametric; X-ray emission; Young stellar clusters

* Corresponding author. Tel.: +1 818 825 8078.
E-mail address: smhpci@rit.edu (S.M. Hojnacki).

1572-3127/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2008.02.008

1. Introduction

Chandra X-ray Observatory (*Chandra*; see Ref. [1]) studies often involve X-ray source populations with hundreds or thousands of members. Significant *Chandra* surveys to date include: deep galactic surveys [2], the galactic center [3,4], the *Chandra* Multiwavelength Project (ChaMP; see Ref. [5]), and nearby galaxies [6]. In addition, the XMM-Newton European Photon Imaging Camera (EPIC) Catalogue (2XMMp 2006) already contains over 100,000 sources.

Statistics will play an important role in analyzing these data. Automated statistical algorithms can systematically sift through this volume of high quality X-ray data. Objective, model-independent methods may be used to identify X-ray spectral classes in the vast population of sources detected in these and other large scale surveys to:

- group X-ray emitting objects with like spectra into classes for further study and possible analysis en masse,
- search for trends in X-ray and stellar properties (e.g., X-ray luminosity and X-ray temperature with stellar photospheric temperature [7]),
- efficiently identify groups of X-ray sources that have no visible counterparts, including sources that may be very young protostars,
- find additional pre-main-sequence (PMS) accretors exhibiting sustained accretion bursts to study the relationship between hard X-ray emission and optical emission for these sources [8, 9],
- guide or fine-tune downstream automated fitting of physical source models to individual source spectra by first sorting young stars into natural groupings.

Furthermore, the resulting classes may be used to discriminate spectroscopically between several known emission classes (i.e., stellar coronal emission, OB winds, jets, and extragalactic contaminants).

Few techniques exist for classifying X-ray sources (see a brief review in Ref. [19]). Our spectral classification algorithm groups X-ray sources based on their X-ray spectra and can handle large quantities of data [10]. Spectral source models and a priori knowledge concerning the nature of the sources (e.g., young stars, interacting binaries) are not required by the algorithm. Our technique employs Principal Component Analysis (PCA; see Refs. [11–13]) to reduce the dimensionality of spectral bands for subsequent clustering. PCA provides a means to automatically define optimal spectral band definitions from the data set itself, rather than imposing bands on the data, such as in hardness ratios.

Using PCA followed by an ensemble classifier, our algorithm groups similar discrete sources of X-ray emission by placing the X-ray sources in a three-dimensional spectral sequence and then grouping the ordered sources into clusters based on their spectra. X-ray sources exhibiting unusual spectra fall outside this 3-D locus. This approach could reveal classes of sources that do not fit any existing spectral models and/or X-ray sources that are extreme spectral outliers.

Here, we describe in detail our statistical method for X-ray source classification. We summarize its application to the young stellar cluster in the Great Nebula in Orion and present preliminary results for the NGC 1333 star formation region.

2. COUP data preparation

The algorithm was developed and tested using a subset of the long integration Chandra Orion Ultradeep Project observation (COUP; see Ref. [14]). The COUP observation of the Orion

Nebula Cluster (ONC), an ∼838 ks exposure obtained over a 13-day period in January of 2003, represents the most sensitive and comprehensive description of X-ray emission from a young star cluster [15,16]. Getman et al. [14] performed fundamental data reduction on the COUP data set, including extracting valid events, detecting X-ray emitting sources, and flagging potential source detection problems and artifacts.

A subset of the 1616 COUP sources was selected for use in developing and testing the classification algorithm described here. Sources with less than 300 counts were eliminated, to restrict the sample to those with higher signal-to-noise CCD spectra. Over 70% of the sources in our subset had counts more than twice this lower limit. Sources that were flagged as a double source, a piled-up source, and/or in a source extraction region crossing a bright source readout trail were also eliminated from the sample. The remaining sample contained 444 sources with high quality CCD imaging spectra.

The 0.4–8.2 keV spectra were rebinned into a number of non-overlapping spectral bands. Band locations and widths were defined based on hardware constraints, with care taken not to bisect key emission features (see Appendix of [19]). The source spectra were normalized for brightness level and a nonlinear monotone transformation [17] was performed prior to dimension reduction solely in an attempt to remove nonlinearities in the data.

3. X-ray spectral classification algorithm

The spectral classification algorithm consists of the following steps: PCA, agglomerative hierarchical clustering [18], and K-means clustering [18].

PCA was 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. Several statistical stopping rules were applied to determine the number of principal components (PCs) to retain for the clustering analysis [10]. These tests indicated that two to four PCs are sufficient to represent the variability in the data set. We chose to retain four PCs for the ensuing clustering steps. The variance explained by these components is 62.3%, 30.7%, 2.2%, and 1.9% respectively, 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 Fig. 1. The x-axis of the eigenvector plots corresponds to the 42 X-ray spectral bands (see the Appendix of [19]). The y-axis corresponds to the eigenvector coefficient (arbitrary units).

The first eigenvector peaks sharply at ∼1 keV and is 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 correspond to the hardest sources. The peak in the second eigenvector (∼1.5 keV) is shifted to a slightly higher energy and is broader than that of the first eigenvector; 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 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.

An ensemble classifier consisting of two unsupervised clustering methods – agglomerative hierarchical clustering and K-means clustering – was applied following PCA in an attempt to determine a "natural" partitioning of the data set into a number of relatively homogeneous clusters of X-ray sources. First, an agglomerative hierarchical clustering method based on Euclidean distance and complete linkage [18] was used on the first four PCs generated from

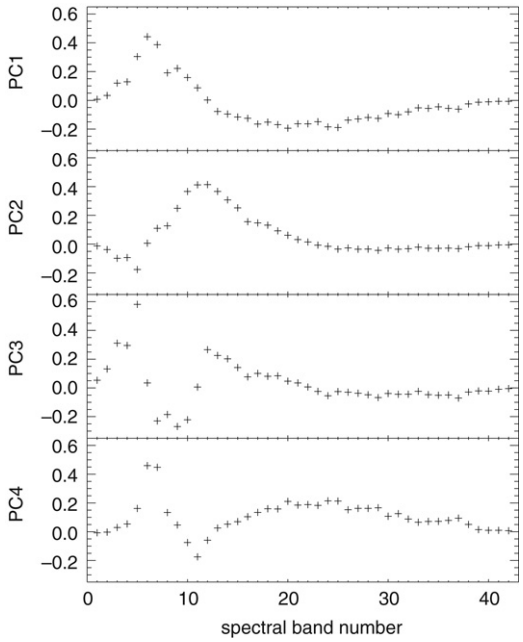

Fig. 1. Eigenvector plots for the four principal components.

the X-ray source spectra. The method begins by assigning each X-ray source to its own cluster and calculating the Euclidean distance between each cluster and all other clusters. In the first amalgamation step, the closest two (i.e., at smallest Euclidean distance) clusters are joined. In the next step, either a third source joins the two that were joined in the first step or two other sources are joined together. Close groups are successively merged in this hierarchical or nested fashion, based on the Euclidean distances between pairs of clusters. Cluster merging continues until there is only one large cluster containing all the sources. 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 [13]. The resulting dendrogram (see Fig. 2) indicates where and how the source sample was broken into groups. The branches of the dendrogram show not only how the clusters were formed but also indicate the distance between clusters. This procedure yielded 17 groups in the final partition for our 444-source COUP subsample. The horizontal dashed line in Fig. 2 illustrates where the dendrogram was cut. A disadvantage of hierarchical clustering is that it cannot transfer a source from one cluster to another if the source was grouped incorrectly in an earlier step. Therefore, the K-means algorithm was used to fine-tune the clusters obtained from the hierarchical clustering algorithm.

K-means does not assume or require multivariate normality of the input data. It groups observations into a collection of K clusters, where K must be identified in advance. The hierarchical clustering algorithm was used to obtain a value of K and the clusters obtained by the hierarchical clustering algorithm were used as the initial partition for the K-means algorithm. Each source was examined and assigned to the group with the nearest centroid (multi-dimensional

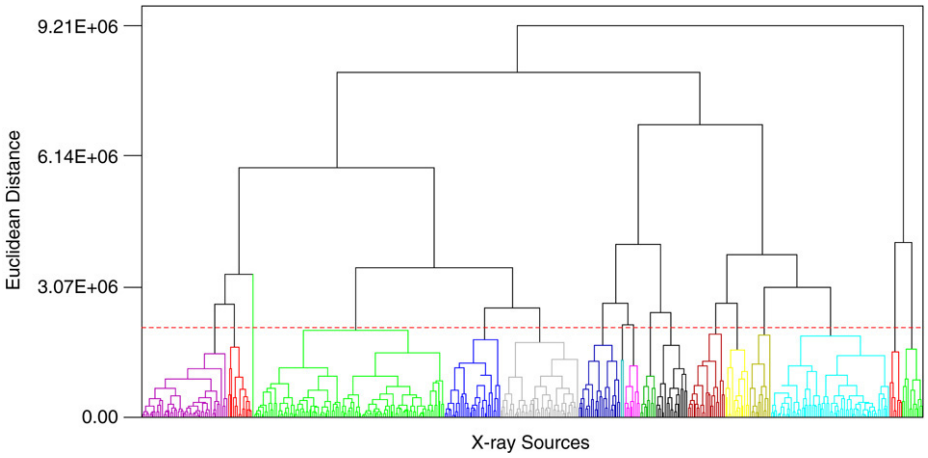

Fig. 2. Dendrogram resulting from hierarchical clustering on the COUP subset. The dashed line shows where the dendrogram was cut, resulting in 17 classes. Each class is represented by a different color.

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. The resulting distance between each source and the centroid of its spectral class was smaller than the distance between the centroids of the spectral classes. We also ran the algorithm with different numbers of clusters in the output of the agglomerative hierarchical clustering step. The distance results were worse, as well as the final cluster membership, when examined manually.

4. Analysis of results for the ONC

Our analysis of the sample of 444 X-ray emitting objects resulted in 17 X-ray spectral classes of sources (see Table 2 in Ref. [19]). A plot of the first two PCs for each source is shown in Fig. 3. The plot is color-coded by class. These colors do not match the colors in the dendrogram in Fig. 2 because the classes in the dendrogram show an intermediate state of the ensemble classifier: after hierarchical clustering and before K-means clustering.

Average spectra of some of the final X-ray spectral classes are shown in Fig. 3. The progression of classes moving clockwise around the horseshoe-shaped plot forms a sequence of decreasing spectral hardness. Strong similarities between spectra of a given class were found, whereas fundamental spectral differences were found between different classes. However, the higher numbered classes do not appear to be physically distinct in this 2-D score plot. The separation in the classes is more readily seen when this PC2 vs. PC1 plot is rotated about the PC2 axis (see Fig. 4). There is curvature in the third dimension (PC3) that can be seen only in the plots in Fig. 4. PC3 breaks a degeneracy between parameters apparently mixed on the PC2 vs. PC1 plot for many of the classes. However, the higher numbered classes still form more of a sequence along the 3-D locus. Nearly all of the X-ray sources in the input sample lie in a single spectral sequence in Fig. 3 with the notable exception of class 17. This 3-D locus of points contains three principal regions corresponding to different groups of X-ray spectral energy distributions (SEDs) present in the 444-source COUP sample. 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-shaped plot. Classes 7–11 display SEDs with strong peaks at

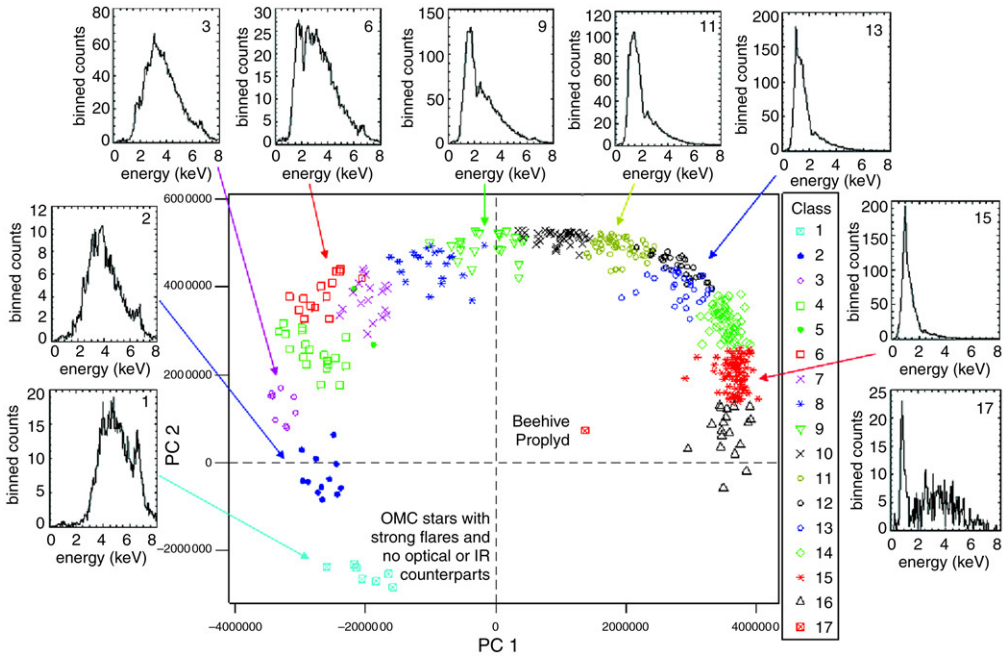

Fig. 3. Plot of only the first two principal components showing the source classes. The class numbers increase moving clockwise around the horseshoe. Average spectra for some of the classes are shown.

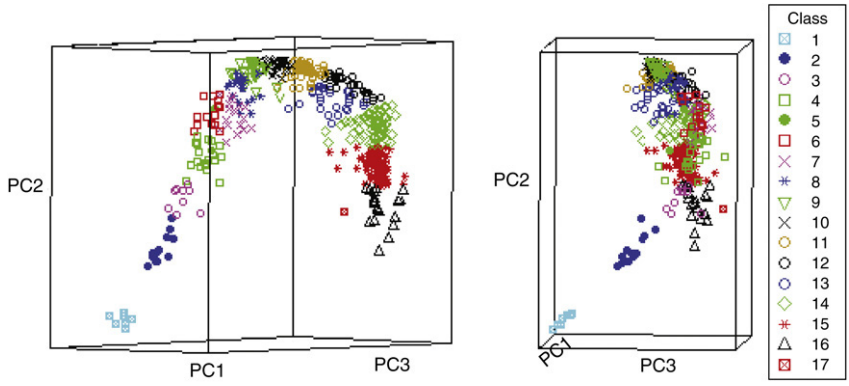

Fig. 4. Rotation of PC2 vs. PC1 plot about the PC2 axis. The plot on the right shows the horseshoe edge-on.

∼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 Fig. 3.

The only source in class 17, the Beehive Proplyd [20], is isolated from the other sources in the plot due to its unique double-peaked X-ray spectrum, indicative of a combination of strong shocks in the jet collimation region and a variable, magnetically active X-ray source. The correct identification of this source as an outlier by our algorithm holds promise for the discovery of

additional examples of such hybrid X-ray spectra among the tens of thousands of PMS stars detected thus far by *Chandra* and XMM-Newton.

The resulting X-ray spectral classes were compared with known optical and near infrared (ONIR) properties of the young stellar population of the ONC [19]. Sources in class 1 are easily identified as a distinct X-ray spectral group by our technique (Fig. 3, Fig. 4). These sources lack ONIR counterparts and exhibit high amplitude, fast-rise X-ray flares [15]. These sources are candidate deeply embedded protostars that show evidence for fluorescent iron line emission arising in cool circumstellar disk material [19]. Trends have been found between X-ray spectral parameters and stellar parameters for very low mass, soft spectra, young sources. Specifically, the hydrogen column density (N_H) decreases monotonically from class 1 to class 16. The large $\log N_H$ value for classes 1 through 8 results in small fractions of ONIR counterparts within those classes. For classes 11 through 15, which have relatively large numbers of ONIR counterparts, our analysis reveals an apparent trend of increasing spectral softness with decreasing T_{eff}. In addition, the near infrared excess is observed to decrease monotonically from class 10 to class 16, suggesting a generally decreasing accretion rate.

To assess the robustness of the classifications to measurement errors, we did the following. Three of the faintest sources from classes 2, 11, and 14 were selected from the COUP subsample to test the effect of random noise fluctuations. Sources were chosen from the "edge" of each class "cloud". The spectrum for each of the three sources was perturbed five times to simulate random noise. Each of these 15 "new" sources was then classified using the algorithm. The algorithm assigned each new source to the same class as that of the original unperturbed source.

To aid in interpretation of the plot in Fig. 3, simulated X-ray source spectra were generated and then processed using the spectral classification algorithm (see [19]). The distribution of the resulting simulated X-ray spectra 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 N_H. This indicates that the position of a COUP source on the horseshoe is determined in large part by the degree of its obscuration by intervening cloud or circumstellar material. The importance of N_H to the membership of the source classes is particularly apparent along the left side of the horseshoe. At the other extreme, the high numbered classes (14–16) appear to form a sequence of decreasing thermal plasma temperature (kT), indicating that this parameter is more important than intervening absorption in determining the positions of sources along the right side of the horseshoe. Simulations with kT greater than 0.26 keV but less than 4 keV bound the majority of the data set.

5. Beyond Orion: Algorithm extension to other star formation regions

The existing X-ray spectral classification algorithm was extended to X-ray sources in other star formation regions. The extended algorithm was tested on several *Chandra* ACIS-I archival observations of star formation regions, to determine whether the relationships between X-ray spectral classes and fundamental stellar parameters found in the ONC generalize to other star forming regions. Preliminary results are shown for NGC 1333, a well-characterized star formation region located at a relatively small distance providing good photon counting statistics for extraction of spectra.

The extended algorithm consists of the following procedure: calculate discriminant functions for each of the existing X-ray classes based on generalized squared distances using PC1 through PC4 and cross-validation; calculate PCs for the additional X-ray sources using the COUP

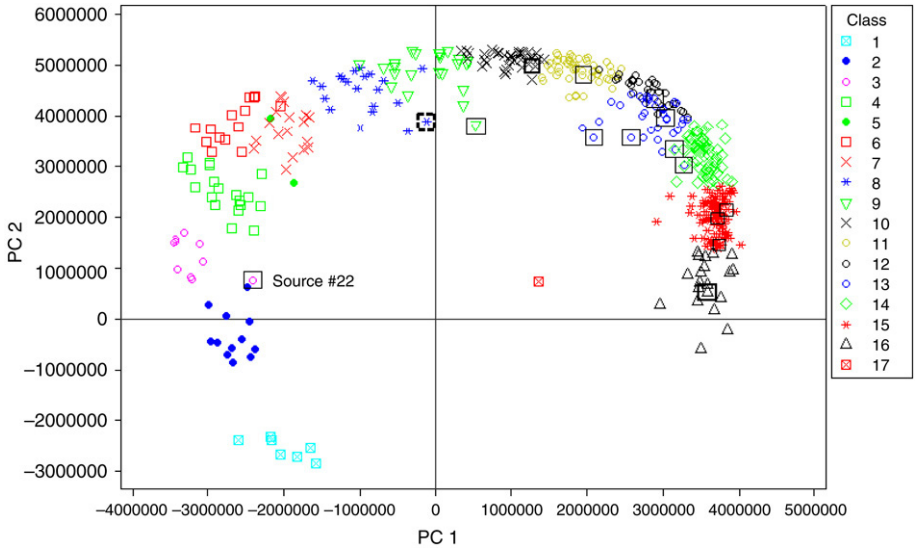

Fig. 5. Plot of the first two principal components showing the NGC 1333 sources overlaid on the COUP principal component plot. Each of the 17 NGC 1333 sources is enclosed within a box. Two NGC 1333 sources with nearly identical RA and DEC are enclosed within a dotted-line box.

eigenvectors; calculate discriminant scores for each additional X-ray source; and assign the additional X-ray sources to an existing X-ray spectral class based on posterior probabilities.

5.1. NGC 1333 data preparation

A web-based database, An Archive of *Chandra* Observations of Regions of Star Formation (ANCHORS; see Ref. [21]), has been developed by the Harvard–Smithsonian Center for Astrophysics. Currently, the database contains X-ray source properties for ∼50 *Chandra* fields totaling 10,000+ sources. The ANCHORS source detection data set was used for the ∼43 ks *Chandra* observation of NGC 1333, a highly active star formation region within the Perseus molecular cloud complex noted for its large population of protostars and young stellar outflows [23]. We then used standard CIAO tools and IDL to extract spectra. Detected sources with fewer than 100 total counts were eliminated, to limit faint sources with poor photon counting statistics. This resulted in a subset of 28 out of 90 detected sources. Class membership probabilities were calculated for the 28 X-ray sources. Seventeen of the sources have class membership probabilities of 0.94 or greater. Five sources have membership probabilities between 0.80 and 0.90, three sources have membership probabilities between 0.70 and 0.79, and the final three sources have membership probabilities less than 0.70. Fig. 5 shows the resulting classification of the 17 NGC 1333 sources with membership probabilities of 0.94 or greater, combined with the plot of COUP sources in Fig. 3.

The 17 NGC 1333 sources are enclosed within boxes. Two NGC 1333 sources with nearly identical RA and DEC are enclosed within a dotted-line box in class 8. One of the two sources is a known, deeply embedded, classical T Tauri star [22]. The other component is an X-ray discovered companion with a projected separation of $2''$ [23]. The two components have nearly identical spectra with log N_H of 22.2, suggesting that they may constitute a physical binary [23]. Source 22 is a young stellar object (YSO) in NGC 1333 that our algorithm correctly placed in

class 3, consistent with object classification obtained from the ONC data set. In Fig. 5, it appears to be located closer to the centroid of class 2 rather than that of class 3. However, this plot only shows the first two PCs. Four PCs were used to classify the sources. When a 3-D plot of the horseshoe is viewed, it is clear that Source 22 is much closer to the centroid of class 3 than that of class 2.

Known optical properties of the 17 NGC 1333 sources [23] fall within the ranges for their corresponding COUP X-ray classes, following the trends found for the ONC. From this data set, it can be seen that the application of our algorithm to NGC 1333 sources demonstrates that the general classification results for Orion are followed closely by a second star formation region.

6. Summary and conclusions

It is clear that X-ray spectra have continuous characteristics and the principal components derived from them are not distributed in completely separated classes in two dimensions. This is not uncommon in astrophysics. For example, the optical stellar spectral classification system of O, B, A, F, G, K, M, each with 10 sub-types, has its basis in the optical spectra of stars and has some ambiguity between sub-types. However, separation of spectra into classes helps in understanding the underlying physics of the objects in each group.

We have developed an algorithm and corresponding input variable definition that is nonparametric and performs without a priori information concerning the nature of the X-ray sources. It objectively sorts and clusters X-ray sources into natural groupings based solely on observables, thereby guiding downstream automated fitting of physical source models. Furthermore, once grouped spectrally, sources within the same spectral group may be sufficiently similar as to be treated identically (and analyzed en masse) for the purpose of further spectral analysis employing physical models.

The algorithm was applied to X-ray spectra of sources detected in the COUP observation of the ONC. The clustering allows for the beginning of the understanding of the changes in the PMS stars over the range of values in the space defined by the four PCs. The apparent diversity of source spectra is reduced to a 3-D locus in PC space. Nearly all PMS stars lie in a single spectral sequence and stars with strong, impulsive flares tend to fall in the lowest numbered classes while stars exhibiting more consistent but lower level coronal activity tend to fall in the higher numbered classes. Some of the X-ray spectral classes have been shown to be unique. In addition, sources with unique and unusual observed spectra are offset from the 3-D locus of sources.

The grouping of the sources into clusters is somewhat a matter of convenience and helps us to discuss the physical properties of a source with respect to its location in the horseshoe-shaped plot. The cluster membership has been determined from the first four PCs, not just PC1 and PC2, and the plot shown in Fig. 3. More analysis needs to be done to determine the meaning of PC3 and PC4 in relation to class membership and how PC3 and PC4 contribute to the cluster shapes and locations.

Preliminary results from applying the algorithm to X-ray sources detected in the shorter exposure observation of NGC 1333 show that young stars in this region fit into the previously established statistical ONC groupings and exhibit the same trends between X-ray spectral parameters and stellar parameters as are exhibited by PMS stars in the ONC. These results demonstrate that the general classification results for Orion are followed closely by a second star formation region.

The results from Orion and NGC 1333 indicate that our non-parametric approach to X-ray spectral classification can yield basic information concerning the nature of X-ray emission from young stars. In addition, this algorithm can be used to efficiently identify very young X-ray sources that lack ONIR counterparts, display strong Fe K $-$ α line emission, and display large amplitude, fast-rise impulsive flares versus more low level activity. Finally, trends between X-ray spectral parameters and stellar parameters have been found for very low mass, soft spectra, young sources. These trends are of significance to the study of star formation in young stellar clusters.

Acknowledgements

This research was supported by NASA under AISRP award number NNG04GQ07G and *Chandra* Award Number AR6-7001X 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.

References

[1] M.C. Weisskopf, B. Brinkman, C. Canizares, G. Garmire, S. Murray, L.P. Van Speybroeck, An overview of the performance and scientific results from the Chandra X-ray observatory, PASP 114 (2002) 1–24.
[2] W.N. Brandt, G. Hasinger, Deep extragalactic X-ray surveys, Annual Review of Astronomy & Astrophysics 43 (2005) 827–859.
[3] M.P. Muno, F.K. Baganoff, M.W. Bautz, et al., A deep Chandra catalog of X-ray point sources toward the galactic center, ApJ 589 (2003) 225–241.
[4] M.P. Muno, F.E. Bauer, R.M. Bandyopadhyay, Q.D. Wang, A Chandra catalog of X-ray sources in the central 150 pc of the galaxy, ApJS 165 (2006) 173–187.
[5] D.-W. Kim, R.A. Cameron, J.J Drake, et al., Chandra multiwavelength project. I. First X-ray source catalog, ApJS 150 (2004) 19–41.
[6] R.E. Kilgard, P.E. Kaaret, M.I. Krauss, J.C. McDowell, A.H. Prestwich, M.T. Raley, A. Zezas, A mini-survey of the X-ray point sources in starburst and non-starburst galaxies, BAAS, 199th AAS Meeting, 33, 2001, p. 1349.
[7] T. Preibisch, Y. Kim, F. Favata, et al., The origin of T Tauri X-ray emission: New insights from the Chandra Orion ultradeep project, ApJS 160 (2005) 401–422.
[8] J.H. Kastner, M. Richmond, N. Grosso, D.A. Weintraub, T. Simon, A. Frank, K. Hamaguchi, H. Ozawa, A. Henden, An X-ray outburst from the rapidly accreting young star that illuminates McNeil's nebula, Nature 430 (2004) 429–431.
[9] J.H. Kastner, M. Richmond, N. Grosso, D.A. Weintraub, T. Simon, A. Henden, K. Hamaguchi, A. Frank, H. Ozawa, V1647 Orionis: The X-ray evolution of a pre-main-sequence accretion burst, ApJ 648 (2006) L43–L46.
[10] S.M. Hojnacki, A source classification algorithm for astronomical X-ray imagery of stellar clusters. Ph.D. Thesis. Rochester Institute of Technology, 2005.
[11] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986.
[12] J.E. Jackson, A User's Guide To Principal Components, John Wiley & Sons, Inc., New York, 1991.
[13] A. Collura, G. Micela, S. Sciortino, F.R. Harnden Jr., R. Rosner, An objective multicolor method for the characterization of low-resolution X-ray spectra, ApJ 446 (1995) 108–114.
[14] K.V. Getman, E. Flaccomio, P.S. Broos, et al., Chandra Orion ultradeep project: Observations and source lists, ApJS 160 (2005) 319–352.
[15] K.V. Getman, E.D. Feigelson, N. Grosso, M.J. McCaughrean, G. Micela, P. Broos, G. Garmire, L. Townsley, Membership of the Orion nebula population from the Chandra Orion ultradeep project, ApJS 160 (2005) 353–378.
[16] E.D. Feigelson, L. Townsley, M. Güdel, K. Stassun, X-ray properties of young stars and stellar clusters, in: B. Reipurth, D. Jewitt, K. Keil (Eds.), Protostars and Planets V, Univ. Arizona Press, Tucson, 2007, pp. 313–328.
[17] J.B. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika 29 (1964) 1–27.
[18] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, fifth edn, Prentice–Hall Inc., New Jersey, 2002.

[19] S.M. Hojnacki, J.H. Kastner, G. Micela, E.D. Feigelson, S.M. LaLonde, An X-ray spectral classification algorithm with application to young stellar clusters, ApJ 659 (2007) 585–598.
[20] J. Bally, D. Licht, N. Smith, J. Walawender, Outflows in the Orion Nebula: HH 540 from the Beehive Proplyd, AJ 129 (2005) 355–362.
[21] B.D. Spitzbart, S.J. Wolk, N.S. Bizunok, An archive of Chandra observations of regions of star formation (ANCHORS), in: Protostars and Planets V, LPI Contribution 1286, 2005, p. 8518.
[22] L.F. Rodriguez, G. Anglada, S. Curiel, The nature of the radio continuum sources embedded in the HH 7–11 region and its surroundings, ApJS 125 (1999) 427–438.
[23] K.V. Getman, E.D. Feigelson, L. Townsley, J. Bally, C.J. Lada, B. Reipurth, Chandra study of young stellar objects in the NGC 1333 star-forming cloud, ApJ 575 (2002) 354–377.