1-1-1998

Three types of gamma-ray bursts

Soma Mukherjee

Eric D. Feigelson

Gutti Jogesh Babu

Flonn Murtagh

Chris Fralev

See next page for additional authors

Recommended Citation

Soma Mukherjee, et. al., (1998) Three types of gamma-ray bursts.Astrophysical Journal508:1314. DOI: http://doi.org/10.1086/306386

THE ASTROPHYSICAL JOURNAL, 508:314–327, 1998 November 20

THREE TYPES OF GAMMA-RAY BURSTS

SOMA MUKHERJEE,[1,2,3] ERIC D. FEIGELSON,[4] GUTTI JOGESH BABU,[5] FIONN MURTAGH,[6,7]
CHRIS FRALEY,[8] AND ADRIAN RAFTERY[8]
Received 1998 February 9; accepted 1998 June 25

ABSTRACT

A multivariate analysis of gamma-ray burst (GRB) bulk properties is presented to discriminate between distinct classes of GRBs. Several variables representing burst duration, fluence, and spectral hardness are considered. Two multivariate clustering procedures are used on a sample of 797 bursts from the Third BATSE Catalog, a nonparametric average linkage hierarchical agglomerative clustering procedure validated with Wilks' Λ^* and other multivariate analysis of variance tests and a parametric maximum likelihood model–based clustering procedure assuming multinormal populations calculated with the Expectation-Maximization algorithm and validated with the Bayesian Information Criterion. The two methods yield very similar results. The BATSE GRB population consists of three classes with the following duration/fluence/spectrum bulk properties: class I with long/bright/soft bursts, class II with short/faint/hard bursts, and class III with intermediate/intermediate/soft bursts. One outlier due to spurious data is also present. Classes I and II correspond to those reported by Kouveliotou et al., but class III is clearly defined here for the first time.

Subject headings: gamma rays: bursts — methods: data analysis — methods: statistical

1. INTRODUCTION

As very few gamma-ray burst (GRB) sources have astronomical counterparts at other wavebands, empirical studies of GRBs have been largely restricted to the analysis of their gamma-ray properties: bulk properties such as fluence and spectral hardness and evolution of these properties within a burst event (Fishman & Meegan 1995). While bursts exhibit a vast range of complex temporal behaviors, their bulk properties appear simpler and amenable to straightforward statistical analyses. Studies fall into two categories: an examination of whether GRB bulk properties comprise a homogeneous population or are divided into distinct classes and a search for relationships between bulk properties. Both types of studies may lead to astrophysical insight, just as the distinction between main-sequence stars and red giants and the measurement of a luminosity-mass relation along the main sequence assisted the development of stellar astrophysics early in the century.

The most widely accepted taxonomy of GRBs is the division between short-hard and long-soft bursts proposed by Dezalay et al. (1992) and Kouveliotou et al. (1993, hereafter K93). K93 noticed a bimodality in the burst duration variable T_{90} (time within which 90% of the flux arrived), suggesting the presence of two distinct types of bursts separated at $T_{90} \simeq 2$ s. The short bursts have systematically harder gamma-ray spectra than longer bursts. The two groups seemed indistinguishable in most other bulk properties, although the larger group of long-soft bursts may have a subclass with a different fluence distribution (i.e., different $\langle V/V_m \rangle$; Katz & Canel 1996), and the groups may have different Galactic latitude distributions (Belli 1997). Other researchers point to small groups of bursts with distinctive properties such as the soft-gamma repeaters (Norris et al. 1991), two possible classes with differing short-timescale variability (Lamb, Graziani, & Smith 1993), fast-rise exponential decay bursts (Bhat et al. 1994), and two types of bursts with different ratios of total fluence and greater than 300 keV fluence (Pendleton et al. 1997).

A variety of relationships between burst properties have also been reported. Norris et al. (1995) find an anticorrelation between T_{90} (calculated after wavelet thresholding) and peak intensity, which is consistent with a cosmological time dilation. However, a positive correlation between T_{90} and total fluence is also seen that does not agree with the simplest cosmological interpretation (Lee & Petrosian 1997). Additional reported relationships include T_{90} correlated with peak heights (Lestrade 1994), peak energy correlated with peak flux (Mallozzi et al. 1995), and pulse duration anticorrelated with gamma-ray energy (Fenimore et al. 1995).

Most of these studies suffer from a failure to treat all of the bulk property variables in an unbiased and quantitative way. Astronomers typically examine univariate or bivariate distributions, sometimes constructing composite variables (such as hardness ratios) with predetermined relationships to include one or two additional variables. But it is quite possible that the complex astrophysics producing GRBs will not manifest themselves in simple bivariate plots, just as the division between short-hard and long-soft bursts is not evident in spectral variables alone (Pendleton et al. 1994). GRB catalogs, like most multiwavelength astronomical catalogs, are multivariate databases and should be treated with multivariate statistical methods that can objectively

[1] Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700 035, India.
[2] Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208-3112.
[3] Present address: LIGO Project, California Institute of Technology, MS 18-34, Pasadena, CA 91125.
[4] Department of Astronomy and Astrophysics, 525 Davey Laboratory, Pennsylvania State University, University Park, PA 16802; edf@astro.psu.edu; to whom correspondence should be addressed.
[5] Department of Statistics, 326 Thomas Building, Pennsylvania State University, University Park, PA 16802.
[6] Faculty of Informatics, University of Ulster, Londonderry BT48 7JL, Northern Ireland, United Kingdom.
[7] Observatoire Astronomique, 11 rue de l'Université, F-67000 Strasbourg, France.
[8] Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322.

and effectively uncover structure involving many variables (Feigelson & Babu 1997). Two previous studies take a fully multivariate approach to understanding GRB bulk properties. Baumgart (1994) constructs a neural network taxonomy of 99 GRBs from the *Pioneer Venus Orbiter* satellite using 26 variables representing both bulk burst properties and detailed temporal characteristics (e.g., number of peaks, fractal dimension, wavelet transform crossings) and finds two or three distinct GRB classes. Bagoly et al. (1998) perform principal components and factor analyses of nine bulk property variables using 625 GRBs from the Third BATSE (BATSE 3b) catalog. They find that the relationships in the database are determined principally by only three variables: an appropriately weighted fluence, a weighted burst duration, and (to a lesser extent) flux in the highest energy bin.

We note, however, that it can be dangerous to look for correlations prior to classifying (or establishing the homogeneity of) the population. Although the anticorrelation between hardness ratio and burst duration seen in full samples (K93) may be the manifestation of a single astrophysical process, it may alternatively reflect differences between distinct processes. The latter possibility is suggested by a reported hardness-duration positive correlation within the long-soft class of bursts (Dezalay et al. 1996; Horack & Hakkila 1997). Most multivariate analyses thus begin with a study of homogeneity and classification, then investigate the variance-covariance structure (i.e., correlations) within each class.

This paper describes a multivariate analysis of GRBs from the BATSE 3b catalog (Meegan et al. 1996). After defining the sample (§ 2), we start with a simple statistical description of the variables and their bivariate relationships for the entire data set (§ 3). We then seek distinct types of clusters in two ways. First, a standard nonparametric agglomerative hierarchical clustering analysis is performed (§ 4) that reveals three distinct classes. The statistical significance of the third cluster is validated, under Gaussian assumptions, with multivariate analysis of variance (MANOVA) tests. Second, a parametric maximum likelihood model–based clustering procedure is adopted that reveals the same three groups and indicates strong evidence for the presence for three rather than two groups (§ 5). The variance-covariance structure of each group is then examined (§ 6). Results are synthesized in the discussion (§ 7).

Throughout the paper, we discuss our mathematical techniques to help the reader understand the complexities of multivariate analysis. From the vast literature in this subject, we recommend the following monographs for interested readers: Johnson & Wichern (1992) and Jobson (1992) for overviews of applied multivariate analysis; Hartigan (1975), Jain & Dubes (1988), and Kaufman & Rousseeuw (1990) for multivariate clustering algorithms; Murtagh & Heck (1987) and, more briefly, Babu & Feigelson (1996) and Feigelson & Babu (1997) for multivariate methodology in astronomy.

2. THE GRB SAMPLE AND STATISTICAL SOFTWARE

Our sample is drawn from the BATSE 3b catalog on board the *Compton Gamma Ray Observatory*. This catalog has 1122 GRBs detected by BATSE between 1991 April 19 and 1994 September 19. The catalog is presented and fully described by Meegan et al. (1996). Our database was extracted from the online database[9] in 1996 May, which provides many properties of each burst. There are roughly 11 variables of potential astrophysical interest: two measures of location in Galactic coordinates, l and b; two measures of burst durations, the times within which 50% (T_{50}) and 90% (T_{90}) of the flux arrives; three peak fluxes, P_{64}, P_{256}, and P_{1024}, measured in 64, 256, and 1024 ms bins, respectively; and four time-integrated fluences, F_1–F_4, in the 20–50, 50–100, 100–300 keV, and 300 + keV spectral channels, respectively. Researchers commonly consider three composite variables: the total fluence, $F_T = F_1 + F_2 + F_3 + F_4$, and two measures of spectral hardness derived from the ratios of channel fluences, $H_{32} = F_3/F_2$ and $H_{321} = F_3/(F_1 + F_2)$. Because of the limitations of available multivariate statistical techniques, we ignore other variables of potential relevance, including the heteroscedastic measurement errors of each quantity (i.e., errors that differ from point to point) and truncation values associated with BATSE triggering operations.

Of the 1122 listed bursts, 807 have data on all the variables described above. Ten bursts listed with zero fluences were eliminated. Our sample thus has 797 BATSE GRBs. For some analyses, we also used a subset of 644 bursts with "debiased" durations, T_{90}^d. Here the durations are modified to account for the effect that brighter bursts will have signal above the noise for longer periods than fainter bursts with the same time profiles (J. Norris 1996, private communication).

Statistical analyses in §§ 3, 4, and 6 were conducted within the Statistical Analysis System SAS/STAT,[10] a very large and widely distributed commercial statistical software package (SAS Institute, Inc. 1989). SAS/STAT procedures CLUSTER, GLM, PRINCOMP, and VARCLUS were used. The analysis in § 5 was performed with the MCLUST software (Banfield & Raftery 1993; Fraley 1998), which is interfaced to the SPLUS statistical package (SPLUS Version 3.4; MathSoft, Inc.) and its extensions.[11]

3. STATISTICAL PROPERTIES OF THE ENTIRE SAMPLE

We are faced with a multivariate database of 797 objects and 15 variables (11 variables from the catalog, three composite variables, and T_{90}^d). Two initial problems are frequently faced in analyses of multivariate databases. First, variables with incompatible units and ranges must be compared. Units can be removed by normalization (e.g., replacing F_1 by $F_1/F_{\rm tot}$), by standardization (e.g., replacing F_1 by F_1/σ where σ is the sample standard deviation), or by taking logarithms. Second, the dimensionality of the problem should be reduced, as many of the variables are closely interrelated either by construction or by astronomical circumstance. Although there are no mathematical rules regulating reduction of dimensionality, it can usefully be guided by a correlation matrix showing bivariate relationships and a principal components analysis showing multivariate

[9] The online BATSE databse may be found on the World Wide Web at www.batse.msfc.nasa.gov/data/grb/catalog.

[10] SAS/STAT is a registered trademark of the SAS Institute, Inc.

[11] Further information is provided at http://stat.washington.edu/fraley/software.html and the StatLib software archive at http://lib.stat.cmu.edu/general/mclust. For multivariate data visualization, we used the XGobi (Swayne, Cook, & Buja 1991) program, available from http://lib.stat.cmu.edu/general/XGobi. Hypertext links to a variety of public domain software for multivariate analysis, classification, and visualization are available at the Penn State StatCodes Web site, http://www.astro.psu.edu/statcodes.

TABLE 1

AVERAGE GRB PROPERTIES FOR THE ENTIRE SAMPLE

Variable	Mean	Standard Deviation[a]
log T_{50} (s)	0.55	0.92
log T_{90} (s)	0.96	0.92
log F_{tot} (ergs cm^{-2})	−5.61	0.76
log P_{256} (photons s^{-1} cm^{-2})	0.16	0.45
log H_{321}	0.25	0.33
log H_{32}	0.48	0.30

[a] Sample standard deviation.

relationships that are mainly responsible for structured variance in the data. Scientific reasoning can also be used to eliminate consideration of variables. We conducted a preliminary examination of data representations, correlation matrices and bivariate plots, and principal components analyses to facilitate choice of variables. When no mathematical preference arose, we selected variables most commonly used by previous researchers to facilitate comparison of results.

Our choices were as follows. We use log variables rather than normalized or standardized variables. We kept information on burst fluence and spectra through F_{tot} and hardness ratios rather than through the original fluences F_1–F_4. We initially eliminated P_{64} and P_{1024} from consideration and later eliminated P_{256} when we found that its main contribution to the clustering process was noise. We chose to remove the location variables (l, b) already established by other researchers to be random for the entire sample but use them later to test for isotropy of subsamples. The debiased T_{90}^d is used only in special tests. Our analysis was thus performed in six or fewer dimensions using log T_{50}, log T_{90}, log F_{tot}, log P_{256}, log H_{321}, and log H_{32}.

Tables 1 and 2 give basic statistics for these six variables: means, standard deviations, and bivariate values of Pearson's linear correlation coefficient r. For $N = 797$ and assuming bivariate normal populations, any $|r| > 0.013$ implies that a correlation between the two variables exists at a two-tailed significance level $P < 0.001$ (Beyer 1968, pp. 389, 283). But from an astrophysical perspective, we might consider any relationship with $|r| \lesssim 0.1$ to be of little interest. Figure 1 shows the bivariate scatter plots.

The correlation structure of the entire sample (Table 2) shows that the two measures of duration and the two measures of spectral hardness have correlation near unity, indi-

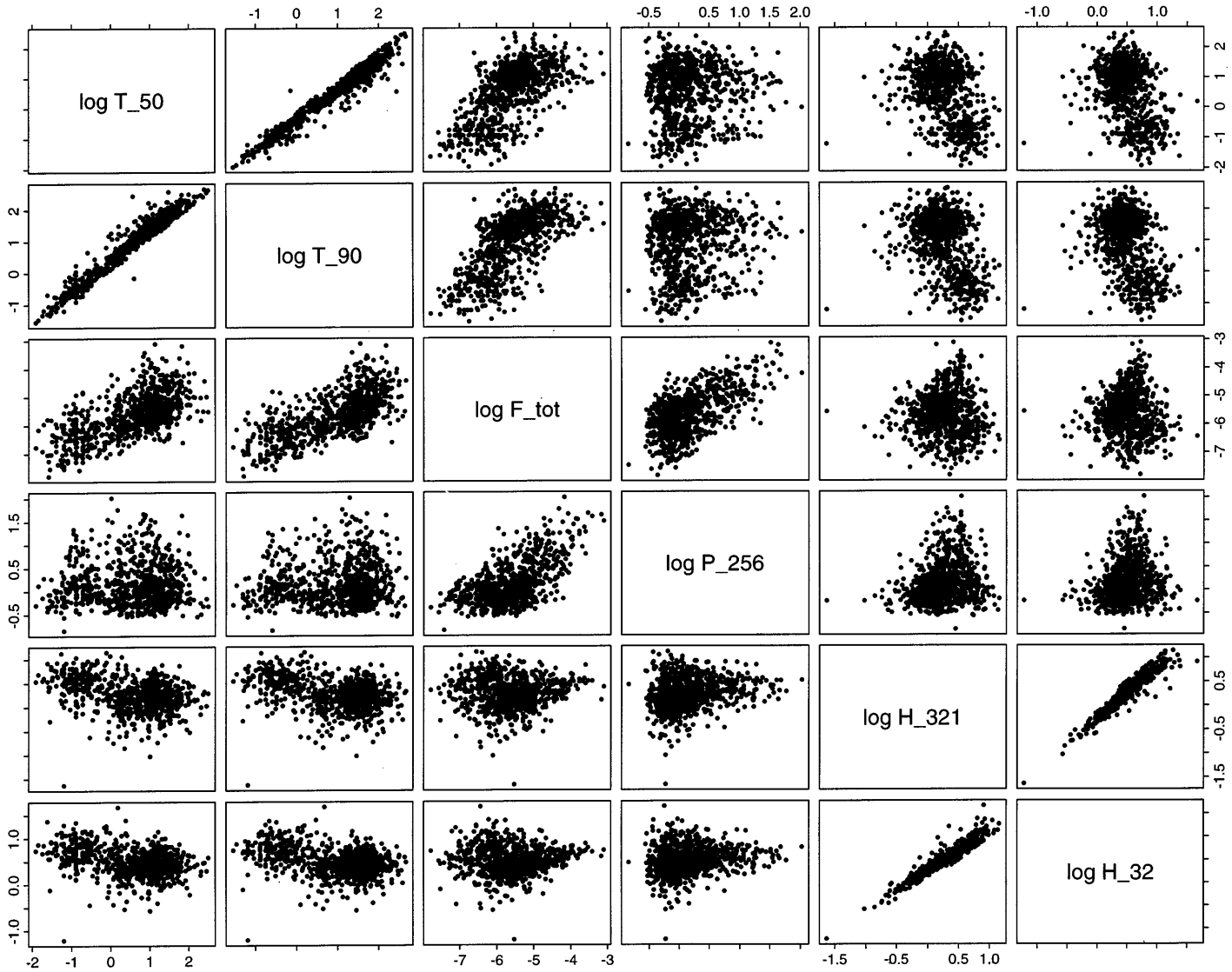

FIG. 1.—Mosaic of scatter plots of six bulk properties for the 797 GRBs from the BATSE 3b catalog used in this study

TABLE 2

CORRELATION COEFFICIENTS FOR THE ENTIRE SAMPLE

	$\log T_{50}$	$\log T_{90}$	$\log F_{\rm tot}$	$\log P_{256}$	$\log H_{321}$	$\log H_{32}$
$\log T_{50}$	1.00	...	...	...	...	...
$\log T_{90}$	0.97	1.00	...	...	...	...
$\log F_{\rm tot}$	0.63	0.66	1.00	...	...	...
$\log P_{256}$	−0.01	0.04	0.59	1.00	...	...
$\log H_{321}$	−0.36	−0.36	0.02	0.24	1.00	...
$\log H_{32}$	−0.35	−0.35	−0.00	0.19	0.96	1.00

cating that they are nearly redundant. $F_{\rm tot}$ and P_{256} are quite dissimilar: $F_{\rm tot}$ shows a strong correlation with burst duration (e.g., Lee & Petrosian 1997) and no relation to hardness, while P_{256} shows no relation to duration but is mildly correlated with hardness (Mallozzi et al. 1995). Burst duration is anticorrelated with hardness (K93; Fenimore et al. 1995). The cosmological anticorrelation between duration and peak flux reported by Norris et al. (1995) is statistically significant but accounts for only a few percent of the variance between these variables. The correlation matrix based on the debiased T_{90}^d values yields very similar results.

However, the scatter plots (Fig. 1) show a more complex story. First, many plots show inhomogeneous distributions inconsistent with the unimodal multinormal (i.e., multivariate Gaussian) population assumed by Pearson's r. The distributions often seem bimodal with asymmetrical non-Gaussian shapes. One outlier burst is also seen in several projections. We therefore consider the hypothesis that the sample consists of two or more distinct classes and proceed to find the "clusters" using well-established methods.

4. NONPARAMETRIC HIERARCHICAL CLUSTER ANALYSIS

4.1. *Methodological Background*

Agglomerative hierarchical clustering is a procedure based on the successive merging of proximate pairs of clusters of objects. It produces a clustering tree or dendrogram starting with N clusters of one member (or a coarse partition based on prior knowledge) and ending with one cluster of N members. Unfortunately, there are many possible ways to proceed; mathematics provides little guidance among the choices and no probabilistic evaluation of the results without the imposition of additional assumptions. The scientist must make four decisions to fully define the clustering procedure:

1. Creating *unit-free variables* is essential for meaningful treatment of objects in multivariate space (§ 3). A favorite choice by statisticians is standardization, where each variable is normalized by the standard deviation of the sample. Astronomers more commonly make logarithmic transformations or construct ratios of variables sharing the same units. We follow the tradition of GRB researchers by measuring spectral hardness with ratios of fluences having the same units and making logarithmic transformations of all variables.

2. The *metric* defines the meaning of proximity between two objects or clusters. Common choices are the simple Euclidean distance between unit-free variables and the squares of Euclidean distances. We chose the former option for most of the analyses in this section.

3. Several *merging procedures* can be used. One might begin by merging the clusters with the nearest neighbors. This is called single linkage clustering and is most familiar in astronomy, where it is frequently called the friends-of-friends algorithm. It tends to produce long stringy clusters and is equivalent to a well-known divisive clustering procedure known as pruning the minimal spanning tree. Complete linkage proceeds by maximizing the distance between clusters and leads to evenly bifurcating dendrograms. For most of our analysis, we choose average linkage, where the distance between two clusters is the average of the distances between pairs of observations, where each member of the pair comes from a different cluster. This is a compromise between single and complete linkage and tends to give compact clusters. Specifically, the distance between clusters K and L is given by (e.g., SAS Institute, Inc., 1989, pp. 17–529; Johnson & Wichern 1992, pp. 226–584)

$$D_{\rm KL} = |\,\bar{x}_{\rm K} - \bar{x}_{\rm L}\,|^2 + \frac{W_{\rm K}}{n_{\rm K}} + \frac{W_{\rm L}}{n_{\rm L}}\,, \tag{1}$$

where the bar indicates an unweighted mean, $W_{\rm K} = \sum_{i=1}^{n_k} |\,x_i - \bar{x}\,|^2$, and n_k is the number of members of the kth cluster. Another popular choice is Ward's minimum variance criterion where the distance between the two clusters is the analysis of variance sum of squares between two clusters added up over all variables (Ward 1963),

$$D_{\rm KL} = |\,\bar{x}_{\rm K} - \bar{x}_{\rm L}\,|^2 \Big/ \left(\frac{1}{n_{\rm K}} + \frac{1}{n_{\rm L}}\right). \tag{2}$$

If the sample is generated by a mixture of multinormal (i.e., multidimensional Gaussian) distributions where each distribution has covariance matrix of the form $\Sigma^2 I$, this method joins clusters to maximize the likelihood at each level of the hierarchy and so is a special case of the model-based clustering methodology to be discussed in § 5.

4. As the procedure gives a hierarchy from N clusters with one object down to one cluster with N objects, the user must choose *how many clusters to report* as scientifically important clusters. This choice can be assisted by examination of two statistics. The squared correlation coefficient, R^2, states the fraction of the total variance accounted for by a partition into g clusters,

$$R^2 = 1 - \frac{\sum_{j=1}^{g} W_j}{\sum_{i=1}^{N} |\,x_i - \bar{x}\,|^2}\,. \tag{3}$$

The squared semipartial correlation coefficient, $R_{\rm sp}^2$, measures the difference in the variance between the resulting cluster and the immediate parent clusters normalized by the total sample variance,

$$R_{\rm sp}^2 = \frac{W_{\rm M} - W_{\rm K} - W_{\rm L}}{\sum_{i=1}^{N} |\,x_i - \bar{x}\,|^2}\,. \tag{4}$$

TABLE 3

AVERAGE LINKAGE HIERARCHICAL CLUSTER ANALYSIS

Level	Merger	Members	R^2_{sp}	R^2
A. Six-dimensional Analysis				
8.....	10 + 15	506	0.08	0.65
7.....	14 + 137	93	0.00	0.65
6.....	8 + 7	599	0.10	0.55
5.....	9 + 266	188	0.00	0.55
4.....	5 + 26	190	0.00	0.55
3.....	6 + 12	606	0.01	0.54
2.....	3 + 4	796	0.53	0.01
1.....	2 + 616	797	0.00	0.00
B. Five-dimensional Analysis				
6.....	15 + 21	107	0.01	0.70
5.....	10 + 8	486	0.01	0.69
4.....	7 + 20	203	0.01	0.68
3.....	6 + 5	593	0.10	0.58
2.....	3 + 4	796	0.58	0.00
1.....	2 + 616	797	0.01	0.00

R^2 thus tells how much of the scatter is explained by a given level of clustering, and R^2_{sp} tells how much improvement is achieved between levels.

We emphasize again that there is no mathematically "best" choice, although extensive experience with problems in many fields has led to a preference for certain combinations (e.g., standardized variables and Ward's minimum variance criterion). We conducted extensive experiments with different choices.

4.2. *Results*

The last several levels of the clustering tree for the 797 GRBs using the six unit-free variables shown in Table 1, average linkage, and a Euclidean metric are shown in Figure 2 (*left panel*) with details in Table 3A. The action taken at each level is indicated in column (2) of Table 3, which may refer to a level higher in the tree that (for brevity) is not shown here. Two types of mergers are seen: the incorporation of "twigs" of one or a few GRBs into a large preexisting "trunk" (levels 1, 3, 4, 5, and 7) and the union of two substantial branches into a single larger trunk (levels 2, 6, and 8). The first type has little effect on the variance of the sample with $R^2_{sp} \leq 1\%$. The single GRB brought into the main trunk at level 1 is the distant outlier seen in several panels of Figure 1. The level 2 merger of clusters with 190 and 606 members is clearly the most important structure, accounting for roughly 53% of the variance of the entire sample. This is the bifurcation of the sample into two classes easily seen in Figure 1 and noted by K93 and others. The principal finding that is not immediately obvious from Figure 1 is the structure indicated at level 6. The main trunk of 599 bursts (plus a few twigs to be merged later) is divided into groups of 93 and 506 bursts. This division accounts for 10% of the total variance of the sample, which is indicated in both the R^2 and R^2_{sp} values.

We found that the twigs in the tree structure disappear if the peak flux P_{256} variable is omitted and the analysis is made in five-dimensional space (Fig. 2, *right panel*, and Table 3B). Here the largest cluster of 593 members is formed by the union of clusters with 107 and 486 bursts, again accounting for 10% of the sample variance. It is possible that P_{256} is a nuisance variable irrelevant to the basic astrophysics of GRBs, producing noisy "twigs" seen in Table 3A and Figure 2 (*left panel*).

We tested many variants of hierarchical clustering. We replaced average linkage hypothesis with complete linkage, single linkage and Ward's minimum variance criterion. The Ward's criterion computation, for example, gave three clusters with 468, 184, and 145 bursts. We clustered using nonparametric density estimation based on the 100 nearest neighbors and clustered using the principal components rather than the observed variables. Various methods were tried with both the observed T_{90} values and debiased T^d_{90} values with little effect on the results. All methods showed two strong clusters and the outlier, but in some cases the third cluster appeared only weakly.

To proceed further, we choose a single clustering structure for detailed study: the five-dimensional average linkage analysis (Table 3B) with three clusters: class I with 486 bursts, class II with 203 bursts, and class III with 107 bursts. Class IV, consisting of the single outlier, is ignored because of independent evidence that its properties are attributable to data of poor quality (§ 7). The membership of these clusters is given in Table 4, and four projections of the clusters

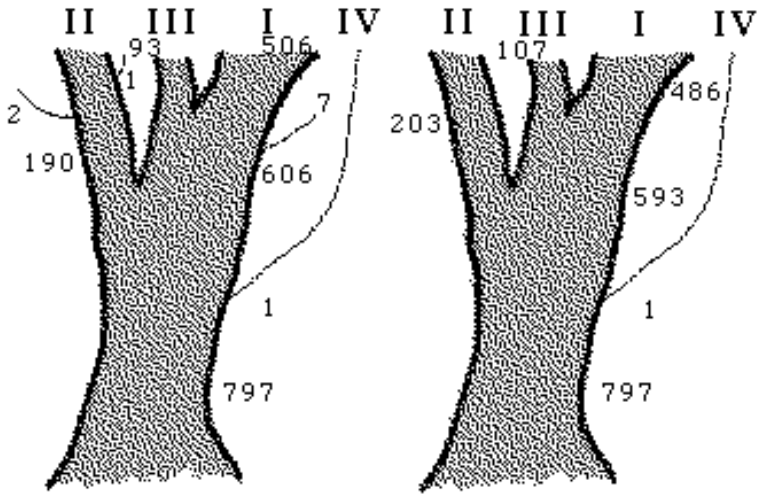

FIG. 2.—Diagram of the base of the dendrogram of average linkage hierarchical clustering procedures in six dimensions (*left panel*) and five dimensions (*right panel*). The number of members in each branch is indicated (see Table 3). Class IV is the spurious outlier.

TABLE 4

BURST CLASSES FROM AVERAGE LINKAGE CLUSTERING (3B TRIGGER NUMBER)

A. Class I (486 Bursts)														
107	472	816	1159	1533	1657	1982	2138	2304	2431	2551	2703	2877	3003	3109
109	473	820	1192	1540	1660	1989	2140	2306	2432	2560	2706	2889	3005	3110
110	503	824	1196	1541	1661	1993	2143	2309	2435	2569	2709	2890	3011	3115
111	540	825	1197	1546	1663	1997	2148	2310	2436	2570	2711	2891	3012	3119
114	543	829	1200	1551	1667	2018	2149	2311	2437	2581	2725	2894	3015	3120
121	548	840	1213	1552	1676	2019	2151	2315	2438	2586	2727	2897	3017	3128
130	549	841	1218	1558	1683	2037	2156	2316	2440	2589	2728	2898	3026	3129
133	559	867	1235	1559	1687	2041	2181	2321	2441	2593	2736	2900	3029	3130
143	563	869	1244	1561	1700	2044	2187	2324	2443	2600	2749	2901	3032	3131
148	577	907	1279	1567	1704	2047	2188	2325	2446	2603	2751	2913	3035	3132
160	591	927	1288	1574	1709	2053	2189	2328	2447	2606	2753	2916	3039	3134
171	594	938	1291	1578	1711	2061	2190	2329	2450	2608	2770	2919	3040	3135
204	606	946	1303	1579	1712	2067	2191	2340	2451	2610	2774	2922	3042	3136
211	630	973	1318	1580	1714	2069	2193	2344	2452	2611	2775	2924	3055	3138
214	647	999	1384	1586	1717	2070	2197	2345	2472	2619	2780	2925	3056	3142
219	658	1009	1385	1590	1730	2074	2202	2346	2476	2620	2790	2927	3057	3143
222	659	1025	1390	1601	1731	2077	2203	2362	2477	2628	2793	2929	3067	3153
223	660	1036	1396	1604	1733	2079	2204	2367	2482	2634	2797	2931	3070	3155
226	673	1039	1406	1606	1734	2080	2211	2371	2484	2636	2798	2932	3071	3156
235	676	1042	1419	1609	1740	2081	2213	2373	2495	2640	2799	2947	3072	3159
237	678	1046	1425	1611	1742	2083	2219	2375	2496	2660	2812	2948	3074	3164
249	685	1085	1432	1614	1806	2087	2228	2380	2500	2662	2815	2950	3075	3168
257	686	1086	1440	1623	1807	2090	2232	2383	2505	2663	2825	2953	3076	3171
288	692	1087	1446	1625	1815	2093	2233	2385	2508	2664	2831	2958	3080	3174
332	704	1122	1447	1626	1819	2101	2230	2387	2510	2665	2843	2961	3084	
351	717	1123	1449	1628	1830	2102	2244	2391	2511	2671	2852	2984	3085	
394	741	1126	1452	1642	1883	2106	2252	2392	2519	2681	2853	2985	3091	
398	761	1141	1456	1646	1885	2110	2253	2394	2522	2688	2855	2992	3093	
404	764	1148	1458	1651	1886	2111	2267	2405	2528	2691	2856	2993	3100	
408	773	1150	1467	1652	1922	2112	2276	2419	2530	2695	2857	2994	3101	
451	795	1152	1468	1653	1924	2119	2277	2428	2533	2696	2862	2996	3102	
467	803	1156	1472	1655	1956	2122	2287	2429	2537	2697	2863	2998	3103	
469	815	1157	1515	1656	1967	2133	2298	2430	2541	2700	2864	3001	3105	
B. Class II (203 Bursts)														
138	512	856	1154	1635	2003	2146	2291	2384	2523	2693	2846	2975	3094	
185	537	878	1211	1636	2040	2155	2312	2395	2529	2701	2849	2977	3113	
207	547	906	1223	1659	2043	2159	2317	2434	2536	2715	2851	2978	3114	
218	551	909	2389	1662	2049	2161	2320	2448	2564	2748	2860	2987	3118	
229	568	936	1308	1665	2068	2163	2326	2449	2583	2755	2873	2988	3121	
254	575	1051	1359	1680	2095	2167	2327	2454	2585	2788	2879	2995	3137	
289	603	1073	1404	1694	2099	2201	2330	2463	2597	2795	2892	3027	3152	
297	677	1076	1453	1719	2103	2205	2332	2464	2599	2800	2896	3037	3173	
298	729	1088	1461	1736	2115	2206	2352	2485	2614	2801	2910	3038		
432	788	1096	1463	1741	2117	2217	2353	2487	2615	2810	2918	3043		
444	799	1097	1481	1760	2125	2220	2357	2502	2623	2814	2933	3051		
474	809	1102	1518	1791	2126	2265	2360	2504	2632	2821	2952	3066		
480	830	1112	1553	1851	2132	2268	2365	2512	2649	2823	2964	3073		
486	836	1128	1566	1953	2142	2273	2372	2513	2679	2828	2966	3078		
491	845	1129	1588	1968	2145	2288	2377	2514	2690	2834	2973	3087		
C. Class III (107 Bursts)														
105	493	752	1120	1298	1492	1974	2207	2381	2458	2750	2880	3028	3160	
108	501	753	1125	1306	1634	2035	2230	2382	2460	2760	2917	3068	3166	
179	516	755	1145	1346	1637	2056	2254	2393	2515	2776	2944	3088	3167	
228	526	834	1153	1382	1664	2105	2283	2401	2633	2830	2945	3096		
373	555	914	1167	1416	1679	2114	2347	2423	2641	2844	2951	3127		
401	680	942	1190	1435	1693	2129	2349	2424	2677	2848	2980	3139		
414	690	974	1204	1439	1701	2133	2358	2442	2680	2850	2986	3144		
465	734*	1114	1221	1443	1747	2152	2368	2453	2719	2861	2990	3146		
D. Class IV (1 Burst)														
2757														

NOTE.—All bursts are placed into class III by Gaussian model-based clustering procedure except burst marked "*", which is placed into class II.

onto two-dimensional scatter plots are shown in Figure 3. These are frames from the "grand tour" movie of the five-dimensional data set provided by the XGobi software where each cluster is "brushed" with a different symbol. Note that, in general, there is no reason why classification structure should be most evident in projections parallel to the variable axes shown in Figure 1. It is more important that the clusters show cohesion in many projections of the data set. The grand tour of the 797 GRBs shows that classes I, II, and the outlier are very distinct in most projections. Class III often lies between classes I and II (e.g., Fig. 3, *top panels*), but in other projections is offset from the line between classes I and II (e.g., Fig. 3, *bottom panels*). It also appears elongated along some projections, while the larger classes I and II appear roughly hyperspherical.

This analyses described here provide considerable evidence for three major clusters and an outlier. But as some nonparametric clustering procedures did not find a strong third cluster, there is some worry that class III is simply a group of bursts with properties intermediate between

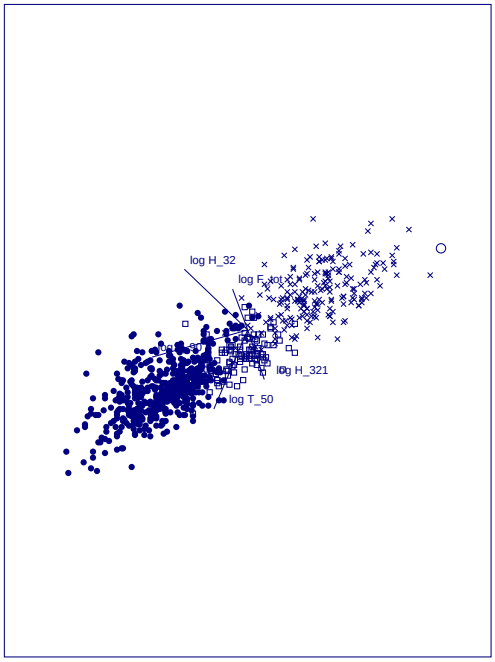

FIG. 3*a*

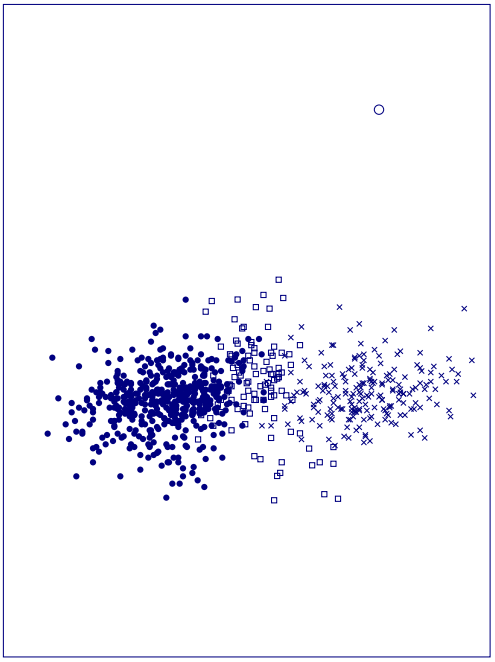

FIG. 3*b*

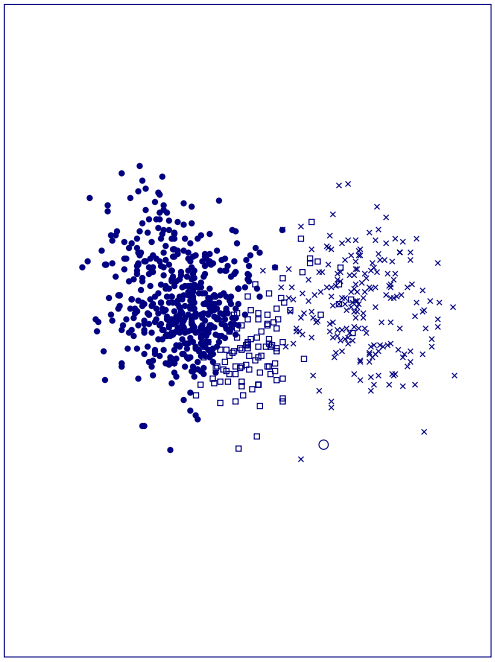

FIG. 3*c*

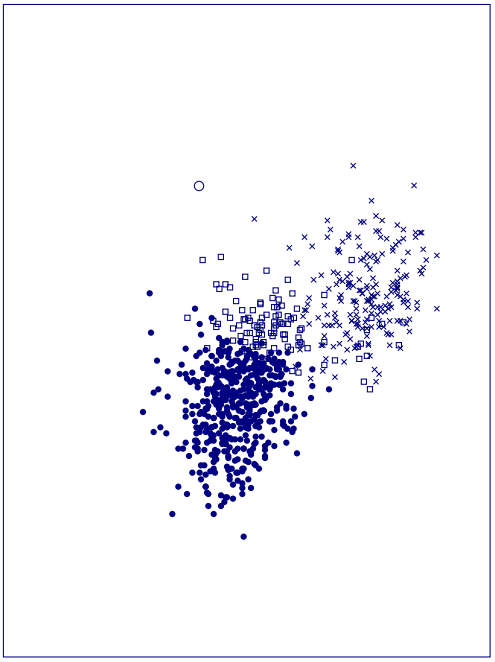

FIG. 3*d*

FIG. 3.—Four snapshots from the XGobi grand tour of the five-dimensional database with bursts brushed according to the nonparametric average linkage clustering results: class I (*filled circles*), class II (*x*), class III (*open squares*), and class IV (*open circles*). Panel (*a*) shows the projections of the five axes, which are suppressed in the other panels.

TABLE 5

MULTIVARIATE ANALYSIS OF VARIANCE STATISTICAL TESTS

CLASSES	WILKS' Λ*	PILLAI'S TRACE	HOTELLING'S TRACE	F	WILKS' Λ*		
					Number dof	Dendrogram dof	Probability
I, II, III.....	0.153	0.934	4.96	245.	10	1578	<0.0001
I, II	0.159	0.840	5.27	722.	5	683	<0.0001
I, III	0.515	0.485	0.94	111.	5	587	<0.0001
II, III	0.301	0.699	2.32	141.	5	304	<0.0001

classes I and II. While nonparametric hierarchical clustering methods cannot address this question, it can be investigated with parametric methods.

4.3. *Validation of the Classification*

Mathematically well-founded methods for evaluating the statistical significance of a proposed multivariate classification scheme are available under the assumption that the population is a multinormal mixture; that is, the objects of each class are drawn from multivariate Gaussians. All relationships between the variables must thus be linear (as in Table 2), although the relationships may differ between clusters. There is no requirement of sphericity, so that clusters may have shapes akin to pancakes or cigars with arbitrary orientations in multidimensional space. The separate existence of each of the postulated subpopulations can be tested using MANOVA.

The model can be expressed as follows (e.g., Johnson & Winchern 1992, pp. 246–584). For a p-dimensional data set of g clusters each with n_l members, the ith GRB in the jth cluster gives a p-dimensional vector

$$X_{ij} = \mu + \tau_j + \epsilon_{ij} , \tag{5}$$

where μ is the overall population mean, τ_j is the offset of the jth cluster mean from μ, and ϵ_{ij} are independent normal variables with zero mean representing the scatter of individual points about the mean. We test the null hypothesis

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0 \tag{6}$$

that the cluster means are not offset from each other. We construct two matrices of sums of squares and cross-products as follows:

$$B = \sum_{l=1}^{g} n_l (\bar{x}_l - \bar{x})(\bar{x}_l - \bar{x})^T$$

$$W = \sum_{l=1}^{g} \sum_{j=1}^{n_l} (\bar{x}_{lj} - \bar{x}_l)(\bar{x}_{lj} - \bar{x}_l)^T , \tag{7}$$

where $()^T$ is the vector transpose. Three test statistics have been proposed to test the null hypothesis (e.g., SAS Institute, Inc., 1989, pp. 17–529):

Wilks' lambda $\Lambda^* = \det(W)/\det(B + W)$,

Pillai's trace $V = \text{trace}\,[B(B + W)^{-1}]$,

and

Hotelling-Lawley's trace $U = \text{trace}\,(W^{-1}B)$.

The distributions of these statistics have been determined mathematically. For example, for large $N = \sum_{l=1}^{g} n_l$, the quantity $-[N - 1 - (p + g)/2] \ln \Lambda^*$ has approximately a χ^2 distribution with $p(g - 1)$ degrees of freedom (Wilks 1932; Bartlett 1938). More generally, the distributions are related to the noncentral F-distribution. For the two-sample case, Hotelling-Lawley's trace is commonly known as the Mahalanobis D^2 statistic. One can thus accept or reject the null hypothesis that the clusters have the same mean location at a chosen level of statistical significance.

The results of our MANOVA calculations are summarized in Table 5. The columns give the values of the three MANOVA statistics followed by details for the Wilks' Λ*: the corresponding value of the F-statistic, the numerator and denominator degrees of freedom for that F-value, and the resulting P-value. Details for Pillai's and Hotelling-Lawley's traces are omitted but give similar results in all cases. The first row tests the null hypothesis that the classes I, II, and III have the same mean, the second row tests the equality of classes I and II, and so forth. The F-values are very high in all cases, indicating that the clusters are different with extremely high statistical significance ($P \ll 10^{-4}$).[12] This is a clear *quantitative* demonstration that at least two clusters exist among GRBs (for a univariate test, see Ashman, Bird, & Zepf 1994), which was *qualitatively* reported by Delazay et al. (1992) and K93. The other rows in Table 5 test the hypotheses that each proposed class has the same mean as each other class.

One problem with these MANOVA tests is that they are conditional on the classification that has been found using the clustering algorithm. Because the clustering algorithm is constructed to find groups that are different from one another, tests such as these tend to be biased toward finding structure, perhaps where none exists. Although the MANOVA results seem to indicate very strong evidence of structure, they cannot be taken as definitive for this reason. Tests arising from model-based clustering can overcome this problem, as is discussed in the following section.

5. MODEL-BASED MAXIMUM LIKELIHOOD CLUSTERING ANALYSIS

5.1. *Methodological Background*

In the previous section, we conducted a hierarchical clustering analysis without making assumptions regarding the shapes of the clusters but needed the parametric assumption of normality to estimate the statistical significance of the resulting classification. It is reasonable to conduct the entire analysis, both clustering and validation, within a model-based framework. We report here an analysis of this type again, assuming that the GRB population consists of a mixture of multivariate Gaussian classes. Early development of this model for clustering is discussed in McLachlan & Basford (1988); we use more recent developments here. First, an initial classification for each possible number of clusters is found via agglomerative hierarchical clustering

[12] Note that it is not meaningful to quote probabilities like $P = 10^{-8}$, as the tails of the distribution are poorly determined unless the sample size is extremely large.

(Murtagh & Raftery 1984; Banfield & Raftery 1993; Fraley 1998). Next, the Expectation-Maximization (EM) algorithm is used to refine partitions obtained from hierarchical clustering (Celeux & Govaert 1995; Dasgupta & Raftery 1998). Finally, the Bayesian Information Criterion (BIC) is used to select the "best" partition among those associated with different numbers of clusters (Dasgupta & Raftery 1998).

In the model considered here, the p-dimensional observations x_i are drawn from g multinormal groups, each of which is characterized by a vector of parameters θ_k for $k = 1, \ldots, g$. Our goals are to determine the number of GRB types, g; to determine the cluster assignment of each burst; and to estimate the mean μ_k and covariance matrix Σ_k for each cluster. Following Fraley (1998), the density of an observation x_i from the kth subpopulation is expressed as follows:

$$f_k(\boldsymbol{x}_i \mid \theta_k) \sim \mathrm{MVN}\,(\mu_k, \Sigma_k)\,, \quad k = 1, \ldots, g\,, \tag{9}$$

where MVN means multivariate normal. We estimate the parameters using the principle of maximum likelihood. In the hierarchical clustering phase, we use the classification likelihood

$$L_C(\theta, \gamma \mid \boldsymbol{x}) = \prod_{i=1}^{N} f_{\gamma_i}(\boldsymbol{x}_i \mid \theta_{\gamma_i})\,, \tag{10}$$

where $\boldsymbol{x} = \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ represents the observations and $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_N\}$ is the cluster assignment: $\gamma_i = k$ when x_i comes from the kth group. Equivalently,

$$L_C(\mu_1, \ldots, \mu_g\,; \Sigma_1, \ldots, \Sigma_g \mid \boldsymbol{x}) = \prod_{k=1}^{g} \prod_{i \in I_k} \times (2\pi)^{p/2} \mid \Sigma_k \mid^{-1/2} \exp\left[-\tfrac{1}{2}(\boldsymbol{x}_i - \mu_k)^T \Sigma_k^{-1} (\boldsymbol{x}_i - \mu_k)\right]\,, \tag{11}$$

where $I_k = \{i : \gamma_i = k\}$ is the set of indices corresponding to observations belonging to the kth group.

The method used here for maximizing the likelihood (Fraley 1998) and implemented in the MCLUST code involves parameterization of the Σ_k matrices in terms of their eigenvectors and eigenvalues (analogous to a principal components analysis) and iterative relocation of the clusters using the EM algorithm. The EM algorithm (Dempster, Laird, & Rubin 1977), one of the most successful methods in modern statistics, is a procedure for iteratively maximizing likelihoods in a wide variety of circumstances. For example, the Lucy-Richardson algorithm in astronomical image restoration is the EM algorithm. In the present application, we apply EM to the mixture likelihood

$$L_M(\theta, \gamma \mid \boldsymbol{x}) = \prod_{i=1}^{N} \sum_{k=1}^{g} \tau_k f_k(\boldsymbol{x}_i \mid \theta_k)\,, \quad \sum_{k=1}^{g} \tau_k = 1\,, \tag{12}$$

where τ_k are mixing probabilities associated with each group. For a given number g of components in the mixture, we use EM to estimate the conditional probability that observation x_i belongs to the kth group for each i and selected k via maximum likelihood. Although the computational procedure has some limitations (e.g., convergence of the EM iterations is not guaranteed; clusters cannot be extremely small), it is generally efficient and effective for Gaussian clustering problems when started from reasonable partitions such as those produced by hierarchical agglomeration.

We use the Bayes factor to assess the evidence for a given number of clusters against a different number of clusters. The Bayes factor, defined in the context of Bayesian statistics, is the posterior odds for one model against the other when the prior odds are equal to one (i.e., when one does not favor one model over the other a priori). Kass & Raftery (1995) review the use of Bayes factors in adjudicating between competing scientific hypotheses on the basis of data. The Bayes factor for a model M_2 against a competing model M_1 (say, for three vs. two classes of GRBs) is defined as

$$\text{Bayes factor} = \frac{p(\boldsymbol{x} \mid M_2)}{p(\boldsymbol{x} \mid M_1)}\,, \tag{13}$$

where $p(\boldsymbol{x} \mid M_j)$ for $j = 1, 2$ is obtained by integrating the likelihood times the prior density over the parameters of the model. It can be viewed as a likelihood ratio, but it differs from the usual frequentist ratio that underlies the likelihood ratio test in that the latter is obtained by maximizing (rather than integrating) the likelihood over the model parameters.

Twice the logarithm of the Bayes factor can be approximated by the BIC (Schwarz 1978),

$$\mathrm{BIC} = 2(l_1 - l_2) - (m_1 - m_2) \log N\,, \tag{14}$$

where l_1 is the likelihood and m_1 is the number of parameters for one mixture model and similarly for l_2 and m_2. The BIC measures the balance between the improvement in the likelihood and the number of model parameters needed to achieve that likelihood. While the absolute value of the BIC is not informative, differences between the BIC values for two competing models provide estimates of the evidence in the data for one model against another. Conventionally, BIC differences less than 2 represent weak evidence, differences between 2 and 6 represent positive evidence, 6–10 strong evidence, and greater than 10 very strong evidence (Jeffreys 1961, Appendix B; Kass & Raftery 1995). The use of the BIC in choosing clusters in a mixture or clustering model is discussed by Roeder & Wasserman (1997) and Dasgupta & Raftery (1998).

Bayes factors and BIC have the advantage that they can be used to assess the evidence *for* a null hypothesis, unlike standard significance tests that can only reject a null hypothesis. They can also easily be used to compare nonnested models, again unlike standard significance tests that require competing models to be nested.

5.2. *Results and Validation*

To reduce the dimensionality of the problem and the complexity of the calculation, we eliminated the highly redundant T_{50} and H_{32} variables (see Fig. 1) and considered only the three variables T_{90}, F_{tot}, and H_{321} for the sample of 797 BATSE GRBs. The MCLUST model-based clustering procedure described above was run for trials of $g = 1, 2, \ldots,$ 24 groups. The resulting values of BIC(g) are plotted in Figure 4. The maximum BIC is achieved for three classes. Most importantly, the BIC value for $g = 3$ is $\simeq 68$ units above that for $g = 2$. This corresponds to strong evidence indeed for the presence of three groups rather than two. This result strongly confirms the analysis in § 4 indicating the existence of three clusters and this time is free of the problem that the MANOVA tests are conditional on the estimated partition. The result here takes account of the fact that the partition is not known in advance.

We have also calculated the BIC for $g = 1, \ldots, 9$ with various constraints on the covariance matrix Σ such as hypersphericity and uniformly shaped ellipsoids. Spherical clusters give poor fits. Uniform ellipsoids give good fits with

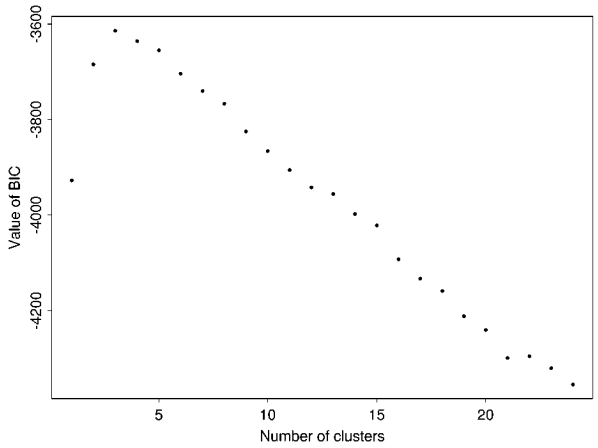

FIG. 4.—BIC value vs. assumed number of clusters for the model-based Gaussian mixture maximum likelihood clustering analysis.

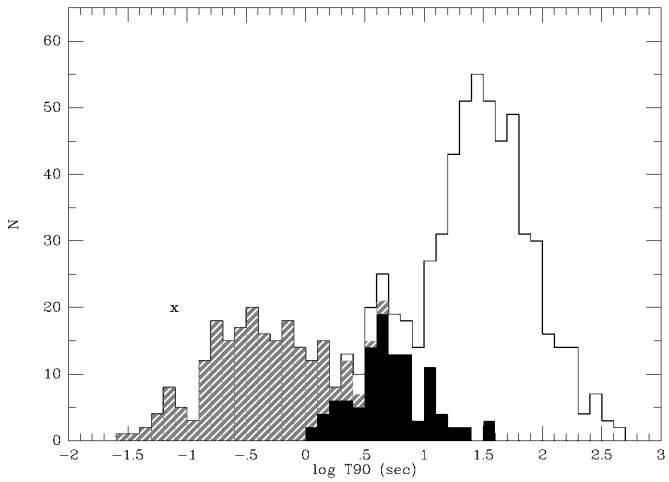

FIG. 5.—Burst types and T_{90} burst duration based on five-dimensional average linkage clustering procedure. Class I = white, class II = gray, and class III = black. The class IV outlier is shown as an "x".

4 and 8 clusters. But in all cases, the maximum likelihood assuming 2 clusters is much lower than the likelihood of ≥ 3 clusters.

The cluster assignment vector γ for the $g = 3$ model with unconstrained Σ is given in Table 4. Over 85% of the assignments are the same as those obtained from the nonparametric hierarchical clustering procedure in § 4 so that we note only differences between the two clustering results using asterisk and dagger markings. All but one of the 96 assignment differences move bursts from classes I and II into class III. The close agreement between the cluster assignments in the two methods reinforces confidences in the conclusions from both of them.

6. CLUSTER PROPERTIES

We can now examine the properties of GRBs within each cluster with reasonable confidence that the populations are distinct from each other but internally homogeneous. These properties become inputs to astrophysical theories seeking to explain GRB bulk properties. Table 6A lists the means and standard deviations of the principal variables for each cluster based on both the nonparametric and model-based clustering procedures. The two methods give very similar results. The three types are well separated in the burst duration variables: cluster I bursts have the longest durations, around 10–20 seconds, cluster II bursts have the shortest durations, below 1 second, and cluster III bursts have intermediate durations, around 2–5 s. This is shown clearly in Figure 5, which projects each class onto the univariate T_{90} axis. Cluster III bursts are also intermediate in their fluences, although their fluence distribution overlaps that of

TABLE 6

CLASS PROPERTIES

VARIABLE	METHOD[a]	CLASS I	CLASS II	CLASS III
A. Means and Standard Deviations				
log T_{50}	NP	1.13 ± 0.44	-0.80 ± 0.41	0.33 ± 0.27
log T_{90}	NP	1.55 ± 0.40	-0.42 ± 0.44	0.71 ± 0.32
	MB	1.22 ± 0.39	-0.91 ± 0.35	0.29 ± 0.41
log F_{tot}	NP	-5.21 ± 0.59	-6.37 ± 0.57	-6.11 ± 0.37
	MB	-5.13 ± 0.58	-6.46 ± 0.54	-5.93 ± 0.47
log H_{321}	NP	0.19 ± 0.27	0.51 ± 0.27	0.08 ± 0.40
	MB	0.21 ± 0.26	0.52 ± 0.28	0.16 ± 0.40
log H_{32}	NP	0.43 ± 0.23	0.70 ± 0.26	0.35 ± 0.39
B. Isotropy				
$\langle \cos \theta \rangle$	NP	0.015	−0.041	0.010
$\langle \sin^2 b - 1/3 \rangle$	NP	−0.012	−0.025	0.028
Rayleigh-Watson	NP	0.39	1.28	1.80
Bingham	NP	2.02	7.32	1.95
C. Summary				
Number	NP	486	203	107
	MB	426	170	201
Duration	...	long	short	intermediate
Fluence	...	bright	faint	intermediate
Spectrum	...	intermediate	hard	soft

[a] NP = nonparametric clustering analysis in five dimensions (§ 4); MB = model-based clustering analysis in three dimensions (§ 5).

TABLE 7

CORRELATION COEFFICIENTS WITHIN CLASSES

Variable	log T_{50}	log T_{90}	log F_{tot}	log H_{321}	log H_{32}
A. Class I: Nonparametric Clustering ($N = 486$) ($\|r\| > 0.15$ Corresponds to $P < 0.001$ Significance Level)					
log T_{50}	1.00	...	...	...	...
log T_{90}	0.88	1.00	...	...	...
log F_{tot}	0.10	0.22	1.00	...	...
log H_{321}	−0.11	−0.08	0.39	1.00	...
log H_{32}	−0.11	−0.08	0.38	0.97	1.00
B. Class I: Model-based Clustering ($N = 426$) ($\|r\| > 0.16$ Corresponds to $P < 0.001$ Significance Level)					
log T_{50}	1.00	...	...	...	...
log T_{90}	N/A	N/A	...	...	...
log F_{tot}	N/A	0.01	1.00	...	...
log H_{321}	N/A	−0.01	0.06	1.00	...
log H_{32}	N/A	N/A	N/A	N/A	N/A
C. Class II: Nonparametric Clustering ($N = 203$) ($\|r\| > 0.23$ Corresponds to $P < 0.001$ Significance Level)					
log T_{50}	1.00	...	...	...	...
log T_{90}	0.89	1.00	...	...	...
log F_{tot}	−0.05	−0.02	1.00	...	...
log H_{321}	−0.00	−0.08	−0.21	1.00	...
log H_{32}	−0.00	−0.08	−0.26	0.96	1.00
D. Class II: Model-based Clustering ($N = 170$) ($\|r\| > 0.25$ Corresponds to $P < 0.001$ Significance Level)					
log T_{50}	1.00	...	...	...	...
log T_{90}	N/A	N/A	...	...	...
log F_{tot}	N/A	−0.03	1.00	...	...
log H_{321}	N/A	−0.05	0.02	1.00	...
log H_{32}	N/A	N/A	N/A	N/A	N/A
E. Class III: Nonparametric Clustering ($N = 107$) ($\|r\| > 0.32$ Corresponds to $P < 0.001$ Significance Level)					
log T_{50}	1.00	...	...	...	...
log T_{90}	0.86	1.00	...	...	...
log F_{tot}	0.02	0.06	1.00	...	...
log H_{321}	−0.24	−0.34	−0.16	1.00	...
log H_{32}	−0.22	−0.32	−0.22	0.95	1.00
F. Class III: Model-based Clustering ($N = 201$) ($\|r\| > 0.23$ Corresponds to $P < 0.001$ Significance Level)					
log T_{50}	1.00	...	...	...	...
log T_{90}	N/A	N/A	...	...	...
log F_{tot}	N/A	0.03	1.00	...	...
log H_{321}	N/A	−0.01	0.07	1.00	...
log H_{32}	N/A	N/A	N/A	N/A	N/A

the fainter class II bursts. The hardness ratios of all three clusters overlap considerably; class III bursts have similar or slightly softer spectra than class I bursts. We can thus classify the types in the three principal dimensions duration/fluence/spectrum (Table 6C): class I is long/bright/soft, class II is short/faint/hard, and class III is intermediate/intermediate/soft.

A major constraint for the astrophysical interpretation of GRBs has been the remarkable isotropy of their spatial distribution in the celestial sphere. It is possible that, while the bulk of GRBs are isotropic and have an inferred extragalactic origin, some class of GRBs have significant anisotropy that would reflect a Galactic origin (see Lamb 1995). We apply four statistical tests for isotropy discussed by Briggs (1993) and applied by Briggs et al. (1996) to various subsamples of the BATSE 3b catalog of GRBs. The statistics are: $\langle \cos\theta \rangle$, where θ is the angle between a burst and the Galactic center; $\langle \sin^2 b - \frac{1}{3} \rangle$, where b is the Galactic latitude; Rayleigh-Watson $\mathscr{W}$; and Bingham $\mathscr{B}$. $\langle \cos\theta \rangle$ tests the dipole moment around the Galactic center, $\langle \sin^2 b - \frac{1}{3} \rangle$ tests the quadrapole moment with respect to the Galactic plane, $\mathscr{W}$ tests the dipole moment around any point in the celestial sphere, and $\mathscr{B}$ tests the quadrapole moment around any plane or two poles. The expected values for the four statistics assuming random isotropic distribution on the sphere are 0, 0, 3, and 5, respectively. The asymptotic distributions of these statistics are known.

Table 6B shows the results of this analysis for clusters I–III, which were kindly calculated for us by Michael Briggs. No deviations from isotropy are found. The $\langle \cos\theta \rangle$ and $\langle \sin^2 b - \frac{1}{3} \rangle$ values lie within one standard deviation of the expected value for a random distribution. The $\mathscr{W}$ and $\mathscr{B}$ values must be larger than the expected value to indicate anisotropy. The only such case, class II with $\mathscr{B}$ 7.32, has a

deviation with very low significance (probability < 0.2). We thus do not confirm Belli's (1997) report of significant differences in spatial distributions of burst classes I and II, although we did not specifically test the Galactic latitude distribution.

In principle, the relative populations of the three classes may be an important constraint on astrophysical theory. We find that class I contains more than half of the bursts with the remainder divided between class II and class III (Table 6C). But we do not believe our analysis gives a precise census for two reasons. First, the exact assignments of individual bursts to clusters depend on the detailed assumptions of the clustering algorithms. For example, class II is larger than class III in the five-dimensional nonparametric procedure but is smaller in the three-dimensional model-based procedure. Second, the numbers of weaker bursts in classes II and III are strongly dependent on the details of the BATSE instrument's burst triggering process that may produce a complicated truncation bias for fainter bursts.

We look for structure within each of the clusters by computing correlation coefficients similar to those in § 3 for the entire sample. Results are given in Table 7. Here we see a systematic difference between the two clustering methodologies: nonparametric average linkage clustering tends to give stronger correlations between the variables than the model-based clustering. For example, in the nonparametric analysis we find significant positive correlations between total fluence and hardness in classes I and II and a correlation between duration and fluence in class II. However, we attach more credence to the model-based results for this purpose than the average linkage results because the former method is specifically designed to provide optimal estimates of the within-group covariances given the clustering model. The model-based results do not give strong evidence for any nonzero correlations between variables, suggesting that the partition into three clusters explains all of the correlation between variables in the full data set.

7. DISCUSSION

We thus find, using multivariate clustering and validation methods with different mathematical underpinnings, that three classes of GRBs are present in our large subset of the BATSE 3b catalog. Most of the structure can be found using three fundamental burst properties, duration/fluence/spectrum. The class properties and relation to previous research can be briefly summarized as follows:

1. *Class I*—These long/bright/intermediate bursts correspond to the well-known populous long-soft class of K93 and others. Within this group, we do not confirm a hardness-duration correlation reported by Dezalay et al. (1996) and Horack & Hakkila (1997).

2. *Class II*—This short/faint/hard group corresponds to the short-hard burst type of K93 and others. Fluence-duration and fluence-hardness correlations may tentatively be present within the class. Note that while the mean location of this type is consistent in the two clustering schemes, its size and population (e.g., $\frac{1}{2}$ or $\frac{1}{4}$ that of class I) differs between clustering algorithms.

3. *Class III*—The discovery of this group with intermediate/intermediate/soft properties is the principal result of this study. The group is easily distinguished in the projections of Figure 3 but can also be discerned in some panels of Figure 1. For example, it lies between classes I and II in the $T_{50} - H_{32}$, $T_{90} - F_{\rm tot}$, and $T_{90} - H_{321}$ scatter plots. In the univariate T_{90} distribution shown in Figure 5, class III accounts for most, but not all, of the bursts in the small peak around $2 < T_{90} < 5$ s between the major short and long duration peaks. It is possible that our class III is related to the class of no-high-energy (NHE) bursts and peaks discussed by Pendleton et al. (1997). These bursts have unusually weak F_4 emission, soft 50–300 keV spectra, and low $F_{\rm tot}$. However, the NHE class does not appear to exhibit a clear duration segregation from other bursts as we find for class III. Class III does not appear to be the third cluster found by Baumgart (1994; see his Table 3), but the high dimensionality of his analysis prevents a simple comparison with our low dimensionality study.

4. *Outlier*—BATSE trigger event 2757, burst 3B 940114, is the outlier in the nonparametric analysis of § 4 and is clearly visible in many projections in Figures 1 and 3.[13] After this study was complete, the BATSE group reanalyzed the satellite data for this burst and found the published data were incorrect because of a processing error (C. A. Meegan 1998, private communication). The unusual properties of this burst are thus illusory.

The multivariate analysis described here is not comprehensive and may not have uncovered all of the structure in the BATSE 3b catalog of bulk GRB properties. Our reduction of dimensionality may have been too severe, omitting, for example, the potentially important F_4 as a distinct variable (Pendleton et al. 1997; Bagoly et al. 1998). Many methodological options were not exercised. For example, it would be valuable to repeatedly apply the k-means partitioning algorithm to the database under the assumption that three clusters are present (see Murtagh 1992 for an astronomical application of this method), check for skewness or kurtosis in the clusters, and undertake an oblique decision tree analysis to give analytical formulation to hyperplanes separating the clusters (see White 1997).[14]

However, the efforts described here are far more capable of finding and quantifying clustering in the database than most previous analyses (§ 1). Most studies have been based on qualitative rather than quantitative procedures for identifying structures and provide no statistical validation of their claims.[15] It is thus not surprising that we uncovered structure missed by previous researchers. In particular, our confidence in the presence of a third cluster, class III, is strong. Two completely independent mathematical procedures (§ 4 and § 5) found very similar structure, each validated with high statistical confidence.

It is possible that the clustering reported here is indeed present in the database but has an instrumental rather than astrophysical origin. We have investigated two plausible manifestations of such problems. First, some properties,

[13] The model-based analysis of § 5 cannot locate clusters with very few members and assigned this event to class II. An extension of model-based clustering that models outliers as Poisson noise can do this (see Banfield & Raftery 1993; Dasgupta & Raftery 1998), but it does not seem necessary in this application.

[14] Codes for these and many other multivariate techniques are publicly available through the Web metasite StatCodes at www.astro.psu.edu/statcodes.

[15] After this work was completed, Horvath (1998) independently reported the existence of class III. His result was based on a Gaussian mixture model similar to that presented in § 5 but using only the univariate distribution of T_{90} (Fig. 5).

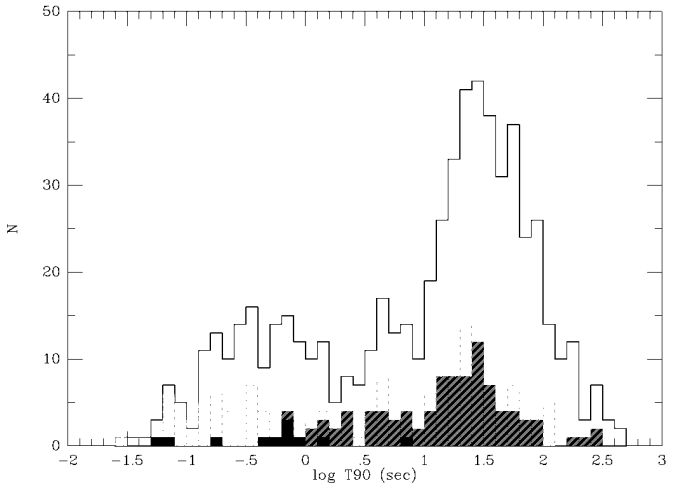

FIG. 6.—BATSE trigger and T_{90} burst duration for 612 bursts. 64 ms trigger only = light gray; 256 ms trigger only = black; 1024 ms trigger only = dark gray. The remainder of the bursts were triggered at two or all three timescales.

such as burst duration, may be systematically biased for weak bursts compared with strong bursts. However, our hierarchical clustering using the debiased durations T_{90}^{d} showed the same structure as seen using T_{90} (§ 4.2). Second, the burst triggering mechanism of the BATSE instrument is activated on three timescales (64, 256, and 1024 ms), which conceivably may produce the trimodal distribution seen in Figure 5. We show in Figure 6 the T_{90} distribution for the 612 bursts with published triggering information. It shows the expected effect that bursts triggered at 64 ms tend to have shorted durations than those triggered at 1024 ms. But neither this diagram nor XGobi grand tours showing more trigger combinations in multivariate space show any correlation between trigger timescales and the class III group of bursts around $T_{90} \simeq 3$–10 s. We conclude that, to the extent we can test with current data sets, there is no evidence that the three classes have an instrumental origin.

We conclude that the BATSE 3b catalog shows three statistically significant types of bursts (duration/fluence/spectrum): class I GRBs are long/bright/soft, class II GRBs are short/faint/hard, and class III GRBs are intermediate/intermediate/soft. These types are likely to be astrophysically real (rather than instrumental artifacts) and their existence should be considered an important input into astrophysical theories for GRBs. For example, the three types may reflect different types of external environments and internal shocks in relativistic fireball models (Mészáros & Rees 1993; Panaitescu & Mészáros 1998). Note that statistical anlaysis is unable to determine whether burst types represent fundamentally different astrophysical processes or distinct conditions within a single astrophysical model.

Our results can be confirmed and extended in two fashions. First, the analysis described here can be validated with several hundred more bursts collected by BATSE since the 1994 September cutoff in the database used here. Second, following Baumgart (1994), the dimensionality of the problem can be enlarged to include detailed characteristics of the burst temporal behaviors. Burst smoothness versus peakiness, characteristic wavelet scales, spectral evolution, and other parameters can be included. With this enlarged database, one can perform both an unsupervised exploratory cluster analysis similar to that described here and MANOVA-type analyses that assume the existence of the three groups to determine whether the clusters have distinctive temporal properties.

This work was supported by NASA grant NAGW-2120 and NSF grant DMS 96-26189 funding astrostatistical research at Penn State and ONR grants N-00014-96-1-0192 and N-00014-96-1-0330 funding statistical research at Washington. S. M. would like to thank the Departments of Astronomy & Astrophysics and Statistics at Penn State for hospitality while this work was underway. We are very grateful to J. Norris (NASA-GSFC) for providing debiased duration values, M. Briggs (NASA-MSFC) for performing isotropy calculations, and C. Meegan for investigating the outlier burst. We also thank Drs. Briggs, Norris, and P. Mészáros (Penn State) for sharing many valuable insights into GRB astronomy. S. M. was supported by NSF grant PHY9503084 at NWU.

REFERENCES

Ashman, K., Bird, C. M., & Zepf, S. E. 1994, AJ, 108, 2348
Babu, G. J., & Feigelson, E. D. 1996, Astrostatistics (London: Chapman & Hall)
Bagoly, Z., Mészáros, A., Horváth, I., Balázs, L. G., & Mészáros, P. 1998, ApJ, 498, 342
Banfield, J. D., & Raftery, A. E. 1993, Biometrics, 49, 803
Bartlett, M. S. 1938, Proc. Cambridge Philos. Soc., 34, 33
Baumgart, C. W. 1994, in SPIE Proc. 2243, Applications of Artificial Neural Networks V, ed. S. K. Rogers & D. W. Ruck (Bellingham: SPIE), 552
Belli, B. M. 1997, ApJ, 479, L31
Beyer, W. H. 1968, Handbook of Tables for Probability and Statistics (2d ed.; Boca Raton: CRC Press)
Bhat, P. H., Fishman, G. J., Meegan, C. A., Wilson, R. B., Kouveliotou, C., Paciesas, W. S., Pendleton, G. N., & Schaefer, B. E. 1994, ApJ, 426, 604
Briggs, M. S. 1993, ApJ, 407, 126
Briggs, M. S., et al. 1996, ApJ, 459, 40
Celeux, G., & Govaert, G. 1995, Pattern Recognition, 28, 781
Dasgupta, A., & Raftery, A. E. 1998, J. Am. Stat. Assoc., 93, 294
Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, J. Royal Stat. Soc. B., 31, 1
Dezalay, J.-P., Barat, C., Talon, R., Sunyaev, R., Terekhov, O., & Kuznetsov, A. 1992, in AIP Conf. Proc. 265, Gamma-Ray Bursts, ed. W. S. Paciesas & G. J. Fishman (New York: AIP), 304
Dezalay, J.-P., Lestrade, J. P., Barat, C., Talon, R., Sunyaev, R., Terekhov, O., & Kuznetsov, A. 1996, ApJ, 471, L27
Feigelson, E. D., & Babu, G. J. 1998, in IAU 179, New Horizons from Multiwavelength Sky Surveys, ed. B. J. McLean et al. (Dordrecht: Kluwer), 363
Fenimore, E. E., in't Zand, J. J., Norris, J. P., Bonnell, J. T., & Nemiroff, R. J. 1995, ApJ, 448, L101
Fishman, G. J., & Meegan, C. A. 1995, ARA&A, 33, 415
Fraley, C. 1998, SIAM J. Sci. Stat. Comput., in press
Hartigan, J. A. 1975, Clustering Algorithms (New York: Wiley)
Horack, J. M., & Hakkila, J. 1997, ApJ, 479, 371
Horvath, I. 1998, preprint, astro-ph/9803077
Jain, A. K., & Dubes, R. C. 1988, Algorithms for Clustering Data (Englewood Cliffs: Prentice Hall)
Jeffreys, H. 1961, Theory of Probability (3d ed.; Oxford: Clarendon)
Jobson, J. D. 1992, Applied Multivariate Data Analysis, 2 vols. (New York: Springer)
Johnson, R. A., & Wichern, D. W. 1992, Applied Multivariate Statistical Analysis (3d ed.; Englewood Cliffs: Prentice Hall)
Kass, R. E., & Raftery, A. E. 1995, J. Am. Stat. Assn., 90, 773
Kaufman, L., & Rousseeuw, P. J. 1990, Finding Groups in Data (New York: Wiley)
Katz, J. I., & Canel, L. M. 1996, ApJ, 471, 915
Kouveliotou, C., Meegan, C. A., Fishman, G. J., Bhat, N. P., Briggs, M. S., Koshut, T. M., Paciesas, W. S., & Pendleton, G. N. 1993, ApJ, 413, L101 (K93)
Lamb, D. Q. 1995, PASP, 107, 1152
Lamb, D. Q., Graziani, C., & Smith, I. A. 1993, ApJ, 413, L11

Lee, T. T., & Petrosian, V. 1997, ApJ, 474, L37
Lestrade, J. P. 1994, ApJ, 429, L5
Mallozzi, R. S., Paciesas, W. S., Pendleton, G. N., Briggs, M. S., Preece, R. D., Meegan, C. A., & Fishman, G. J. 1995, ApJ, 454, 597
McLachlan, G., & Basford, K. 1988, Mixture Models: Inference and Applications to Clustering (New York: Marcel Dekker)
Meegan, C. A., et al. 1996, ApJS, 106, 65
Mészáros, P., & Rees, M. J. 1993, ApJ, 405, 278
Murtagh, F. 1992, in Statistical Challenges in Modern Astronomy, ed. E. D. Feigelson & G. J. Babu (New York: Springer), 449
Murtagh, F., & Heck, A. 1987, Multivariate Data Analysis (Dordrecht: Kluwer)
Murtagh, F., & Raftery, A. E. 1984, Pattern Recognition, 17, 479
Norris, J. P., Bonnell, J. T., Nimiroff, R. J., Scargle, J. D., Kouveliotou, C., Paciesas, W. S., Meegan, C. A., & Fishman, G. J. 1995, ApJ, 439, 542
Norris, J. P., Hertz, P., Wood, K. S., & Kouveliotou, C. 1991, ApJ, 366, 240
Panaitescu, A., & Mészáros, P. 1998, ApJ, 492, 683
Pendleton, G. N., et al. 1997, ApJ, 489, 175
Pendleton, G. N., Paciesas, W. S., Briggs, M. S., Koshut, T. M., Fishman, G. J., Meegan, C. A., Wilson, R. B., Harmon, A. B., & Kouveliotou, C. 1994, ApJ, 431, 416
Roeder, K., & Wasserman, L. 1997, J. Am. Stat. Assn., 92, 894
SAS Institute, Inc. 1989, SAS/STAT User's Guide, Version 6, 2 vols., (4th ed.; Cary: SAS Institute, Inc.)
Schwarz, G. 1978, Ann. Stat., 6, 461
Swayne, D. F., Cook, D., & Buja, A. 1991, User's Manual for XGobi, a Dynamic Graphics Program for Data Analysis Implemented in the X Windows System, BellCore Tech. Memo.
Ward, J. H. 1963, J. Am. Stat. Assn., 58, 236
White, R. L. 1997, in Statistical Challenges in Modern Astronomy II, ed. G. J. Babu & E. D. Feigelson (New York: Springer), 135
Wilks, S. S. 1932, Biometrika, 24, 471