

## CLASSIFYING X-RAY BINARIES: A PROBABILISTIC APPROACH

GIRI GOPALAN<sup>1</sup>, SAEQA DIL VRTILEK<sup>2</sup>, AND LUKE BORNN<sup>1,3</sup>

<sup>1</sup>Harvard University Department of Statistics, One Oxford Street, Cambridge, MA 02138, USA

<sup>2</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

<sup>3</sup>Department of Statistics and Actuarial Science, Simon Fraser University, 888 University Drive, Burnaby, BC V5A 1S6, Canada

Received 2015 May 17; accepted 2015 July 9; published 2015 August 7

### ABSTRACT

In X-ray binary star systems consisting of a compact object that accretes material from an orbiting secondary star, there is no straightforward means to decide whether the compact object is a black hole or a neutron star. To assist in this process, we develop a Bayesian statistical model that makes use of the fact that X-ray binary systems appear to cluster based on their compact object type when viewed from a three-dimensional coordinate system derived from X-ray spectral data where the first coordinate is the ratio of counts in the mid- to low-energy band (color 1), the second coordinate is the ratio of counts in the high- to low-energy band (color 2), and the third coordinate is the sum of counts in all three bands. We use this model to estimate the probabilities of an X-ray binary system containing a black hole, non-pulsing neutron star, or pulsing neutron star. In particular, we utilize a latent variable model in which the latent variables follow a Gaussian process prior distribution, and hence we are able to induce the spatial correlation which we believe exists between systems of the same type. The utility of this approach is demonstrated by the accurate prediction of system types using *Rossi X-ray Timing Explorer* All Sky Monitor data, but it is not flawless. In particular, non-pulsing neutron systems containing “bursters” that are close to the boundary demarcating systems containing black holes tend to be classified as black hole systems. As a byproduct of our analyses, we provide the astronomer with the public R code which can be used to predict the compact object type of XRBs given training data.

*Key words:* methods: data analysis – methods: statistical – pulsars: general – stars: black holes – stars: neutron – X-rays: binaries

### 1. INTRODUCTION

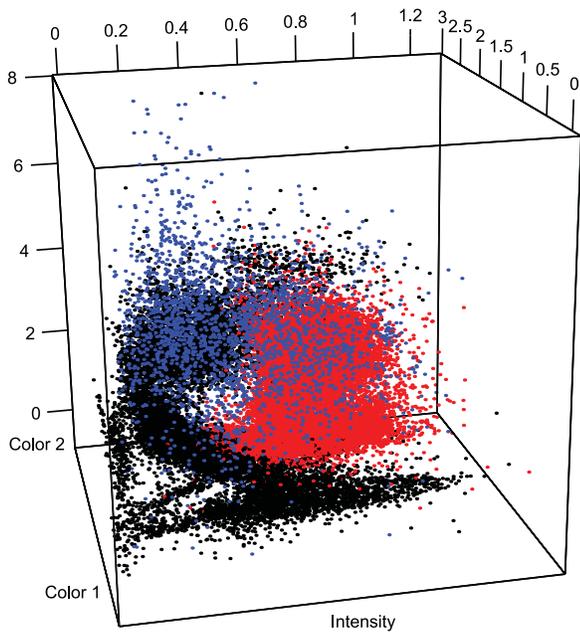
As our ability to acquire and archive data in all fields rapidly grows, the tools for searching these data for pattern, order, and, ultimately, meaning, need to grow commensurately. A critical issue in this ongoing paradigm shift is that of multivariate data with complex, hidden geometric structure. Color-color, or CC, diagrams (which provide spectral information over different energy ranges) and color-intensity, or CI, diagrams (which show brightness variations for a given color) are common, easily obtained measurements that have long been used to classify X-ray binary types. White & Marshall (1984) plotted all of the X-ray binaries (XRBs) observed by the *HEAO-1* satellite on one CC plot; they found that systems containing black holes clustered in one corner of their diagram and pulsars clustered in an opposing corner. While they found significant overlap of several classes of object in the center, they were able to use this clustering to identify new BHC candidates. In Vrtilek & Boroson (2013; hereafter VB13) we show that when CC and CI are combined into a three-dimensional CCI plot, different types of XRBs separate into complex but geometrically distinct volumes. VB13 model the volumes crudely by computing a centroid and constructing an ellipsoid around the centroid that contains 50% of all points while minimizing the volume of the ellipsoid. We suggest that these diagrams provide an easily used, model-independent way to separate classes of systems, in particular systems containing black holes from those containing neutron stars or systems that can produce jets from those that cannot. As a next step toward understanding the physical mechanisms behind this separation of compact object types, we have developed a probabilistic (Bayesian) model which provides a supervised learning approach: unknown classifications of XRBs are predicted

given known classifications. We provide the astronomer with the R code which takes as input CCI data and outputs the estimated probabilities of a system being a black hole, pulsar, or non-pulsing X-ray binary system, in addition to standard errors for these estimates. This software provides the astronomer with more information than an off-the-shelf machine learning solution because such a method typically produces point estimates for the classes of the observations, as opposed to an entire distribution for the classes of the observations.

In Section 2, we describe the data used. In Section 3, we specify the models we have used for estimating the probabilities that the compact object type of an X-ray binary system is a black hole, non-pulsing neutron star, or pulsar. In Section 4, we present our results and their implications and we conclude with a summary and future directions for this work in Section 5.

### 2. DATA

Data on XRBs were obtained by the All Sky Monitor (ASM; Levine et al. 1996) on board NASA’s *Rossi X-ray Timing Explorer* (*RXTE*) which operated continuously for nearly 15 years. Due to as-yet uncalibrated gain changes in the instrument over the last two years of its life, we only use data obtained within the first 13 years. The MIT ASM team provides data in three energy bins (1.3–3.0; 3.0–5.0; 5.0–12.0 keV.) sampled 4–8 times a day. We take one day averages and define our colors as ratios of the mid- to low-energy range (C1) and high- to low-energy range (C2). The sum of the three energy bands is used to represent the intensity of the source; this value is normalized by dividing the total counts by the average of the top 1% of the data for any given source. These form a



**Figure 1.** Visualization of *RXTE* ASM data for 24 XRBs over 13 years including only the  $5\sigma$  threshold values. Each individual point is the one day average of the CCI data from one of the 24 systems. Black points are observations from black hole XRB systems, red points are observations from non-pulsing neutron star XRB systems, and blue points are observations from pulsing neutron star XRB systems. The general pattern is that observations from different system types separate geometrically in this CCI coordinate system. (Note that since CCI coordinates are defined in terms of ratios of counts, they are unitless.)

three-tuple consisting of the features, or background covariates in statistical terms, for each of the observations. Note that since these features are defined in terms of ratios of counts, they are unitless. We also restrict ourselves to detections that have a signal to noise of at least five, where signal to noise is defined as the ratio of the number of counts in a particular bin to the error on the number of counts. Figure 1 shows an example of a CCI diagram constructed with three types of XRBs.

The classification of XRBs is not simple and different authors tend to use different criteria. For our training set, we used 24 systems whose classifications are robust, that is, they are consistent for numerous authors. In particular, we first considered the classifications from the catalogs of Liu et al. (2001, 2006). We then used Remillard & McClintock (2006) identifications of confirmed black hole systems, Homan et al. (2010) identifications of Z and Atoll sources, and Bildsten et al. (1997) identifications of accreting pulsars. We excluded any of the above types that were also identified as bursters as these vary from author to author. This left us with nine systems containing confirmed black holes (Cyg X-1, LMC X-1, J1118+480, J1550-564, J1650-500, J1655-40, GX 339-4, J1859+226, GRS 1915+105), nine confirmed pulsars (J0352+309, J1901+03, J1947+300, J2030+375, J1538-522, Cen X-3, Her X-1, SMC X-1, Vela X-1), and six non-pulsing neutron star systems (Sco X-1, Cyg X-2, GX 17+2, GX 349+2, GX 9+1, GX 9+9).

We test our model by predicting the compact object type of three groups of systems. The first contains six systems whose type is unambiguously classified: one confirmed black hole (LMC X-3); one non-pulsing neutron star system (GX 5-1); and four pulsing neutron star systems (1744-28, 0656-072, 0535+262, 0115+634). The second group of systems contain

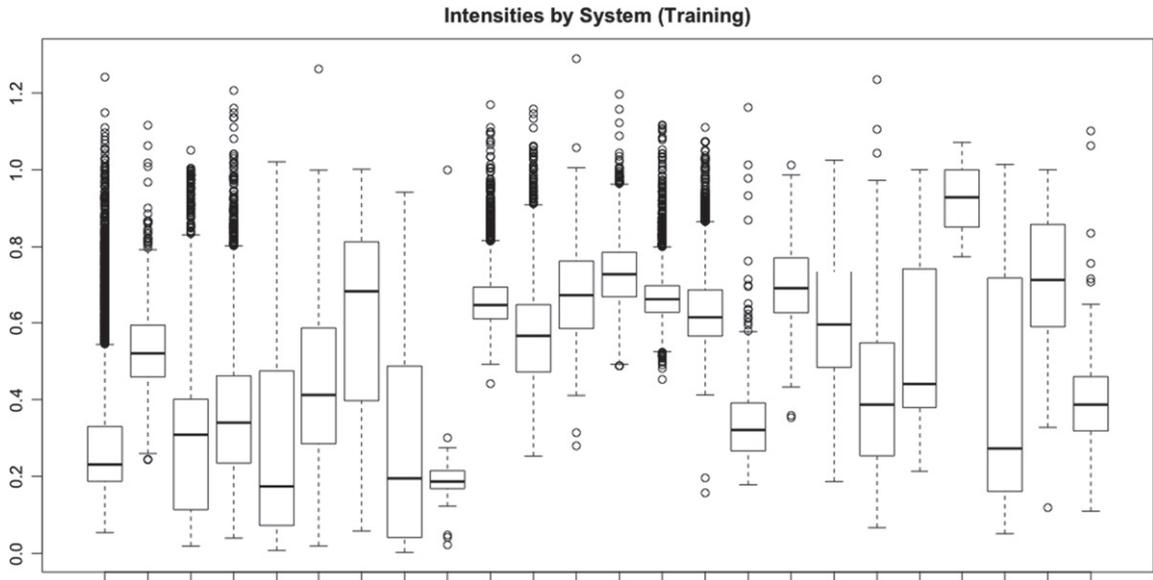
stars that are classified as both burster and atoll sources (Ser X-1, Aql X-1, 1916-053, 1608-522, 1254-69, and 0614+091), and hence possibly have ambiguous classifications. The third group of systems are sources that are either unclassified or have multiple classifications across various authors (1900-245, GX 3+1, 1701-462, 1636-53, and 1700-37).

The training data set consists of 40,857 observations after the preprocessing steps delineated above, of which 13098 come from black hole systems, 25,366 come from non-pulsing neutron star systems, and 2393 come from pulsing star systems. The imbalance in observations between different compact star types is due to the fact that brighter sources are more likely to be detected than weak sources, yet the brightness of the sources is not uniform among star systems. See Figure 2 comparing brightness for various systems.

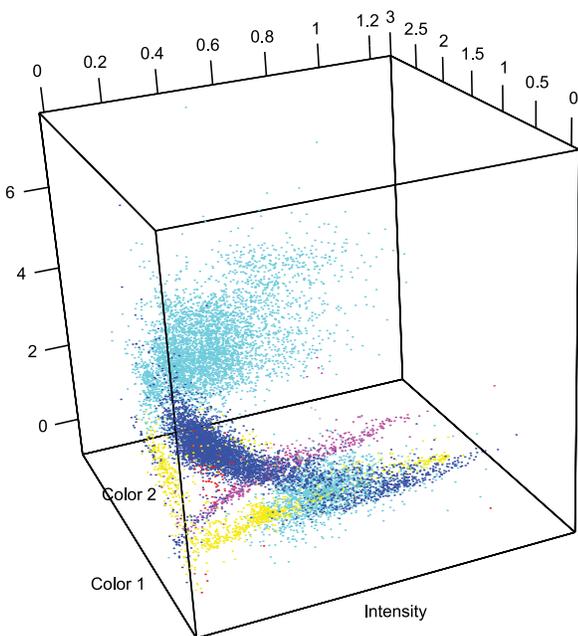
Because observations that are less bright are inherently more likely to be below the signal-to-noise threshold which we utilized in preprocessing, the data are not missing at random (Little & Rubin 2002). The imbalance in observation types by system is problematic because visual inspection of systems of the same compact object type indicates substantial variability between systems; hence, it is prudent to ensure that the true variation of CCI values for each compact object is accurately reflected in the training set. For instance, see Figure 3 for a visualization of the variance between systems that are black holes. Additionally, the model we employ for generating multiple imputations, discussed in the next section, involves a Gaussian process, and hence can be quite computationally expensive to work with, generally scaling with computations that take  $O(N^3)$  time, where  $N$  is the number of data points in the data set. One solution to mitigate both of these issues is to subsample the training set where the probability that a particular observation is selected for inclusion in the smaller training set is inversely proportional to the total number of observations of its system in the entire training set. We sample 10% of the training data in this manner, without replacement, to achieve a final training set consisting of 4085 observations, of which 1486 come from black hole systems, 1465 come from non-pulsing neutron star systems, and 1134 come from pulsing neutron star systems. The histograms in Figure 4 show the balance between various systems before and after subsampling.

### 3. MODELS AND ALGORITHM

Our approach estimates the probabilities that the compact object type of an XRB is a black hole, non-pulsing neutron star, or pulsar from posterior predictions of the compact object type associated with CCI observations within the system. This approach is similar to the multiple imputations methodology developed in the context of survey analysis where the proper handling of missing data is crucial for obtaining valid statistical inferences, which is discussed in further detail by Rubin (1996). In summary, the multiple draws from the predictive distribution of compact object type allow the astronomer to make a judgement about what the compact object type of a given system is, and provide more information than a point prediction. The salient property of this approach is that it takes into account the inherent uncertainty in the prediction for each individual observation, and a concrete illustration of the output of this methodology is displayed in the Results section. For more on applied Bayesian data analysis, see Gelman et al. (2013).



**Figure 2.** Visualization of intensities for observations by each of the 24 systems within the training data set illustrating substantial variability. Systems 1–9 are black hole systems, 10–15 are non-pulsing neutron star systems, and 16–24 are pulsing neutron star systems. This may explain the wide variability for the number of observations of each system above the signal-to-noise threshold, since fainter measurements tend to be noisier.



**Figure 3.** Illustration of the wide variability in CCI data between different systems that contain black holes, where each color represents data from a different system.

The astronomer’s probability model for the compact object of observations from a particular system is a trinomial model where the relevant estimands are the probabilities that the compact object type is a black hole, non-pulsing neutron star, or pulsar. The astronomer’s objective is to estimate these probabilities given the compact object type for all observations within the system. Note that by employing a model which is not a constant, the astronomer is implicitly modeling “imputer noise”: in principle, the compact object type of a system should be constant for all observations from that system, and so the probability model employed by the astronomer is not physically correct, but instead a pragmatic solution to allow

for the (inevitable) mistakes made by an imputer. Indeed, it would be unrealistic to assume any probabilistic model to be accurate all of the time.

Next, we describe the probabilistic (Bayesian) model used to generate predictions for the compact object type of CCI observations. We denote the training set as the 2-tuple  $(X_{\text{train}}, Y_{\text{train}})$ , where  $X_{\text{train}}$  is an  $N_{\text{train}}$  by three matrix consisting of the three CCI values of the training points, and  $Y_{\text{train}}$  is a length  $N_{\text{train}}$  vector with the labels 1, 2, or 3 corresponding to the compact object type of the system each individual data point comes from: more precisely, 1 represents a black hole system, 2 represents a non-pulsing neutron star system, and 3 represents a pulsing system. The test set is denoted by  $X_{\text{pred}}$ , which is an  $N_{\text{pred}}$  by three matrix that contains the CCI values of observations from a single system for which we would like to predict the compact object type, and we write  $Y_{\text{pred}}$  to indicate the labels for the observations from the system. The model aims to predict the unknown vector  $Y_{\text{pred}}$ , and from this estimate the probability of the compact object type being a black hole, pulsar or non-pulsar. From this point of view,  $Y_{\text{pred}}$  is the inferential object of interest and the remaining parameters discussed are nuisance parameters, meaning that they exist for the mathematics of the probabilistic model employed but are not of ultimate inferential interest.

We introduce three independent latent variables for each compact object type (black hole, non-pulsar, and pulsar) whose marginal distribution is a Gaussian process with mean 0 and covariance matrix  $\Sigma$  which has a squared exponential kernel, a standard choice in the computer experiments and machine learning literature, which is discussed in detail by Rasmussen & Williams (2005), i.e.,

$$\Sigma_{ij} = \sigma^2 \exp(-\|X_{i,\cdot} - X_{j,\cdot}\|_2^2 / \phi) \quad (1)$$

where  $X_{i,\cdot}$  denotes the  $i_{\text{th}}$  row of the data matrix  $X$  for which the first  $N_{\text{train}}$  rows are the rows of  $X_{\text{train}}$  and the rows from  $N_{\text{train}} + 1$  to  $N_{\text{train}} + N_{\text{pred}}$  contain the rows of  $X_{\text{pred}}$ . Each latent variable is tied to the compact object type through a multinomial logistic

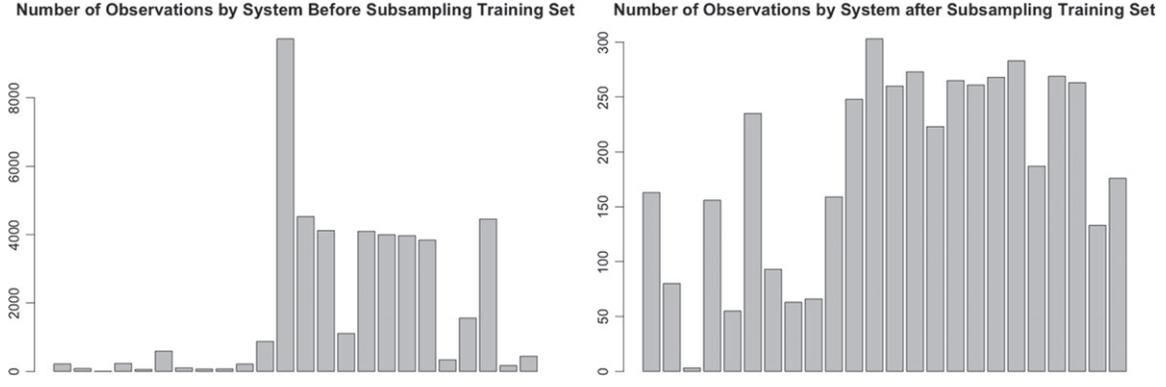


Figure 4. Comparison of the distribution of the number of observations by system in the training set before and after subsampling.

Table 1  
Probability Estimates and Predictions for Compact Object Type of Previously Classified XRBs

System	$P(\text{BH})$	SE	$P(\text{Nonpulsar})$	SE	$P(\text{Pulsar})$	SE	Prediction	Actual Class
GX 5–1	0.0853	0.0519	0.8285	0.0756	0.0863	0.0259	NP	NP
1744–28	0.2626	0.0978	0.1244	0.0570	0.6129	0.1072	P	P
0656–072	0.0782	0.0565	0.0659	0.0472	0.856	0.0842	P	P
0535+262	0.2164	0.0728	0.1312	0.0562	0.6525	0.0883	P	P
0115+634	0.2179	0.0780	0.1425	0.0539	0.6396	0.1057	P	P
LMC X–3	0.8585	0.0648	0.0762	0.0347	0.0654	0.0379	BH	BH

**Note.** We demonstrate the estimated probabilities and predictions for six XRBs whose classifications have been established in the literature, along with their standard errors. All of these systems are properly classified. The predicted class is that with the maximum estimated probability, which is equivalent to the mode of the posterior predictive draws; see Section 4 for further discussion of the justification of this choice.

link function: the probability of a particular observation  $k$  to be of a class  $l$  is proportional to  $\exp[\alpha_l + \beta_l Z_{kl}]$ , where  $\alpha_l$  and  $\beta_l$  have independent unit normal marginal distributions, and  $Z_{.,l}$  is the latent variable drawn corresponding to type  $l$ . Note that  $Z_{.,l} \in \mathbb{R}^{N_{\text{train}} + N_{\text{pred}}}$ , where the first  $N_{\text{train}}$  elements correspond to the training points,  $Z_{.,t}$  for short, and the remaining elements correspond to the prediction test points,  $Z_{.,p}$  for short. The parameter  $\alpha_l$  is the mean offset for the latent Gaussian process corresponding to the compact object type  $l$ , and the parameter  $\beta_l$  is indicative of the marginal effect that each latent variable has on the probability of a data point to be of type  $l$ . It is important to remember that since we are interested in predicting the vector of compact object types  $Y_{\text{pred}}$ , these additional parameters introduced ( $\alpha_l$  and  $\beta_l$ ) into our model are ultimately marginalized out (i.e., forgotten) as nuisance parameters. The likelihood of the observed labels given the latent variables and remaining nuisance parameters takes a multinomial form:

$$L(\alpha, \beta, Z_t; Y_{\text{train}}) = \prod_{k=1}^{N_{\text{train}}} \exp[\alpha_{Y_{\text{train}_k}} + \beta_{Y_{\text{train}_k}} Z_{t_i, Y_{\text{train}_k}}] / N_k \quad (2)$$

where the normalizing constant  $N_k$  is given by

$$N_k = \sum_{l=1}^3 \exp[\alpha_l + \beta_l Z_{t_{k,l}}]. \quad (3)$$

For notational brevity, we denote the posterior distribution of the vector  $Y_{\text{pred}}$ ,  $p^*(Y_{\text{pred}} | Y_{\text{train}}, X_{\text{train}}, X_{\text{pred}})$  as  $p^*(Y_{\text{pred}})$ . Then, by the definition of conditional probability and marginalization,

we have the following integral representation for  $p^*(Y_{\text{pred}})$ :

$$p^*(Y_{\text{pred}}) = \int_{Z_p, Z_t, \alpha, \beta} p(Y_p, Z_p, Z_t, \alpha, \beta | Y_t, X_t, X_p) \times dZ_p dZ_t d\alpha d\beta \quad (4)$$

$$= \int_{Z_p, Z_t, \alpha, \beta} p(Y_p | Z_p, \alpha, \beta, -) p(Z_p | Z_t, -) p(Z_t, \alpha, \beta, | Y_t, X_t, X_p) dZ_p dZ_t d\alpha d\beta. \quad (5)$$

Note that we purposely overload the  $p(\cdot)$  notation and use the dash symbol to represent “all other variables,” for less clutter. This decomposition suggests the iterative algorithm described in the appendix for sampling from the posterior distribution of  $Y_{\text{pred}}$ .

## 4. RESULTS

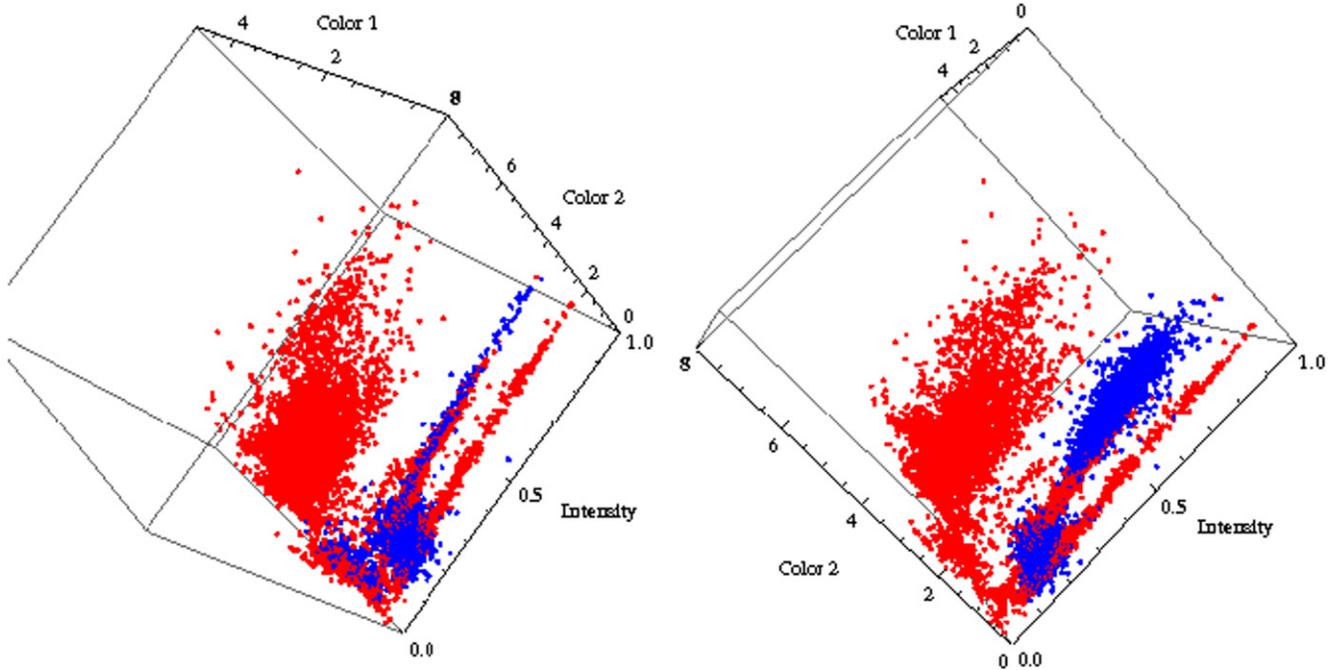
Here, we present our predictions for the compact object types of those systems in the test set discussed in Section 2 using the model and algorithm discussed in Section 3. The predicted class of an XRB is that which presents the maximum estimated probability, an approach which can be justified from a theoretical point of view because the posterior mode is a Bayes’ estimator under a 0–1 loss function, which is a reasonable loss function for a classification problem. The class with the maximum estimated probability is mathematically equivalent to the class with the maximum number of posterior predictive draws for the class labels: i.e., the posterior predictive mode.

Table 1 lists predictions for the six systems with known classifications as well as probability estimates and associated standard errors. Using this scheme, there are no

**Table 2**  
Probability Estimates and Predictions for Compact Object Type of “Burst” XRBs

System	$P(\text{BH})$	SE	$P(\text{Nonpulsar})$	SE	$P(\text{Pulsar})$	SE	Prediction	Actual Class
Ser X–1	0.0441	0.0081	0.9341	0.0133	0.0218	0.0077	NP	NP
Aql X–1	0.5869	0.0692	0.3093	0.0507	0.1038	0.0411	BH	NP
1916–053	0.7683	0.1328	0.1010	0.0810	0.1307	0.0765	BH	NP
1608–522	0.4919	0.0678	0.3535	0.0492	0.1545	0.0310	BH	NP
1254–69	0.2472	0.0231	0.6330	0.0246	0.1200	0.0181	NP	NP
0614+091	0.7062	0.0560	0.1720	0.0441	0.1217	0.0392	BH	NP

**Note.** We demonstrate the estimated probabilities and predictions for six XRBs classified as “bursters” along with their standard errors. Four of these six “bursters” are improperly classified as black holes while the rest are correctly classified. The predicted class is that with the maximum estimated probability, which is equivalent to the mode of the posterior predictive draws; see Section 4 for further discussion of the justification of this choice.



**Figure 5.** Left: example of a burster system in blue, Aql X–1, improperly classified as a black hole system by the algorithm with a comparison to black hole training data in red. Right: example of a burster system in blue, Ser X–1, properly classified as a non-pulsing system by the algorithm with a comparison to the black hole training data in red.

misclassifications for this group of X-ray binary systems. Additionally, in Table 2, we include the predictions and probability estimates for “burster” non-pulsing systems, for which there are a number of wrong predictions: four out of six systems. In all cases, these systems are mistakenly thought to contain black holes. Visual inspection of the data is consistent with this result because the region these systems occupy interferes with the region defined by systems containing black holes: for instance, consider Figure 5 which compares a burster system misclassified as a black hole system with a burster system that is not misclassified, where there appears to be significantly more overlap with the black hole system training data for the misclassified system. It is also possible that some of these systems are misclassified as non-pulsing neutron star systems in the literature, yet significantly more scientific investigation must be performed in order to verify this possibility. Finally, in Table 3, we include predictions and probability estimates for unclassified or ambiguously classified XRB systems, of which notably GX 3+1 has a reasonably high estimated probability of being a non-pulsing neutron star system: 0.7674 with a standard error of 0.0326.

It is important to note that the entire distribution of posterior predictive draws provides significantly more spatial information than a point estimate for a compact system type. For instance, consider the systems Ser X–1 and Aql X–1, the first of which is a properly classified non-pulsing system, and the second of which is one that is improperly classified as a black hole. From Figure 6, there seems to be no question that Ser X–1 is indeed a non-pulsing neutron star system since the proportion of posterior predictive draws is 0.9341 with a standard error of 0.0133. On the other hand, while Aql X–1 is not properly classified, as is evident in Figure 7, we do see some evidence for a non-pulsar signal from the 0.3093 estimated probability of being a non-pulsing system, with a standard error of 0.0507.

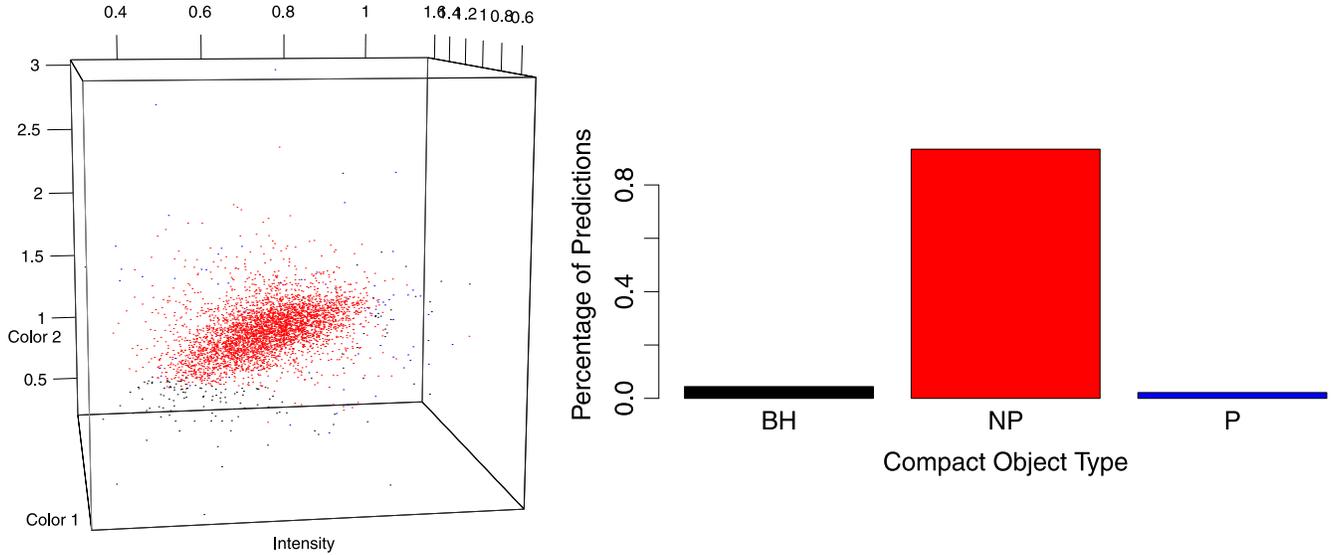
## 5. SUMMARY AND FUTURE DIRECTIONS

The main objective of this work has been to develop a probabilistic model for predicting the compact object type of CCI observations from an XRB and to use the predictions generated from this model to estimate the probabilities that the

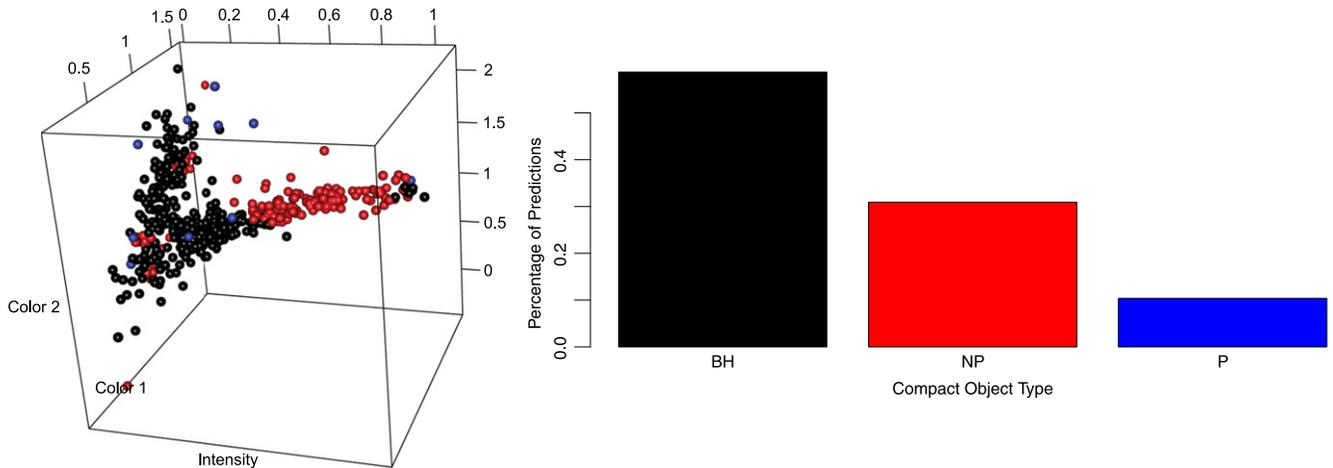
**Table 3**  
Probability Estimates and Predictions for Compact Object Type of Unclassified Or Ambiguously Classified XRBs

System	$P(\text{BH})$	SE	$P(\text{Nonpulsar})$	SE	$P(\text{Pulsar})$	SE	Prediction
1900–245	0.4858	0.0989	0.2334	0.0736	0.2808	0.0846	BH
GX 3+1	0.0757	0.0284	0.7674	0.0326	0.1569	0.0210	NP
1701–462	0.5025	0.1107	0.2179	0.0642	0.2796	0.0698	BH
1700–37	0.1574	0.0650	0.1497	0.0503	0.6926	0.0749	P
1636–53	0.3619	0.0304	0.5411	0.0235	0.0969	0.0263	NP

**Note.** We demonstrate the estimated probabilities and predictions for five XRBs whose classifications are unknown or ambiguous, along with their standard errors. The predicted class is that with the maximum estimated probability, which is equivalent to the mode of the posterior predictive draws; see Section 4 for further discussion of the justification of this choice.



**Figure 6.** Example of a non-pulsing neutron star system Ser X–1 that is properly classified by the classifier. The left panel indicates all of the observations from Ser X–1 and the associated predictions of each observation, with the mode taken to be the prediction. The histogram on the right illustrate the estimated probabilities of Ser X–1 being of each of the three classes.



**Figure 7.** Example of a non-pulsing neutron star system Aql X–1 that is improperly classified by the classifier and mistaken for a black hole system. There appears to be some signal for a non-pulsar system, however.

compact object is a black hole, non-pulsing neutron star, or pulsar. We have shown that the model we developed works reasonably well for this purpose based on the accurate classification of well known XRBs, but note that the model seems to make mistakes when classifying bursters that are close

to the boundary between black hole systems and non-pulsars in the CCI coordinate system. This suggests the benefit of further investigation of these systems as well as a refinement of our approach, including the data sampling, models, and algorithms employed. It is also possible that some of these “burster” systems

are inappropriately classified, but further scientific investigation is required before such a claim can be vigorously asserted.

In order to improve the predictive accuracy of our classification scheme, we can extend the imputation model by imposing a distribution on the Gaussian process parameters or using a cross validation approach to fine tune these parameters. There is a growing body of literature on Gaussian process prediction which attempts to bypass the associated computational impediments, and so we would like to investigate these methods and their potential application to this problem further, e.g., with the INLA method introduced by Rue et al. (2009). Additionally, the *RXTE* ASM data contain a large fraction of missing data due to the signal-to-noise threshold we employ, and so we may consider applying the framework of Little & Rubin (2002) to model this missing data. The primary advantages of this approach are that it utilizes a Bayesian model for generating imputations, which is consistent with our model for compact object prediction, and it may lead to more plausible predictions for the unknown compact object types. Additionally, we may want to consider different subsampling schemes besides the one we employ to ensure that they do not corrupt the inherent structure in the data set. Finally, in addition to predictive accuracy, so as to make the model more scientifically relevant, we may want to include physically meaningful parameters; the inference of such parameters may explain the scientific reasons for the separation of observations into different regions by compact object type.

The CCI method, by definition, uses measurements of X-ray intensity and color in two X-ray bands. This information will generally not only reflect on the properties of the source but also on the absorption of the intrinsic spectrum by the interstellar medium. ISM absorption will clearly affect the lowest energy band the most, and thus the soft color. However, at higher column densities, the hard color will be affected as well. We are developing a general method for the correction of CCI plots given the sensitivity curve of an X-ray monitoring telescope and likely models of the spectral shape (B. S. Boroson et al. 2015, in preparation). The eROSITA telescope developed at the Max Planck Institute for Extraterrestrial Physics, due to be launched in 2016, has 20 times the sensitivity of the *ROSAT*/ASM in the low-energy band and will be particularly beneficial for studying the ISM (Merloni et al. 2012).

We can extend our long-range study using data from past and present large field of view X-ray instruments such as MAXI (Matsouka et al. 2009), the *HETE*-WXM (Yoshida et al. 1995), and BeppSAX-WFC (Boella et al. 1997). Current and planned X-ray telescopes with high sensitivity, such as *Chandra* (Weisskopf et al. 2002), *XMM* (Mason et al. 1995), and eROSITA (Merloni et al. 2012), will enable us to apply our methodology to XRBs of much lower luminosity.

Finally, we reiterate that the R code we have written to make predictions for this analysis ought to be applicable to other CCI data sets quite easily, and so we have provided it for public use.

We acknowledge the Harvard ICHASC for their helpful feedback. Additionally, G.G. and S.D.V. acknowledge partial support through a Smithsonian Institution CGPS grant to SDV.  
Facility: *RXTE*.

## APPENDIX ALGORITHM DESCRIPTION

As noted in detail by Gelman et al. (2013), prediction in the Bayesian paradigm essentially follows the following iterative scheme: first, draw from the posterior distribution of model parameters and latent variables through a Monte Carlo simulation, and then draw from the predictive distribution of interest, which in our case is that of  $Y_{\text{pred}}$ , conditional on these draws. Adapting this general strategy for our problem, Equations (4) and (5) from Section 3 suggest the following iterative algorithm for sampling from the posterior predictive distribution for the compact object type.

1. Sample from the posterior distribution  $p(Z_{\text{train}}, \alpha, \beta, |X_{\text{train}}, X_{\text{pred}}, Y_{\text{train}})$  using elliptical slice sampling due to the joint multivariate normal distribution of  $(\alpha, \beta, Z_{\text{train}})$ . The method of elliptical slice sampling was introduced by Murray et al. (2010).
2. Sample the posterior latent variables at the prediction points,  $Z_{\text{pred}}$ , using the conditional multivariate normal distribution of  $p(Z_{\text{pred}} | Z_{\text{train}}, -)$ , which has mean  $\Sigma_{\text{pred,train}} \Sigma_{\text{train,train}}^{-1} Z_{\text{train}}$  and covariance matrix  $\Sigma_{\text{pred,pred}} - \Sigma_{\text{pred,train}} \Sigma_{\text{train,train}}^{-1} \Sigma_{\text{train,pred}}$  due to fundamental properties of conditional MVN distributions.<sup>3</sup>
3. Sample from  $Y_{\text{pred}}$  from a multinomial distribution conditional on the posterior latent draw of  $Z_{\text{pred}}$ ,  $\alpha$ ,  $\beta$  where, as mentioned previously, the probability of  $Y_{\text{pred},k}$  being of type  $l$  is proportional to  $\exp[\alpha_l + \beta_l Z_{\text{pred},k,l}]$ .

Additionally, we set  $\sigma^2 = 1$  and  $\phi = 0.1$ .<sup>4</sup>

Elliptical slice sampling is a Monte Carlo algorithm developed to simulate a posterior probability distribution where the prior distribution is jointly multivariate normal, a condition that holds in our model, as discussed in Section 3. As explained by Murray et al. (2010), this is a scenario where traditional Monte Carlo methods applied within a Bayesian context, such as Gibbs sampling or Metropolis–Hastings, perform poorly. Routines to implement elliptical slice sampling and draw from the posterior distribution of  $Z_{\text{pred}}$  and  $Y_{\text{pred}}$  were written in the R programming language using the Rcpp, RcppEigen, and RcppArmadillo packages for the efficient inline implementations of linear algebraic routines in C++ (Bates & Eddelbuettel 2013; Eddelbuettel 2013; Eddelbuettel & Sanderson 2014; R Core Team 2015; Sklyar et al. 2015). Additional packages used in the testing and development of this code were mvtnorm and MASS (Venables & Ripley 2002; Genz et al. 2014). As discussed by Murray et al. (2010), the computational impediments of elliptical slice sampling stem primarily from determining the Cholesky decomposition of and inverting a multivariate normal covariance matrix. RcppEigen and RcppArmadillo provide efficient implementations for determining the Cholesky decomposition and performing matrix inversion that can be conveniently included directly within R code. This code, along with the *RXTE* ASM data, is freely available at <https://github.com/gopalan/XRay-Binary-Classification>.

<sup>3</sup> Note that  $\Sigma_{\text{pred,train}}$  is the submatrix consisting of the rows  $N_{\text{train}} + 1$  to  $N_{\text{pred}} + N_{\text{train}}$  and columns 1 to  $N_{\text{train}}$  of  $\Sigma$ . The other submatrices of  $\Sigma$  used in the previous formula are analogously defined.

<sup>4</sup> More judicious ways of selecting these parameters are discussed in Section 5.

## REFERENCES

- Bates, D., & Eddelbuettel, D. 2013, *Journal of Statistical Software*, **52**, 1
- Bildsten, L., Chakrabarty, D., Chiu, J., et al. 1997, *APJS*, **113**, 367
- Boella, G., Butler, R. C., Perola, G. C., et al. 1997, *A&AS*, **122**, 299
- Eddelbuettel, D. 2013, *Seamless R and C++ Integration with Rcpp* (New York: Springer)
- Eddelbuettel, D., & Sanderson, C. 2014, *Computational Statistics and Data Analysis*, **71**, 1054
- Gelman, A., Carlin, J., Stern, H., et al. 2013, *Bayesian Data Analysis* (3rd ed.; London: Taylor and Francis)
- Genz, A., Bretz, F., Miwa, T., et al. 2014, *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-2
- Homan, J., van der Klis, M., Fridriksson, J. K., et al. 2010, *ApJ*, **719**, 201
- Levine, A. M., Bradt, H., Cui, W., et al. 1996, *ApJ*, **469**, 33
- Little, R., & Rubin, D. 2002, *Statistical Analysis with Missing Data* (2nd ed.; New York: Wiley)
- Liu, Q. Z., van Paradijs, J., & van den Heuvel, E. P. J. 2001, *A&A*, **368**, 1021
- Liu, Q. Z., van Paradijs, J., & van den Heuvel, E. P. J. 2006, *A&A*, **455**, 1165
- Mason, K. O., Bignami, G., Brinkman, A. C., & Peacock, A. 1995, *AdSpR*, **16**, 41
- Matsouka, M., Kawasaki, K., Ueno, S., et al. 2009, *PASJ*, **61**, 999
- Merloni, A., Predehl, P., Becker, W., et al. 2012, arXiv:1209.3114
- Murray, I., Adams, R., & MacKay, D. 2010, in *Proc. of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, *JMLR W & CP*, **9**, 541
- R Core Team 2015, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria
- Rasmussen, C., & Williams, C. 2005, *Gaussian Processes for Machine Learning* (Cambridge, MA: MIT Press)
- Remillard, R., & McClintock, J. 2006, *ARA&A*, **44**, 49
- Rubin, D. 1996, *JASA*, **91**, 434
- Rue, H., Martino, S., & Chopin, N. 2009, *JRSSB*, **71**, 319
- Sklyar, O., Murdoch, D., Smith, M., et al. 2015, inline: Functions to Inline C, C++, Fortran Function Calls from R. R package version 0.3.14.
- Venables, W. N., & Ripley, B. D. 2002, *Modern Applied Statistics with S* (4th ed.; New York: Springer)
- Vrtilek, S. D., & Boroson, B. S. 2013, *MNRAS*, **428**, 4
- Weisskopf, M. C., Brinkman, B., Canizares, C., et al. 2002, *PASP*, **114**, 1
- White, N. E., & Marshall, F. E. 1984, *ApJ*, **281**, 354
- Yoshida, A., Matsuoka, M., Kawai, N., et al. 1995, *Ap&SS*, **231**, 463