Machine Learning from Hard X-ray Surveys: Applications to Magnetic Cataclysmic Variable Studies

A thesis submitted to the University of Southampton
for the degree of Doctor of Philosophy in the
Faculty of Engineering, Science and Mathematics
Department of Physics & Astronomy

Simone Scaringi
Astronomy Group

2009

University of Southampton

Abstract

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

School of Physics & Astronomy

Doctor of Philosophy

Machine Learning from Hard X-ray Surveys: Applications to Magnetic Cataclysmic Variable Studies

by Simone Scaringi

Within this thesis are discussed two main topics of contemporary astrophysics. The first is that of machine learning algorithms for astronomy whilst the second is that of magnetic cataclysmic variables (mCVs). To begin, an overview is given of ISINA: *INTEGRAL* Source Identification Network Algorithm. This machine learning algorithm, using random forests, is applied to the IBIS/ISGRI data set in order to ease the production of unbiased future soft gamma-ray source catalogues. The feature extraction process on an initial candidate list is described together with feature merging. Three training and testing sets are created in order to deal with the diverse time-scales encountered when dealing with the gamma-ray sky: one dealing with faint persistent source recognition, one dealing with strong persistent sources and a final one dealing with transients. For the latter, a new transient detection technique is introduced and described: the *transient matrix*. Finally the performance of the network is assessed and discussed using the testing set and some illustrative source examples. ISINA is also compared to the more conventional approach of visual inspection. Next mCVs are discussed, and in particular the properties arising from a hard X-ray selected sample which has proven remarkably efficient in detecting intermediate polars and asynchronous polars, two of the rarest type of cataclysmic variables (CVs). This thesis focuses particularly on the link between hard X-ray properties and spin/orbital periods. To this end, a new sample of these objects is constructed by cross-correlating candidate sources detected in *INTEGRAL*/IBIS observations against catalogues of known CVs. Also included in the analysis are hard X-ray observations from *Swift*/BAT and *SUZAKU*/HXD in order to make the study more complete. It is found that most hard X-ray detected mCVs have $P_{spin}/P_{orb} < 0.1$ above the period gap. In this respect, attention is given to the very low number of detected systems in any band between $P_{spin}/P_{orb} = 0.3$ and $P_{spin}/P_{orb} = 1$ and the apparent peak of the P_{spin}/P_{orb} distribution at about 0.1. The observational features of the P_{spin} - P_{orb} plane are discussed in the context of mCV evolution scenarios. Also presented is evidence for correlations between hard X-ray spectral hardness and P_{spin} , P_{orb} and P_{spin}/P_{orb}. An attempt to explain the observed correlations is made in the context of mCV evolution and accretion footprint geometries on the white dwarf surface.

Contents

List of Figures

List of Tables

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning. This thesis represents the results of three years of research performed by the Author. However, this work was not completed in isolation. In particular the IBIS survey team is an multi-national collaboration and as a consequence some contributions from IBIS survey members may be found within this thesis.

Publications

Material found within this thesis has been published or accepted for publication. The articles resulting from this work are:

- "*ISINA: INTEGRAL Source Identification Algorithm*", Scaringi, S., Bird, A.J., Clark, D.J., Dean, A.J., Hill, A.B., McBride, V.A., Shaw, S.E., 2008, *Monthly Notices of The Astronomical Society*, volume 390, page 1339.

- "Hard X-ray properties of magnetic cataclysmic variables", Scaringi, S., Bird, A.J., Norton, A.J., Knigge, C., Hill, A.B., Clark, D.J., McBride, V.A., Barlow, A.J., Bassani, L., Bazzano, A., Fiocchi, M., Landi, R., 2009, accepted for publication in *Monthly Notices of The Astronomical Society.*

- "The 4^{th} IBIS/ISGRI soft gamma-ray survey catalog", Bird, A.J., Bazzano, A., Bassani, L., Capitanio, F., Fiocchi, M., Hill, A.B., Malizia, A., McBride, V.A., Scaringi, S., Sguera, V., Stephen, J.B., Ubertini, P., Dean, A.J., Lebrun, F., Terrier, R., Renaud, M., Mattana, F., Gotz, D., Rodriguez, J., Belanger, G., Walter, R., Winkler, C., 2009, accepted for publication in *The Astrophysical Journal Supplement Series.*

Copyright

Dedication

A una delle persone piu' importanti della mia vita.

Resterai per sempre nei miei migliori ricordi.

Nicola Mich

15 Gennaio 1985 - 22 Novembre 2009

Acknowledgements

I would like to express my thanks to everyone without whom this thesis would have not been possible. In particular:

- Dr. A.J. Bird, which has supported me as a great supervisor for the past 3 years. It's hard to deal with a such hard-headed italian for that long, however he has managed to do more than that with his excellent patience and insightful supervision. I am also extremely grateful to him whose vision and belief in me has allowed me to study for my PhD during the past 3 years. Cheers Tony!

- Dr. C. Knigge, who has supported me in the past by allowing me to undertake my M.Phil, which eventually lead me to begin my PhD. I am also grateful to him for the fascinating insights and discussions regarding statistics, data analysis and cataclysmic variables which has fueled my interest for these topics even further.

- The IBIS survey team (or part of it at least!) for the hard work that has been devoted in the production of the 4^{th} IBIS/ISGRI soft gamma-ray survey catalog, which has also enhanced and complemented part of this thesis

- My office mates, Dr. D.J. Clark, Dr. V.A. McBride and Dr. A.B. Hill

(at least for the first few years) who have supported me during the many varied discussions and conversations regarding astronomy and less related matters.

- All of the Astronomy Group in Southampton, with whom I spent the last 4 years greatly enjoying myself with.

- The Science and Technology Research Council (STFC, previously PPARC), which have provided the valuable funding required for me to complete my research and PhD.

- Finally the whole astronomy community for all the research which has been undertaken in the past, helping to bring the next discovery into light.

Prologue

"Astronomy has been among the first scientific disciplines to experience this flood of data. The emergence of data mining within this and other subjects has been described as the *fourth paradigm*. The first two are the well-known pair of theory and observation, while the third is another relatively recent addition, computer simulation. The sheer volume of data not only necessitates this new paradigmatic approach, but the approach must be, to a large extent, automated. In more formal terms, we wish to leverage a computational machine to find patterns in digital data, and translate these patterns to useful information, hence *machine learning*. This learning must be returned in a useful manner to a human investigator, which hopefully results in human learning."

– *Nicholas M. Ball 2009.*

Chapter 1

Introduction

"A mathematician is a device for turning coffee into theorems."

– *Paul Erods.*

One of the main issues facing astronomy in the coming century will be the exploration and exploitation of extremely large amounts of data gathered by various observatories. This scenario will essentially be inevitable given the ever increasing capabilities of astronomical observing facilities. The amount of information gathered in astronomy in the coming years will help tackle many of the current problems in contemporary astrophysics, however novel methods are required to deal with such huge and diverse amounts of data. Exploratory data mining of large astronomical datasets is thus the main concern of this thesis. In particular the application of machine learning algorithms for the introduction of new science in contemporary astronomy by exploiting large area surveys such as the IBIS/ISGRI gamma-ray survey performed as part of the European Space Agency's *INTEGRAL* space observatory mission. To this end the thesis introduces ISINA (*INTEGRAL* Source Identification Network Algorithm), a machine learning algorithm constructed in order to identify the IBIS real source population against the fake one caused by both the large

statistical dataset and the highly systematic noise [Scaringi et al., 2008]. The algorithm will be described in detail, tested and applied to the new IBIS dataset in preparation for the future releases of gamma-ray source catalogues.

Analysing the IBIS/ISGRI dataset for the production of future catalogues has also reintroduced a somewhat overlooked population, that of magnetic cataclysmic variables. Because of the excellent survey capabilities of *INTEGRAL*, the number of detections of this source population has grown in the hard X-ray/soft gamma-ray regime. This has allowed us for the first time to study the global properties of these systems, and in particular has allowed us to compare the hard X-ray selected sample of mCVs against the global one. This kind of analysis was not possible before given the very low number statistics of these objects and as one would expect, new analysis will bring forward new results as we will see in the last chapters of this thesis.

Chapter 2 introduces ISINA, a semi-automated algorithm for the identification of sources found within the IBIS/ISGRI images created in order to ease the production of future gamma-ray source catalogues. ISINA has been built keeping in mind the main issues encountered when creating catalogues through visual inspection, and tries to overcome these issues by creating unbiased candidate lists based on more homogeneous criteria. In order to construct a reliable algorithm for this task we will also have to take into account the origin of the systematic noise found within the IBIS/ISGRI images caused by the coded mask imaging technique. This systematic noise, correlated with the real source population, is particularly hard to characterise and will result in an excess of candidates selected by ISINA.

Given the dynamical timescale encountered when observing the gamma-ray sky ISINA will have to be trained on different source populations, defined by their timescale of activity. More specifically ISINA will be trained to recognise

faint persistent sources, strong persistent sources and transients independently of each other. The accuracy of the algorithm is then analysed using the testing set which uses as a reference the published IBIS/ISGRI catalogue 3 [Bird et al., 2007]. This will help us understand the possible pitfalls of ISINA in preparation for the next chapter which will see ISINA being applied for the construction of the catalogue 4 release.

Chapter 3 takes the ISINA algorithm and applies it to the construction of the IBIS/ISGRI catalogue 4. Contrary to Chapter 2, we will not have a reliable testing set to compare our results against. Moreover, catalogue 4 has also been constructed in parallel using the more conventional approach of visual inspection. This method is also described, and will allow us to compare the ISINA result to the more "human" approach. This comparison will shed light on some additional pitfalls introduced by ISINA, and also some introduced by the visual inspection method implying that at the moment the best result will be obtained using a combination of both methods.

Chapter 4 diverges slightly from the application of machine learning algorithms to astronomical classification and focuses more on the analysis of one of the IBIS/ISGRI source populations: magnetic cataclysmic variables (mCVs). This faint persistent population has yielded some very interesting results when analysing the global properties of the hard X-ray selected sample. Moreover the more general properties of the whole mCV population is reviewed in the context of some contemporary accretion models for mCVs. Particular emphasis is given to the P_{orb} - P_{spin} plane of the global mCV population. We will find that all but a few of the hard X-ray selected mCVs occupy the low synchronicity ($P_{spin}/P_{orb} < 0.1$) region of this parameter plane, in agreement with model predictions. Finally the spectral properties of the hard X-ray selected sample are analysed in the context of the systems orbital and spin parameters. We

find that P_{spin}, P_{orb} and P_{spin}/P_{orb} all show evident correlations with hardness ratios defined to be $Flux_{30-60keV}/Flux_{17-30keV}$. The faster the white dwarfs (WDs) in mCVs spin the harder their spectra. This new result is discussed and speculations are brought forward in order to try and explain this phenomenon.

Chapter 5 ends this thesis with some conclusions and thoughts about future developments in both the fields of machine learning applied to astronomy and future mCV studies.

Chapter 2

Learning IBIS data with ISINA

"My CPU is a neural net processor; a learning computer."

– *TERMINATOR*

THIS chapter will present the creation of a semi-automated algorithm for source identification within IBIS images, ISINA: *INTEGRAL* Source Identification Network Algorithm. ISINA has been created using as a reference IBIS catalogue 3 data [Bird et al., 2007], in preparation for the upcoming catalogue 4 release. This will hopefully enable future catalogue releases to be far more objective and consistent in the future.

Within this chapter we will have to review the problems encountered when dealing with IBIS data, and the nature of the classification task we have to pursue. As we will see, the creation of an effective classifier will highly depend on the parameters used, and these have to be chosen in the context of the classification task we have to pursue.

We will begin by introducing in more depth the *INTEGRAL* satellite, the IBIS detector and associated coded mask techniques for imaging. This will also lead us to describe the imaging problems related with such systems. Moreover we will also describe the intrinsic behaviour of high energy sources as observed

by IBIS, as this will also play a crucial role in understanding how to build a reliable classifier. The dataset is then introduced, together with a brief description of mosaic creation. Having understood the dataset and related problems, together with the kinds of objects observed with IBIS, we will then introduce the ISINA algorithm.

IBIS maps will be searched for excess above background and flagged as possible candidates. Features describing temporal behaviour, shape and significance will be defined and extracted for each excess candidate. Next we will create training and testing sets for ISINA to learn on, composed of both real and fake sources. In particular we will also describe our choice to construct three independent classifiers within ISINA in order to deal with the dynamic temporal nature of the gamma-ray sky. One classifier will be trained on faint persistent sources, such as AGNs, and is built to recognise such objects. A second classifier will deal with strong persistent objects and the final classifier will deal with transients. Using three classifiers in this way will allow us to recover the majority of real sources in the testing set, keeping the number of false positives relatively low. This is discussed in the final sections of this chapter, which prepares the reader for the next chapter describing the application to ISINA for the recovery of objects within IBIS catalogue 4 data. We note that all of the algorithm has been developed from independently, without the use of external software, except where stated.

2.1 The *INTEGRAL* satellite

The *INTEGRAL* satellite (Figure 2.1) is an ESA mission launched on October 17, 2002, on board a Russian proton rocket from Baikinour and placed in a 66 hour high elliptical orbit with an apogee of 153,000 km, a perigee of 690 km,

and an inclination of 51.6^o. It has since been fully operational and with its extended lifetime is expected to remain operational until 2012. The mission is particularly dedicated to fine imaging and spectroscopy of gamma-ray sources in the energy range from 17 keV to 10 MeV. The mission is also complemented with imaging in the X-ray and optical bands with additional instruments on board. In total, *INTEGRAL* consists of 4 instruments:

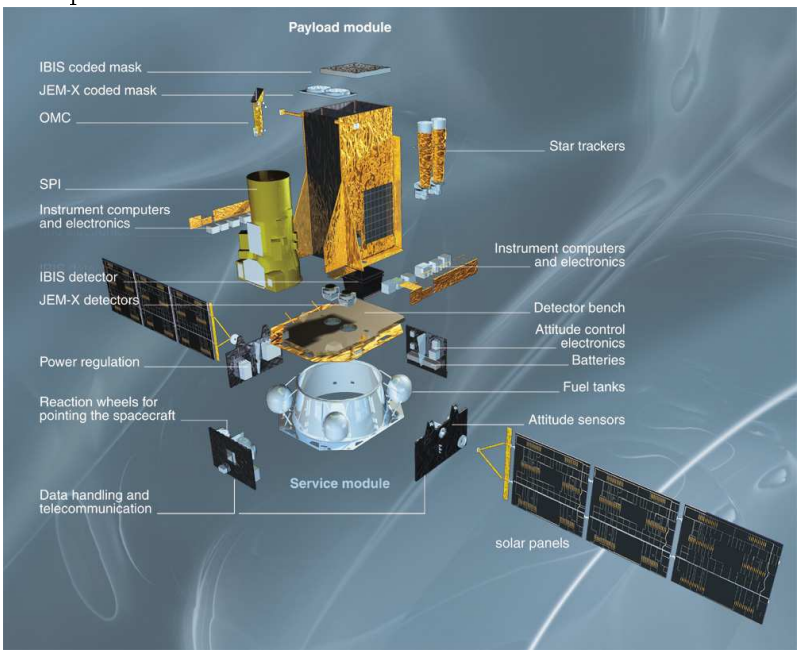
- IBIS is the gamma-ray imager operating between 17 kev to 10 MeV, and has been specifically designed for Galactic surveys. It possesses a large field of view of 30^o with an angular resolution of $12'$ FWHM. It is composed of two detector layers. The first called ISGRI detects photons ranging from 17 keV up to about 300 keV and is composed by an array of 128 $\times$ 128 CdTe detector elements. The bottom layer, called PICsIT is instead responsible for the detection of photons in the range 175 keV - 10 MeV and is composed of 4096 CsI(Tl) elements.

- SPI is the gamma-ray spectrometer and has been optimised for high spectral resolution (3 keV @ 1.7 MeV) and high sensitivity, at the cost of having a poor angular resolution (2.5^o). It is composed of 19 high purity germanium crystal detectors with a total area of $\sim 508 cm^2$. The great spectroscopic capabilities of SPI allows for the detection and study of nucleosynthesis, specifically close to supernovae remnants. Key gamma-ray lines which SPI is able to observe are ^{22}Na, ^{26}Al, ^{60}Fe and the 511 keV annihilation line.

- JEM-X is the onboard X-ray monitor observing in the range 3-35 keV band. It is designed to give contemporaneous measurements in the X-ray band, and helps refine the positional accuracy of IBIS detected sources.

- OMC is the onboard optical camera which takes images in the V band.

Figure 2.1: The *INTEGRAL* satellite showing the different instuments and satellite components.



Similarly to JEM-X, the OMC aids during the localisation of IBIS and SPI detected sources.

All of the instruments on board of *INTEGRAL*, except the OMC, use the coded mask technique for imaging. This is because focusing photons at such high energies is not an easy task, and thus indirect imaging techniques have to be employed. Coded masks however inherently create artefacts in the processed images, which will make the task of identification and/or classification of sources with an automated algorithm particularly non-trivial. This is explained in more detail in the next section, with particular emphasis on the IBIS/ISGRI detector, the instrument on which data ISINA is based on.

2.2 Introducing IBIS/ISGRI

Before we begin describing the identification algorithm, some consideration on the kind of data to be used needs to be addressed. In particular the *INTEGRAL*/IBIS imaging system uses the coded mask technique in order to produce images of the gamma-ray sky. To do this, a shadowgram of the mask pattern is recorded in the IBIS/ISGRI detector plane. This shadowgram is then deconvolved with the mask pattern (Fig. 2.2) in order to produce an image of the sky. This method inherently produces mirror images (or *ghosts*) of real gamma-ray sources together with structures related to the mask pattern, which are then removed by the data reduction software. However a good model of the telescope is required, which takes into account background radiation, and a source list of where the gamma-ray objects are. Because no perfect model of *INTEGRAL* exists, and because we do not necessarily know in advance where all the gamma-ray sources are, image artefacts are created. These are the result of bad ghost and mask pattern subtraction, where the source PSF has not been modelled and subtracted correctly from the image artefacts. In order to best interpret the IBIS/ISGRI images, four different data products are produced by the deconvolution software. These are an flux image, a detection significance image, a variance image displaying the errors per pixel and lastly a residual image displaying the errors associated with the deconvolution per pixel.

Most of these artefacts, or noise, are highly systematic in that they are correlated with the real source population and have the same characteristics (spectral and temporal) as real sources. This will introduce a problem when trying to discriminate the real source population against the fake one as, in some cases, the characteristics for both will be exactly the same. This problem

Figure 2.2: The *INTEGRAL*/IBIS tungsten mask

will be examined later in this chapter in more detail, whilst here we only describe the origin of the systematic noise.

In Figure 2.3, we show the final mosaic image in the 17-30 keV band centred on a bright extragalactic AGN. This source is one of many to create the systematic structure artefacts being described. These are essentially mirror images of the real source, but with apparent reduced flux, situated at about 10.54° from the real source (minor ghosts are also found at other distances). This particular object has been observed with *INTEGRAL* always having the same orientation. This is particularly bad since the ghosts will always appear in the same place and will be enhanced further by mosaicking the data. Contrary to this observing strategy *INTEGRAL* is sometimes rotated before observing the same object, smearing out the ghost structures, at the cost however of producing ring-like structures at the same distance as where the ghosts would have been. This is illustrated in Figure 2.4 for a bright galactic source.

In both cases the artefacts produced are of high significance (i.e. compa-

Figure 2.3: IBIS ghost artefacts caused by the coded mask imaging system. These artefacts are always located at the same distance from the source they have been created and are quite easy to recognise visually. The green ring is there for reference and is centered on the source creating the ghosts with a radius of 10.54°. The problem resides in constructing a good classifier for these structures since, by definition, these artefacts have the same characteristics as that of the real source population.

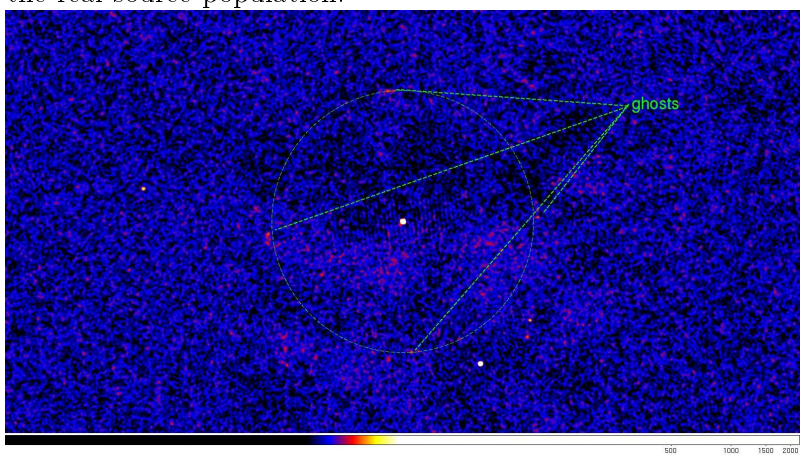
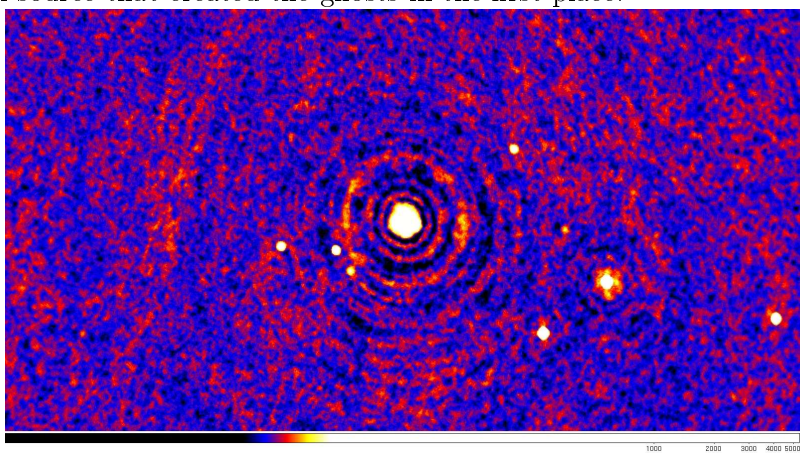


Figure 2.4: IBIS ring artefacts caused by the coded mask imaging system. These artefacts are of the same kind to that of ghosts (see Figure 2.3). The reason for the apperance of ring like structures resides in the *INTEGRAL* observing mode. In the figure displayed below the central object has been observed with different orientatations by *INTEGRAL*. After mosaicking the images the ghosts appear to be smeared in a ring-like structure centered around the real source that created the ghosts in the first place.



rable to the faint sources within the field) and highly correlated with the real source population. This will make source discrimination a particularly hard task for both a machine learning algorithm and a person. Keeping this in mind, we will attempt in this chapter to produce a reliable identification algorithm in order to aid the correct identification of objects found within the IBIS images and coded masks in general.

2.3 Machine learning algorithms for IBIS

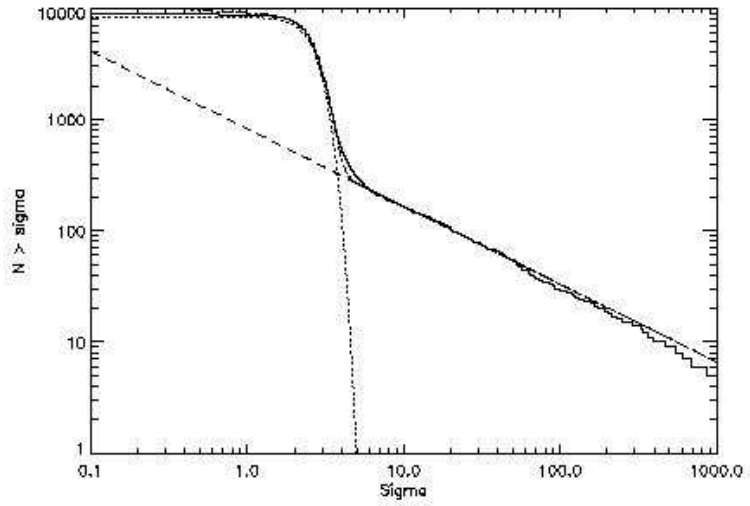
One of the main challenges when choosing a machine learning algorithm for the purpose of identification and/or classification is to first understand as best as possible the dataset. This, together with the kind of classification to be pursued, is crucial in building a reliable algorithm. For example, in a very simplistic scenario, we would never hope to correctly identify and classify a star given only its luminosity. In order to achieve a respectable classification rate we would at least need two parameters namely, the luminosity and temperature (colour difference). Moreover, imagine a scenario where the measurements have been taken with different CCDs, and that each one has a systematic uncertainty which is different from all the others. Then the problem for the classifier won't only be to use the correct parameters for this particular classification task, but also to "learn" how to take different errors into account. This is clearly not an easy task, and we also point out that, even if luminosity and temperature would be enough to correctly classify all stars, we, as the creators of the algorithm, would have to train our classifier based on these two parameters, which until now we have assumed we just knew in advance. But one of the main problems in constructing reliable classifiers is just that: what parameters are needed to correctly classify a particular set? In some cases this

question is very easily answered, like for example determining the star class of a particular candidate, however the task might be much more challenging if we are not sure what parameters are best for a particular set, as is the case in most realistic circumstances.

Identifying real candidates from IBIS data is essentially affected by all the problems described above. There has been extensive mass modelling for the spacecraft and the detector in the past years [Dean et al., 2008, Ferguson et al., 2003], unfortunately however the data products are still suffering a lot from systematic noise. In particular, because of the imperfect implementation of a correct model for the *INTEGRAL* spacecraft, the noise is highly correlated with the real source population and highly resembles real objects. This is an issue which will highly affect any classifier (and indeed humans too!), and we have to be very cautious and aware of the problem. Clearly choosing reliable parameters will be a hard task. By not being aware of this one can easily mislead the algorithm into thinking a particular candidate is real when in reality it is not. An analogy with our previous simplistic example would be to try and classify star types using luminosity and distance. Obviously no sensible classification relevant to star types could be obtained with these two parameters, but it does illustrate the importance of parameter selection for the given classification task.

IBIS catalogues in the past have been constructed through visual inspection. During this process, a set of candidates, mainly selected on significances, (see Figure 2.5) are inspected by eye by a team of experienced astronomers to try and recognise if a particular candidate is real or fake. This process obviously also relies on the astronomer to understand the data thoroughly, so that his/her correct identification rate is high. In essence, each astronomer has created in his/her mind a set of rules which have to be satisfied in order

Figure 2.5: In order to identify an excess as a source it is necessary first to identify the significance level at which the source population dominates over the noise distribution. To this end catalogue 3 has been produced using log-log plots of the number of excesses detected by SExtractor above a specific significance as a function of that significance. This is shown below for the 30-60 keV all-sky mosaic. Two distributions are fitted to the plots, a gaussian representing the noise population and a power law for the real source population. A 1% false positive accuracy is adopted in order to determine a reliable threshold for source identification.



to claim with high confidence that a candidate is real. These rules might be related to the shape of the FWHM of the candidate in question, or even the local signal-to-noise. There are many problems however using this method. The obvious one is that as the data volume for IBIS (or any other instrument) increases, this method requires more time and more "inspectors" to do the job. As an example catalogue 1 was based on solely 5 maps, whilst the latest catalogue 4 uses $\sim 11,500$ maps. There is no alternative to this issue, and only a larger workforce can overcome this. The second more subtle but more relevant problem is that each astronomer has in mind his/her own idea of what a real source should be or look like. This is because, similarly to machine learning algorithms, the astronomer can only rely on past examples in order to make a decision on a new candidate. This essentially means that the same astronomer might choose to classify a particular object as real today and fake later on in the future, after having inspected more cases and having changed his/her mind. The problem gets worst when we introduce many different astronomers, since each one would have inspected a slightly different set, and will have created in his/her own mind a slightly different set of rules for what he/she considers to be real. In some circumstances having many "opinions" on a particular candidate might be useful, however we can think of cases where this is definitely not optimal when most of the astronomers make the wrong decisions. In fact we would expect a group of astronomers to usually select more sources than are real, mainly because our brain is prone to find patterns, making the astronomer "see" a real source when in effect it is not. This has the consequence of causing high rates of false positives. Making our selection bias even worse is the fact that most astronomers do not only visually inspect candidates (luckily!) but also have their favourite objects in the sky they enjoy studying. Even if very subtly, this can bias the selection of real objects further

when we imagine the astronomer believing too many of his/her own favourite objects. This bias would be caused by someone who, for example, studies AGNs and is, even if unconsciously, trying to raise the number of catalogued AGNs for further studies. Obviously the problem is not restricted to AGNs only, and can be turned the other way round by someone very conservative about any object, biasing the selection the other way.

In order to address most of the issues discussed above on the IBIS dataset we have chosen what we think is a flexible algorithm which can undertake most problems intrinsic to the classification task. In the following sections we will describe the creation of an algorithm similar to Random Forest (Brieman et al. 1984) which we call ISINA. This is an algorithm that has the potential to deal with redundant parameters (i.e. parameters which do not help or confuse the algorithm in making a decision) so that we are allowed to choose many more parameters than needed without affecting the final result. This is particularly useful as we do not know in advance what features to use for our classifier, so we will decide to include many more than we actually think we need. The other promising feature of ISINA, and Random Forests in general, is that it is structured in a similar way to the visual inspection process. Essentially we will build many classifiers using different features and let each classifier decide on each candidate independently. We will then merge the results at the end, similarly to what happens during visual inspection. The great advantage however is that each of our individual classifiers will not be biased, and will exclusively be built using the training set. This leaves very little space for "opinion", as should be the case for scientific classification. However before we introduce the algorithm in full we first have to understand the data and classification scope better in order to select appropriate features for our classifier to work on.

2.4 The dataset

The data are collected with the low-energy array ISGRI (*INTEGRAL* soft gamma-ray imager, Lebrun et al. [2003]), consisting of a pixelated 128 × 128 CdTe solid-state detector that views the sky through a coded mask. The instrumental details and sensitivity can be found in Lebrun et al. [2003] and Ubertini et al. [2003]. IBIS/ISGRI generates images of the sky with 12' [full width at half-maximum (FWHM)] resolution and ≈3 arcmin source location accuracy over a 19° fully coded field of view in the energy range 15-1000 keV. The data set used for the creation of ISINA is the same as the one used in the production of the third IBIS/ISGRI soft gamma-ray survey catalogue [Bird et al., 2007], which uses image data for the first 3.5 yr of IBIS/ISGRI core programme and public observations. The data set used here ensures that $> 70\%$ of the sky is observed with at least 10ks exposure. This yields a data volume of ≈5 Tb of raw data and ≈10 Tb of processed data.

Each *INTEGRAL* pointing is referred to as a Science Window (ScW). In particular each IBIS/ISGRI ScW image will have an exposure of about 2000 s and can produce different images for different energy ranges. This will have to be taken into account in our classifier as different types of objects have different spectral shapes and might only appear in some band images and not others.

The IBIS dataset is not only composed of ScW images, but also of mosaics specifically created for survey studies. These mosaics are created in 5 different energy bands using as much of the ScW data as possible. It is important, however, to remove a small fraction of images for which the image deconvolution process has not been successful. These mainly include data taken during or following severe solar activity or near spacecraft perigee passage when the background modeling is difficult due to the spacecraft passing close or within

the South Atlantic Anomaly or the Van Allen radiation belts.

The image rms[1] was determined for each significance map (at ScW level), and the distribution of the image rms statistics for all science windows was determined. The mean and variance of this distribution was then found in order to define what can be considered a "good" image rms. An acceptance threshold was then set at 2σ above the mean image rms, and any individual images with higher rms than this were discarded. Typically, this resulted in any image with an rms greater than 1.08 (after removal of sources) being rejected (depending on energy range). Of the ≈24,000 ScW processed, ≈20,000 ScW were retained in the final ScW list. In addition, science windows acquired in "staring" mode, and data taken during the instrument performance verification (PV) phase (for simplicity, this was taken as up to and including the calibration activities in revolution 45) were removed from the main ScW lists due to their potential adverse effect on the final mosaic quality.

The ScWs were mosaicked using a tool developed in Southampton, optimized to create all-sky galactic maps based on several thousand input ScWs. However given the long timebase spanned by this dataset, we additionally require mosaics composed of only a subset of ScWs in order to locate transients. These will be objects that have an increase in flux above the noise level only in specific ScWs, and require mosaicking only a subset of ScWs to be significant enough and be considered for identification. In order to compensate for this problem, we constructed mosaics over three timescales. Maps were created for each revolution that contained valid data. This is optimized to detect sources active on timescales of the order of a day[2]. We identified 26 sequences of consecutive revolutions that had similar pointings. Thus, these revolution sequences could best be analyzed as a single observation, and sensitivity for

[1] Root Mean Squared in units of Significance
[2] 1 revolution = 3 days

sources on longer timescales than revolutions (i.e., order of weeks) could be optimized. Ultimately, persistent sources can best be detected in an all-archive accumulation of all available high-quality data.

Maps were created for each of these timescales, in five energy bands (20-40 keV, 30-60 keV, 20-100 keV, 17-30 keV and 18-60 keV), these being chosen to provide both coverage of the most sensitive energy range for ISGRI and sensitivity to various typical source emission profiles. For each energy band and time period all-sky mosaics were made in four projections: centered on Galactic center, centered on Galactic anticentre, north Galactic polar, and south Galactic polar. The purpose of these multiple projections is to present the automatic source detection algorithms with source PSFs with the minimum possible distortions.

2.5 The Algorithm

In this section we describe how ISINA is built, trained and tested. The process is relatively long to describe and, in order to make it clearer to the reader we give here a brief description of the process. The steps involved after having created the mosaics are as follows:

1. **Locating candidates:** Here we describe the steps involved in creating an initial candidate list for ISINA to work on. We will search all the mosaic maps for possible candidates in a very un-conservative way so as to make sure no real candidates are missing. ISINA will then learn to discriminate the majority of fake sources.

2. **Filtering candidates:** Given the previous step we are faced with an incredibly large candidate list, where many candidates are in actual fact

the same but detected in multiple mosaic maps. Here we describe the procedure employed in order to merge the initial excess list.

3. **Feature extraction:** Having produced an excess list we will define some features that will represent what a human astronomer inspects before deciding whether a candidate source is real or fake. These features will then be extracted for all candidates in all the ScWs and mosaic maps available. These are the features, or parameters, ISINA will work on in order to decide whether candidate sources are real or fake.

4. **Feature merging:** Having extracted features for all candidates in all possible images we are faced with a data redundancy problem. In particular, the features extracted for transients or variable sources will not all be useful, and in fact will confuse ISINA in deciding whether the candidate is real or fake. In other words, if a transient is on for a very small portion of the observed time, we run the risk of not identifying it properly if we do not use only the observations where it is active. The quiescent phase of the source will confuse the algorithm. This subsection will describe the method we employ to overcome this problem by creating three independent classification networks: one based on faint persistent sources, one based on strong persistent sources and one on transients. This will hopefully allow the correct recovery of the majority of IBIS detected sources.

5. **Training and testing sets:** Having extracted and merged the features for all our candidates we need to produce a reliable training set for ISINA to learn on. Moreover a reliable testing set needs to be produced so that we can later assess the performance of ISINA. Again this is done by creating three independent training and testing sets as described in the

previous subsection.

6. **Random forest algorithm:** The main engine of the identification algorithm is described here. This is where the training takes place for the three independent networks.

7. **Results:** Finally, after training, the testing set is passed to ISINA and identification tags are given to the testing candidates. Here we assess the performance on this testing set and discuss possible improvements for the algorithm.

2.5.1 Locating candidates

The first stage in ISINA is to look in all images (mosaics, revolution mosaics and revolution sequence mosaics) in order to locate potential candidates for further analysis. We do this by simply running the source searching algorithm SExtractor [Bertin and Arnouts, 1996] and recording all excesses above 4.5σ. This threshold might be too optimistic given the level of systematic noise in the maps; however, we will show how this is not a problem as the network will be able to learn and discriminate the fake candidates from the real ones. On the contrary, the threshold is too conservative for some maps where systematic background noise is very low; however, at this stage it is best to have more fake candidates at the cost of recovering most of the real ones. We note that this was the global threshold employed in creation of catalogue 3 [Bird et al., 2007]. The source position measured by SExtractor relies on calculating first order moments of the source profile (referred to by SExtractor as the barycentre method). At the faintest levels source detectability will be limited to background noise; however, this can be improved by applying a linear filter to the data. Moreover, in crowded regions of the sky, confusion

can be avoided by applying the SExtractor *mexhat*. This filter convolution alters the significance of sources in the original mosaics by increasing it, deblending two (or more) close candidates. The drawback of this filter is that it sometimes creates extra ring-like candidates around apparent or real excesses, which will be extracted as possible candidates by SExtractor, and later fed to ISINA. Doing this yields an excess list of 58,603 candidates, where most are in common between ScWs however. The next section will describe the adopted method for source filtering and merging.

2.5.2 Filtering candidates

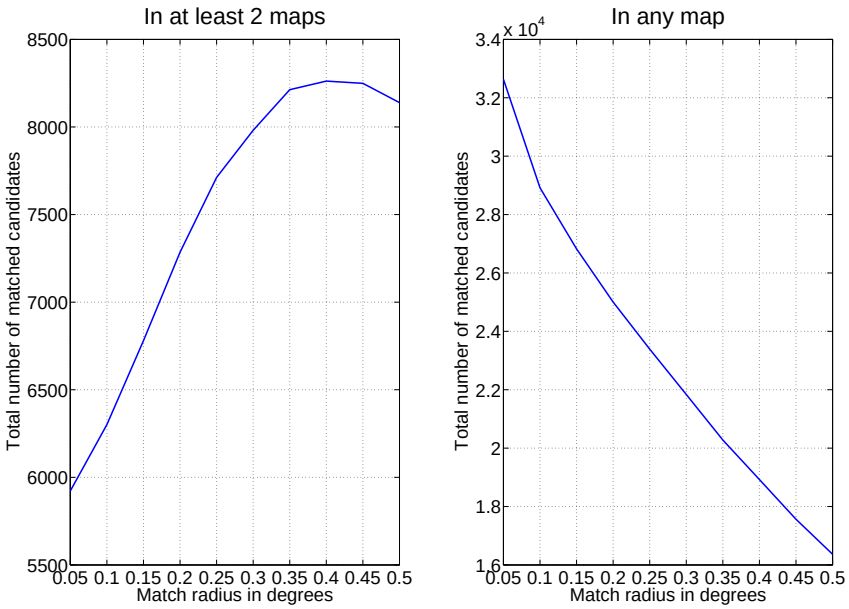
An initial list of 58,603 excesses was extracted as described above. We need to employ some sort of filter in order to discriminate against duplicates and to remove the most obvious fake excesses in common between ScWs. We do this by merging excesses from multiple maps by assuming sources within 0.2° (the IBIS/ISGRI angular resolution) from each other are actually the same, beginning from the highest significant excess. The 0.2° merge radius might seem too large; however, this has been chosen as a trade-off between keeping the number of false positives caused by instrumental artefacts low, while still retaining the majority of objects in catalogue 3. By decreasing the merge radius we allow for more fake excesses caused by the imaging system. For example bright sources in the IBIS maps tend to have propeller- like and/or ring-like structures (see Figure 3.9) around them sometimes extending 0.5° from the source centre, and these are extracted with SExtractor. By decreasing the merge radius we allow for these to be treated as independent candidates; however, by increasing the radius we allow for the candidates to be merged with the bright source from which they were created in the first place. We note that the merging process must start from the brightest excess first and

subsequently merge the candidates in brightness order. This has to be done in order to avoid mismatches between real sources and fake candidates. If, for example, we had to merge the candidates starting randomly from our excess list, we run the risk of merging a fake candidate created from a real source with other fakes. This would not be the case if we start with the real candidate first (brighter than the corresponding fakes associated with it), the fakes created from it would be merged into the real one.

We also eliminated all excesses that appeared only in one mosaic. This additional criterion was introduced in order to minimize the number of false positives in the final candidate list and was also the basis of the creation of catalogue 3, thus no real sources are missed by employing this cut but a high number of fakes are discarded. This is best illustrated in Figure 2.6 displaying the number of recovered objects as a function of match radius for two different methods. On the right an excess was considered as a candidate if it appeared in any map, whilst on the left an excess was considered as a candidate if it appears in at least 2 maps. The difference in recovered objects is very significant. We note that the graph on the left keeps increasing until $\approx 0.4^{\circ}$. This is because, even though there are fewer "propeller" candidates as the search radius increases, many more extra-galactic fake candidates are introduced.

The final coordinates of the candidates are then taken from the highest significance excess. Thus, the initial excess list reduces to 7221 candidates, which are shown in Figure 2.7. Out of the 421 sources identified by Bird et al. [2007] only 13 were not recovered with these filters. Of these, five were observed before revolution 46, and therefore are not present in our initial excess list. The remaining eight were excluded due to the $0.2'$ merge radius and reside very close to a real source. It is possible that human intervention could recover them in

Figure 2.6: Graphs displaying the total number of recovered objects as a function of match radius. On the left an excess was considered as a candidate if it appeared in at least two maps, whilst on the right if it appeared in any map.

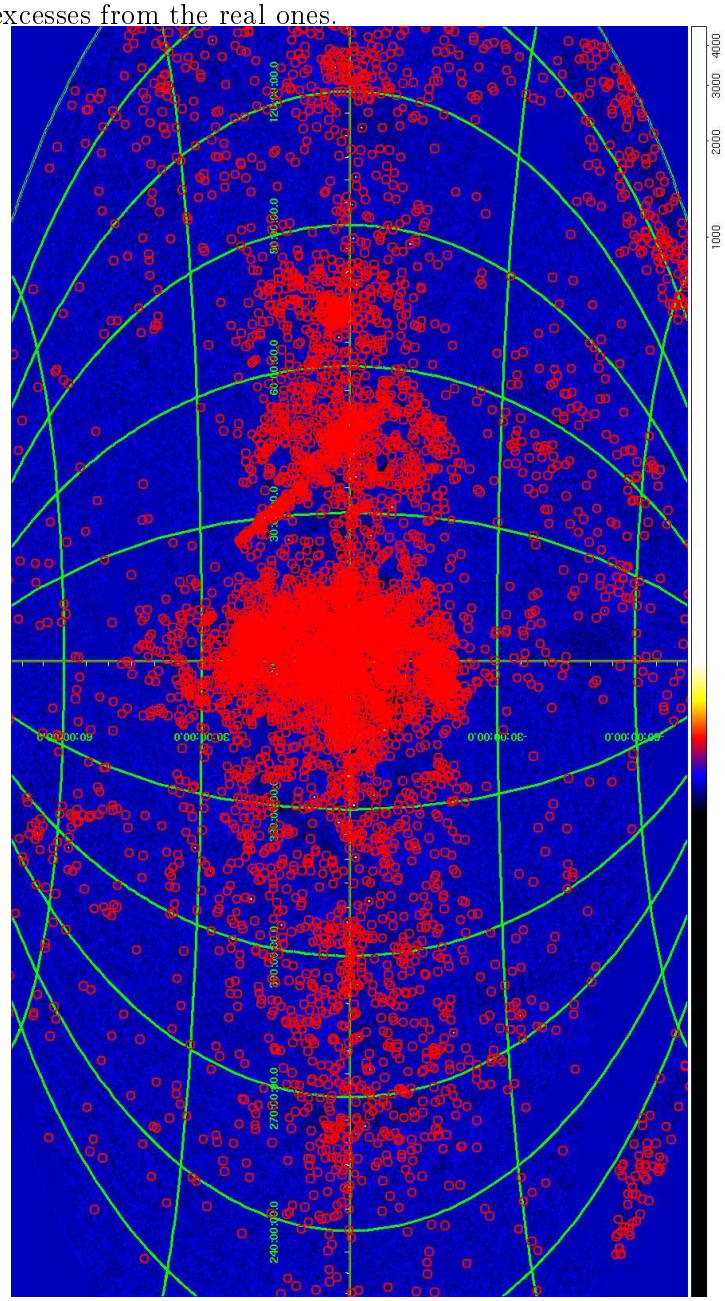


the final inspection phase given they are so close to real sources, however new methods are being investigated in order to localise them for the production of future catalogues. We point out that other candidate merging methods could be developed in the future without affecting ISINA.

2.5.3 Feature selection and feature extraction

When the input data set to a classification algorithm is too large and/or suspected to be significantly redundant, as is the case for the IBIS/ISGRI images, then the input data will be transformed into a reduced representation set of features (also referred to as a feature vector). As a trivial example one feature could be the significance value for a particular candidate on a particular timescale. This process is called feature extraction, or more generally dimensionality reduction.

Figure 2.7: Galactic centre in the 18-60 keV band as seen by IBIS/ISGRI with the reduced subset of 7221 candidates overlayed. Candidates tend to follow the artefacts caused by the detector system. Moreover most candidates have been detected using SExtractor deliberately to a very low detection threshold, and most will turn out to be fake. The network will try to learn how to discriminate the false excesses from the real ones.



Feature extraction and parameter selection are the most important steps in building a reliable classification algorithm. By feature extraction we mean producing a set of variables, extracted from the IBIS/ISGRI sky images, whilst by parameter selection we mean combining these variables in order to best represent the objects we are trying to classify. Even the most perfect classification network will not perform well if the wrong parameters are passed to it. This is why in a general scenario one has to answer the question "what are we trying to classify?" in order to decide what features best describe the given classes. In our case we are trying to discriminate between real sources and fake candidates within the IBIS/ISGRI images. Our features need to provide the maximum possible discrimination between real and fake sources. A feature that describes a real source is of no use if it describes a fake one in the same way. In fact features that only apply to fake sources are equally useful. Moreover they also have to take into account the nature of the artifacts caused by the imaging system, in our case the ISGRI layer on IBIS and coded masks together with the temporal nature of the gamma-ray sky. Here we explore the methods employed in order to extract reliable features to be passed to the network(s) for classification.

In the context of IBIS/ISGRI source identification we have decided to use the following features from past experience in manually creating survey catalogues. First, a 2D Gaussian is fitted to all ScWs where candidates might be present. We allow the Gaussian to be fitted in a 9×9 pixel (40×40 arcminutes) window centred at the candidate's coordinate. The following features are then extracted from the intensity, significance, variance and residual images described in Section 2.2:

1. Distance between Gaussian centre and original candidate coordinate: Too large deviations in this parameter might suggest we are actually

looking at structure as real sources should not move around more than their point source location accuracy.

2. Fitted Gaussian peak (amplitude): This parameter will help us discriminate between high and low significance detections.

3. Local standard deviation: Will help the algorithm to determine the various levels of local signal-to-noise.

4. FWHM difference in two perpendicular directions: We would expect this parameter to be very close to 0 for real sources as their PSF is meant to be circular.

5. FWHM ratio: Similarly to feature 4, we expect this feature to be close to 1 for real sources

6. Significance value at candidate position: This again is similar to feature 2, however this value is read at the candidate's position rather than where the fitted Gaussian peak value is.

7. Intensity value at candidate position: Same as feature 6, but for intensity.

8. Variance at candidate position: Noise level indicator.

9. Residual at candidate position: Again a noise level indicator mainly related to how the deconvolution process performed for the particular ScW in question.

Features 1-5 are extracted from both intensity and significance images in four energy ranges. A conservative cut is employed by ignoring all extracted features where the centre of the fitted Gaussian is offset by more than 2.5 pixel (30 arcmin) from the original candidate coordinate. In these cases the candidate is likely to be not observable in the ScW and the Gaussians were fitted

to background structure within the candidate region. Similarly astronomers would tend to emply a similar cut where any articular candidate would be excluded if the resulting Gaussian fit was substantially offset from the nominal position. In addition to the above we also extract all nine features from the final significance mosaic maps as these will prove useful in identifying the faint persistent population. Obviously parameters such as the FWHM will depend on the kind of projection (galactic centre, galactic anticentre, north and south polar) from which the feature is extracted. This is not appropriate as the network will then be discriminating projections rather than real FWHM. In order to deal with the problem we extract the features from the projection which has its centre closest to the candidate position, optimally minimizing the distortions caused by the projections.

On average, with large scatter, each candidate has a total of ≈600 ScW pointings used in the extraction process, yielding more than 10,000 features. The feature extraction process takes just about more than 7 days on 5 1.8GHz CPUs. It is clear that for many objects, in particular transients, most of the ≈10,000 features will be redundant and not useful, suggesting that a further step has to be employed in order to further reduce the dimensionality of our data set. The next subsection will deal with this process called feature merging. Once a set of relevant and reliable parameters have been chosen the problem becomes one of pattern recognition. Essentially one has created a multidimensional parameter space, where some variables will have a greater discriminatory power than others, whilst on the other hand some combinations of two or more would be more efficient. The problem is that we are not sure which (if any) of the features are best for class discrimination and this is why one employs classification networks for pattern recognition. We therefore need to reduce our dataset in a sensible manner and merge our features.

2.5.4 Feature merging

In order to reliably train a classification network, the nature and behaviour of the objects one is trying to classify needs also to be taken into account. In the case of the gamma-ray sky this behaviour is very diverse, and one has to define coherent subclasses that any classification network can deal with separately. After all a network which is very well trained at recognizing the Crab, a bright, constant flux source, would not necessarily perform well at recognizing a faint active galactic nucleus (AGN). The most obvious separation is that of faint persistent versus strong persistent. By strong persistent we mean any objects which would be observable in one ScW pointing. On the other hand a faint persistent source might not be observable in one ScW pointing; however, it's signal will still be present, and will show up in the final mosaic, for example, after having increased the exposure time on that part of the sky. To be more precise for the IBIS/ISGRI detector, a source will be observable in one ScW pointing if its flux is greater than ≈10 milliCrab with a ≈2000s exposure. Everything with a lower flux will need longer exposures to be observable, even though its signal will still be present in any one pointing. This is the case for most AGN and cataclysmic variables (CVs).

Another source behaviour that must be taken into account when dealing with the gamma ray sky is that of transients. These objects are usually X-ray binaries (XBs) but include a diverse set of objects as well (gamma-ray burst, supernovae). These will vary on a huge range of time-scales, from being observable in only one ScW to being observable only by mosaicking several orbits of data. As one might expect these are tricky to detect as it is not known in advance what sort of time-scale to expect from these objects and in particular when, in a series of pointings, to extract features from them. This is also a big problem for "human" searches too. If we do not make a mosaic

map on the right timescale, we can never hope to find the source, unless we bias the search and look for known sources.

From here on we will refer to the definitions just described when referring to our three different source behaviour types: faint persistent, strong persistent and transients. Each one of these subclasses needs to be treated independently when training as the time-scales and features of each subclass vary enormously. We therefore have to tell the network what features are relevant for classification of a given subclass of sources. The danger of this approach is that we train for specific characteristics, and the detection of new source types may be inhibited. Balancing this, our subclasses are as generic as possible, which reduces the risk with specific subclasses. In the next three subsections we will explain how the extracted features are merged in order to produce a set of merged features per network together with their respective training sets.

Faint persistent sets

In order to deal with the faint persistent population we decide to merge the candidate features (Section 2.5.3) by simply taking the average of, or combinations of features (see Table 2.1). After all from our definition of faint persistent, all ScW pointings will have a signal, even if a small one. It might occur that the level of noise in any particular ScW will be much higher than the signal. As described in Section 2.5.3, features get discarded if the Gaussian fit is offset by more than 2.5 pixels, suggesting we are looking at a "bright" noise structure. In our approach 12 features are used which were extracted from an ScW level and averaged as described above. We also included six features extracted from the final mosaic level. For a list of used features refer to Table 2.1. It should be noted that these features have been chosen to try and mimic what an expert astronomer would consider when assessing source credibility. For example, in

Table 2.1: Summary of the features used within the three networks as described in Section 2.5.4. Each column has a Yes for used features and a No for dropped features for the particular network in question. TM stands for Transient Matrix.

Description	Faint	Persistent	Transient
ScW significance features			
Fitted Gaussian amplitude	Yes	Yes	Yes
FWHM difference	Yes	Yes	Yes
FWHM ratio	Yes	Yes	Yes
Significance / Local background	Yes	Yes	Yes
Fitted Gaussian peak / Significance	Yes	Yes	Yes
ScW intensity features			
Fitted Gaussian peak	Yes	Yes	Yes
FWHM difference	Yes	Yes	Yes
FWHM ratio	Yes	Yes	Yes
Intensity / Local background	Yes	Yes	Yes
Fitted Gaussian peak / Intensity	Yes	Yes	Yes
ScW general features			
Variance	Yes	Yes	Yes
Residual	Yes	Yes	Yes
Maximum significance from TM	No	No	Yes
Significance mosaic features			
Fitted Gaussian peak	Yes	Yes	No
FWHM difference	Yes	Yes	No
FWHM ratio	Yes	Yes	No
Significance	Yes	Yes	No
Significance / Local background	Yes	Yes	No
Fitted Gaussian peak / Significance	Yes	Yes	No

assessing the Gaussian fit we would look at the difference between the fitted Gaussian peak and the respective pixel value for a candidate source. For good fits we would expect the value to be very close to zero, whilst the value will be high for bad fits. Also note the energy bands used. For the faint persistent class we decided to use three energy ranges: 20-40 keV, 20-100 keV and 18-60 keV. This is because most faint persistent objects are AGN and appear in these bands from experience in compiling previous catalogues. This might inhibit the correct identification of CVs, another subclass considered to be faint persistent

but spectrally different. However, INTEGRAL has not yet detected enough of these systems for them to be treated independently within the context of a source identification network. So in summary for each candidate in the faint persistent network we will have 12 + 6 features merged from each used energy band, giving a total of 54 features.

Strong persistent sets

The second subclass is that of strong persistent objects. This subclass has to be treated separately from the previous, as training a network on strong persistent objects will not necessarily recover faint objects (and vice versa). The features used for this subclass are the same as for the faint persistent subclass with the only addition of features from the 17-30 keV band. This is because we think that strong persistent sources, mainly populated by XBs, are detectable through a wider spectral range. Moreover XBs are much brighter and will be detected in more energy bands. However, we realize that both are persistent and that is why we essentially use the same feature time-scales for both, but as we will describe later, the training sets for these will be substantially different.

Transient sets and the Transient Matrix

The final subclass, transients, is the least trivial to train for, as the features are harder to define and show most variations from source to source. For this task we introduce what we call a "transient matrix" (TM) for the selection of ScW pointings to use. Essentially the aim of this technique is to locate a timescale which maximizes the significance detection for transient candidates. This is important for feature merging as it will give us the features we need to average (rather than averaging all features as in the previous networks). Suppose the intensity light curve I of a particular candidate contains N points. Moreover

assume each point *I(i)* in our light curve has a variance *V(i)* associated with it. We can define weights for each point in the light curve *w(i) = 1/V(i)*. We will then create an upper diagonal $N \times N$ matrix T. For each row i in T we compute

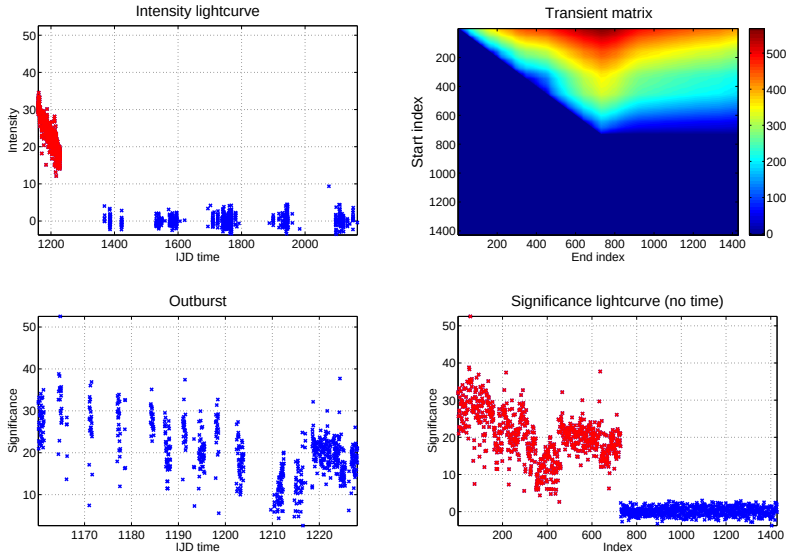
$$T(i,j) = \frac{\sum_{k=i}^{j} I(k)w(k)}{\sqrt{\sum_{k=i}^{j} w(k)}}, \forall j \geq i \tag{2.1}$$

where j denotes the column value. The best significance timescale is then identified by locating the row r and column c with the maximum value in matrix T. This translates to a subset of the light curve I beginning at *I(r)* and ending at *I(c)*. Having located the beginning and end of the brightest burst/excess, we can take the mean of the features in a similar way as for the other subclasses; however, this time only average those in the interval between pointings r and c. In addition to the already defined 12 features we decide to add, for this particular network, the value *T(r, c)*. This will be an indicator of the maximum significance achievable from the light curve. By definition the *TM* method will always locate a "burst" even if one is not there, even for faint persistent sources with no outburst. The method is meant to maximize significance, and as a result it will select all of the light curves for faint persistent sources and usually only select a small fraction of the light curve for fake excesses. For this reason one might think the method is biased; however, we note that this method is only employed to create an additional time-scale on which to merge the features; the classification of the excess will happen later in the network, which will discriminate between the real and fake excesses. The length of the outburst is not used as a feature and the coordinates i and j are not linear in time.

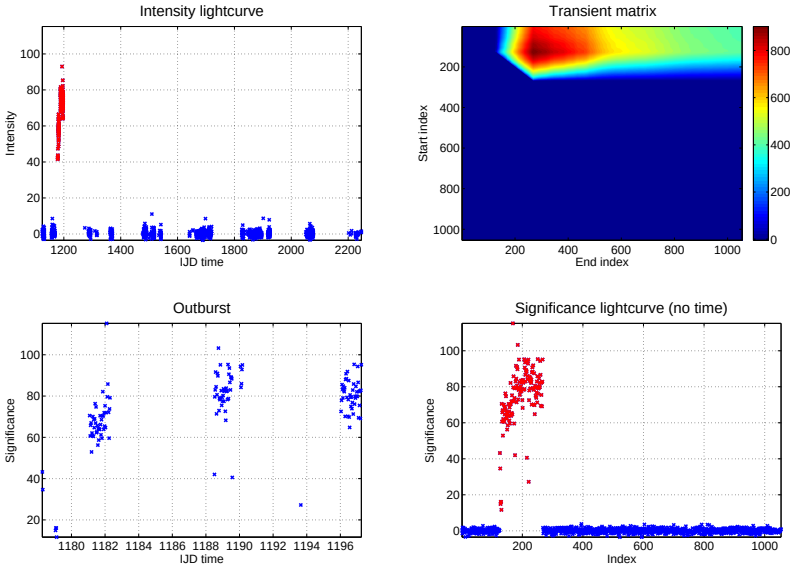
In Figures 2.8, 2.9 and 2.10 we show some examples of the transient matrix

Figure 2.8: TM applied to two transients.

4U1901+03



XTE J1550-564



technique applied to various different objects. Figure 2.8 presents the technique applied to two transients 4U1901+03 and XTE J1550-564, a high mass X-ray binary and a low mass X-ray binary respectively. For each object we show in the top left panel the lightcurve of the object and on the top right the corresponding transient matrix. The indices of the maximum value in the transient matrix are recorded in order to locate the beginning and end of the outburst. The selected datapoints are shown in red in the lightcurve. In the bottom left we show the selected outburst only, and for reference in the bottom right we again show the whole lightcurve, however without any time information (i.e. all data points are equally spaced in terms of there indices). In both cases the transient matrix performs extremely well at selecting all datapoints from the outbursts. The next examples shown in Figure 2.9 display again two transients, however this time recurrent ones. The first example, IGR J17464-3213, seems from looking at the lightcurve, as if there have been three main outbursts with smaller ones in between. This is also evident in the transient matrix, however we note that only the first one was selected by the method. This is because by definition, the technique locates the sequence of consecutive points that maximises significance, which in this case comprises only the first outburst. Also in Figure 2.9 we show the technique applied to yet another LMXB, Aql X-1. This source also has multiple outbursts as seen in both the lightcurve and the transient matrix, however, contrary to the previous example, the best significance here is obtained by using data from two outbursts with a quiescent stage in the middle. Both examples in Figure 2.9 have been chosen particularly because of their multiple outburst behaviour. It shows the potential of the transient matrix technique in locating multiple recurring outbursts, and not just single ones. One can then imagine how this method might also lead to the discovery of fixed timescales for outbursts in

some objects, and maybe more challenging period determination.

Finally in Figure 2.10 we show the method applied to a faint persistent AGN source, IGR J21247+508, and a fake candidate. In the case of IGR J21247+508, the method has selected all of the light curve since each data-point contributes to increase significance. On the other hand the technique applied to the fake source only selects a subset of the lightcurve. Both of these are typical examples of how the method performs on faint persistent sources and fake ones respectively. We note that for the AGN, the transient matrix smoothly increases to its maximum at about 37σ, whilst for the fake source the increase to the maximum value is very erratic, and the significance range for this candidate is only up to 6σ.

Having understood how our features will be merged we will now describe the creation of the testing and training sets before describing the classifier.

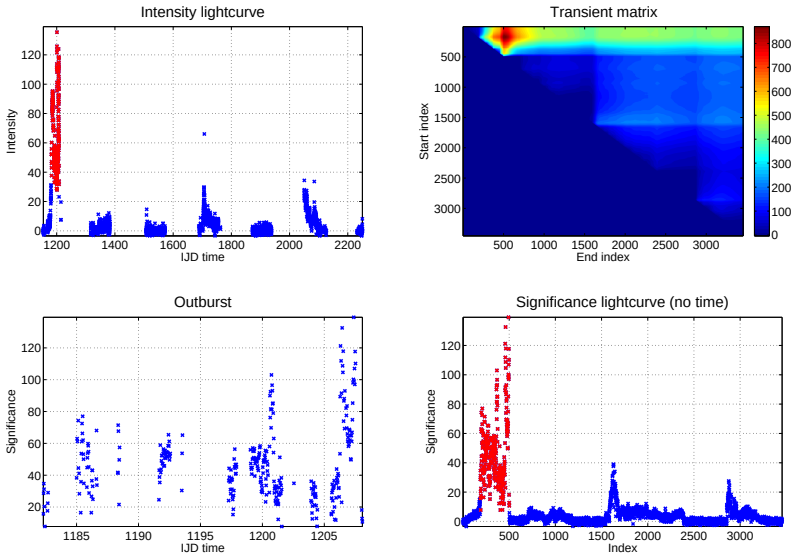
2.5.5 Training and Testing sets

Another important issue in building a reliable classification network is the choice of reliable training and testing sets. One has to make sure that neither of these are biased towards a particular type of subclass, for example, lots of faint AGN or lots of bright XBs or even worse not having any transients. This is one of the main reasons why we produce a training set and a corresponding classification network for each subclass. In this section we describe the methods employed to achieve this. Recall that after candidate filtering we end up with 7221 candidates of which 408 are present in catalogue 3. We now have to split our candidate list into training and testing sets. We have two reasonable options for producing unbiased sets.

1. Use the published second IBIS/ISGRI survey catalogue objects [Bird et al., 2007], with 209 sources for our training together with an extra ≈200 fake

Figure 2.9: TM applied to two recurring transients.

IGR J17464-3213



Aql X-1

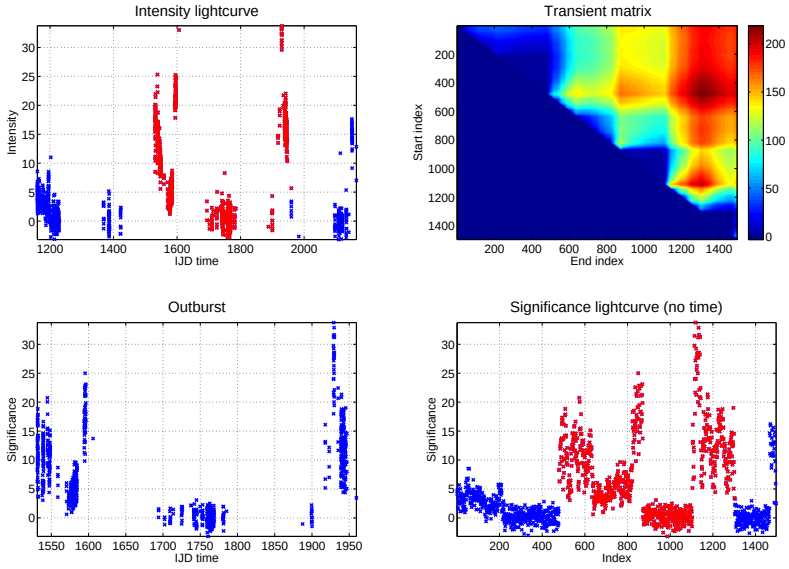
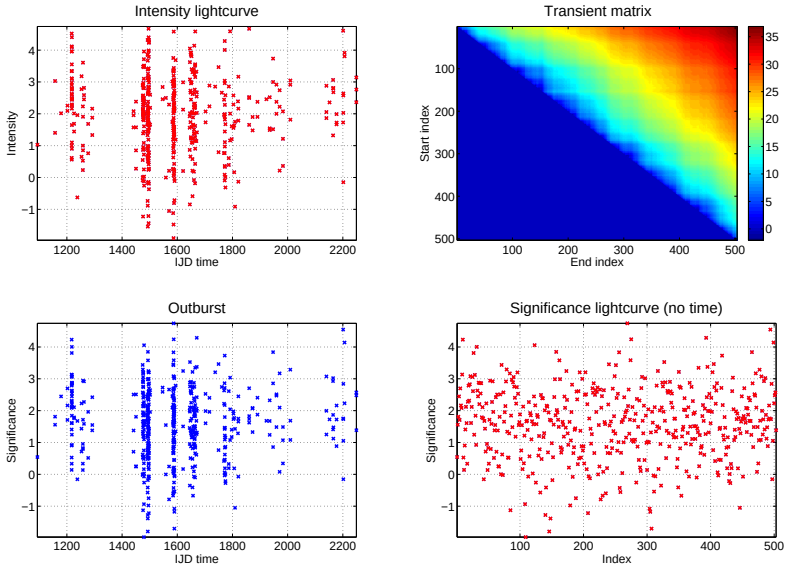
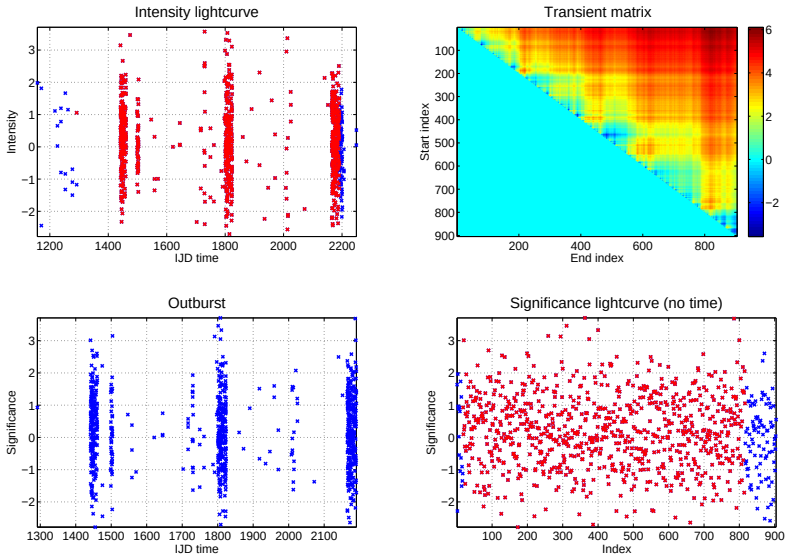


Figure 2.10: TM applied to a faint source and a fake candidate.

IGR J21247+508



Fake candidate

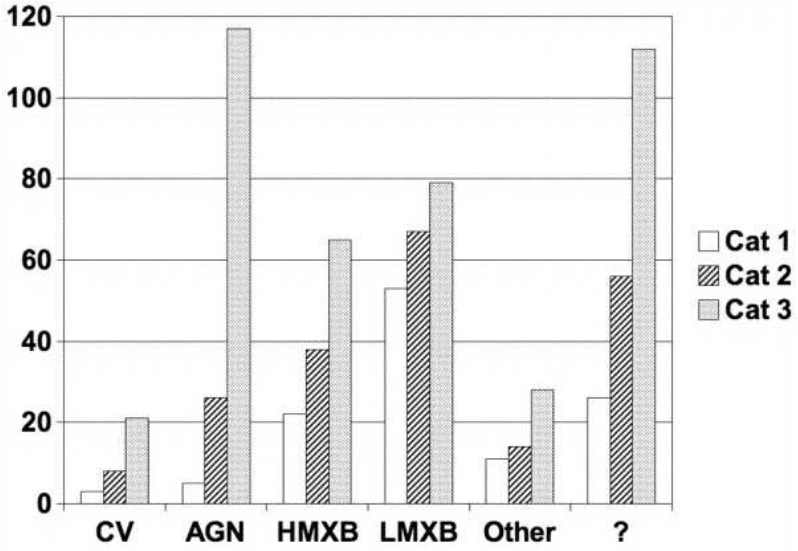


candidates and then evaluate how the network performs in recovering the published catalogue 3 objects.

2. Split the sky into two halves in galactic coordinates and use one half for training and the other for testing. In this case we would use catalogue 3 as a reference for what is a real source.

We decided to use option (2) as this will include some faint persistent objects only detected for catalogue 3 due to the longer exposure times and greater sky coverage compared with catalogue 2. Moreover, the source types between the two catalogues are not all the same (see Figure 2.11). Option (1) will be heavily biased towards detecting more luminous sources and this will bias the network too. Moreover, by employing option (2) we can train the network for the future creation of catalogue 4. More explicitly we would expect in future to use cat_{n-1} in the training for cat_n; however, we require a testing set to assess the network performance and only option (2) allowed for this in an unbiased fashion. From our initial 7221 candidate list we now have 220 real sources and 3114 fake candidates for our training from the western half of the galactic sky ($0^\circ < l < 180^\circ$) together with 188 real sources and 3699 fake candidates for testing the other half of the sky. We have chosen to split the galactic sky into west/east rather than north/south halves due to a greater similarity in the exposure times in the former case. One note of caution still needs to be addressed: some unknown fraction of what a human astronomer would consider real might actually turn out to be fake with future catalogue releases (and vice versa). This misclassification will affect the training set and therefore affect the final classification on the testing set too. Unfortunately one cannot know in advance what is real and what is not, therefore the only way to deal with this problem is having multiple iterations of the network to try and

Figure 2.11: Numbers of sources in the first, second, and third IBIS/ISGRI catalogues, classified by type



reduce the number of these false positives and true negatives. We note that including sources that we are certain are real would bias against the fainter sources. The extra iteration step is not performed at this stage, but will be considered for future catalogue releases.

2.5.6 The Random Forest Algorithm

As mentioned before, when building a classification network one has to take into account the nature of subclasses present in our general sets. For example one would have very limited success in correctly identifying transients if the training set only consists of faint AGN and vice versa. We have therefore decided to build three independent random forests with the three sets of features described in Section 2.5.4. One will be trained only on AGN using the faint persistent set of features in order to recover faint persistent objects. A second set of forests will be trained only on XBs using the strong persistent set of features to recover bright objects, and a third set of forests, only trained on

transients, using features selected by the transient matrix method. This should then allow us to recover in one or more forests all other types of sources that do not necessarily fall into the AGN/XB/transient subclasses. We reiterate that the algorithm is meant for source identification only, but as shown later, will turn out useful in source classification as well.

2.5.7 How to build a Random Forest

Classification tree methods are a good choice when the data mining task is classification or prediction. The goal of any single tree is to generate discriminatory rules that can be easily understood. Trees are constructed through a process known as binary recursive partitioning, an iterative process of splitting the data into partitions, and then splitting it up further on each of the branches [Brieman et al., 1984]. We employ various classification trees in what are called random forests, devised by Brieman et al. [1984]. Essentially we build many classification trees, each tree casting a "vote" on a particular object. We will build three sets of random forests using the features and training sets described previously. The final judgment as to what particular class the object is in will then be decided by the number of votes it received in each of the three forests. We reiterate that our goal is not actually to classify source types, but to maximize the efficiency for real/fake decision making.

If our training set consists of M input variables (features), we will randomly choose for each tree a value $m \ll M$ of variables such that each tree will be grown using only those m variables. The value of m is held constant for all trees grown for each subclass and is one of only two variable parameters in the network. It is responsible for two things. Increasing m increases the correlations between any two trees in the subclass forest, thus decreasing its recognition strength. On the other hand increasing m increases the strength of

any one individual tree. A tree with a low error rate is a strong classifier; however, increasing the strength of the individual trees increases the forest error rate. Reducing m reduces both the correlation and the strength. Somewhere in between is an "optimal" range of m which is usually quite wide. The other variable parameter with random forests is the number of trees to be grown. This has to be quite large in order to be able to use all M variables through bootstrapping. There is no limit on how many trees we build in the forest as the algorithm does not overfit [Brieman et al., 1984].

There are several reasons why random forests were used for our classification purpose. When building the network one of the main concerns was with dealing with very large data sets. Even though he IBIS/ISGRI data set is not so large (yet!), the method presented here can deal with much larger sets. Further reasons are listed below.

- It can handle thousands of input variables.
- Generated forests can be saved for future use on other data.
- These capabilities can be extended to unlabeled data, leading to unsupervised clustering, data views and outlier detection.
- It has the potential to give estimates of what variables are important in the classification.

Once a random forest has been built for a particular subclass of objects we can classify the testing set by asking how many trees in the forest will "vote" for that particular excess. Using this voting scheme allows us to have a feel for how confident the random forest is at assessing a particular candidate (as will be shown in the results section). If any particular excess gets enough votes

in any of the networks, then it will be considered as a good candidate worth inspecting.

Training

As mentioned before, our aim is to build three independent random forests in order to identify each subclass of objects separately. Recall that we have 220 real sources and 3114 excesses available for training (one half of the galactic sky). To build each one of the three training sets we use the classification types of the real objects published in the IBIS/ISGRI third catalogue. In the faint persistent case we use the 73 AGN in the western galactic hemisphere together with 3114 fake excesses for our training. We cannot use all the fake candidates for a single tree or else it would bias our classification. Instead, for each tree grown in the forest, we keep the same training set of 73 AGN for our real sources and randomly pick 73 fake excesses from our pool of 3114. This ensures that no individual tree is biased towards recognising too many fake excesses, while still incorporating a wide range of them. By having this "pool" of fake excesses to choose from, we essentially ensure no two grown trees are the same, avoiding overfitting. As mentioned in Section 2.5.7, the only variables in our random forests are the values m and the number of trees. Thus for each available set of features for a particular energy band we will choose a value m together with the number of trees to grow. For example in the faint persistent network we mentioned already the use of three energy bands and two sets of features per energy band (ScW average merging and final mosaic features). We have chosen the number of trees per set to be 200 in this case, yielding a random forest with $3 \times 2 \times 200 = 1200$ trees. The value of m (the random subset of features used per tree) was set to 8 for the average features and set to 3 for the mosaic features. These values were achieved through trial and error

Table 2.2: Summary of the number of trees used within ISINA.

	Faint	Strong	Transient
Number of trees	200	200	500
Number of energy ranges	3	4	4
Number of sets of features	2	2	1
Total number of trees	1200	1600	2000

by maximizing the accuracy of the final output given by the testing set.

In the XB case we use 46 XBs, again from the western galactic hemisphere and use the same technique as for AGN in dealing with the fake excess training set. In this case we chose the same value for m and number of trees; however, for this network we decided to include one extra energy band, yielding $4 \times 2 \times 200 = 1600$ trees. Similarly for transients we use 32 transients for training. This network however was chosen to have a value $m = 7$ and the number of trees grown per set was set to 500. This might seem very large but was used in order to have more bootstrapping from the fake candidates given the low number of transients in the training set. This yields $4 \times 1 \times 500 = 2000$ trees. We point the reader to Table 2.2 for a table showing the parameters used.

Testing

Recall that in our testing set we have 3887 candidates, of which 3699 are fake excesses and 188 real sources. In this section we will inspect how these candidates perform within the three independent networks. Note that all three networks had exactly the same testing set. In order to assess the recovery performance of each of the networks we will look at how many trees voted for a particular source within the forest. If a candidate achieves 50 per cent or more of the votes then it will be considered as "recovered". For clarity, the analysis described here is illustrated in the flow diagram in Figure 2.12, which includes the number of candidates in both training and testing sets for the

three networks.

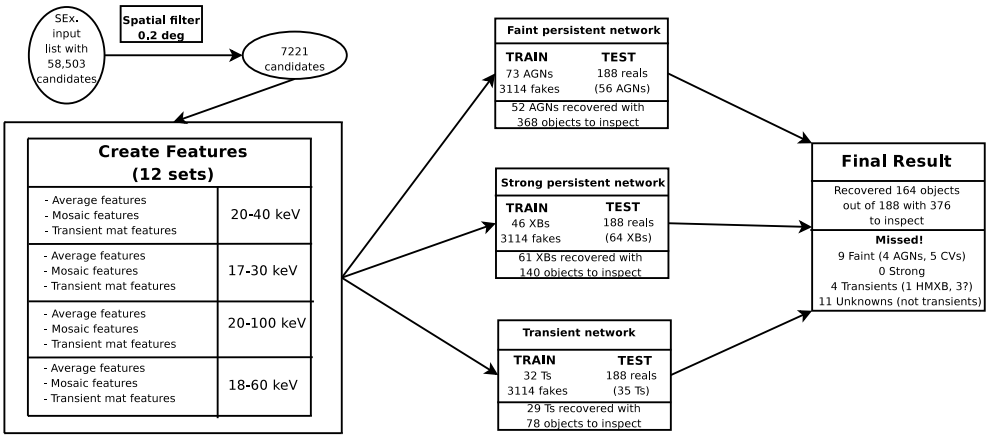
The testing set for the faint persistent network contained 56 AGN of which 52 were recovered with the 50 percent cut. The missing four AGN were marginally below the recovery threshold in the faint persistent network. Moreover, by definition this is the network that recovers most objects. In fact a 50 per cent threshold yields 368 candidates out of the initial 3887. A lot of these will be strong persistent sources; however, most will be unidentified faint persistent objects.

The strong persistent network on the other hand performed slightly better in that it achived a lower ratio of false positives. This however is not surprising as bright sources are more easily discriminated against faint ones. Out of 64 XBs in the testing set, only three were not recovered within this network; however, as we will discuss later, these get recovered in the faint persistent network. The number of candidates to inspect with the 50 per cent threshold is 140, approximately half of that produced by the faint persistent network.

Finally, the network producing the lowest candidate list to inspect is the transient network. This yields 78 candidates to inspect with the usual threshold. Out of 35 transients in our testing set, six were not recovered in this network. Of these, two were recovered in the faint persistent network.

The final box in Figure 2.12 shows the break down of missed objects. Clearly most are unidentified; however, CVs are also poorly recovered. As will be discussed later, this can easily be caused by not training a network for these specific source types, or they are some of the faintest and/or narrowest spectral range.

Figure 2.12: Graphical flow diagram of the steps involved in the classification network from the extraction of the initial source candidate list to the final result. SExtractor is run on all revolutions, revolution sequences and final mosaic images. A 0.2° merge radius is applied reducing the list to 7221 candidates. Features are extracted for these on an ScW level and on the final mosaic. The features are then merged in three different ways and passed to three different networks accordingly for faint and strong persistent and transients sources. In each of the network boxes we display the number of objects in the training and testing sets. In brackets we have the number of objects from the testing set in the respective subclass. Below each network box we show the result on the testing set using a 50% cut on the tree votes. The final result box also applies on the testing set with a 50% global cut. There we also show the missed objects and their break down.



2.6 Results

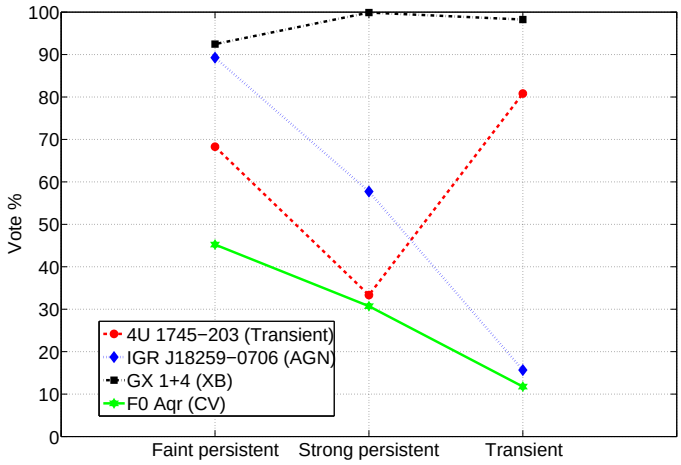
This section will explore the results from ISINA when applied to the testing set. We first describe a few individual examples which will also show how some of the metadata produced by ISINA can be a useful aid in the final inspection phase and then move on to describe the global properties of the whole sample and its efficiency.

2.6.1 Individual examples

While Section 2.5.7 described the performance of the network in terms of retrieval of catalogue 3 sources, here we try to quantify the network performance in more detail. First we show in Figure 2.13 some examples taken from different source class categories to illustrate some of the outputs from ISINA. For each candidate we show the vote percentage obtained in the three networks. Note that a candidate is considered recovered by ISINA if it obtains more than 50% votes in any network.

The first example is the faint persistent AGN IGR J18259-0706. This is a new source detected in the third IBIS/ISGRI catalogue with a maximum detection significance of $\approx 5.1\sigma$ in the 18-60 keV band and a relatively high 1570 ks exposure time. This puts it firmly in the faint persistent category. The blue curve in Figure 2.13 shows the percentage of votes as recorded by the three networks. It can clearly be seen that this particular example receives a greater "confidence" from the faint persistent network, where the curve peaks at 90 per cent. This will be the source's "global" vote as explained later on in this section. We note that at this point the network can also be interpreted as giving information about class characteristics and not just identification. In this particular example it is clear that IGR J18259-0706 is classified as a faint

Figure 2.13: Voting percentage obtained by four real sources within the three different networks. Black dot-dashed line: LMXB GX 1+4, blue dotted line: AGN IGR J18259-0706, red dashed line: LMXB 4U 1745-203, green solid line: CV FO Aqr.

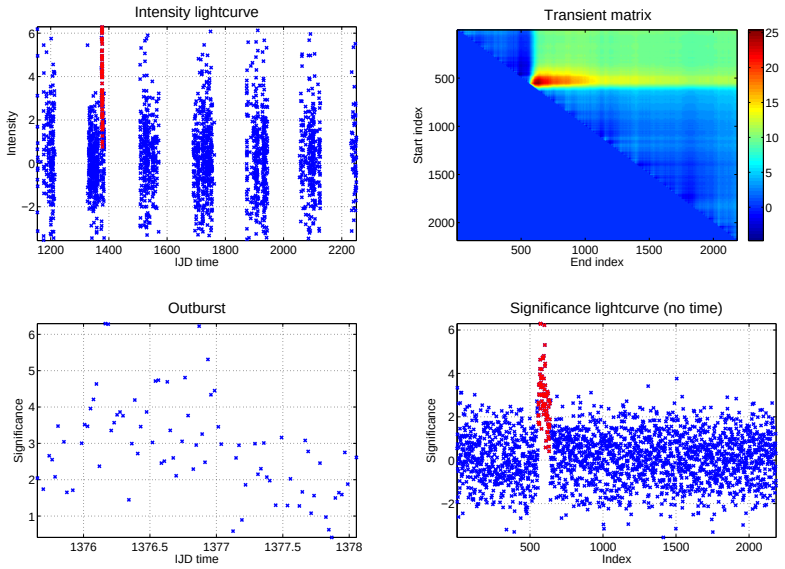


persistent source, as the network which uses all the information available has achieved the largest number of votes: the faint persistent network.

The next example chosen is the strong persistent low-mass X-ray binary (LMXB) GX 1+4 shown in black in Figure 2.13. The system was detected in the third catalogue with a maximum significance of 544σ in the 18-60 keV final mosaic. Note the voting percentage difference between strong and faint persistent and transient network is very small. This is the case for most strong sources but, as observed in the previous example, not for the faint ones. In fact really strong sources tend to have high vote percentages in all three networks essentially because they are detectable on any time-scale and in all energy bands. Realistically, we only need to identify persistent versus transient. Information on how bright they are is best obtained with other methods.

Another example chosen is the LMXB 4U 1745-203. The source was detected in the third catalogue at a significance of 20.7σ in the 20-40 keV band

Figure 2.14: Graphical view of the TM method applied for 4U 1745-203.



mosaic for revolution 120. Again, just by inspecting its corresponding red curve in Figure 2.13, we can get an idea of what type of object we are dealing with had we not known in advance. The system obtains the highest score in the transient network with >80 per cent. For this particular example we also show its transient matrix in Figure 2.14. It can be seen that the outburst has a relatively low detection significance in any individual pointing; however, from the result of the transient matrix, the maximum significance obtained in the selected timescale is 22.4σ. The source was in outburst for $\approx$3 days, reaching a flux of $\approx$115 mCrab. The difference between the two detection significances is due to the fact that the transient matrix has localized as an outburst a subset of the pointings of revolution 120 instead of using them all.

The final example chosen is the CV F0 Aqr, another weak new detection in catalogue 3. The source obtained a significance of 4.8σ in the 20-40 keV band for a 85-ks exposure. This particular candidate did not achieve enough votes

to be included in our "recovered" list; however, it appears from the percentages obtained in the three networks that this is a faint persistent source. This kind of analysis can help identifying new sources even if the vote count has not passed the identification threshold.

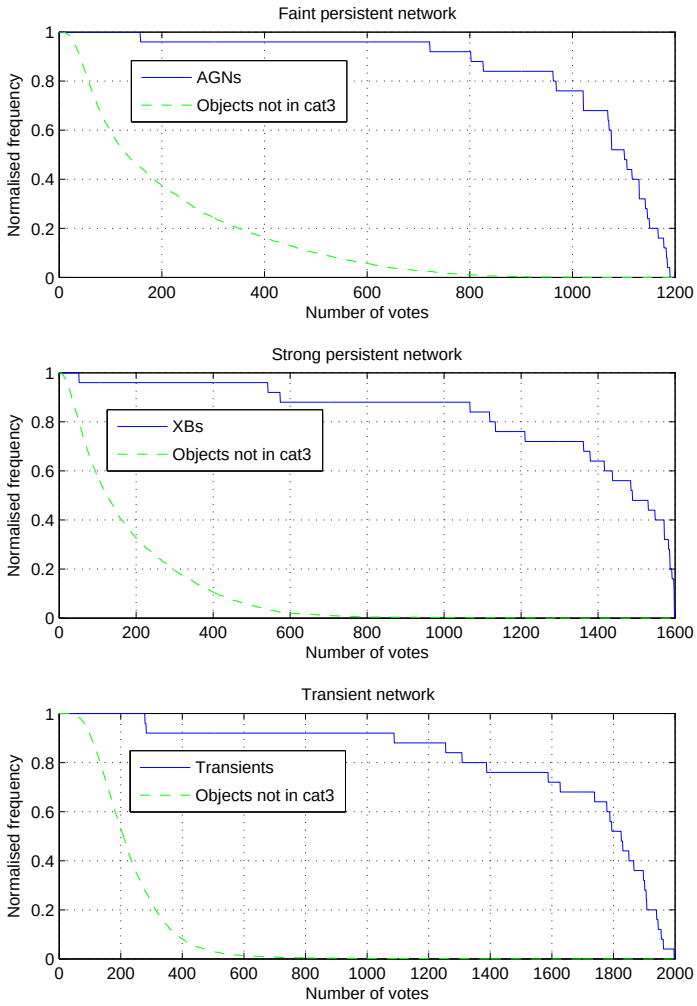
2.6.2 Global results

We will now take a look at the global results from ISINA on the testing set. Figure 2.15 shows the cumulative distribution functions obtained from the three networks on the testing set. On the x-axis we have the number of votes for the network in consideration. A high vote count implies candidates are "real" according to the network, a low vote count rate implies the opposite. As can be seen all three networks perform relatively well in recovering their respective "real" objects; however, contamination from the fake candidates is still present. This can be noted in the worst case scenario for the faint persistent network. This is somewhat expected as the training set for this network is by definition highly populated by low significance sources. However, the transient network has performed quite well in recovering the majority of real sources whilst excluding fake ones much more easily.

In order to assess the overall network performance and compile a candidate catalogue produced produced by ISINA we have to merge the results from the three networks. This is simply done by transforming the vote number for each network into percentages. Once this is done we can merge the results as a function of vote percentage as shown in Figure 2.16. This time the blue line represents any of the initial 188 objects found in catalogue 3 (testing set). For any candidate, the highest percentage in any of the three networks is used as a "global" percentage.

From visual inspection of Figure 2.16 one can see that any object with 90%

Figure 2.15: Cumulative distribution functions for candidates within the three networks as a function of number of votes obtained. Blue solid line showing corresponding subclass objects present in the third IBIS/ISGRI catalogue. Green dashed lines show fake candidates. In order to estimate performance we can draw a line through the graphs at half the total number of votes and infer the number of recovered objects. For example in the faint persistent case, drawing a line at 600 votes would recover over 90% of the real AGNs ant the cost of including 10% of the fake candidate population.

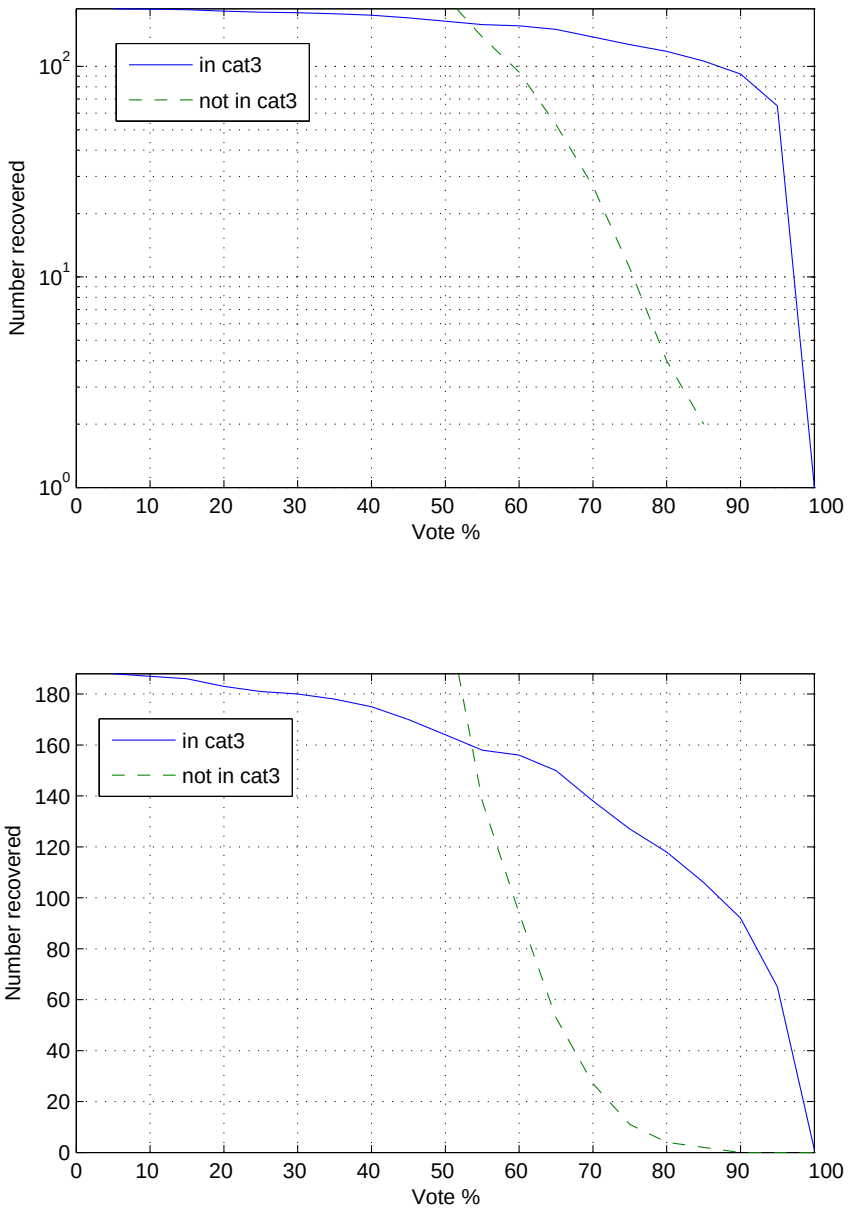


votes in any network or more will certainly be real. This includes $\approx 50\%$ of the 188 real sources in the testing set. We can now query the network for the sake of reducing the amount of visual inspection for the compilation of a new catalogue. For example, if we only visually inspect and assess all candidates between 50% and 90% in Figure 2.16, this yields 284 objects (in addition to the 92 already accepted with $> 90\%$ votes). From these we have 73 belonging to the published catalogue 3. The remaining 23 objects that have less than 50% of the votes will be tricky to locate with this method, as below 50% of the votes the number of fake candidates grows very rapidly. We note however that most of these 23 objects are very low significance, unidentified, sources, which might even turn out to be fake excesses in future catalogue releases. On the other hand we also note that some of the fake excesses with high vote rates might turn out to be real upon further investigation.

2.7 Discussions

We have developed a reliable algorithm to aid the production of future IBIS/ISGRI gamma-ray survey catalogues. The algorithm will help produce less subjective catalogues, unbiased by human intervention. Meant for source identification, ISINA has also turned out to be useful in discriminating source types. We have shown how to automate the task of selecting and reducing a set of candidates from IBIS/ISGRI images. The distribution of recovered objects, sorted by type, together with the objects published in the third catalogue present in the testing set are shown in Figure 2.17. It is clear that the majority of objects are recovered correctly with a 50% global vote threshold. It is interesting to note that the only populations to suffer from a substantial decrease in recovered objects are the CVs and the unknown source types. The drop in the number of

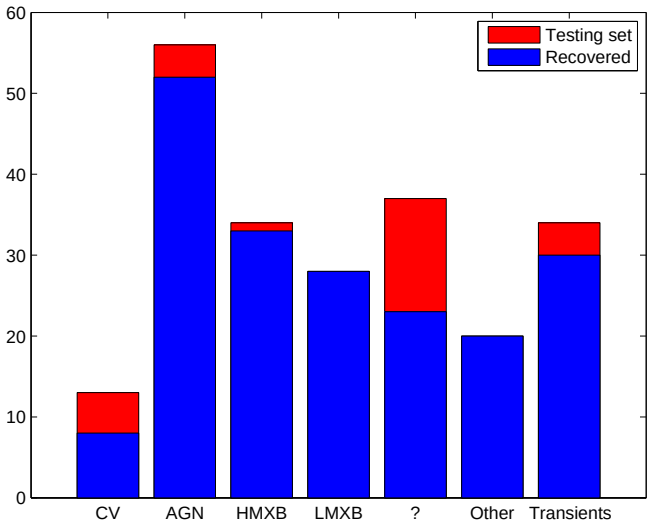
Figure 2.16: Distribution of recovered objects present in the third IBIS/ISGRI catalogue (blue solid line) and recovered objects not present in the latest catalogue (green dashed line) as a function of global vote percentage.



CVs can easily be explained by the fact that most of the non-recovered ones lie in crowded regions, where the systematic noise is greatest. Moreover we point out that, as later mentioned in Chapter 4, these objects have great spectral differences between them, including some very soft sources. However, missing objects in these regions does not cause a big problem for the creation of future catalogues. This is because crowded regions will be the most inspected ones, so that if the network does not recover certain objects, human intervention will. The other population to have a significant drop in the number of recovered objects is the unknown category which is a bit less trivial to assess. This is because, by definition, the only real way to determine their nature is to have longer exposure times for the regions where these are present. We also point out that both the CV population and the unknown one was not part of our training set, which might also explain the relatively low recovered rate for these. This may also have an impact in our final result as the three networks have now specialized in recovering their subclass of objects. One last observation of the general behaviour of the network on the testing set is that despite the fact that the remaining classes perform well within the network, it has to be pointed out that our training and testing sets might have misclassified objects within them (false positives and true negatives). Given the nature of the classification task, the training set will always be biased towards this. However, given the extremely fast data growth the problem can only get better, and these small discrepancies will systematically reduce.

It has to be pointed out the potential of such a network for exploratory data analysis in other wavelengths. The networks described here can be easily tuned to deal with different images, taken from different observatories. The features defined are quite generic, and anyway may be adjusted according to the new data set, probably suffering from different systematic effects than the

Figure 2.17: Number of sources in the testing set classified by type (red) and recovered objects using a 50% global vote cut (blue). Note that objects in the transient category are also present in their respective class types.



ones presented here.

Chapter 3

Building Catalogue 4

"The real danger is not that computers will begin to think like men, but that men will begin to think like computers."

– *Sydney J. Harris*

In this chapter we will describe the creation of the latest IBIS/ISGRI catalogue release, catalogue 4. The steps involved are similar to the production of the previous catalogue 3, however with some differences. In particular transient detection is addressed in a more complete way since the larger time-span of the dataset has grown dramatically. Firstly we describe the mosaic creation stage which, similarly to catalogue 3, includes three timescales: revolution, revolution sequence and all-archive. Additionally a new timescale to identify transient sources has been introduced.

This chapter is layed out into two main sections. One describing the "human" way of constructing catalogue 4 and one describing the "machine" way using ISINA. In the last sections will be some comparisons between the resulting catalogues after using both methods. We point out however, that contrary to chapter 2, where a testing set was available for us to assess the ISINA performance, here we do not have such a set. As a result a clear examination of

the ISINA performance is not possible. Nonetheless we examine the pitfalls of both methods and try and suggest future improvements for ISINA.

3.1 Mosaic Construction

This section will describe how mosaics for the IBIS images have been constructed during the production of catalogue 4. These mosaics will also be used by ISINA as we will see later on and are therefore common to both methods, the "human" method and the "machine" method.

First of all we need to discard a subset of ScW with poor image quality. This can be caused, for example, by a solar flare event during the observations. This is done by using the image rms as an indicator of image quality. Filtering is therefore applied based on the image rms, such that the rms should not exceed 2σ above the mean image rms for the whole dataset. Moreover data taken in "staring mode", even if processed, are not used in the construction of the final sky mosaic image. After removal of high-rms and staring data, approximately 36,000 ScWs remain in the dataset, totaling $\sim$ 70Ms of exposure time.

The selected ScWs are then mosaicked using a proprietary tool developed in Southampton, optimised to create all-sky galactic maps based on a large number of input ScWs. Mosaics were constructed in five energy bands similarly to the catalogue 3 mosaics and four projections. Four timescales were employed for the mosaic construction. Firstly maps were created for each satellite revolution (approximately 3 days). This timescale is optimised to detect sources active on a timescale of one day. Secondly, we identified 32 sequences of consecutive revolutions which had similar pointings. These *revolution sequences* could therefore be best analysed as a single observation, and provide sensitivity for sources on longer timescales than revolutions ($\sim$weeks).

Ultimately, persistent sources will be best detected in an all-archive accumulation of available high-quality data. However the problem of high exposure time and long timebase spanned by this latest dataset has worsened the detection of transient sources. This last search method has been optimised for the detection of persistent flux sources, therefore a highly variable source which would be clearly detected during outburst will have an undetectably low flux when analysing the full dataset. For this reason we searched for the optimum detection timescale for known sources or candidate sources. This is a newly introduced timescale for which maps have only been created with this latest catalogue 4 data. This new timescale search is based on a new metric defined in Southampton, namely the *bursticity index*. This is defined by creating a light curve for each candidate source in the 18-60 keV band, and then scanning a variable-sized time window along each light curve. The window length is varied from 0.5 days to the length of the whole light curve, and all the points within the time window are included in the analysis. The duration and range of times over which the source significance is maximised is recorded. The bursticity index is then defined to be the ratio of the maximum recorded significance over the significance of the whole lightcurve. Thus a bursticity of 1 defines a persistent source, whilst a bursticity greater than 1 implies that the significance of a source can be increased by omitting some observations from the analysis, presumably when the source is in quiescence. Sources with a high bursticity are then selected and mosaics are specifically created using the subset of pointings selected using the bursticity method. This will allow faster transient sources to be identified.

3.2 The Human way

We now move on in describing the process by which the latest catalogue 4 has been constructed. This has been a long job for the whole IBIS survey team, and here we only try and give a brief overview. Firstly we will see how a list of potential candidates is extracted from all the mosaic maps described in the previous section. Then the problem of deciding which candidates are real and which ones are fakes is addressed. Similarly to ISINA, where many trees vote on a particular candidate to decide its classification, many astronomers vote on each candidate to assess its reliability. The only main difference in the voting scheme is that all astronomers have to agree on a candidate source before it is considered as real (although all sources with a mixed vote are reviewed a second time), whilst in the case of ISINA we only require more than 50% of the trees to have the same vote. Having described the catalogue production procedure we will review the main results from the catalogue, and briefly compare this to the previous catalogue releases.

3.2.1 Selecting candidates

Source searching has been performed initially on the mosaics described in Section 3.1. In total 11,500 maps were created at this stage. All mosaics were searched for sources using two methods:

- the SExtractor 2.5 software [Bertin and Arnouts, 1996]. The source positions measured by SExtractor represent the centroid of the source calculated by taking the first order moments of the source profile (referred to by SExtractor as the barycentre method) Source detectability is limited at the faintest levels by background noise and can be improved by the application of a linear filtering of the data. In addition source

confusion in crowded regions can be minimised by the application of a bandpass filter. To this end, the *mexhat* bandpass filter is used in the SExtractor software.

- a proprietary "peakfind" tool developed in Southampton which employs a basic iterative removal of sources technique, combined with an assessment of the local background rms to reduce the false detection of sources in areas of the map with high systematic noise structures (see Section 2.2), mainly in crowded regions and around the brightest sources.

A list of candidate sources was constructed by merging the $> 4\sigma$ excess list from each mosaic, using a merge radius of 0.1°[1]. A source had to be detected by both methods in order to be included in the candidate list. Moreover manual inspection has been performed on the rare occasions where SExtractor fails due to the close proximity of two or more sources, and any additional candidates found were included. Also all previously identified *INTEGRAL* sources were added to the list of candidates. This resulted in 1266 excessesw which were passed on to the next stage of the analysis.

3.2.2 Deciding on candidates

At this stage a list of excesses has been created and we need to discriminate between the real and the fake excess and a number of steps are performed in order to minimise the possibility of false catalogue entries. These methods are designed to take into account both the statistical fluctuations (which we can to some extent assess) in the maps and systematic effects present in the maps, which are much harder to quantify.

Firstly, each source is manually inspected by a number of people (including

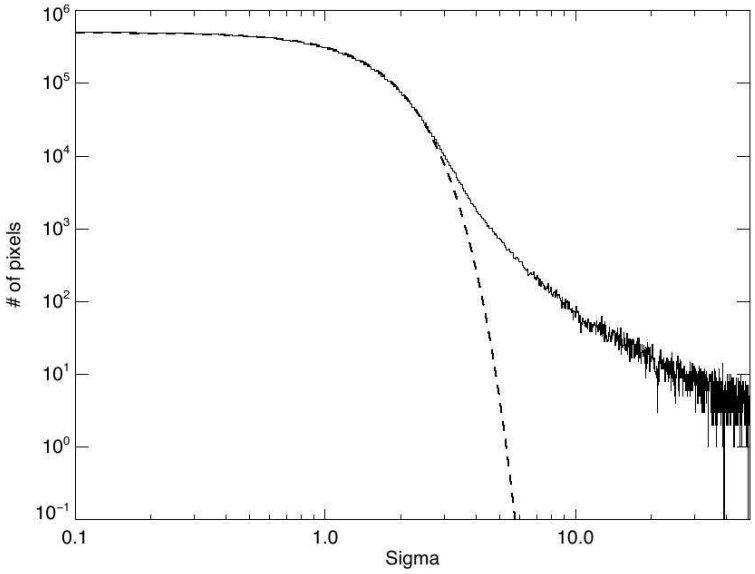
[1] Discussed in more depth in the next section.

myself!) experienced with working on IBIS/ISGRI images. These inspections cover aspects such as PSF shape, consistency across multiple energy bands and the significance of the source relative to the local noise levels in the maps, similarly to the features defined for ISINA. We require an unanimous agreement among many viewers that the candidate excess is a true source, a very conservative approach, but one designed to minimise as much as possible the false detection rate. For reference with ISINA, this stage takes approximately 8 months to complete.

Moreover, a flux-exposure analysis has been carried out in which each detected flux has been compared to the predicted minimum detectable flux for the exposure of the candidate in question. Sources for which the mean flux is much lower than that which could reasonably be detected in a corresponding timescale may have been boosted by systematic effects, or may just be an outlier in the statistical fluctuations of the maps. In either case the excess candidate is rejected.

The last step in creating catalogue 4 is based on the detection significance, similarly to the production of catalogue 3. To this end, a histogram of the individual pixel significances produced for each of the mosaics. One of these histograms is displayed in Figure 3.1. A Gaussian fit with mean 0 and standard deviation of 1 is found to be a good representation of the noise distribution. Looking at the pixel significance distribution across all mosaics we can confidently conclude that $<< 1\%$ of the pixels found at significances above 4.8σ are produced by the statistical noise distribution. Furthermore, in the 18-60 keV all-sky mosaic, of the pixels found between 4.5σ-4.8σ, $< 6\%$ originate from the statistical noise distribution. However we point out that these limits are based on the global properties of the mosaics and maps containing systematic noise, localised to specific regions, and owning the same characteristics as the real

Figure 3.1: Distribution of individual pixel significances found in the 18-60 keV catalogue 4 all-sky mosaic. The solid line represents the data; the dashed line represents a a Gaussian fit to the noise distribution. Image taken from Bird et al. [2009].



source population.

3.2.3 The Final Human Catalogue

The final IBIS/ISGRI catalogue contains in total 723 sources. Figures 3.2 and 3.3 show the evolution of the numbers of sources through the 4 IBIS/ISGRI catalogs. A continuous increase from the first IBIS survey release can be noted, with particular emphasis on extragalactic sources, rising from only 4% of the detected sources in 2005 to 35% in the latest source list (see Figure 3.3). We believe this number will continue to rise once follow up of the currently unidentified sources can be initiated. Clearly the sources dominating the catalogues are strongly linked to the sky coverage. *INTEGRAL* has spent the first 4 years more on the plane and in particular in the region of the Galactic Bulge while more recently the high latitude sky has been exposed more thoroughly.

Figure 3.2: Evolution of source type and number through the 4 IBIS/ISGRI catalogues produced to date. Image taken from Bird et al. [2009].

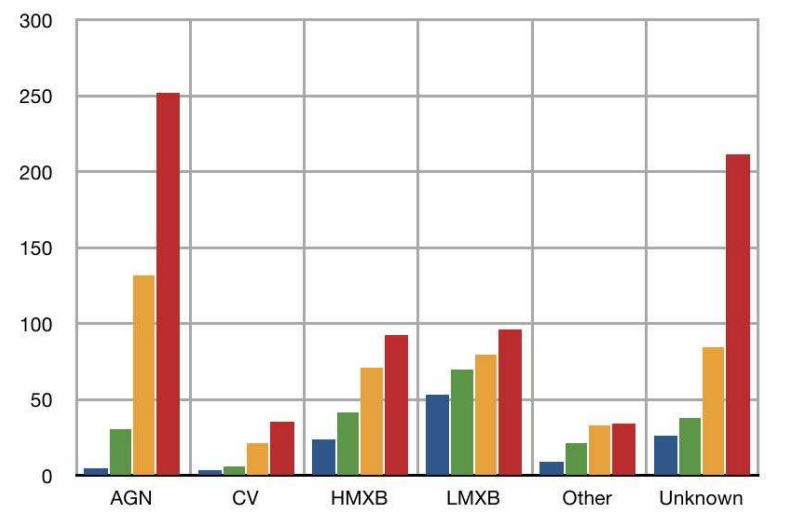


Figure 3.3: Classifications of sources in the 4 IBIS/ISGRI catalogues produced to date. Image taken from Bird et al. [2009].

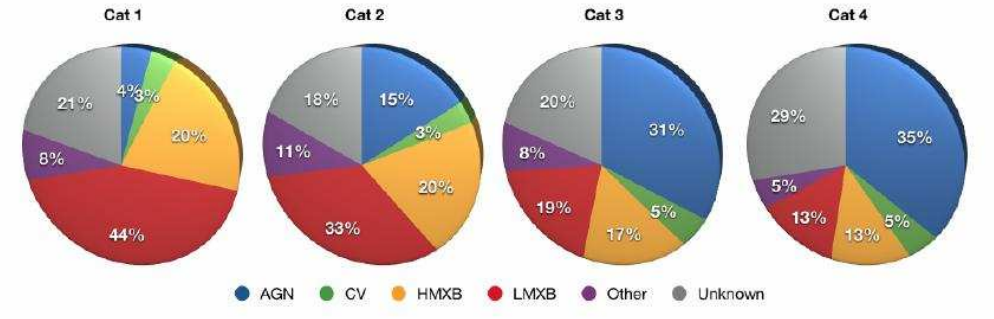
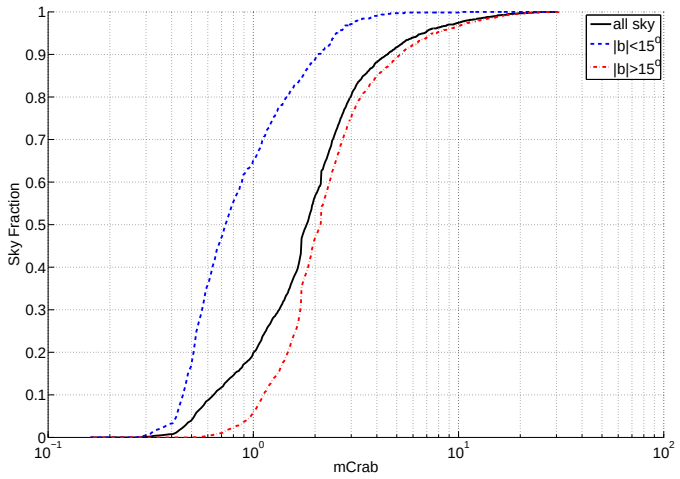


Figure 3.4: Sky fraction as function of minimum detectable flux.



The minimum detectable flux of the survey can be estimated as a function of the sky position (Figure 3.4) based on the accumulated exposures. Due to the non-uniform exposures covered by the mission, the sensitivity of the survey is still strongly biased. In particular the region of the Galactic Plane ($\approx 70\%$ of the sky) is covered to better than 1mCrab sensitivity, while 90% of the extragalactic sky is now covered at the 5mCrab level.

There are now 331 new sources when compared to the third catalog. Of these, ≈ 120 are associated with extragalactic sources, while only ≈ 25 are associated with known Galactic sources, and the remainder are so far unidentified. This could mean that *INTEGRAL* has reached its sensitivity limits, and is now primarily detecting extragalactic objects. However, the sky distribution of new sources (Figure 3.5) shows a rather different picture. Superimposing the new sources onto the delta exposure (i.e. the increase in exposure since the third catalog) shows how the new detections follow the new exposures, still comprising a very significant Galactic component. We are therefore forced to conclude, that while the extragalactic observations are at a sensitivity limit where IBIS

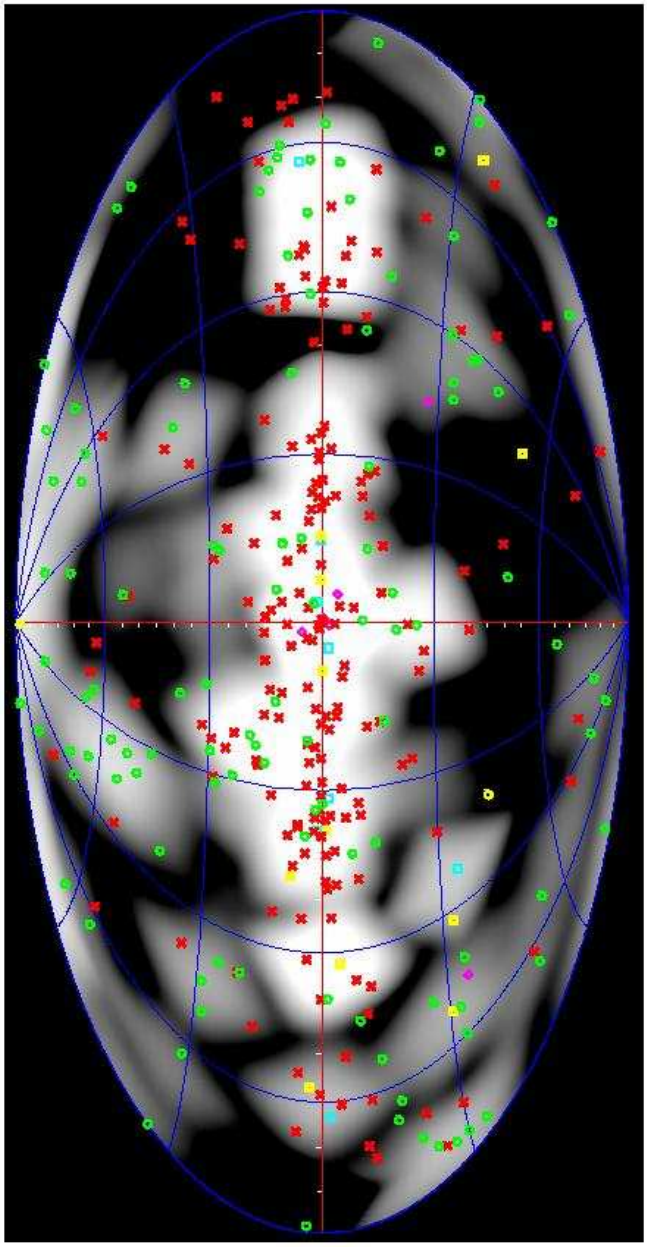
is still re-detecting known objects, the observations near the Galactic Plane have reached a level of depth where previous X-ray observations are no longer always able to provide associations for the new sources. Combined with the variability of the Galactic sources, this is a clear indication that further observations of the Galaxy will continue to uncover new sources, and follow-up of these new sources is of critical importance. However, many of the new sources found in the Galactic Plane by *INTEGRAL* have been identified as AGN, so this separation of Galactic and extra-galactic sources is not a straightforward one.

From this catalog, we can state that the detections above 4.8σ are drawn from an ensemble of maps, all of which show statistical quality that indicates much less than 1% of the excesses above that level will be false detections. Of the 40 sources below 4.8σ, half are associated with known X-ray emitters, and the estimated $\approx 6\%$ false detection rate should result in a total number of false detections in this catalog of no more 10, with the vast majority drawn from the sources detected below 4.8σ.

3.3 The Machine way

Here we explore how ISINA, presented in Chapter 2, can be used for the production of future catalogue releases. We will give an overview of the required inputs for the algorithm, train it based on catalogue 3 (whole sky), and produce a list of candidates. Throughout this section we will not use any of the information accumulated during the visual construction of catalogue 4, as this may bias our training. The comparison between the candidate list created visually and that produced by ISINA will then be explored in the next section.

Figure 3.5: Map of incremental exposure since the third catalog, showing the locations of the new sources found. Key: Green circles = AGN; Cyan squares = HMXB; Magenta diamonds = LMXB; Yellow boxes = CVs; Red crosses = Unknown.



3.3.1 Using ISINA for cat4

The first task in running ISINA is that of locating candidates from the IBIS maps. Recall from the introduction to this Chapter that for catalogue 4, similarly to catalogue 3, mosaics were created for revolutions, revolution sequences, and all data. Additionally burstmaps have been created in order to best maximise the chances of locating transients.

We begin our candidate searching procedure by running SExtractor on all created mosaics. This yielded an initial list of 141,956 excesses. Similarly to Chapter 2, this list will contain many duplicates, and far more fake candidates. We reduce this list in the same way as previously mentioned (see section 2.5.2), by excluding all excesses only present in one map. However now, given the experience on applying ISINA to catalogue 3, we have decided to change our choice on the merge radius, from 0.2° to 0.1°. This will allow us to begin with fewer candidates (see Figure 2.6), however mildly increasing the number of fake sources associated with "propeller"-like structures. The advantage of reducing the merge radius will be explored in more detail later. From the final stage of ISINA, where some visual inspection is required anyway, these fake structures can easily be localised and removed, since they are so close to real sources. The initial excess list thus reduces to 9931 candidates, which includes all of catalogue 3 objects, separated more than 0.1° apart, and is shown in Figure 3.6.

The next stage is that of extracting the features and merging them appropriately for all candidates. All that is required for this stage is the positions of candidates taken from above. From those positions all parameters from Table 2.1 are created in about 7 days on 5 CPUs (1.8GHz). Now some user input is required in order to create 3 reliable training sets: faint persistent, strong persistent and transient. For the running of this particular instance of ISINA

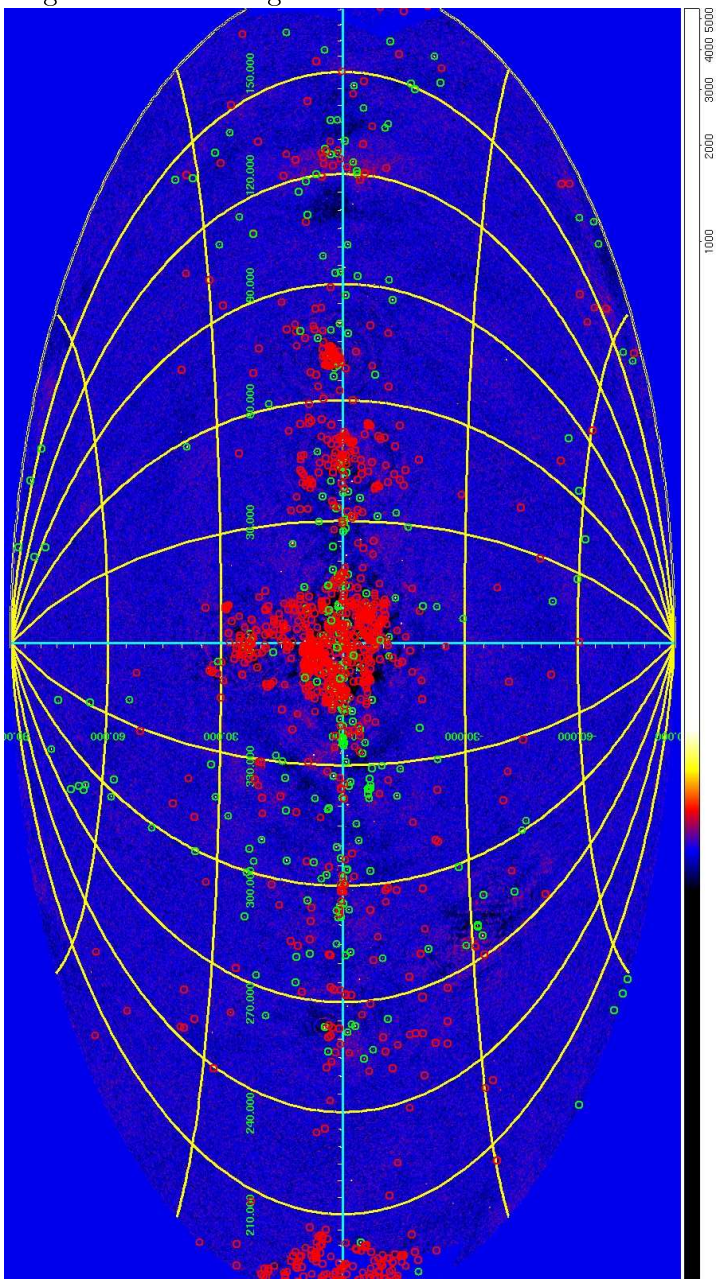
we have decided to create our training sets based on the class tags produced during the production of catalogue 3. So, for training the faint persistent network we will use 129 AGNs, for the strong persistent, 110 XBs, and for the transient, 67 transients. At the same time we located, from a pool of fake 9931 candidates, 1338 which were considered as fakes during the catalogue 3 ISINA run (again using a 0.1° merge radius). These fake candidates were all tagged by ISINA as fakes during the catalogue 3 run, and purposely do not contain any catalogue 3 real objects. Figure 3.6 displays the 18-16 keV band final mosaic with our training set. In green we have our real candidates, whilst in red the selected fakes.

Having created our training sets we are ready to run the classifier, again in a similar fashion to Chapter 2. This takes about an hour per forest, creating ≈2000 trees each. The next section will explore the results from ISINA compared to the results obtained through visual inspection.

3.4 Comparing results

In Chapter 2, Section 2.6, after having trained ISINA on half of the Galactic sky, we were able to easily evaluate our results against the testing set comprised of objects on the other half of the Galactic sky. This was only possible because we were able to obtain a homogeneous testing set, with similar characteristics to the training set. In this Chapter however, we only have a reliable training set comprised of objects identified during the production of catalogue 3. This leaves us with no reliable testing set, since no catalogue for catalogue 4 has been created yet. However, we can still compare the ISINA results with those obtained through visual inspection. We point out that none of the two methods can, at this point, give the perfect answer, especially for low significance

Figure 3.6: 18-60 keV band final mosaic centered on the Galactic centre. The ISINA training set is circled in green for real candidates and in red for fakes.

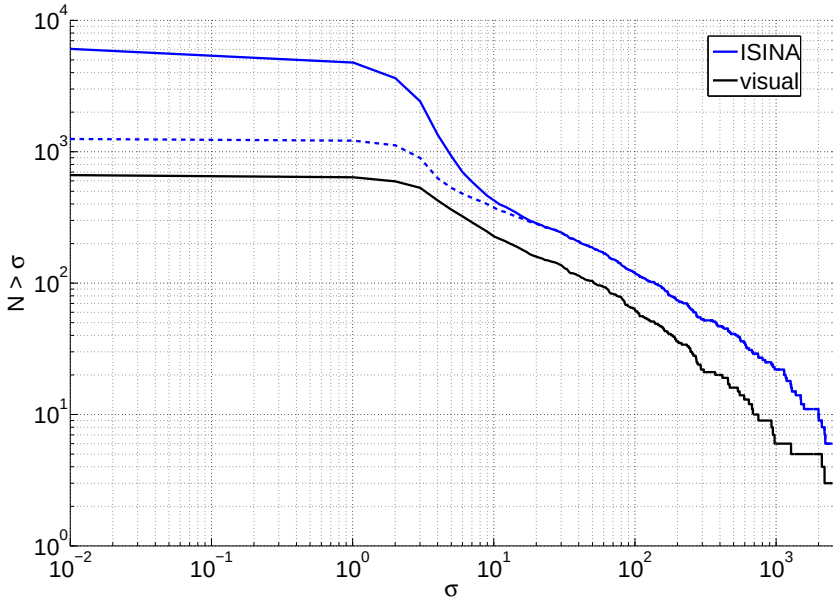


sources.

We believe the best way to compare the results, given the fact that no "correct" answer exists, is that of producing logN - logS distributions for the different methods, and see where they do not agree. This will give us insight into the kind of problems encountered using ISINA or visual inspection. Figure 3.7 displays such a plot for significances obtained in the 18-60 keV map, displaying the number of excesses detected above a specific significance as a function of significance. We will use this energy map through this analysis as it is the most sensitive band for the majority of objects observed with IBIS. The same analysis can however be done in any of the 5 bands available. The blue solid line shows our initial excess list, whilst the dotted blue line the objects recovered by ISINA. To compare our results we show with the black solid line the objects recovered using visual inspection. Before we begin comparing the results, we point out how such logN - logS plots need to be interpreted. Taking the solid blue line as a reference, we can see that it is composed of two superimposed distributions, a powerlaw extending to high significances representing the real source population, and a Gaussian at low significances representing the noise component distribution.

The most obvious discrepancy between ISINA and visually recovered excesses (dashed blue line and solid black line respectively) is a systematic offset in the number of recovered objects. This is also the case when comparing the visually created list with our initial excess list, suggesting we might have an additional noise population which highly resembles that of a real population. In fact, recall that for this ISINA instance we have decided to reduce our initial excess list with a 0.1° merge radius rather than 0.2°. This has been done so as to reduce our initial excess list (see Fig. 2.6), at the expense however of retaining more "propeller"-like structure surrounding real sources. It is indeed

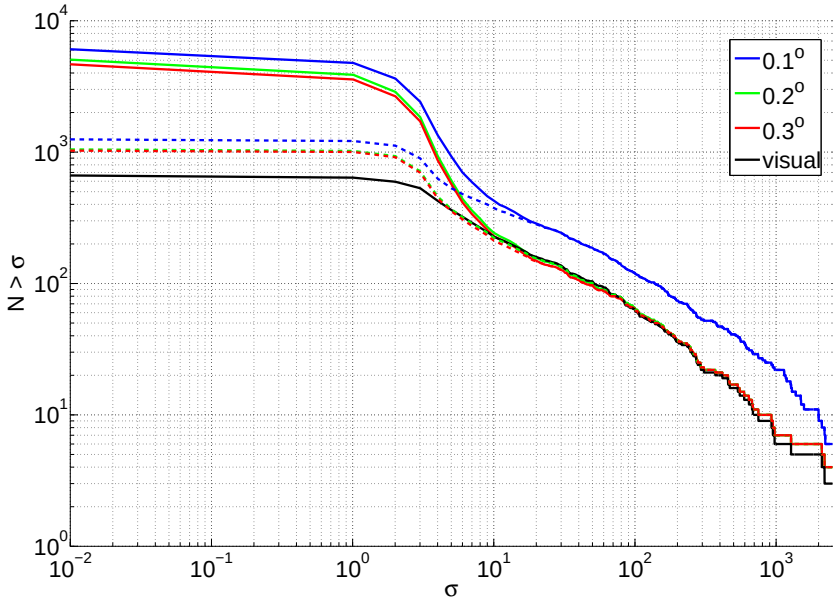
Figure 3.7: Cumulative distribution for candidate detections above a specific significance as a function of that significance. The blue solid line represents all of the ISINA candidates whilst the dotted blue line the recoverd candidates. In solid black we display the candidates recovered through visual inspection only.



this that is the main reason for the systematic offset between the ISINA initial excess list and the visually recovered objects: the initial excess list, merged at 0.1° , includes many "propeller"-like structures related to real sources, giving rise to a logN - logS distribution resembling the real population. This can be demonstrated by further reducing our excess list. We do this by sorting all our excess separated more than 0.1° apart by significance (again in the 18-60 keV band for consistency), and removing any excess within 0.2° from the brightest candidate. Moreover we have performed the same exercise for a 0.3° merge radius. The point of extending the matching to such large radii is to ensure the problem is really that of "propellers"-like structure. We would expect the 0.2° and 0.3° merged lists to produce very similar logN - logS distributions, as we know from previous experience that a 0.2° merge radius is sufficient to eliminate the majority of "propeller"-like structure. This is shown in Figure 3.8. Similarly to Figure 3.7, the solid lines represent the excess lists, and the dashed lines represent the recovered ISINA objects. For reference we also display again the visually recovered objects with the black solid line. It is clear that both the 0.2° and 0.3° merged lists produce very similar results, suggesting that we have removed the additional "propeller" noise excesses associated with the real source population. To demonstrate this effect even further we show in Figures 3.9 and 3.10 two images of the same source with candidates overlayed employing a 0.1° and 0.2° merge radius respectively. The smaller number of fake candidates within the PSF of the real source in the centre is clear when comparing the two images.

We point out that, even though we employ a further reduction based on radius, after features have been extracted, coordinates remain unchanged, and information for the excluded sources based on this last criteria is still retained for further analysis (in order to locate real sources within 0.2° of a brighter

Figure 3.8: Cumulative distribution for candidate detections above a specific significance as a function of that significance. The blue solid line and dotted line and the black solid line are the same as in Figure 3.7. The green and red lines display the same CDFs as for the blue line, but with candidates merged using 0.2° and 0.3° merge radii repectivley.



one).

Having established that we require a further reduction of our excess list using a 0.2° merge radius we will now compare the visually created candidate list with that selected by ISINA. For this we show yet another logN - logS distribution in Figure 3.11. The solid and dashed blue lines display the excess list and ISINA recovered objects respectively employing a 0.2° merge radius, and in black the visually recovered objects. Additionally we show with the green line the distribution of objects selected by ISINA but not selected using visual inspection, the red line on the other hand shows the opposite. It is clear at first that ISINA selects far more objects than visual inspection. This is also noticeable by looking at the increase in the number of recovered objects below $\approx 4\sigma$ for the ISINA selected sample. In fact the green line does indeed

Figure 3.9: 18-60 keV all-sky mosaic image for a bright persistent source. The crossses represent the candidate positions fed to ISINA for identification. The black cross is the correct position for this object. The remaining green positions have been introduced due to the artefacts produced by the bright central source.

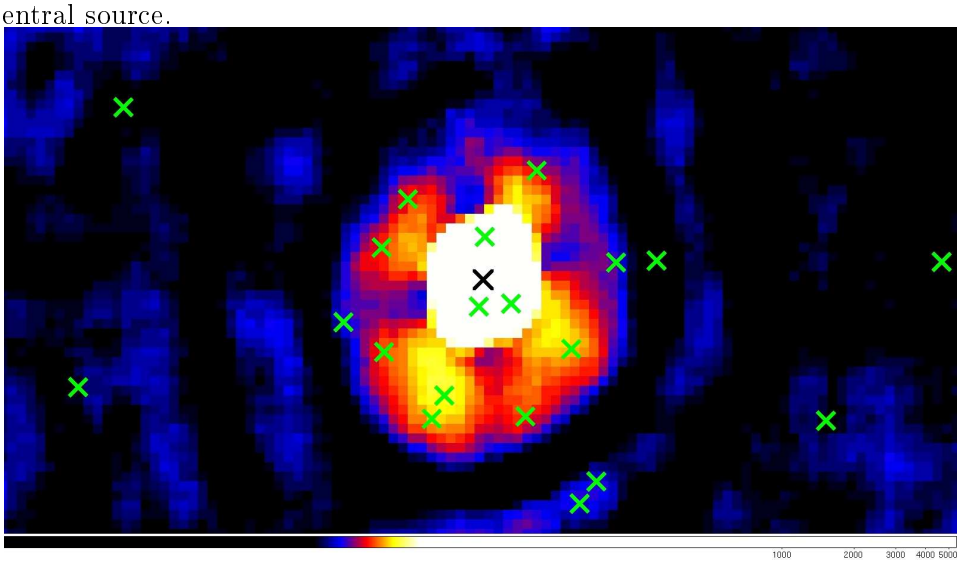
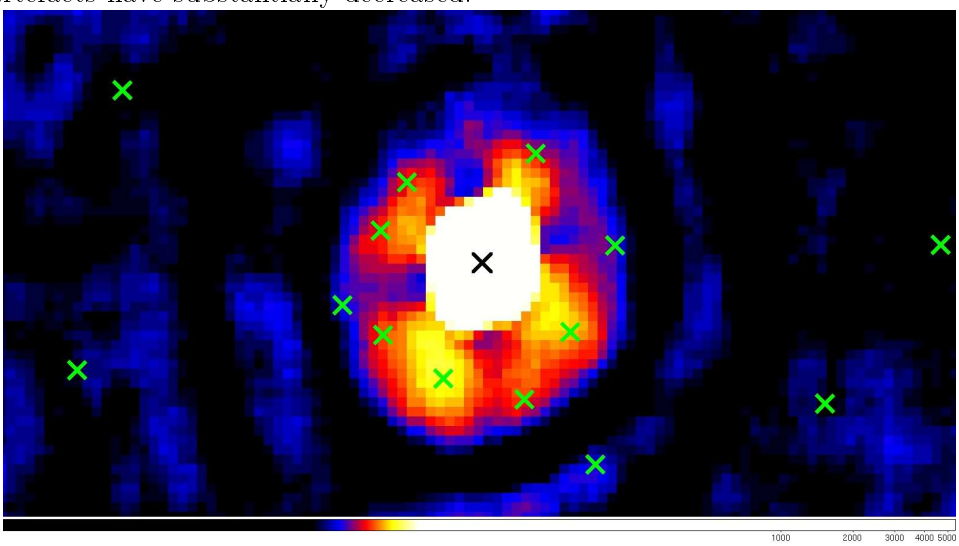


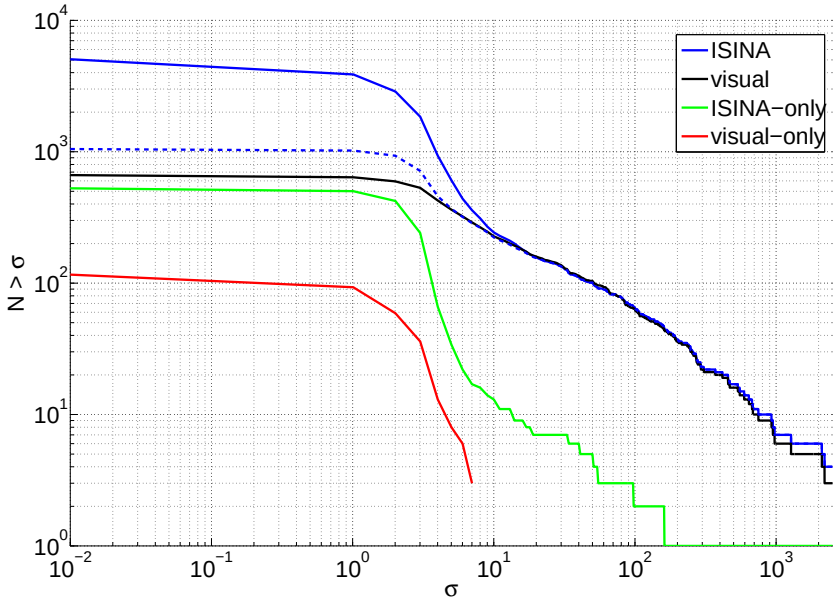
Figure 3.10: Same image as in Figure 3.9 but with candidates merged using a 0.2^o radius. The number of false candidates produced by the bright source artefacts have substantially decreased.



seem to contain a noise distribution below 4 σ, however we also point out that, even though to a lesser extent, the red curve also displays such distribution shape. We note that in both cases the noise distribution could be caused by the presence of transients, not achieving a high enough significance in the final mosaic map. The only way to discriminate such sources is to use the ISINA meta-data in order to aid the the selection of transients. For example, if a candidate is selected by ISINA within the noise distribution (say 3σ), then the particular candidate has the potential to be a transient. This can be checked by querying the meta-data to understand which forest (faint persistent, strong persistent or transient) had the highest number of votes. Moreover we can additionally inspect the transient matrix for transient behaviour, and obviously inspect the source visually if required. The difference compared to the "standard" visual inspection process is that the additional ISINA meta-data can aid the recovery of real sources in a much more objective way, since we know in advance that the algorithm has been created to recognise similar objects to the training set.

Figure 3.12 shows once again the 18-60 keV mosaic map, but now with discrepant candidates between the two methods. With green circles we show the positions of candidates selected by ISINA but not visually, and in red the opposite. From this image we can see where the fake candidates producing the noise distribution in Figure 3.11 are coming from. These are areas of the mosaic where local noise is relatively high, which can be seen by the clustering of ISINA-only selected candidates in some regions of the map. Disregarding the clustered regions we notice, from the mosaic, that the ISINA-only candidates are distributed mainly out of the Galactic plane, whilst the visual-only candidates seem to follow it. The next sections will explore some particular candidates in more detail, and shed light into this peculiar sky distribution,

Figure 3.11: Cumulative distribution for candidate detections above a specific significance as a function of that significance. The blue solid line and dotted line are the same as the red lines in Figure 3.8 and the black solid line is the same as in Figure 3.7. The green and red lines display the CDFs of candidates identified by ISINA and not visually and viceversa, respectivley.



with emphasis on the kind of problems faced by the visual inspection method and ISINA.

3.4.1 Human problems

In this section we will explore some source candidate examples and compare the ISINA selection method with visual inspection. In particular we will focus on some faint persistent source candidates from the extra-Galactic sky ($l > 30°$), and hopefully demonstrate some of the pitfalls of using solely visual inspection for the selection of sources.

For this purpose we have selected from Figure 3.12, 6 candidate excesses, all of which obtained their highest score within the faint persistent network of ISINA. We specifically choose 3 of these candidates to have obtained at least 1 "No" vote from visual inspection, however achieving high enough scores by ISINA to be considered real. Conversely, we also selected 3 candidates achieving 3 "Yes" votes through visual inspection but not high enough score by ISINA to be considered real. These are shown in Figure 3.13 with green and red circles respectively in each column. All chosen examples are relatively faint for all sky mosaics, and some are below the catalogue 4 threshold of 4.5σ. Moreover we highlight that the images have been taken from the best significance mosaic for the particular candidates in question.

At first, all 6 examples look comparable, and indeed they do share very similar characteristics. Beginning from the ISINA selected objects on the left column of Figure 3.13, we will briefly describe each candidate. The first example on the top obtained a significance of 5.2σ in the 18-60 keV mosaic image, but has not been identified through the visual inspection process. The reason for it not being included in the visually selected objects is because this particular excess was never included in the inspection list. This implies that the

Figure 3.12: 18-60 keV band final mosaic centered on the Galactic centre. The only-ISINA recoverd objects are shown in green circles whilst the visual-only recovered objects in red circles.

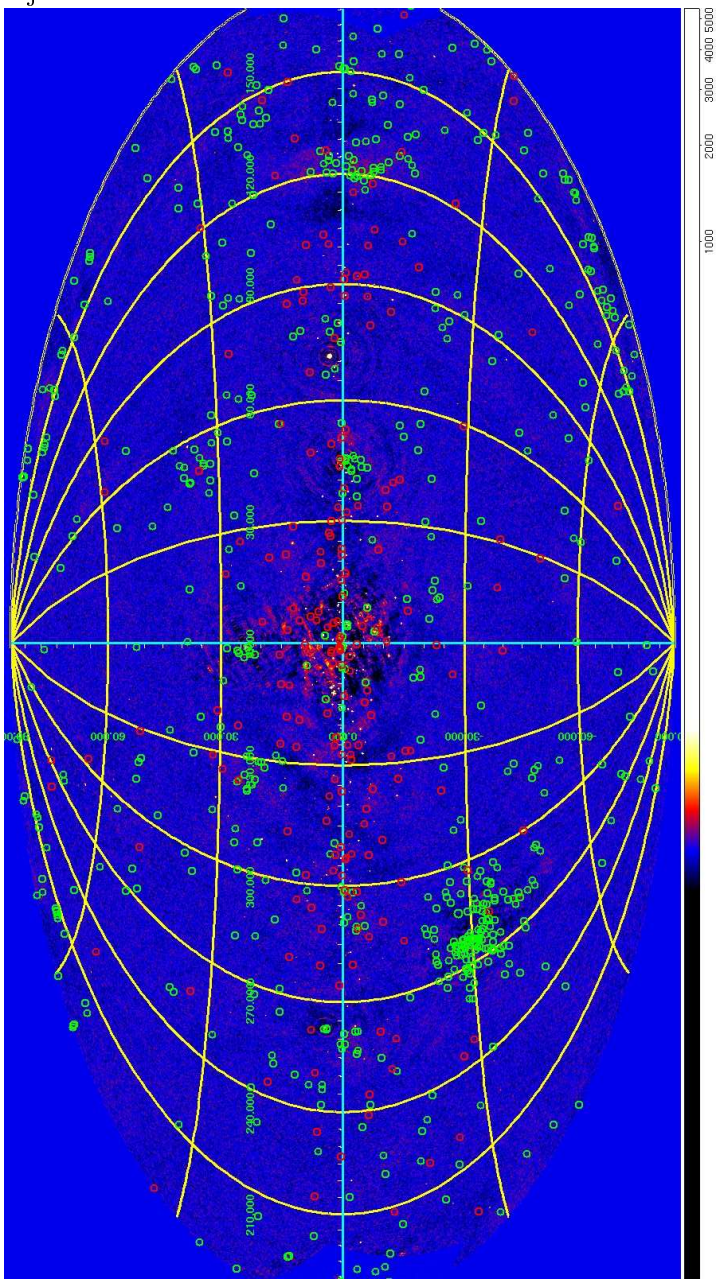
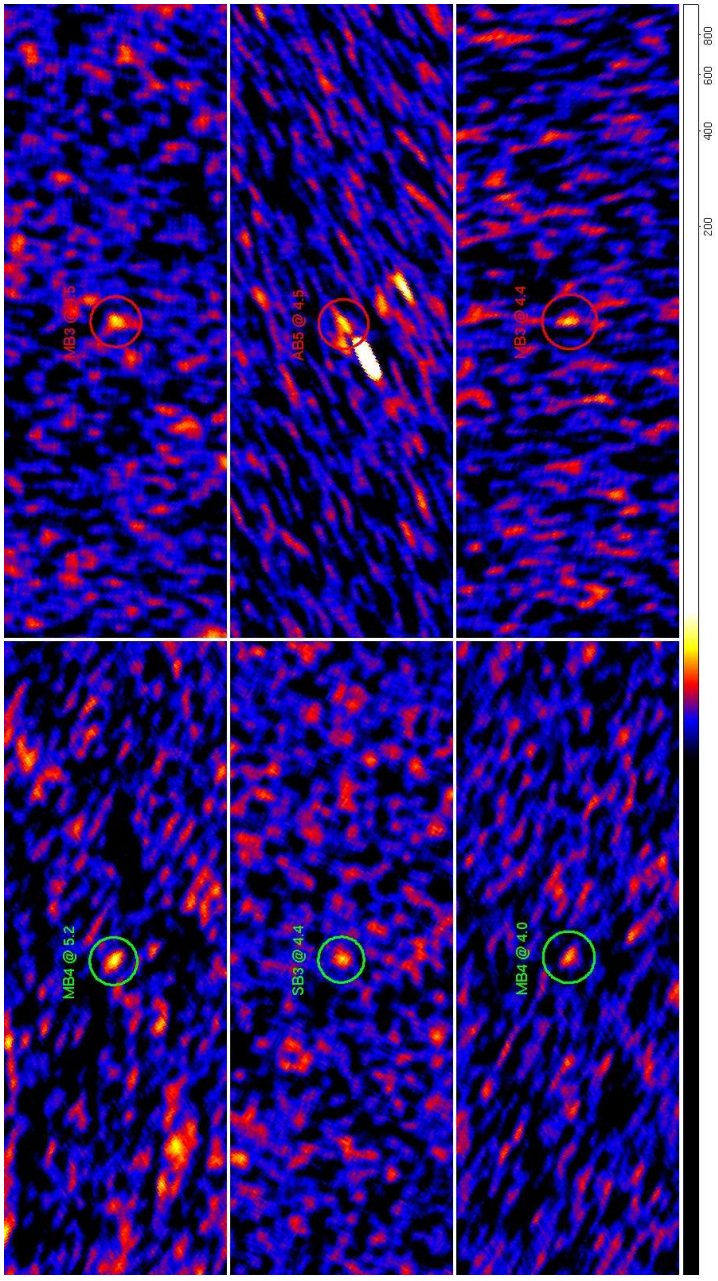


Figure 3.13: Best significance mosaic maps for 6 candidate persistent sources. On the left column, circled in green, are candidates which ISINA believes are real but have obtained at least one No vote visually. On the right column, circled in red, are candidates with three Yes visual votes but ISINA does not identify as real.



standard catalogue creation technique described in Section 3.2 has failed to locate this excess. The reason for this is because the *peakfind* tool has failed to locate this candidate due to the high systematic background within the region Conversely ISINA begins with many more excesses to classify, which included this particular candidate as well. In particular this example demonstrates the advantage of having machine learning algorithms for identification, since we can begin with a very high number of candidates to classify than just visual inspection, without affecting the final timescale of the results.

The next examples in the middle-left and bottom-left panels have been chosen to demonstrate the capability of ISINA to select candidates with "clean" PSFs, even if their significances are relatively low. It is hard to decide if these candidates are real, however we believe that they are consistent with each other meaning they possess very similar characteristics. We therefore expect ISINA to have similar opinions on these two candidates. We note that the bottom-left candidate was not included in the inspection list, similarly to the candidate in the top-left panel.

We now examine the visually selected candidates on the right panels of Figure 3.13. On the top-right and middle-right panels we show two excesses which display very distorted PSFs and owning again relatively low significances. Conversely the bottom-right displays an excess with a reasonable PSF, though a bit less significant. These are all borderline cases, similarly to the excess on the left panels, for which we believe no definite answer exists concerning their identification, yet all have achieved "Yes" votes. The degree of confidence about these candidates raises some concerns, and opens the possibility of systematic false identifications for other candidates too. The excess in the middle-right looks as if it is a possible "propeller-like" structure associated with the bright source close by. It is essentially impossible to discriminate for or against this

claim, and extreme caution has to be taken when classifying such objects. Having said this we believe that including fewer false positives at the cost of recovering fewer real objects is better than the contrary. Moreover we note the high degree of similarity between the quality of the PSFs of the two objects in the bottom panels.

It is hard to imagine how all these candidates have obtained such different visual judgments, suggesting the need for introducing a continuous identification scheme. By this we mean a scheme which would not just include binary results ("Yes" and "No" votes), but also has the capability of producing continuous results. This would not only enable candidates to be classified as real or fakes, but also obtain a level of "realness" and "fakeness" respectively. The improvement using such a scheme would be enormous, and would also give a third party user the capability of making up his/her own mind for a particular candidate.

3.4.2 Machine problems

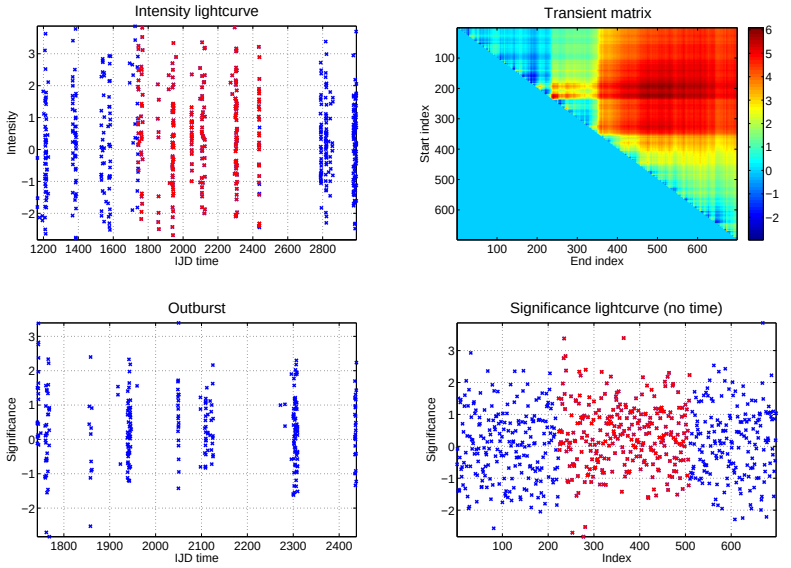
We will now turn our attention to some of the main issues facing ISINA, and in particular look at some examples from Figure 3.12 taken from the Galactic plane. Most of this section will focus on transients, and describe how and why ISINA fails to correctly identify these objects in some cases. For analysis purposes, each discussed example will include its transient matrix panel and, if appropriate, an image from the highest significance map. Firstly, we select four examples which we believe are real transients but ISINA has failed to identify. Analysing these examples will give insight into the reason for ISINA failing to recognise these objects. Secondly, we present four examples which have been selected by ISINA as transients, but turn out to be fakes, shedding light onto some of the problems associated with the transient matrix technique.

Figures 3.14 and 3.15 show the first four real transients not identified by ISINA discussed in this section. For these objects we also show the best significance mosaic map in Figure 3.16, where it is visually clear these candidates are real sources. We believe the main reason for the discrepancy on the recovered objects lies in the new, updated, method for locating such objects in catalogue 4 data. Recall from section 3.1, that a new variable timescale for the identification of transients is introduced for catalogue 4 analysis: the bursticity analysis. These were not produced for the compilation of catalogue 3, mainly because their usefulness is enhanced by large data streams, which catalogue 4 has obtained. This has major consequences for interpreting catalogue 3 results and the results from the ISINA testing set (see section 2.5.7). The first is that catalogue 3 transients requiring a burstmap image for identification might have been missed during catalogue production. More importantly, the transients requiring burstmaps have not been trained for in ISINA. Moreover burstmap, revolution map or revolution sequence map information is not included in the ISINA metadata, since the testing set results on transients in section 2.5.7 were sufficiently reliable, making us believe the available metadata was enough for the particular identification task. Obviously this has now changed, and the burstmap transients have revealed that additional information regarding shape and significance taken from such maps is essential for ISINA if we wish to identify such objects correctly.

In all cases in Figures 3.14 and 3.15, the transient matrix has located an outburst timescale being longer, however a superset, than the selected pointings using the bursticity technique. This is a problem which did not occur during the testing of ISINA in Chapter 2. Nonetheless, the main reason why ISINA has failed to recognise these candidates is due to when/where the parameters are extracted. In the case of transients, these are extracted at the

Figure 3.14: TM panels for two transient sources.

IGR J18014+0202



IGR J15107-5414

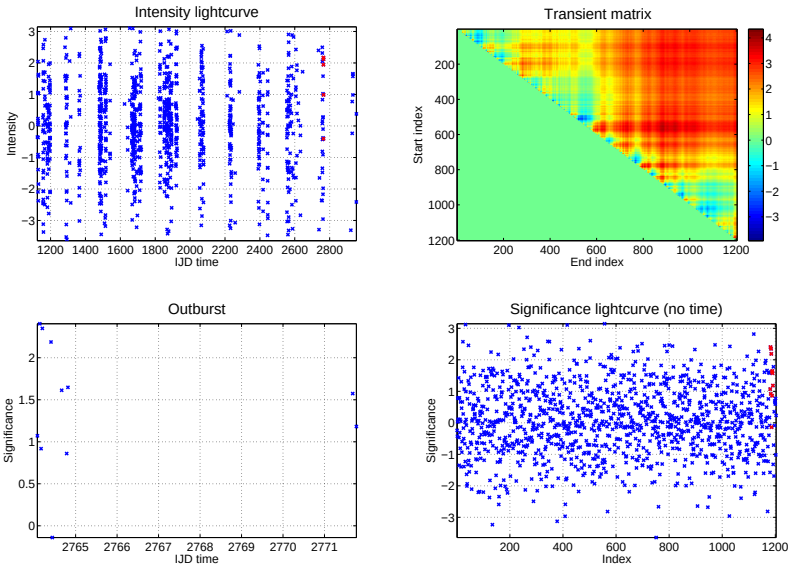
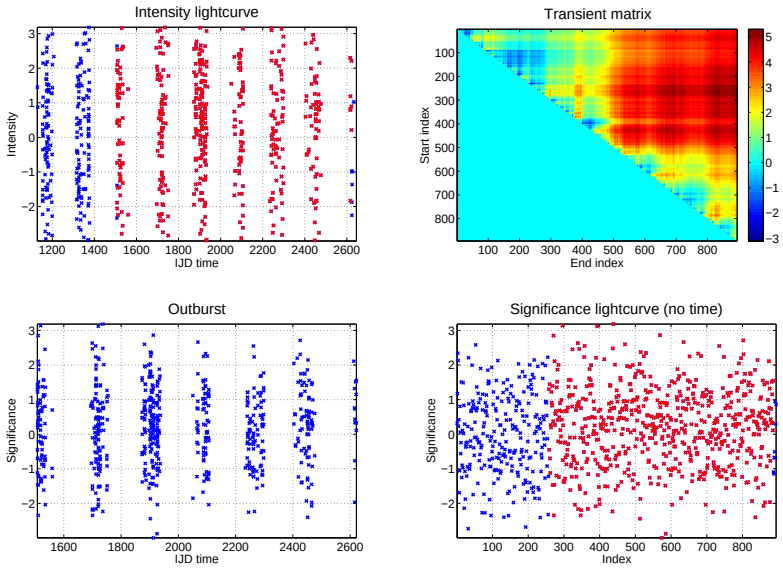


Figure 3.15: Transient matrix panels for two transient sources.

IGR J16291-2937



IGR J21319+3619

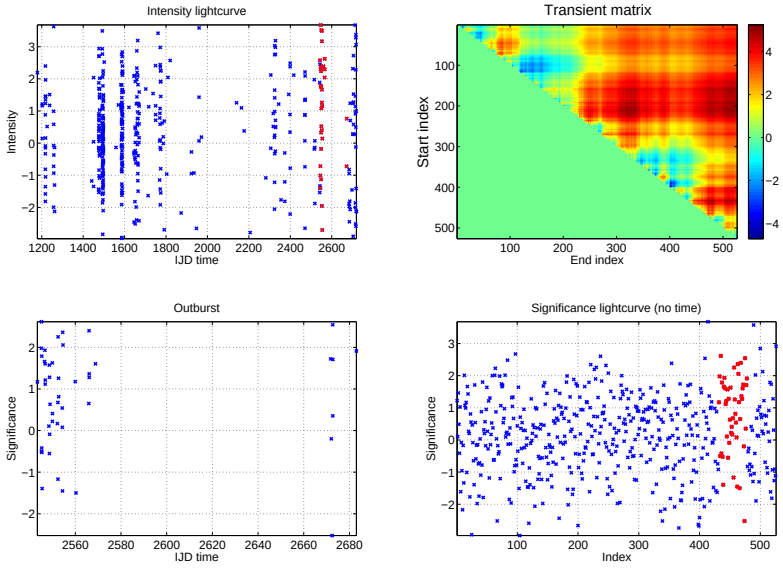
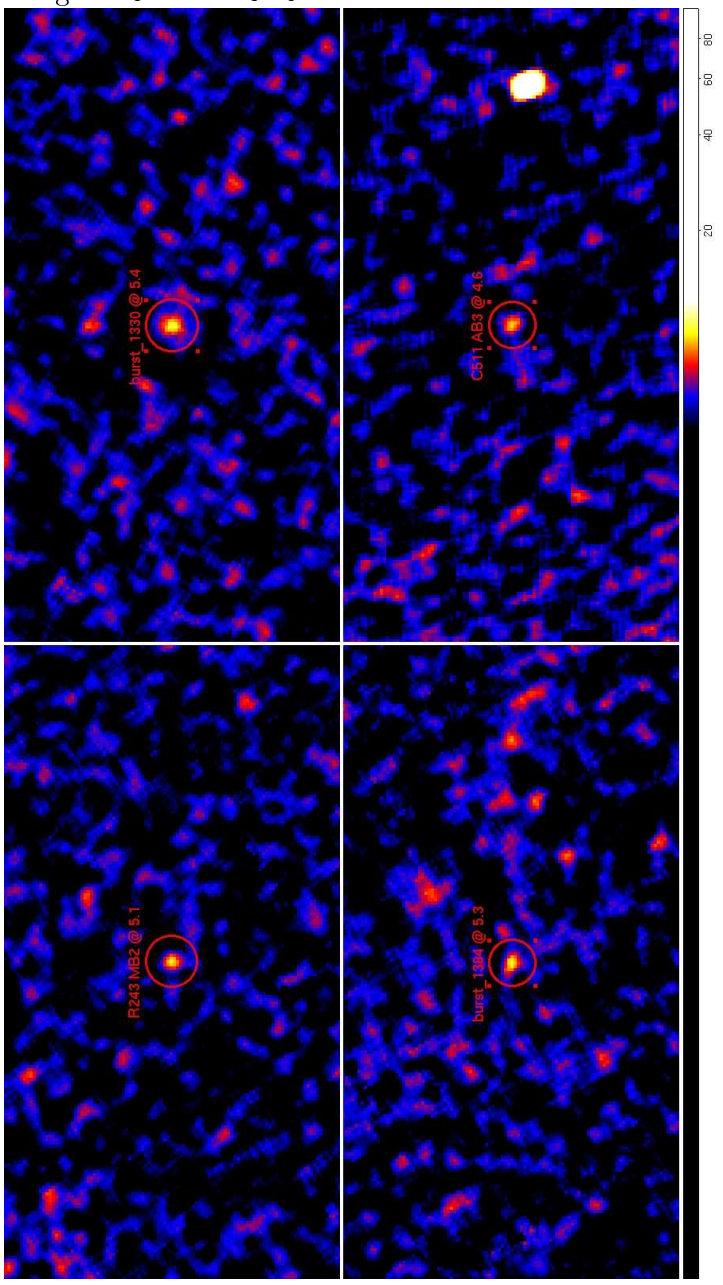


Figure 3.16: Best significance image mosaics for transient objects. From left to right: IGR J18014+0202, IGR J15107-5414, IGR J16291-2937 and IGR J21319+3619. The corresponding transient matrix panel for these sources are displayed in Figures 3.14 and 3.15



ScW level (the ScWs are selected using the transient matrix technique) and then averaged before being passed to the identification algorithm. On the other hand the visual inspection method uses different timescale mosaics to inspect the transient candidates, i.e. burstmaps, revolution maps and revolution sequence maps. In these mosaics the transients will be more evident as the signal from the candidate source is increased. On the other hand we find that averaging the parameters extracted from the ScW level is not enough. In the same way as we extract parameters for persistent sources from both the ScW level and the all-archive mosaic, we should in a future run of ISINA extract parameters for transients from both the ScW level and the respective burstmap mosaic. The problem with this, and also with the "human way" of locating candidates, is that one would first need to create these mosaics for all candidates using the selected ScWs from the transient matrix. This would take a long time for ≈ 9000 candidates. This problem is being investigated in more depth, and ways to select transient candidates are being sought in order to reduce the amount of transient mosaics to be created for a future ISINA run.

We now turn our attention to four fake candidates selected by ISINA as transients, however associated with noise structures within the IBIS images. The transient matrix panels for these fake candidates are shown in Figures 3.17 and 3.18. In all, the maximum significance obtained by the transient matrix is about 5σ, similar to the real objects shown in Figure 3.16. However after inspecting these candidates visually (from the ScW images), it was clear that these candidates were introduced due to the deconvolution software failing to remove artifacts in some ScWs. This implies that within the lightcurve of such candidates a spike will be present where the deconvolution software failed, giving rise to a source-like artifact, resembling very much transients. There

are not many of these cases, however from the transient matrix panels they do look real, and only visual inspection of a few selected ScWs can discriminate against these fake candidates.

One last note of caution regarding ISINA comes from the inspection of the LMC region in Figure 3.12 ($20^\circ \times 20^\circ$, bottom left). This region is overpopulated by ISINA candidates which have been selected due to their source-like characteristics, however being produced by the systematic noise originating from bright LMC sources. Visual inspection can consider a much larger "local" area than ISINA when deciding on such candidates, making their rejection much more trivial. Future improvements to ISINA would include various rms values obtained from different size regions centred on candidates to try and overcome this problem.

It should be clear after this section that if we wish to use ISINA for the correct recovery of transients, then more work needs to be undertaken. In particular it is clear that information regarding burstmaps, revolution maps and revolution sequence maps has to be included in the ISINA metadata during training. Moreover it is also clear that if we wish to use the transient matrix appropriately for low significance transients like the ones presented in Figure 3.16, then additional criteria (maybe similar to the bursticity index), will have to be introduced. This will be discussed in the next and last section of this chapter, together with additional improvements relevant to ISINA, and a possible future application of the algorithm.

3.5 ISINA's future?

We now discuss the future improvements to be made to the ISINA algorithm and a possible future meta-catalogue release. First however the discrepancies

Figure 3.17: TM panels for two fake candidates considered to be real by ISINA.

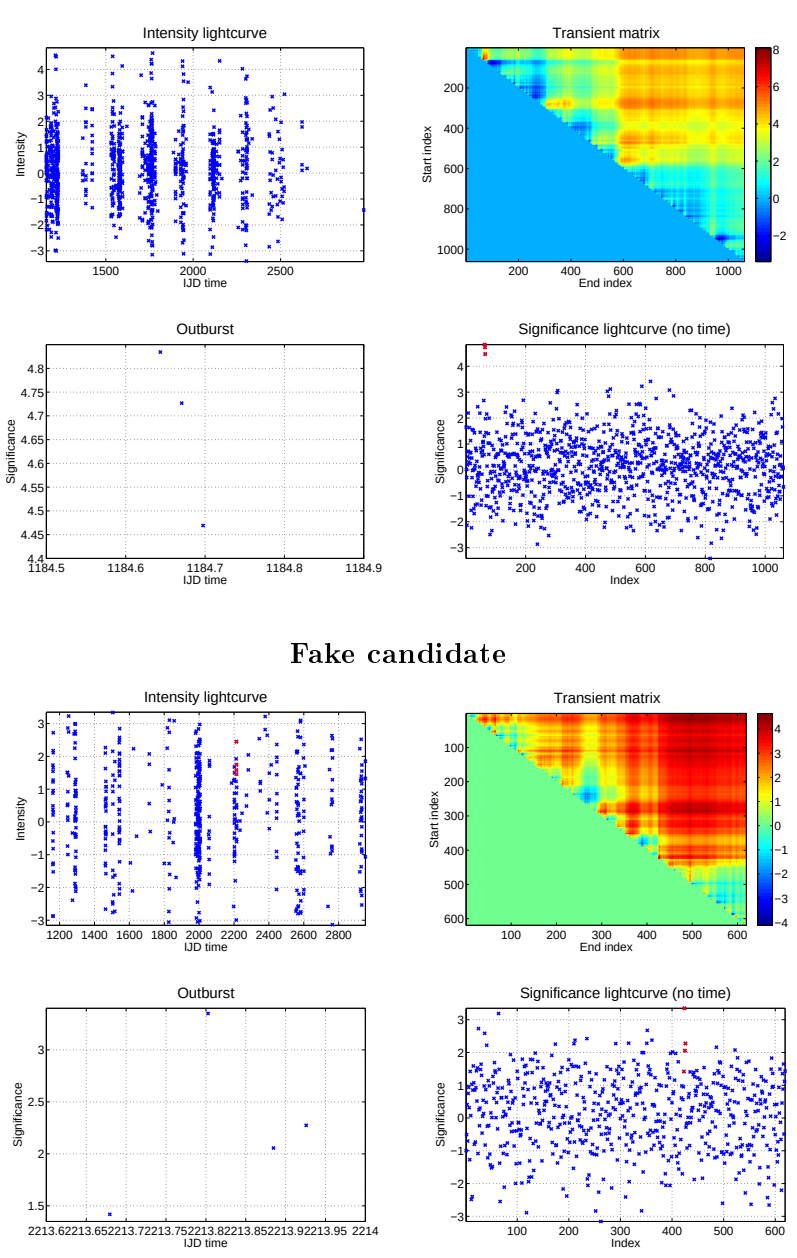
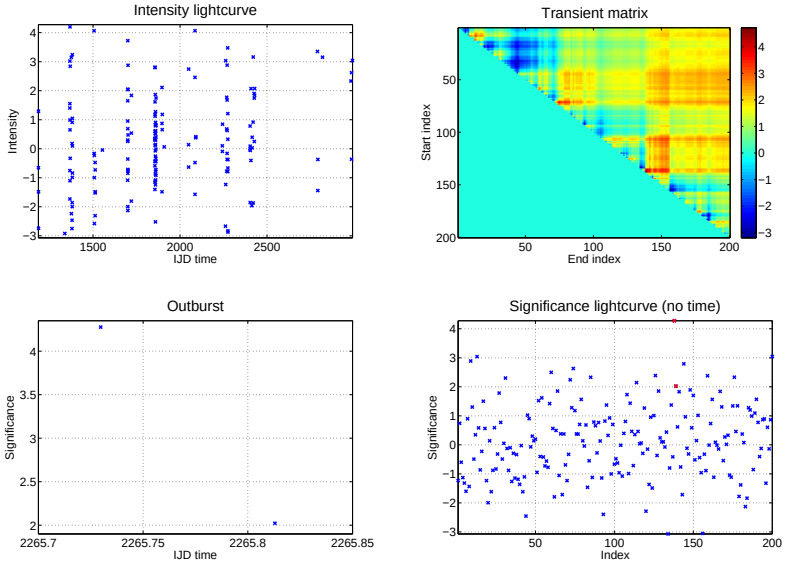
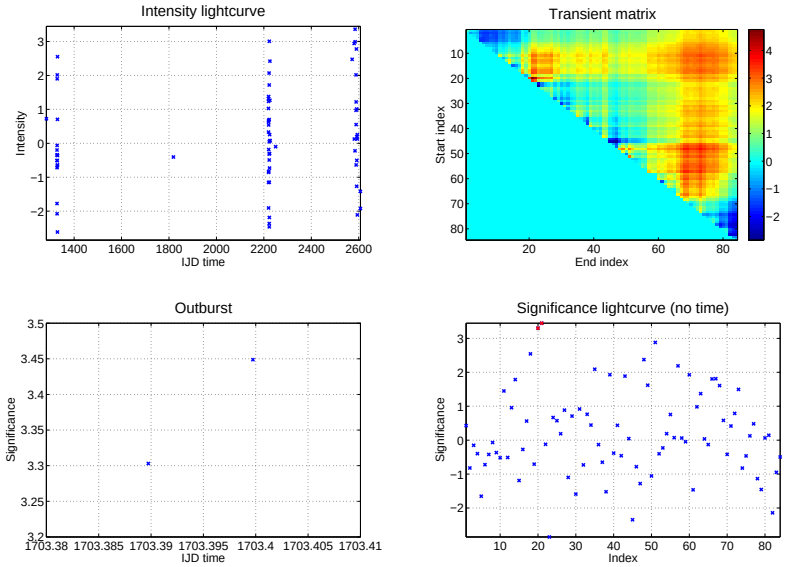


Figure 3.18: TM panels for two fake candidates considered to be real by ISINA.

Fake candidate



Fake candidate



between catalogue 4 and the ISINA catalogue needs to be addressed in more detail. In total, out of 723 objects identified through the visual inspection process, 188 were missed by ISINA (red circles in Figure 3.12). Of these, 33 have definite identifications however most lie in crowded regions of the sky (mainly galactic centre) and it is of no real surprise that ISINA has missed these. 40 have unknown identifications and were only detected in the all-archive mosaic, owning relatively low flux, so could be classed as border line cases. The remaining 115 objects have been detected in burstmaps, revolution maps or revolution sequence maps, making them transient objects. We have already discussed why ISINA fails to recover these correctly and future improvements such as the ones mentioned in the previous section will be implemented in ISINA in the future. This will allow the correct recovery of these objects. We therefore believe that, apart from transient detection, ISINA does relatively well in identifying gamma-ray sources given the run-time of the algorithm of about two weeks (from initial candidate selection to producing a first look catalogue). In particular, of the 723 sources present in catalogue 4, 306 were present in training, leaving 417 for classification, of which 55% (229 sources) have been identified using the ISINA algorithm, where the missing sources are mainly (37%) transients or unidentified sources.

Once we incorporate the transient maps in the algorithm we think ISINA meta-catalogues could be usefully released to the scientific community. These would contain all the initial candidate sources together with the ISINA meta-data associated with each. In particular, each candidate source will have the usual astronomical data associated with it such as, for example, RA and DEC, count rates in the different IBIS energy ranges, exposure times and fluxes in each energy range. Moreover the results from the ISINA algorithm will also be associated with each candidate. This would include:

- the total percentage of votes from the Random Forest which would give an idea of how confident the algorithm is in general
- the percentage of votes obtained from each of the individual Random Forests (see Figure 2.13)
- the transient matrix selected timescale and maximum significance giving the user an idea of any transient behaviour of the candidates
- the percentage of votes obtained within each network in each energy range. This would allow the user to locate objects observable only in one energy range as these objects will have low total percentages.
- some of the parameters used for the identification could also be included such as the average local noise around candidates to give users more of an idea on the detection confidence

With such a meta-catalogue available to the scientific community, each user could then create his/her own catalogue, depending on how confident he/she is about a particular set of candidates. Moreover this meta-catalogue could be useful for cross-correlation studies with objects at other wavelengths. For example, we know already that the visual catalogue might have missed some real sources which could be recovered if some extra information, such as detections in other wavelengths was available.

Ideally, if ISINA had to be used for the production of a future catalogue, it is expected that the production speed improvement for the catalogue release would be enormous. For example, after all the dataset has been reduced and mosaic maps created, the visual inspection method requires an additional 6-7 months for a workforce of about 10 astronomers to visually identify each of the candidate excesses. Given a reliable training set, ISINA can accomplish

a similar task in about 2 weeks, where most of the time is spent extracting the classification parameters. It has been already explained that an additional visual inspection stage is required after ISINA has produced a catalogue, however this stage will be performed on a much smaller candidate list and thus is expected to last about a month for the same workforce. Moreover, as described in chapter 2, candidates with more than 90% of the votes from ISINA can be considered real without any further inspection, again reducing the amount of human load on the task. Also to be pointed out is the fact that ISINA created catalogues will be less biased in that they should contain homogeneously selected candidates, purely on the basis of their parameters and not selected through human intervention which can sometimes be very biased and opinionated.

We conclude this chapter by mentioning the potential usefulness of meta-catalogues, not only from ISINA, but any other catalogue production technique. Given the ever increasing amounts of data and discovered objects, it is inevitable that future catalogue releases will have more spurious false detections and more failed real detections. Meta-catalogues have the potential to overcome this, and more specifically will contain the necessary information to recover the missed objects later in the future, since every possible candidate is recorded. They will also contain the information required to best understand why objects are being missed or false candidates included. It also makes catalogue production more transparent to the scientific community, as all the results from the production stage will be published and readily available.

Chapter 4

mCVs: Back in Business

"Somewhere, something incredible is waiting to be known."

– *Carl Sagan*

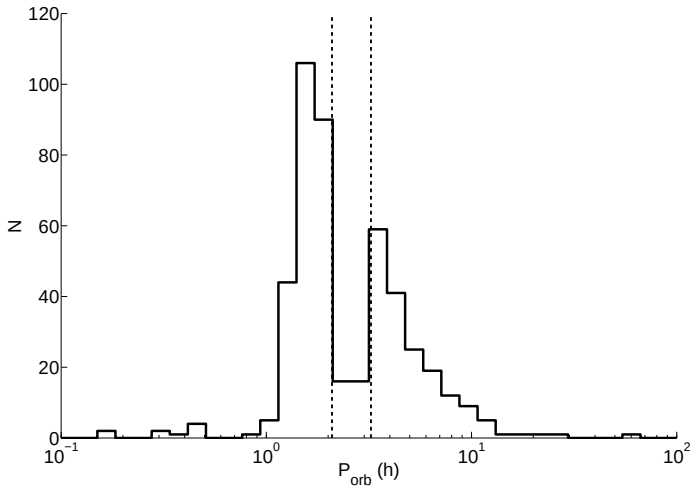
In this chapter we will investigate the properties of magnetic cataclysmic variables detected in the hard X-ray domain. This exotic population, somewhat overlooked in the past, will bring forward some new and exiting results, possibly associated with their accretion mechanisms. Firstly we will introduce mCVs, and briefly describe the types of mCVs found and their corresponding classes. We then will look at the contemporary models for mCV evolution as predicted by numerical simulations so as to prepare the reader for the new upcoming results. Next we present the most up to date hard X-ray observations of mCVs and produce a catalogue of hard X-ray selected mCVs. These are then studied in the context of their orbital and spin periods which will show us how these hard X-ray selected samples only occupy a specific place within the P_{orb}-P_{spin} plane. Finally the chapter will analyse the spectral hardness properties arising from the sample and show how these are well correlated with the orbital, spin and synchronicity parameters of the mCV systems. The chapter will conclude with some discussions and speculations on the origin of

the discovered correlations.

4.1 Cataclysmic variables: a brief overview

Cataclysmic variable stars are compact, interacting, binary systems in which a white dwarf (WD) primary accretes from a low-mass, roughly main sequence donor star. The mass transfer and secular evolution of these systems is driven by angular momentum losses. Systems with long orbital periods ($P_{orb} > 3hr$) are thought to lose angular momentum mainly via magnetic breaking caused by the stellar wind of the secondary star [Verbunt and Zwaan, 1981, Rappaport et al., 1983]. In the canonical scenario, magnetic breaking stops when the secondary becomes fully convective, at about $P_{orb} = 3hr$, at which point the donor star shrinks and detaches from the Roche Lobe, and gravitational radiation (GR) becomes then the only remaining angular momentum loss mechanism [Faulkner, 1971, Paczynski and Sienkiewicz, 1981]. The orbital period will thus continue to shrink, ultimately bringing the secondary back into contact with the Roche Lobe at about $P_{orb} = 2hr$, allowing for mass transfer to resume. The main motivation for this scenario is the presence of the so-called period gap between 2 and 3 hours within CV systems as shown in Figure 4.1. The secondary will keep losing mass to the primary until the mass of the donor becomes sufficiently low to be unable to sustain hydrogen burning, at which point the secondary starts becoming degenerate. The orbital period evolution reverses sign at this stage, implying a minimum observable orbital period within CV systems. This is supported by the accumulation of systems at very low orbital periods, dubbed the "period minimum spike" by [Gänsicke et al., 2009], where the faintest CV populations have been uncovered and found to have orbital periods below 86 minutes.

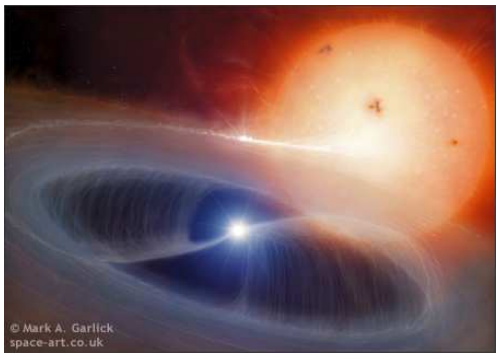
Figure 4.1: Orbital period distribution for non-magnetic cataclysmic variables. The vertical dotted lines mark the period gap. Data taken from Ritter and Kolb [2003].



4.2 Introducing magnetic cataclysmic variables

Magnetic CVs (mCVs) are a small subset of the catalogued CVs ($\approx$ 10% - 20%, Downes et al. 2005, Ritter and Kolb 2003), and fall into two (or possibly three) categories: polars (or AM Her types after the prototype system), intermediate polars (IPs or DQ Her types) and asynchronous polars (APs). The WDs in polars possess such strong magnetic fields that they can synchronise (see King et al. [1990] for the Polar synchronisation condition) the whole system, yielding $P_{orb} = P_{spin}$ (see Figure 4.3). The strong magnetic field in these systems is confirmed by strong optical polarisation. Accretion in polars is thought to follow the magnetic field lines of the WD straight from the L1 point onto the WD magnetic poles, and no accretion disk is expected (for a review of polars, see Cropper 1990). APs on the other hand are out of synchronisation by only a few percent, and it is not known exactly why this is. One suggestion is that these systems are polars which have had a recent

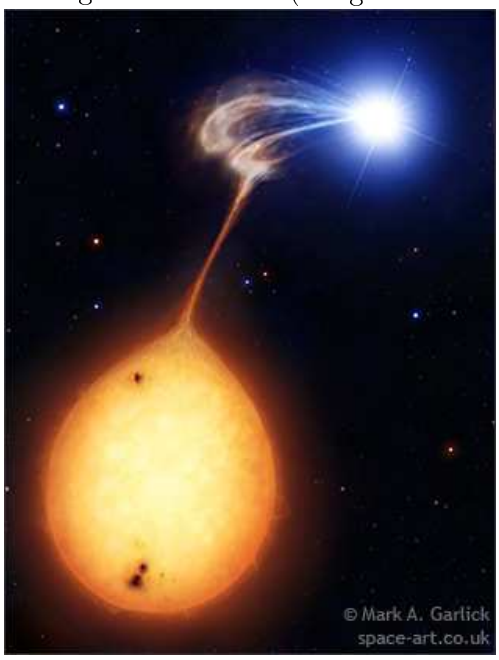
Figure 4.2: Schematic diagram of an intermediate polar. (Image taken from Mark A. Garlick)



nova event, kicking them slightly out of synchronisation [Warner, 2003]. For IPs, the lack of strong optical polarisation implies a much weaker magnetic field, not powerful enough to synchronise the secondary (for a review of IPs, see Patterson 1993). In these systems, material leaving the L1 point usually forms an accretion disc up to the point where the magnetic pressure exceeds the ram pressure of the accreting gas (see Figure 4.2). From this point onwards the accretion dynamics are governed by the magnetic field lines, which channel the material onto the WD magnetic poles. The nature of these systems is confirmed by the detection of coherent X- ray modulations associated with the spin period of the WD.

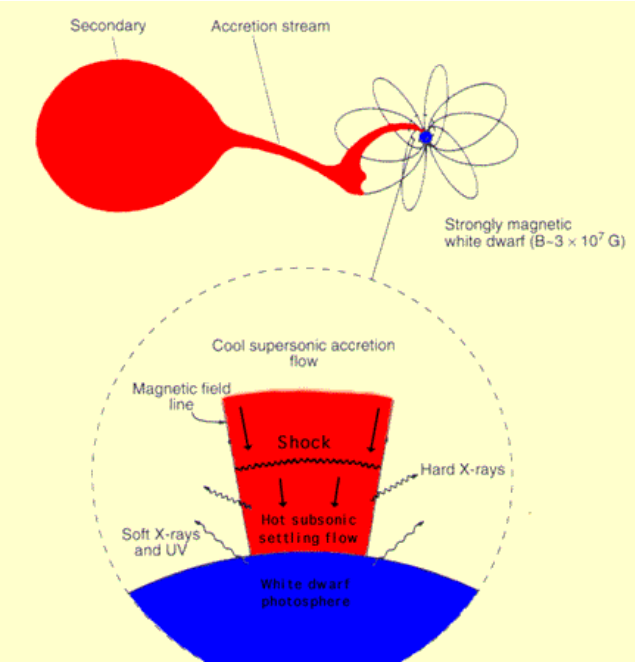
In the simplest scenario for X-ray production in mCVs, the magnetically channeled accretion column impacts the WD poles producing hard X-rays (see Figure 4.4) from thermal bremsstrahlung cooling by free electrons with kT of the order of 10s of keV [Cropper, 1990, Warner, 2003]. The hard X-ray emission is thought to originate in the post-shock region, a region below the shock front created from the impacting accretion column. This is also supported by the expected amounts of X-rays produced by unfalling matter onto a WD, where the kinetic energy of the infalling matter is converted into thermal energy

Figure 4.3: Schematic diagram of a Polar. (Image taken from Mark A. Garlick)



($\frac{3}{2}kT = \frac{1}{2}mv^2$). Softer X-rays are also produced from the absorption and reprocessing of these higher energy photons in the WD photosphere. As a result, both polars and IPs are expected to emit high energy photons, but, discrepancies exist between the observed ratio of soft-to-hard X-rays between polars and IPs, with polars showing an excess of observed soft X-rays. Lamb [1985] and others have reported that the total X-ray luminosities of IPs are greater than those of polars by a factor of ≈ 10, attributed mainly to the higher accretion rates. Moreover it has been proposed that strong magnetic fields in polars produce a more "blobby" flow than in IPs [Warner, 2003]. These high density "blobs" are then able to penetrate within the post-shock region, emitting fewer bremsstrahlung photons and contributing more to the observed X-ray blackbody spectral component, thought to be produced at the base of the post-shock region.

Figure 4.4: WD schematic diagram displaying the regions of X-ray emission



4.2.1 The Accretion flows and evolution of mCVs

In a series of papers [Norton et al., 2008b, 2004], Norton and collaborators have demonstrated, using numerical simulations, that four types of flows are possible for accreting binary mCVs. They have shown that the fundamental observable determining the accretion flow type is the spin-to-orbital period of the system. This section will review their results which later in the chapter will help explain some of the observations presented whilst analysing the global properties of hard X-ray emitting mCVs.

As shown in Norton et al. [2008b, 2004], the mCV orbital and spin parameters evolve towards a spin-to-orbital equilibrium. Essentially, for any given orbital period, mass ratio and magnetic field strength, there exists a spin period that will balance the gain and loss of angular momentum within the system. For example, if an mCV is spinning too fast a lot of the material latching onto

the field lines will be expelled by the fast spinning WD, thus carrying angular momentum and slowing the WD spin. Conversely if the WD is spinning too slowly most of the material will make it to the WD poles, thus giving extra angular momentum to the WD and spinning it up. Somewhere in between is an equilibrium where the WD is in a state of accretion and ejection, therefore in general, mCVs are expected to remain close to this equilibrium when considering long timescales. At any particular instant however the WD may be spinning up or down as shown by Patterson [1993]. The observed spin-up or spin-down rates observed within IPs however correspond to much longer timescales than those expected to reach equilibrium, suggesting the systems are only exhibiting random excursions from their equilibrium, driven by mass loss fluctuations. In fact this phenomenon is also predicted by the numerical simulations carried out by Norton and collaborators. Broadly speaking, they have shown that four types of flows are possible within IP systems, characterised as one of:

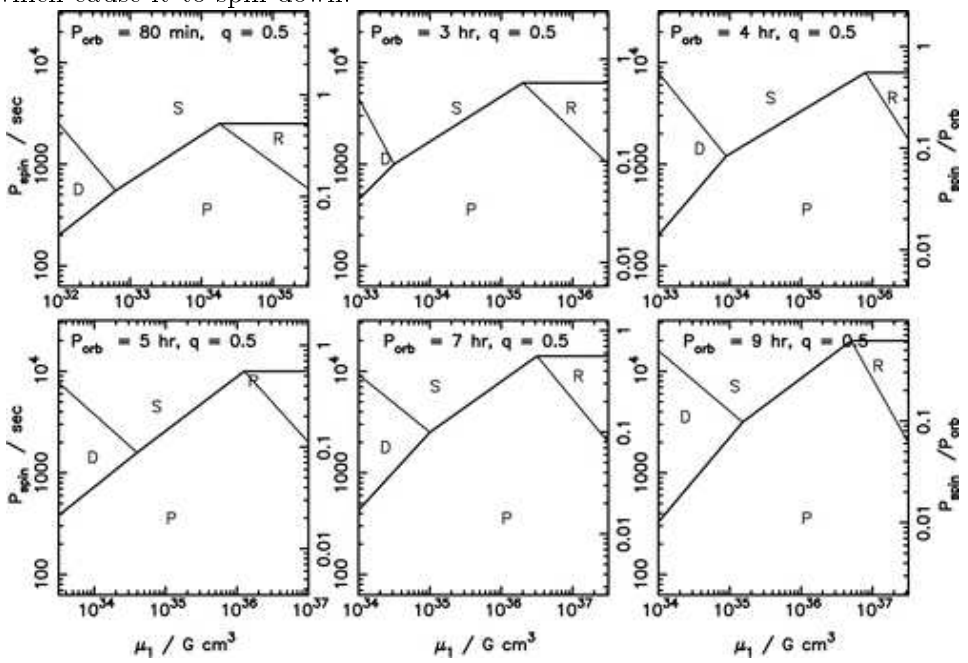
propellers in which most of the transferred material from the secondary is magnetically propelled away from the system by the rapidly spinning magnetosphere of the WD.

discs in which most of the material forms a circulating flattened structure around the WD, truncated at its inner edge by the WD magnetosphere where the material latches to the magnetic field lines before accreting onto the WD surface.

streams in which most of the material latches onto field lines immediately and follows these on a direct path down the WD poles.

rings in which most of the material forms a narrow annulus circling the WD at the outer edge of its Roche lobe, with material being stripped from its

Figure 4.5: The distribution of accretion flow types as a function of orbital period, in the spin period vs. magnetic moment plane, at a mass ratio of q = 0.5 taken from Norton et al. [2008b]. The right hand axes show the spin to orbital period ratio in each case. Approximate regions within which each type of flow is seen are delineated as shown, where D stands for disc accretion, S for stream accretion, R for ring accretion and P for propeller flow. The thick line shows the approximate locus of the equilibrium spin period in each case and marks the boundary between accretion flows that spin-up the WD and those which cause it to spin-down.



inner edge by the magnetic field lines before being channeled down the WD surface.

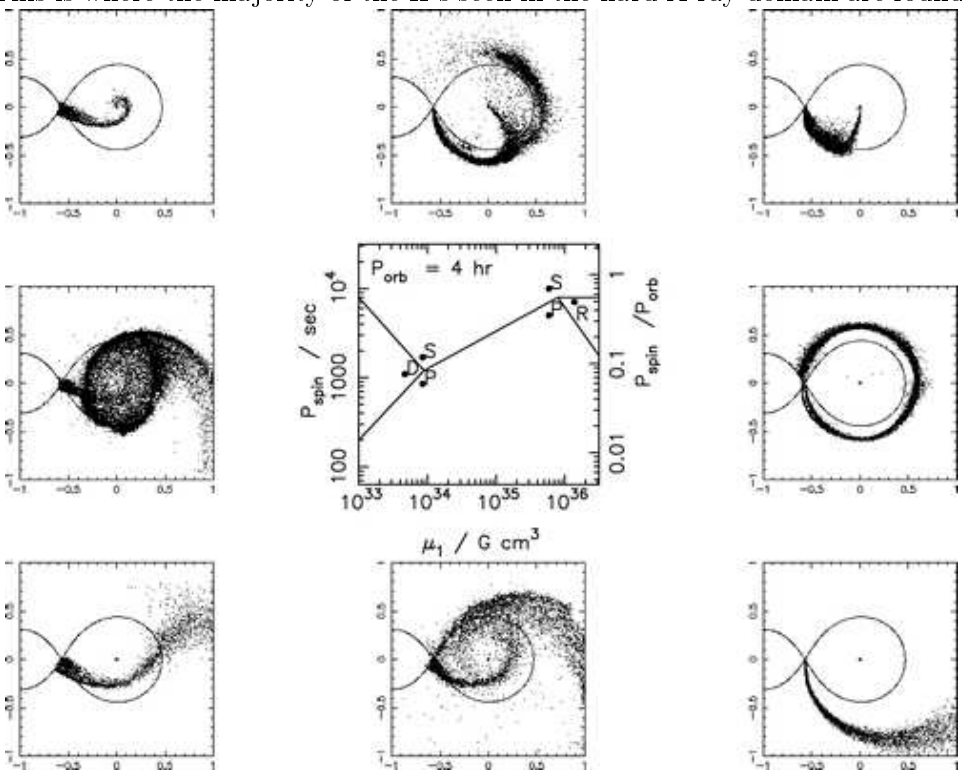
As mentioned before, the main observable in determining the kind of flow a system exhibits is the spin-to-orbital period, however mass ratio and magnetic field strength of the WD also play a role. Figure 4.5, taken from Norton et al. [2008b], shows some of the results from their simulations for systems with mass ratio of $q = 0.5$. Each panel is for a particular orbital period. The drawn boundaries are there for reference and it should be noted that in reality these are quite blurred. Nonetheless the planes all divide into four regions.

Clearly this marks the boundary between accretion flow types that will generally spin-up (streams) and accretion flow types that will generally spin-down (propellers). Broadly speaking, if an IP system is found in a region of the parameter space where it is fed by stream accretion it will spin-up the WD and move it downwards in the plane in Figure 4.5 towards the equilibrium line. On the other hand if an IP system finds itself in a region of parameter space where the flow takes the form of a propeller it will spin-down the WD and so move it upwards towards the equilibrium line. We note that both the ring and stream accretion will keep the WD close to spin equilibrium through a combination of accretion and ejection of material. Moreover from Figure 4.5 we also point out the two triple points of equilibrium which all systems are trying to reach according to simulations. If an IP reaches one of these then it is prone to stay there forever and not become a totally synchronised polar.

To have a better idea of what these accretion flow types might look like we show Figure 4.6 again taken from Norton et al. [2008b]. From this, one can see that close to the stream-disk-propeller triple point (at about $P_{spin}/P_{orb} = 0.1$) and the stream-ring-propeller triple point (at about $P_{spin}/P_{orb} = 0.6$), the equilibrium flows are a combination of the various flow types. In each case the angular momentum accreted by the WD is balanced by an equal amount lost from the system via material magnetically propelled away: the definition of an equilibrium spin period.

Having reviewed some of the relevant results from theoretical simulations we will move on to introduce the contemporary observations of mCVs, with particular emphasis on the *INTEGRAL*/IBIS observations.

Figure 4.6: The variation of the accretion flow in the vicinity of the boundaries between the different flow types. The panels on the left show flows in the vicinity of the stream-disc-propeller triple point, whilst the panels on the right show flows in the vicinity of the stream-ring-propeller triple point. The panel at the bottom, centre is the accretion flow at the stream-disc-propeller triple point and shows characteristics of all three flows at an equilibrium spin period. This is where the majority of the IPs seen in the hard X-ray domain are found.



4.3 Recent hard X-ray observations of mCVs

In recent years, an increasing number of mCVs have been detected and discovered by hard X-ray telescopes such as *INTEGRAL*/IBIS [Barlow et al., 2006, Landi et al., 2009], *Swift*/BAT [Brunschweiger et al., 2009] and *SUZAKU*/HXD [Terada et al., 2008]. This increase has been mainly caused by the observing strategy of hard X-ray observatories focusing on large field of view survey studies.

The *INTEGRAL* satellite, launched in October 2002, has now carried out more than 6 years of observations in the energy range 5 keV to 10 MeV. In particular, the *INTEGRAL*/IBIS survey is one of the main mission objectives. The IBIS (Imager onboard *INTEGRAL* spacecraft) detector [Ubertini et al., 2003, Lebrun et al., 2003] has been optimised for survey work, with a large field of view (30°) and with unprecedented sensitivity in the soft-gamma ray regime, yielding excellent imaging capabilities. It is worth emphasising that the IBIS survey has been optimised to detect faint persistent sources, which mCVs are. The aim of the survey is to expand the current knowledge of the 20-100 keV sky by cataloguing high-energy sources and examining their properties, both individually and globally. The IBIS survey dataset consists of dedicated observations along the Galactic plane and around the Galactic centre. Additionally, a combination of pointed and deep exposure observations are added to the dataset once they become public. As a result, the latest release of the *INTEGRAL*/IBIS survey provides all-sky coverage, albeit with spatially variable sensitivity. The depth of the IBIS survey has increased significantly with each release [Bird et al., 2004, 2006, 2007] and has now reached a peak sensitivity corresponding to a flux limit below 1 mCrab in the 20-100 keV range. The latest catalogue 3 release [Bird et al., 2007] contained a total of

421 objects, of which 20 were catalogued as CVs or were identified as CVs later on.

This work uses a more recent IBIS dataset consisting of over 36,000 individual Science Window (ScW) pointings, covering 6 years of observations (revolutions $\approx$ 46-660). These have been processed with the latest pipeline (OSA 7.0, Goldwurm et al. 2003), and mosaics were created from the deconvolved images in 5 energy ranges. Staring and performance verification (PV) observations have not been used for the mosaic creation, and noisy ScWs have been excluded based on the rms of the individual images. These new survey maps are a considerable improvement on any previous ones, due to the new pipeline software, together with the significantly increased exposure times. It is worth pointing out that this dataset is exactly the same as the one used in the previous chapter, where we have applied ISINA to catalogue 4 data.

Swift has been optimised to locate gamma-ray bursts, and as a consequence the main hard X-ray instrument, the Burst Alert Monitor (BAT), has similar capabilities to IBIS, possessing a large field of view and operating in essentially the same energy range. BAT has also been used for survey work [Tueller et al., 2009], and in particular has also detected a high number of IPs [Brunschweiger et al., 2009]. In order to make our study as complete as possible, we have decided to include the IPs from Brunschweiger et al. [2009] observed with the *Swift*/BAT detector. Moreover, one extra IP has been added, AE Aqr, as observed by *SUZAKU*/HXD [Terada et al., 2008], sampling again a very similar energy range to IBIS. The similarity of energy range allows us to construct a hard X-ray selected sample with minimal biases.

In total, the three telescopes mentioned above have observed $\approx$ 30 mCVs above 17 keV. More than 90% of these are IPs, and there are also two rare asynchronous polars. When compared to the older soft X-ray selected samples

of mCVs the picture is slightly different. First of all, as one would expect, soft X-ray detectors are more sensitive to mCVs and as a consequence will produce larger samples, including equal amounts of polars and IPs. However, hard X-ray observations are consistently revealing a particular subset of fast spinning IPs above the period gap.

4.4 The hard X-ray CV population

4.4.1 *INTEGRAL*/IBIS CVs

Here we describe the catalogue matching procedure adopted in order to identify known CVs in our IBIS dataset. To do this we require the most complete and up-to-date catalogue of such objects, thus we merge the two most complete CV catalogues in the literature: The Catalogue and Atlas of Cataclysmic Variables (Downes et al. 2005, hereafter DWScat) and the Catalogue of Cataclysmic Binaries (Ritter and Kolb 2003, hereafter RKcat). DWScat contains 1830 CVs whilst RKcat contains 731, and we note that 656 CVs are common to both catalogues. The main reason for RKcat having fewer objects is that only CVs with known orbital periods are included in the sample; however, RKcat also includes a few CVs that DWScat does not report. Our final known CV set therefore contains 1905 CVs. Fewer than 10% of the total number of CVs within RKcat are known to be magnetic in nature (included in the catalogue as either DQ Her, AM Her or IP), and only approximately 3% (56 sources) are IPs.

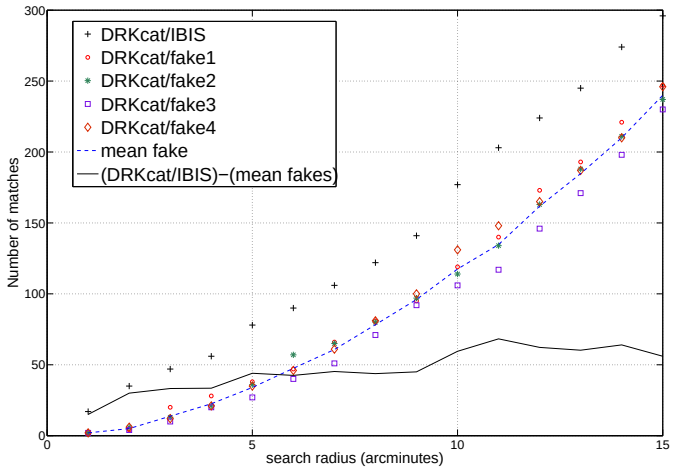
Catalogue matching has been performed between the total CV set produced, which contains 1905 CVs (hereafter DRKcat) and a preliminary IBIS candidate excess list containing real sources constructed in the same way as for the ISINA algorithm (Chapter 3), containing over 9000 excesses constructed

from the public data available for revolutions 46-660. This includes excesses detected in the final mosaics, revolution mosaics and revolution sequence mosaics in any of the 5 main energy bands in order to locate variable candidates too. We define a variable search radius around the IBIS coordinates with a maximum value of 15' (this value is extremely large but will allow us to observe the general trend of the matching procedure). If an object in the DRKcat was found within the search radius, it is flagged as a possible match. Four additional "fake" IBIS catalogues have been created in the following way, following a similar method to Stephen et al. [2006]:

- transposing the IBIS coordinates by one degree in Galactic longitude (fake 1)
- mirroring the IBIS coordinates in Galactic longitude (fake 2),
- mirroring the IBIS coordinates in Galactic latitude (fake 3)
- mirroring the IBIS coordinates in both Galactic latitude and longitude (fake 4).

The results are shown in Figure 4.7. Ideally using this method, one would expect the black solid line in Figure 4.7 to flatten out at the optimal matching radius. This is because, ideally, we expect that after a certain radius the DRKcat-IBIS matches would grow at the same rate as the fake samples (blue-dashed line in Figure 4.7). Clearly however this is not the case, the black-solid line is still increasing at $10'$, an exaggeratedly large radius for cross-matching objects. We associate this effect to the fact that our initial $\approx$9000 candidates which cross-matching is performed against DRKcat is over populated by noise. In particular noise correlated to real objects as discussed in Chapters 2 and 3 , and the galactic plane. This will cause a lot of sporadic matches, in particular

Figure 4.7: The number of matches as a function of search radius between the DRKcat and the IBIS excess list (crosses). Also plotted are the results from correlating the DRKcat with the 4 fake excess lists and the mean number of matches from the fake correlations (dotted line). The solid line is the number of DRKcat/IBIS matches minus the mean of the fake matches.



with very large matching radii. We therefore suggest that one has to be careful in using this method when the initial candidate list is overpopulated by noise correlated with real Galactic sources. Having established the poor reliability of the method in this circumstance we adopted a similar radius to that chosen by Barlow et al. [2006] of $4'$, which also corresponds well with the expected error on faint IBIS detections [Gros et al., 2003]. For a search radius of $4'$ we obtain 56 sources as confirmed or candidate CVs, of which 23 are expected to be false coincidences. We have visually inspected all of the correlated sources and found 33 matches coincide with mainly non-CV globular cluster sources and previously identified X-ray objects, however some are image artefacts related to the Galactic centre region. It is important to note that with a $\approx 2'$ source location accuracy it is very hard to associate a detection with an optical counterpart alone. We have performed the same exercise by increasing the search radius to $5'$ which increases the sample to inspect to 76 candidates. We

find that all the additional matches obtained with increasing the radius are false, with the possible exception of the Dwarf Nova DN V1830 Sgr located 4.8′ away from the IBIS detection in revolution 106 (MJD 53128.8 - 53131.7). This is slightly out of the 90% error radius for a $\approx 6\sigma$ detection and we cannot definitely associate the two at the moment.

Table 4.1 shows the main characteristics of the 23 objects identified from our correlation analysis. We have estimated distances for this sample using the method described by Knigge [2006] based on the evolution of the donor star where 2MASS K-band magnitudes were available [Cutri et al., 2003] and $P_{orb} < 6.2$ hours[1]. In addition we show in Table 4.2 the 9 other IBIS-detected mCVs used in this work. These were not part of the correlation analysis, because they were not present in DRKcat, but have been identified through optical spectroscopy following the IBIS discovery. Of the 23 objects considered to be real matches from our analysis, 17 are previously known *INTEGRAL* detected CVs, whilst 6 sources are new detections. Most of the new objects are of the intermediate polar subclass with the possible exception of TW Pic which is considered by some as a VY Scl star (Norton et al. 2000) and AX J1832.3-0840 which is not identified in full at the moment.

4.4.2 *Swift*/BAT and *SUZAKU*/HXD CVs

Swift/BAT has also observed a large number of mCVs, and it is worth including these in our study for completeness. We decided to include all the 22 BAT detected IPs [Brunschweiger et al., 2009], where 14 have been observed by IBIS as well. Similarly to the IBIS-only IPs, the BAT-only IPs are all placed above the period gap with the exception of EX Hya.

SUZAKU/HXD on the other hand has a different observing strategy com-

[1] The range within the method is applicable

Table 4.1: Results of the IBIS-DRK catalogue matching with $4'$ search.

Name[a]	α, δ[b] (IBIS position)	type[c]	offset[d] ($'$)	Map code[e]	Count rate[f] ($ct\ s^{-1}$)	Exposure (ks)	Flux[g] 20-100 keV	P_{orb} (min)	P_{spin} (s)	Distance[h] (pc)	Refs
1RXS J002258.3+614111	5.739,61.714	IP	1.7	B4(8.7)	0.15 ± 0.01	3767	0.81	241.98	563.53	510	[1,2,3,4]
V709 Cas	7.207,59.303	IP	0.8	B5(54.3)	1.03 ± 0.01	3562	5.53	320.4	312.77	300	[1,2,5]
XY Ari⋆	44.047,19.457	IP	1.1	B5(5.5)	0.53 ± 0.12	119	2.85	363.884	206.298	610	[1,2,6]
GK Per	52.777,43.928	IP/DN	1.7	B5(4.7)	0.26 ± 0.07	277	1.4	2875.4	351.34	-	[1,2]
TV Col⋆	82.357,-32.819	IP	0.1	B4(11.6)	0.68 ± 0.08	248	0.37	329.181	1911	330	[1,2,7,8]
TW Pic⋆	83.766,-57.998	IP?/VY Scl?●●	2.5	B1(5.8)	0.3 ± 0.07	363	1.61	-	-	-	[1,2,9,10,11]
BY Cam	85.728,60.842	AP	1.2	B5(5.1)	0.69 ± 0.11	162	3.7	201.298	11846.4	140	[1,2]
MU Cam	96.316,73.567	IP	0.6	B4(5.4)	0.24 ± 0.06	548	1.29	283.104	1187.24	440	[1,2,12]
SWIFT J0732.5−1331⋆	113.13,-13.513	IP	1.6	B3(6.1)	0.39 ± 0.06	409	2.09	336.24	512.42	-	[1,2,13]
V834 Cen	212.260,-45.290	P	0.9	B1(5.4)	0.16 ± 0.03	1675	0.86	101.51712	6091.0272	70	[1,2]
IGR J14536−5522	223.421,-55.394	P	2.0	B4 (11.9)	0.27 ± 0.03	2658	1.45	189.36	11361.6	140	[1,2,14]
NY Lup	237.052,-45.481	IP	0.5	B5(49.1)	1.17 ± 0.03	3141	6.28	591.84	693.01	-	[1,2,15]
V2400 Oph	258.173,-24.279	IP	2.2	B1(33.4)	0.68 ± 0.02	4453	3.65	204.48	927.6	180	[1,2,16]
1H 1726−058	262.606,-5.984	IP	0.7	B5(22.8)	0.85 ± 0.04	1449	4.56	925.27	128	-	[1,2,17]
V2487 Oph	262.960,-19.244	IP/N●●	2.3	B3(9.1)	0.18 ± 0.02	4562	0.97	-	-	-	[1,2,18]
AX J1832.3−0840⋆	278.083,-8.721	?	3.1	B4(5.5)	0.07 ± 0.03	3090	0.38	-	-	-	[1,2,19,20]
V1223 Sgr	283.753,-31.153	IP	0.8	B5(52.2)	1.45 ± 0.03	2358	7.79	201.951	746	150	[1,2]
V1432 Aql	295.052,-10.421	AP	0.2	B5(10.8)	0.69 ± 0.07	429	3.7	201.938	12150.4	240	[1,2,21,22]
V2069 Cyg	320.906,42.279	IP●	1.8	B5(6.2)	0.21 ± 0.03	1648	1.13	448.824	743.2	-	[1,2,23,24]
1RXS J213344.1+510725	323.446,51.122	IP	0.3	B5(25.8)	0.65 ± 0.03	2207	3.49	431.568	570.82	-	[1,2,25]
SS Cyg	325.698,43.582	DN	0.6	B5(23.0)	0.7 ± 0.03	1674	3.76	396.1872	-	-	[1,2,26]
FO Aqr	334.514,-8.354	IP	1.7	B4(6.1)	0.65 ± 0.2	54	3.49	290.966	1254.45	250	[1,2,27]
AO Psc⋆	343.815,-3.194	IP	1.3	B4(4.8)	0.43 ± 0.11	108	2.31	215.461	805.2	200	[1,2,28,29]

References: [1]Ritter and Kolb [2003]; [2]Downes et al. [2005]; [3]Bonnet-Bidaud et al. [2007]; [4]Masetti et al. [2006a]; [5]Bonnet-Bidaud et al. [2001]; [6]Norton and Mukai [2007]; [7]Hellier [1993]; [8]Augusteijn et al. [1994]; [9]Buckley and Tuohy [1990]; [10]Chen et al. [2001]; [11]Norton et al. [2000]; [12]Araujo-Betancor et al. [2003]; [13]Butters et al. [2007]; [14]Masetti et al. [2006b]; [15]de Martino et al. [2006]; [16]Hellier and Beardmore [2002]; [17]Gänsicke et al. [2005]; [18]Hernanz and Sala [2002]; [19]Muno et al. [2004]; [20]Sugizaki et al. [2000]; [21]Watson et al. [1995]; [22]Geckeler and Staubert [1997]; [23]Thorstensen and Taylor [2001]; [24]de Martino et al. [2009]; [25]Bonnet-Bidaud et al. [2006]; [26]Friend et al. [1990]; [27]Marsh and Duck [1996]; [28]Kaluzny and Semeniuk [1988]; [29]van Amerongen et al. [1985];

[a] ⋆ indicates new hard X-ray detections

[b] Right ascension and declination in degrees, J2000

[c] IP=intermediate polar, P=polar, AP=asynchronous polar, N=nova, DN=dwarf nova. All are confirmed except for ●:probable, ●●: possible

[d] Angular distance between the DRKcat catalogue positions and the IBIS coordinate

[e] IBIS detection only. Map with maximum significance: (B1) 20-40 keV; (B2) 30-60 keV; (B3) 20-100 keV; (B4) 17-30 keV; (B5) 18-60 keV; in brackets appears the significance value.

[f] Determined between 20-100 keV

[g] The flux is calculated assuming a power law spectra with index of -2.9, the average index for IPs (Barlow et al. 2006) in unitis of $10^{-11} erg\ cm^{-2}\ s^{-1}$

[h] The distances have been computed with 2MASS K band magnitudes (Cutri et al. 2003) using the method presented by Knigge [2006] based on the evolution of the secondary.

Table 4.2: Additional mCVs detected by IBIS not included in DRKcat.

Name	α, δ	type	Map code	Count rate[a] ($ct\ s^{-1}$)	Exposure (ks)	Flux 20-100 keV	P_{orb} (min)	P_{spin} (k)	Distance (pc)	Refs
IGR J08390−4833	129.705,-48.524	IP••	B5(6.3)	0.08 ± 0.02	3072	0.43	–	–	–	[1]
XSS J12270−4859	187.004,-48.906	IP••	B5(9.9)	0.42 ± 0.04	955	2.26	–	–	–	[2,3,4]
IGR J15094−6649	227.406,-66.823	IP	B5(11.0)	0.18 ± 0.03	3265	0.97	353.40	809.42	600	[3,5]
IGR J16167−4957	244.140,-49.974	IP••	B4(20.7)	0.4 ± 0.02	3466	2.15	–	–	–	[3]
IGR J16500−3307	252.491,-33.114	IP/DN	B1(13.3)	0.33 ± 0.03	3353	1.77	217.02	597.92	270	[5,6]
IGR J17195−4100	259.906,-40.997	IP	B5(23.3)	0.54 ± 0.02	4279	2.9	240.30	1139.55	120	[2,3,5]
IGR J18173−2509	274.353,-25.158	IP••	B1(14.9)	0.27 ± 0.01	5744	1.45	–	–	–	[7]
IGR J18308−1232	277.696,-12.532	IP••	B5(7.1)	0.18 ± 0.03	3265	0.97	–	–	–	[8]
IGR J19267+1325	291.670,13.425	IP•	B5(7.5)	0.15 ± 0.03	2963	0.81	–	–	–	[9,10]

All classifications are correct except for, •: probable classification ••: possible classification. References:[1]Kniazev et al. [2008]; [2]Butters et al. [2008]; [3]Masetti et al. [2006b]; [4]Saitou et al. [2009]; [5]Pretorius [2009]; [6]Masetti et al. [2008a]; [7]Masetti et al. [2008b]; [8]Parisi et al. [2008]; [9]Steeghs et al. [2008]; [10]Evans et al. [2008].

[a]Determined between 20-100 keV

Table 4.3: Additional IPs used in this work detected with *Swift*/BAT and *SUZAKU*/HXD with known spin and orbital periods.

Name	α, δ	type	Detection	P_{orb} (min)	P_{spin} (s)	Refs
V1062 Tau	75.615,24.756	IP	BAT	597.60	3726	[1]
TX Col	85.834,-41.032	IP	BAT	343.15	1909.7	[1]
V405 Aur	89.897,53.896	IP	BAT	249.12	545.455	[1]
BG CMi	112.871,9.940	IP	BAT	194.04	847.03	[1]
PQ Gem	117.822,14.740	IP	BAT	310.80	833.4	[1]
DO Dra	175.910,71.689	IP	BAT	238.139	529.31	[1]
EX Hya	193.102,-29.249	IP	BAT	98.257	4021.62	[1]
AE Aqr	310.038,-0.871	IP	HXD	592.785	33.0767	[2]

References:[1]Brunschweiger et al. [2009]; [2]Terada et al. [2008].

posed of small field of view pointings. This does not allow the telescope to produce survey data like IBIS or BAT, however mCVs have been detected and observed with HXD. In particular HXD had observed the IP AE Aqr [Terada et al., 2008], which has not been observed with either IBIS or BAT, and therefore is included in our study as well. We caution however that the possible origin of the hard X-rays in AE Aqr could be non-thermal [Terada et al., 2008]. This however has not been shown in full, and given the power-law model fit to the 3-25 keV spectra of AE Aqr yielding an index of 2.10 [Terada et al., 2008] not far from the indices found in fast spinning mCVs [Landi et al., 2009]. Moreover, Mauche [2009] has very recently brought forward additional evidence, using *Chandra*/HETG, that the high energy X-ray excess in AE Aqr is of thermal nature, so we believe this object should still be included in our analysis.

All the IPs observed with BAT and HXD used in this study are presented in Table 4.3, together with their orbital and spin periods. It is worth pointing out, that as CVs are a very local population, we expect to see them as an isotropic distribution, which favours IBIS and BAT, but particularly BAT which has a

more uniform sky coverage.

4.5 The P_{orb} - P_{spin} plane

Theoretical simulations on the evolution of mCVs have been performed by Norton et al. [2008b, 2004]. One interesting prediction of these models is that of different accretion flows for IPs depending on where they are in their evolutionary stage. In particular these models also predict the existence of spin equilibria among IPs, one at about $P_{spin}/P_{orb} \sim 0.1$ and a second one at $P_{spin}/P_{orb} \sim 0.6$ (depending on mass ratio).

Figure 4.8 shows the P_{orb} - P_{spin} plane for mCVs together with their orbital, spin and synchronicity (P_{spin}/P_{orb}) distributions. The hard X-ray detected systems used in this work are shown with stars. Also plotted for reference is the period gap and three "synchronicity" lines showing $P_{spin} = P_{orb}$ (polars), $P_{spin} = 0.1P_{orb}$ and $P_{spin} = 0.3P_{orb}$. It is clear that most of the hard X-ray detected systems are above the period gap and have $P_{spin} \leq 0.1P_{orb}$. The only IP system outside this range is EX Hya which is closer than 100 parsec [Brunschweiger et al., 2009]. Similarly to the two detected polars, it is of no real surprise that hard X-ray detectors can see this close object.

The fact that no IP has yet been observed with hard X-ray telescopes above the period gap with $P_{spin}/P_{orb} >> 0.1$ suggests that these IPs may have different accretion flows, yielding different emission mechanisms compared to IPs with $P_{spin}/P_{orb} < 0.1$. This idea is supported by the fact that the distribution of all known mCVs does indeed seem to peak at about $P_{spin}/P_{orb} \sim 0.1$ in the bottom left panel in Figure 4.8, where a spin equilibrium has been predicted. As IPs evolve from low synchronicity ($P_{spin}/P_{orb} << 0.1$) towards their $P_{spin}/P_{orb} \sim 0.1$ equilibrium, their accretion flows resemble those of propeller

systems, where a lot of the material incoming from the secondary is actually propelled away and does not reach the pole of the WD. So in essence most IPs with $P_{spin}/P_{orb} << 0.1$ have yet to reach their equilibrium (regardless of field strength), and their accretion flows are different from any other IP elsewhere in the P_{spin} - P_{orb} plane [Norton et al., 2008b, 2004]. As a consequence one would not necessarily expect the hard X-ray emission mechanisms to be the same for all IPs.

Another interesting observational feature of this plane is that only one unconfirmed IP system, V697 Sco, lies within what we call the synchronicity gap: a region in the P_{orb} - P_{spin} plane within $P_{spin} > 0.3P_{orb}$ and total synchronicity above the period gap. The low number of IP systems with high synchronicity above the period gap is also predicted by the models. As explained by Norton et al. [2008b, 2004], as mCVs evolve, both their mass ratios and orbital period decrease. These trends individually cause opposite shifts in the spin-to-orbital ratio at which the spin-equilibrium occurs for a given magnetic field. As a result, typical IPs with magnetic field strength of a few MG will evolve from being disc-like accretors at long orbital period (where $P_{spin}/P_{orb} \sim 0.1$), to ring-like accretors at short orbital period (where $P_{spin}/P_{orb} \sim 0.6$), providing that they do not synchronise along the way and become polars. The two spin-to-orbital equilibria are determined approximately by two conditions. The first $P_{spin}/P_{orb} \sim 0.1$, is given by the condition that $R_{co} \sim R_{circ}$ [King and Wynn, 1999], where R_{co} is the corotation radius, at which matter within the accretion disk corotates with the magnetic field lines, and R_{circ} the circularisation radius at which point the Kepler specific angular momentum equals that of the matter being accreted through the L1 point. The second equilibria at $P_{spin}/P_{orb} \sim 0.6$ is given by the condition $R_{circ} \sim b$, where b is the distance of the L1 point to the WD. Both these conditions come from the interaction of the magneto-

spheric radius with the formed accretion disk, and more information can be found in King and Wynn [1999]. If the magnetic field of the systems is of the order of a few hundred MG, on the other hand, then they will be prone to synchronise and become polars, crossing the synchronicity gap fairly quickly, helping explain the low number of systems in this region. This is the likely history of EX Hya and systems neighboring it in the P_{orb} - P_{spin} plane.

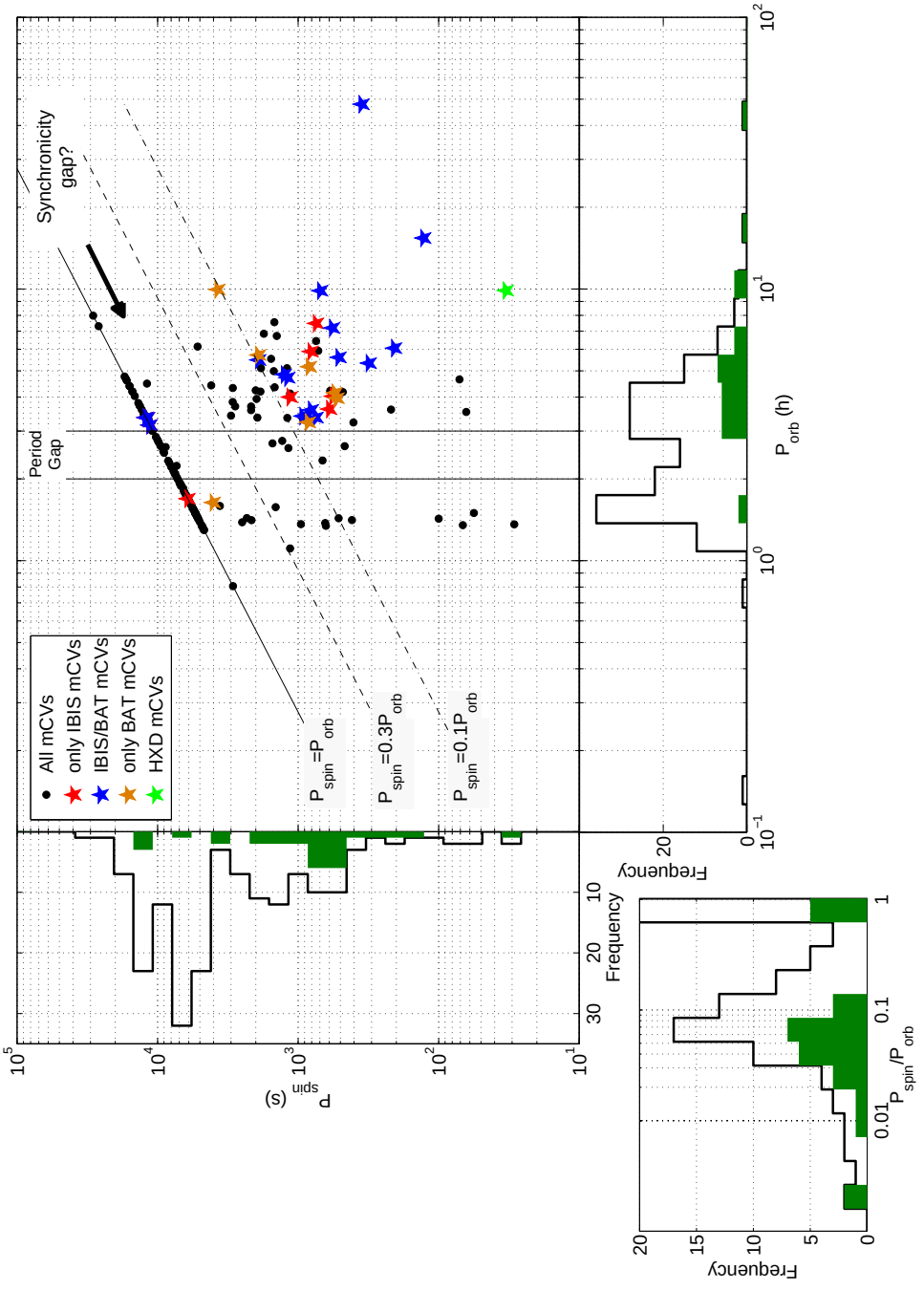
We therefore predict that, given the already long exposure times accumulated with IBIS and BAT, the IPs within $P_{spin} = 0.1P_{orb}$ and $P_{spin} = 0.3P_{orb}$ region (synchronicity gap) will not be detected in significant numbers. Conversely we expect most of the IPs with $P_{spin} < 0.1P_{orb}$ to be observable with longer exposure times and better sensitivity.

4.5.1 Are hard X-ray mCVs different?

It should be clear by now that hard X-ray telescopes are more sensitive at detecting the IP population rather than the polar one. However it is not yet clear whether hard X-ray telescopes are producing populations which are consistent with being drawn from the general mCV population (mostly identified through soft X-rays). In order to assess this we have performed a Kolmogorov-Smirnov test (KS test) on all the P_{spin}, P_{orb} and P_{spin}/P_{orb} distributions of hard X-ray selected samples versus the known mCV population taken from RKcat. In all cases the test rejects the null hypothesis that the distributions are drawn from the same parent with 99.99% confidence.

As mentioned before, the difference in distributions within P_{orb} (bottom panel in Figure 4.8) between these sets could be caused by the fact that all mCVs below the period gap are intrinsically less luminous given their lower accretion rates. However it does not exclude the possibility that the X-ray emission mechanism for mCVs below the period gap is substantially different

Figure 4.8: P_{orb} vs. P_{spin} for mCVs taken from RKcat. Stars indicate mCVs detected at hard X-ray energies. Also plotted is the period gap and "synchronicity" lines. For reference we also display the orbital, spin and synchronicity distributions. The shaded green areas represent hard X-ray selected mCVs.



than the mCVs above the gap. Both of these effects can be regarded as systematic, however the possibility of hard X-ray missions detecting homogeneous IP samples has already been suggested by Gänsicke et al. [2005], and here we bring forward more observational evidence for this. It is clear however that if mCVs below the period gap emit hard X-rays, then current telescopes are not sensitive enough to detect them without deeper exposures.

Moreover, from the P_{spin}/P_{orb} KS test (bottom left panel in Figure 4.8) result we can also comment on the fact that hard X-ray telescopes are not sensitive to high synchronicity systems. These do not necessarily have to live below the period gap, and in fact about half of the polar population lives above the gap. This suggests that hard X-ray surveys are very insensitive to the polar population as well as mCVs below the period gap. We can also look at this result and suggest that, on the other hand, hard X-ray surveys are extremely sensitive to IPs with long P_{orb} and to systems with low P_{spin}/P_{orb}.

4.6 Hard X-ray properties of mCVs

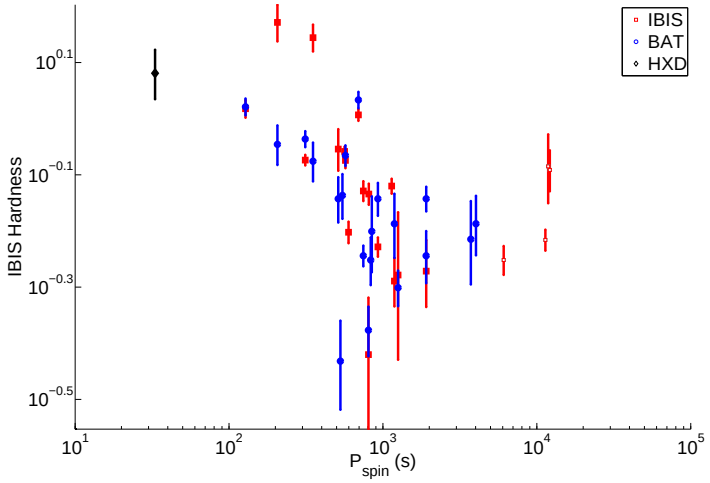
As revealed in Barlow et al. [2006] and Landi et al. [2009], the vast majority of IBIS detected CVs are magnetic (the only exception being SS Cyg). The total number of objects with known spin and orbital periods in our sample is 30 systems, of which 22 are IBIS detections (18 IPs). It is interesting to note the relative incidence of polar systems and intermediate polars. Optically selected samples favour the former, with the number of known polars being twice that of IPs. However, in our hard X-ray selected sample, the picture is very different, with only 4 polars being included in the sample (2 of which are APs). This result is expected, as IPs are known to produce ≈ 10 times more hard X-rays than polars due to their higher mass transfer and intrinsically

harder spectrum [Warner, 2003].

Only two synchronous polars have been detected in our sample, both at relatively close distance and in regions of the sky with high exposure times. This leads us to conclude that the flux in the hard X-ray range for these systems is much lower when compared to IPs or APs. We note from Table 4.1 that the two detected polars are among the closest objects in our sample, and hence it is probably for this reason that IBIS (and BAT, [Tueller et al., 2009]) can see them. We have also checked this by inspecting where other polars sit in the *INTEGRAL* exposure map and conclude that more deep observations are required before more of these systems are detected in the hard X-ray range.

Two of the four confirmed asynchronous polars (APs) are also included in our sample. At first one might think that these systems should have properties resembling the polar class. However, as shown by Schwarz et al. [2005], their accretion rates are $\approx 10 - 20$ times greater than that of polars. Moreover both *INTEGRAL* detected APs are 2-3 magnitudes brighter than the two non-detected APs in the K-band. This, together with our distance estimates suggests that our efficiency at detecting APs in the hard X-ray range is $\approx$ 50%, similar to the IPs. We note that the two APs not detected are CD Ind [Schwope et al., 1997] and V1500 Cyg [Lance et al., 1988] and have exposures of 3ks and 1987ks respectively. We further note for the record that neither the *ROSAT* all-sky bright or faint source catalogues do not contain V1500 Cyg, whilst they do contain the other 3 APs, which might suggest a low flux in the X-ray range for this source and thus explain the non detection by *INTEGRAL* at the moment, despite the 2Ms exposure..

Figure 4.9: 30-60/17-30 keV hardness versus spin period for the hard X-ray selected mCVs used in this work. Polars and APs are shown in empty squares.



4.7 Hardness plane correlations

The production of hard X-ray photons from mCVs is thought to originate in the post-shock region of the WD by bremsstrahlung cooling of free electrons. This is somewhat different to softer X-rays (< 2 keV) seen from mCVs, which can originate from a blackbody component close to the WD surface. In fact in recent years many medium resolution X-ray spectra have been obtained for different kinds of mCVs [Masetti et al., 2008a, Schwarz et al., 2005, Butters et al., 2008, Evans et al., 2008, Done and Magdziarz, 1998, Landi et al., 2009] and have been fitted with a soft blackbody component ($kT \approx 80$eV) plus a hard component characterised by the stratified accretion column of Cropper et al. [1999]. For those mCVs that are detected in the hard X-ray range the detection only samples the bremsstrahlung component. In particular the hard X-ray energy range (> 17 keV) is telling us about the temperature distribution of components within the multi-temperature bremsstrahlung emission, not the ratio of hard-to-soft X-ray components.

Figure 4.10: 30-60/17-30 keV hardness versus orbital period for the hard X-ray selected mCVs used in this work. Polars and APs are shown in empty squares.

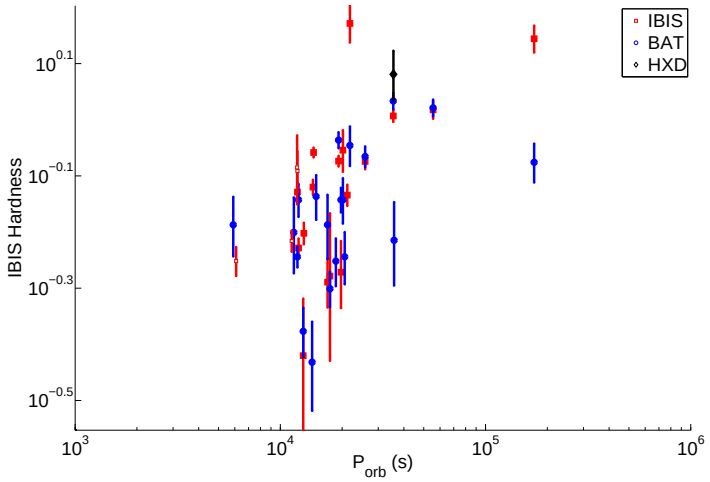
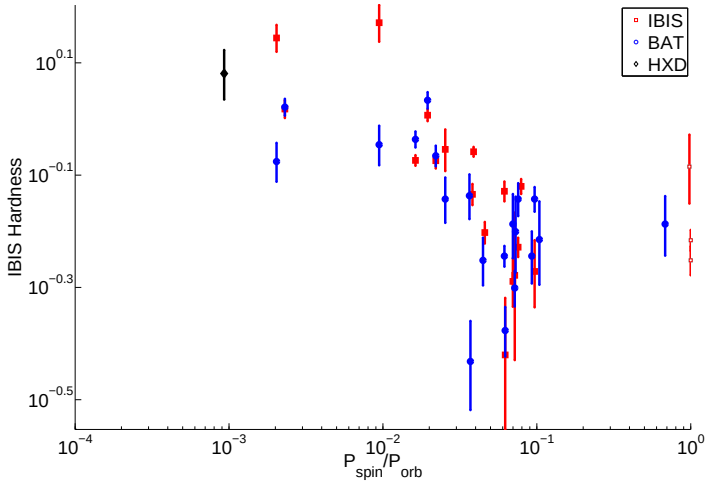


Figure 4.11: 30-60/17-30 keV hardness versus P_{spin}/P_{orb} for the hard X-ray selected mCVs used in this work. Polars and APs are shown in empty squares. We note the evident correlation for IPs with $P_{spin}/P_{orb} < 0.1$.

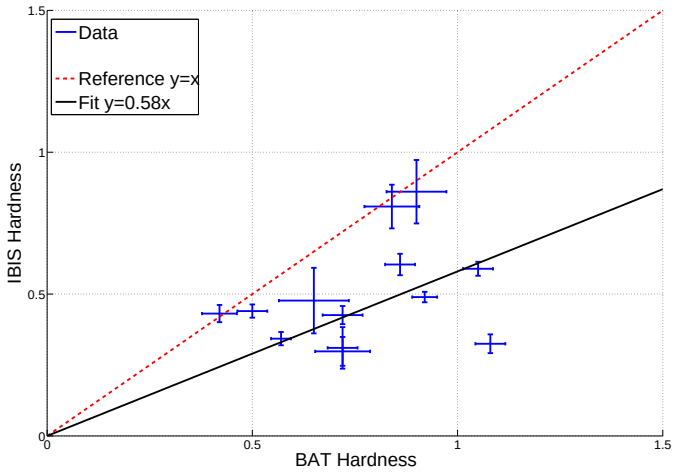


Keeping this in mind, Figures 4.9, 4.10 and 4.11 shows the scatter plots for hardness, defined as the count ratio in the 30-60 keV and 17-30 keV bands, versus P_{spin}, P_{orb} and P_{spin}/P_{orb} respectively for all hard X-ray detected mCVs used in this work. In red, we show all mCVs seen by IBIS, in blue, BAT-detected mCVs and in black, we show the only IP used in this work observed by HXD, AE Aqr. In order to obtain hardness ratios for BAT and HXD mCVs, we have reproduced their bremsstrahlung spectra (power law for AE Aqr) using the temperatures (or photon index) provided by Brunschweiger et al. [2009] and Terada et al. [2008], respectively. We then extracted the hardness ratio from the spectra taking into account errors[2] (symmetric for all) by also reproducing the hottest and coldest spectra using the published errors for each source and computing the hardness. Before being able to add the BAT points to Figures 4.9, 4.10 and 4.11 it is necessary to remove systematic cross-calibration differences between the IBIS hardness and the BAT ones. Figure 4.12 shows the BAT hardness vs. the IBIS hardness for a sample of 13 IPs in common to both. In red we display the one-to-one line where most of the data should sit in the absence of systematic differences between the IBIS and BAT calibrations. However, it is easily noticeable that the BAT extrapolated hardness's are systematically harder than the IBIS ones. This could be caused by the different response matrices of the detectors. We compensate approximately for this by fitting a straight line through the datapoints and the origin (black line in Figure 4.12). We then apply a correction to all the BAT points before plotting them in Figures 4.9, 4.10 and 4.11.

All three plots show evident signs of correlations with hardness ratio when considering IPs alone. In particular, when considering IPs with $P_{spin}/P_{orb} < 0.1$ (systems which are on their way to equilibrium at $P_{spin}/P_{orb} = 0.1$ and

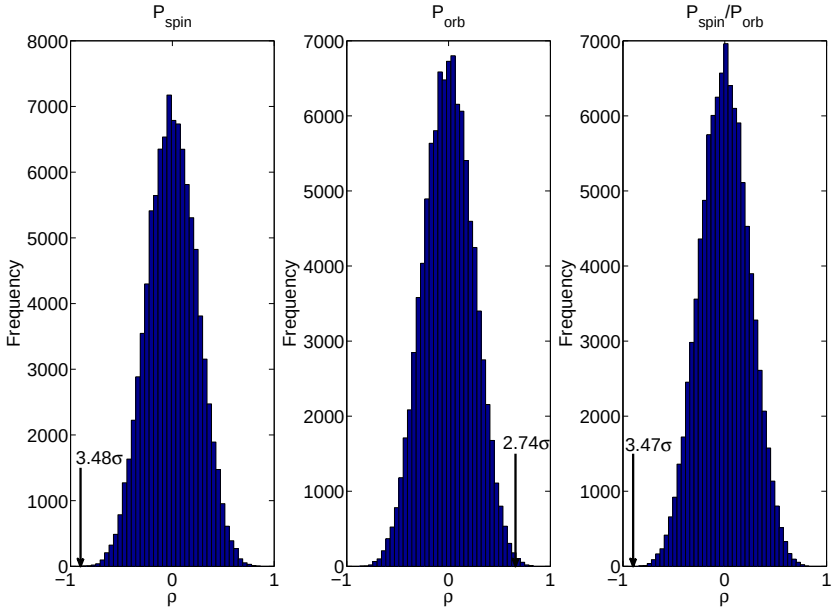
[2]We note that normalisation constants are not required when inferring hardness ratios from single model spectra

Figure 4.12: BAT extrapolated hardness vs. IBIS measured hardness. In blue is the data. In red we show a one to one line for reference. The cross-calibration fit is displayed in black. We note that both BAT and IBIS errors were taken into account when fitting.



are therefore exclusively in the propeller stage), then the correlations become even more evident. In order to obtain a significance for these correlations observed for the IPs thought to be in the propeller stage, we performed a Spearman rank test using a Monte-Carlo scheme. At first we decided to test the IBIS observations only, as we believe these are the measurements with lowest systematic errors. For example, in order to test if the correlation between hardness and P_{spin} is significant, we created 100,000 mock data sets containing the same number of points but shuffling the P_{spin} values randomly each time. Moreover, in order to take the hardness uncertainties into account, we replaced each hardness value with a random variable drawn from a normal distribution whose mean is equal to the observed hardness and whose standard deviation is equal to the error on the observation. We then calculated the Spearman rank coefficients, ρ. The distributions of the coefficients for all P_{spin}, P_{orb} and P_{spin}/P_{orb} are shown in Figure 4.13. Also displayed in each panel is the

Figure 4.13: Results from the Monte-Carlo simulation for estimating the correlation significances of the IBIS IPs only. The distributions display the calculated ρ coefficients for our mock datasets. We display with an arrow the calculated coefficient for the real set.

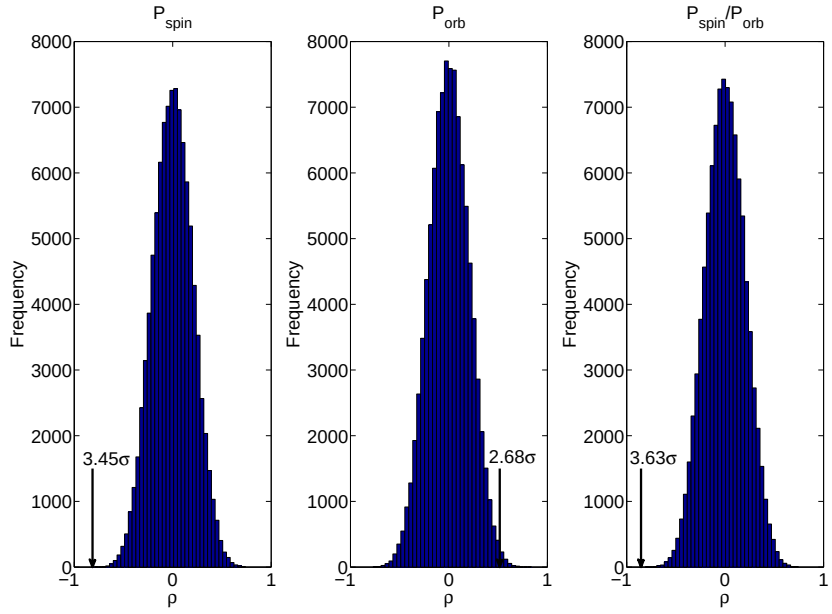


significance for the calculated Spearman rank value of the real data. These are 3.48σ, 2.74σ and 3.47σ for P_{spin}, P_{orb} and P_{spin}/P_{orb} respectively.

The fact that the correlation is apparent in all three plots is somewhat expected, since P_{spin} and P_{orb} are not independent, but expected to evolve together (Norton et al. 2008b, 2004). At this stage we perform the same exercise as for Figure 4.13, but this time including the additional IPs above the period gap observed by BAT only, and AE Aqr as observed by HXD. The results for this simulation are presented in Figure 4.14.

As a final step we decided to extend our analysis further for the observed correlation in P_{spin}/P_{orb} vs. hardness, given that, from an evolutionary perspective, it is expected to be the most relevant parameter [Norton et al., 2008b, 2004]. We produce a linear fit in log-log space to the P_{spin}/P_{orb} vs. hard-

Figure 4.14: Results from the Monte-Carlo simulation for estimating the correlation significances of the IBIS, BAT and HXD IPs. The distributions display the calculated ρ coefficients for our mock datasets. We display with an arrow the calculated coefficient for the real set.



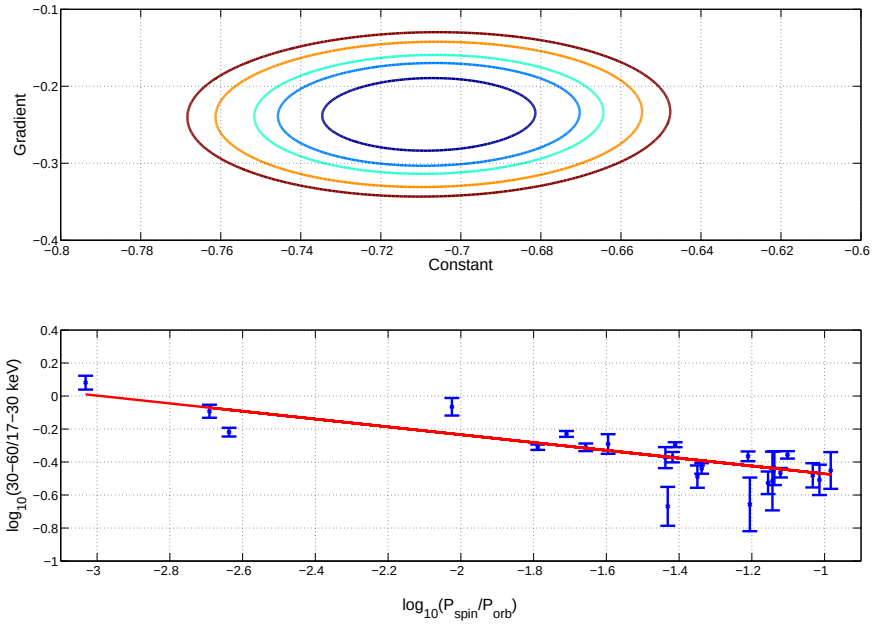
ness plot in Figure 4.15 for the hard X-ray selected mCVs used in this work[3]. Again we chose to perform the analysis on the IBIS-only IPs for consistency. When this is done, a non acceptable value of 4.5 is obtained for a reduced χ^2. However, following a similar procedure to Tremaine et al. [2002] and McHardy et al. [2006], we introduce a small intrinsic dispersion of 0.0046 to all our datapoints in linear space. This corresponds to $\approx 1\%$ for the soft IPs in Figure 4.15 and $\approx 0.5\%$ to the hardest IPs. There may be many reasons why such a small error might be introduced, ranging from a $\approx 1\%$ error in the IBIS response, to any small spectral variability intrinsic to the observed systems. The addition of this intrinsic dispersion lowers the reduced chi-squared to unity and results in more conservative errors on the fit parameters. In Figure 4.15 we display the contour plots for our linear fit in the top panel, and the fit itself on the bottom. Note that the contours represent lines of σ =1, 2, 3, 4, 5. Again one can see that a simple constant value straight line fit is not consistent with the data. The resulting equation to the fit can be expressed as 30-60/17-30 keV = $((P_{spin}/P_{orb}) - 0.0259)^{-0.21\pm0.05} - 10^{0.09\pm0.03}$, and may prove useful for modeling these systems and observations in the future. However at this stage, this is a purely empirical model.

4.8 Discussion

IBIS has so far detected 32 CVs (23 spatially correlated with known CVs + 9 new, optically confirmed discoveries). The majority are intermediate polars, but IBIS has also detected the bright dwarf nova SS Cyg and a few polars. This sample is an extension of the previously presented sample of IBIS CVs by Barlow et al. [2006], which also showed that the spectral characteristics of

[3] We have tried various polynomials but these all worsened the fit

Figure 4.15: Top panel: Contours for a linear fit to the datapoints in the bottom panel. Lines display contours for $\sigma = 1, 2, 3, 4, 5$. Bottom panel: IBIS IP hardness as a function of synchronicity. The equation resulting from the fit is 30-60/17-30 keV $= ((P_{spin}/P_{orb}) - 0.0259)^{-0.21\pm0.05} - 10^{0.09\pm0.03}$



these objects in the 20-100 keV range are actually quite similar and compare well with previous high-energy spectral fits. Moreover BAT has observed a similar number of objects (with many in common), which shows that modern hard X-ray surveys are able to consistently observe mCVs at higher energies than before.

The high incidence of IPs in our sample is not unexpected. Many authors [Kuijpers and Pringle, 1982, Warner, 2003] have suggested that the lower levels of hard X-rays/soft gamma-rays emission from polars may well be related to the low accretion rate and stronger magnetic fields compared to IPs. It has also been suggested that the strong magnetic fields in polars are able to produce a more "blobby" flow. These high density "blobs" are then able to penetrate deeper within the post-shock region and contribute more to the blackbody component of the broadband X-ray spectrum, and less to the bremsstrahlung component, making the broadband X-ray spectrum of IPs harder, and hence more luminous in the hard X-rays.

Two out of four known APs are observed in our sample. The remaining two happen to be in a region of low IBIS exposure and are likely more distant. Schwarz et al. [2005] have already shown that BY Cam, one of the observed APs, has different properties from those of normal polars, in particular having a much higher accretion rate. We therefore tentatively conclude that IBIS has not yet seen the two missing APs due to their distance/exposure. Our sample only includes 2 definite synchronous polars, and we do not expect many of these systems to be observed in the future with higher sensitivities above 17 keV.

Of the many observational characteristics presented here, one feature in the $P_{spin} - P_{orb}$ plane has stood out since the first study by Barlow et al. [2006]: a very low number of IPs below the period gap are detected with hard X-ray

telescopes. The only exception in our study is the very nearby IP EX Hya. We can compare this result to the theoretical models of Norton et al. [2008b, 2004]. It is believed that the IPs below the period gap have accretion flows which are very different to the IPs above the period gap. We would therefore not necessarily expect these systems to behave in the same way as the systems above the period gap, and, in particular, we would not necessarily expect them to emit such high energy photons. This is still an open question, and, as mentioned by Norton et al. [2008a], only deeper hard X-ray exposures will reveal if this subclass of IPs displaying ring-like accretion is able to produce similar amounts of hard X-rays as those observed in disk-fed systems at longer orbital periods. It has already been mentioned [Norton et al., 2008b] that as mCVs evolve through the period gap the magnetic field of the WD may be able to resurface when accretion stops. This can allow the system to synchronise, and we would then expect a system jumping from $P_{spin} \approx 0.1 P_{orb}$ above the period gap to $P_{spin} \approx P_{orb}$ below the period gap. Moreover we add to this that any system with $P_{spin}/P_{orb} \geq 0.6$ will never reach equilibrium until it reaches total synchronisation [Norton et al., 2008b, 2004].

All of the hard X-ray detected IP systems display $P_{spin} < 0.1 P_{orb}$, whilst none have been observed with $P_{spin} > 0.1 P_{orb}$. This may be more observational evidence for different kinds of accretion flows within the IP class, supporting the models of Norton et al. [2008b, 2004]. More evidence for these models comes from the non-detection of objects in any wavelength range within the synchronicity gap (a region above the period gap within $P_{spin}/P_{orb} > 0.3$ and $P_{spin}/P_{orb} < 1$). Such systems are predicted not to be very rare since IPs tend to evolve within the $P_{spin} - P_{orb}$ plane towards their equilibrium spin rate at $P_{spin} \approx 0.1 P_{orb}$ above the period gap or at $P_{spin} \approx 0.6 P_{orb}$ below the gap. Norton et al. 2008b, 2004 have predicted that low synchronisation mCVs

have propeller accretion flows and are all trying to reach equilibrium moving towards $P_{spin} \approx 0.1 P_{orb}$. This equilibrium arises due to the WD trying to balance angular momentum with the surrounding blobs. We believe this is the case for most hard X-ray selected IPs, as relatively few have yet been found above $P_{spin} = 0.1 P_{orb}$ where the accretion flow is thought to take the form of a stream.

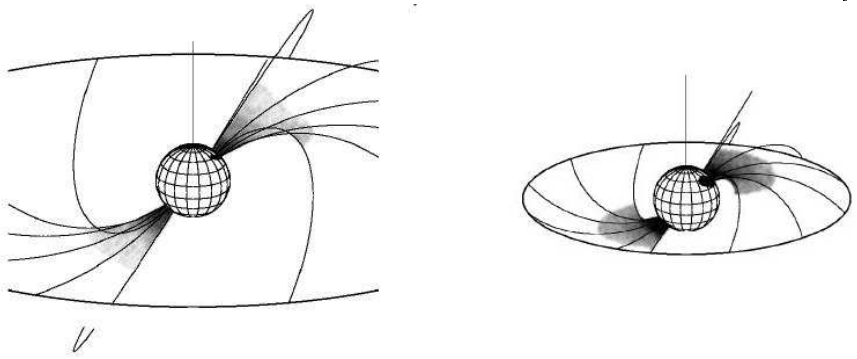
Perhaps the most interesting result of this study is the discovery of a correlation between 30-60/17-30 keV hardness and spin/orbital parameters for IPs. No similar correlation has been reported before, probably because previously measured X-ray hardness ratios of IPs were generally restricted to the range of approximately $\sim 0.5 - 10$ keV. Such ratios sample the lower end of the bremsstrahlung spectrum and the upper end of the blackbody spectrum, without fully measuring either component. In contrast, we note that the spectral hardness variations we have measured in our hard X-ray detected IPs span the energy range$\sim 17 - 60$keV and are *only* sampling the bremsstrahlung component of the spectrum. Therefore these observations tell us nothing about the relative contributions of the bremsstrahlung component emitted by the cooling plasma below the accretion shock and the blackbody component arising from the heated WD surface. Instead, they are directly sampling the relative contributions of the multi-temperature bremsstrahlung components that arise in the plasma below the shock front (the plasma cools as it settles towards the WD surface).

In order to explain the correlation between X-ray spectral hardness and spin-to-orbital period ratio in the hard X-ray detected IPs, we propose that the WDs in IPs are mostly accreting close to their equilibrium spin rates [Norton et al., 2004]. Hence, their spin-to-orbital period ratios are an indication of their magnetic field strength (see Figure 6 in Norton et al. 2008b).

Broadly speaking, for smaller P_{spin}/P_{orb}, the surface magnetic field strength is smaller. So these systems will have smaller magnetospheric radii, and the material will attach onto field lines closer to the WD. This will give rise to a larger footprint area in those systems with smaller values of P_{spin}/P_{orb}. That is to say, faster spinning WDs will have larger accretion footprints beneath a wide but low accretion curtain, as suggested in Norton et al. [1999]. Evidence for this also comes from the observed double-peaked pulse profiles observed in fast spinning WD (hence owning a small magnetic field) as described in Norton et al. [1999], resulting from the optical depths across and along the accretion curtains as the WD rotates.

By spreading the material over a larger area, we suggest that the resulting bremsstrahlung X-ray emission may have a harder spectrum, possibly because the accretion shock is closer to the WD surface and there is less distance for the plasma to travel as it cools within the post-shock region towards the WD surface and so there is less contribution from cooler bremsstrahlung components. In contrast, the systems with a relatively slowly spinning WD have a larger P_{spin}/P_{orb} value, so their magnetic field strength is larger, their magnetospheric radius is larger, and their accretion footprint is smaller and sits beneath a tall but narrow accretion curtain. We suggest that this geometry gives rise to a softer bremsstrahlung spectrum, possibly because the accretion shock is further from the WD surface and so there is a greater distance for the plasma to travel within the post-shock region and cool as it falls towards the WD surface. This interpretation also helps explain the low detection number in the hard X-ray domain of EX Hya-like systems below the period gap with high P_{spin}/P_{orb} which are thought to display ring-like accretion. In these cases the magnetospheric radius extends to a very large distance from the WD implying a very small footprint area. If this then means a very tall shock height

Figure 4.16: Figure illustrating the footprint geometry on the WD surface. The image on the left displays the outcome of a high P_{spin}/P_{orb} system where accreted material is latched onto the field lines far from the WD. This will yioeld a tall but narrow accretion column on the WD poles. The figure on the right displays the footprint geometry for a low P_{spin}/P_{orb} system (and therfore fast P_{spin}). Accreted material is latched close to trhe WD creating a short but wide accretion footprint on the WD. Figures taken from Norton et al. [1999].



(in line with the interpretation above), then the plasma will have a long distance over which to cool as it travels towards the surface, and the spectrum will be dominated by softer photons. As far as we are aware, no-one has modelled whether the multi-temperature bremsstrahlung spectrum is different for a wide, low accretion curtain compared with a tall, narrow accretion curtain, but we suggest this would be a worthwhile test to carry out.

4.9 Conclusions

This chapter has presented a catalogue and analysis of a sample of CVs detected in the hard X-ray range ($>$ 17keV) with IBIS, BAT and HXD. As with previously compiled high-energy samples of CVs, it is shown that most systems are magnetic. Moreover, some of the detected systems are very rare types of objects (2 APs). The sample is dominated by intermediate polars, with only 2 synchronous polars. This suggests the broadband X-ray/gamma-ray spectrum of IPs is harder and more luminous than that of polars. This could be

the effect of accretion rate and magnetic field strength, where IPs have higher accretion and weaker magnetic fields relative to polars, depositing a somewhat smoother accretion stream onto the poles. By contrast to the polars, the accretion flow in IPs may therefore not bury itself deep within the post-shock region, producing a harder broadband X-ray/gamma-ray spectrum.

We have shown that only IPs with $P_{spin}/P_{orb} < 0.1$ are consistently found by hard X-ray surveys. Moreover, we have examined the observational properties of mCVs in the P_{orb}- P_{spin} plane. We find that the observations are consistent with the theoretical models of Norton et al. [2008b, 2004] for mCV evolution, where IPs tend to cluster at about $P_{spin}/P_{orb} \approx 0.1$, and none have yet been observed in the hard X-ray regime above $P_{spin}/P_{orb} \approx 0.1$. Also observed and predicted is the observation that a very low number of IPs are found in any wavelength range within the synchronicity gap: a region between $P_{spin}/P_{orb} \approx 0.3$ and $P_{spin}/P_{orb} = 1$.

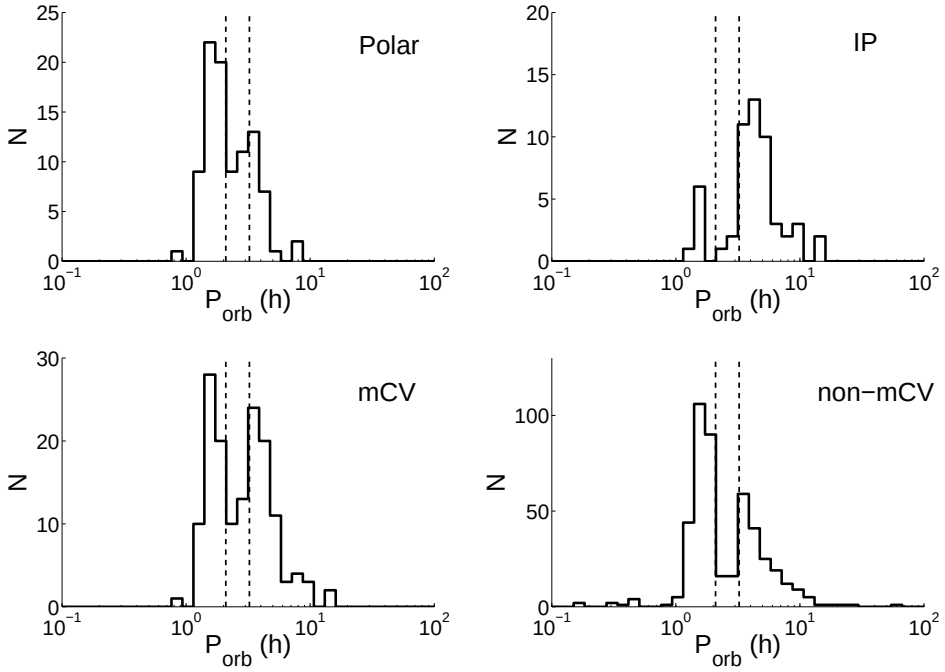
The chapter has also presented the first observed correlations between the $Flux_{30-60}/Flux_{17-30}$ keV hardness and P_{spin}, P_{orb} and P_{spin}/P_{orb}. The correlations have been statistically tested using Monte Carlo simulations.

In an attempt to explain our result we have suggested that hard X-ray selected IPs are spinning towards their equilibrium, so that their spin period is an indicator of magnetic field strength. This in turn will give rise to a short, but wide, post-shock region for fast spinning WDs (and therefore possessing a relatively weak magnetic field) making their hard X-ray spectra harder. In contrast slowly spinning WDs will have a tall but narrow post-shock region (possessing a relatively high magnetic field), yielding a cooler bremsstrahlung component in the hard X-rays.

All of the observations presented in this chapter are consistent with mCV evolution models. It is very likely that hard X-ray missions will continue to

increase this sample of mCVs, and it is also expected that more unidentified hard X-ray sources will be identified as IPs with more optical follow-ups. In particular, more observations will allow us to establish if any IPs are detected below the period gap and if any IPs will ever get detected in the IBIS energy range above $P_{spin}/P_{orb} \approx 0.1$ in order to establish if the Norton accretion models are a plausible explanation to the observed systems. This also implies that hard X-ray selected samples could have their own biases, however more analysis will have consequences on evolution studies of these exotic magnetic systems. Figure 4.17 shows the orbital distribution of various CV subclasses and gives a taste of what can be learned from mCVs now that the number of systems has grown to a statistically useful number. All distributions in Figure 4.17 have been tested with a KS-test and none of them appear to be consistent with each other. This is already strong evidence that the magnetic field strength of the WD has a great impact on the whole evolution of the binary systems. More work and analysis will have to be undertaken in order to best understand the properties from the various orbital distributions, and how these relate to the evolution of the systems.

Figure 4.17: Orbital period distributions for CV subclasses. The dotted lines marks the period gap seen in non-magnetic CVs.



Chapter 5

Conclusions

"As order exponentially increases, time exponentially speeds up."

– *Ray Kurzweil.*

WITHIN this final chapter we will comment on the results obtained from this study and mention possible future developments and studies both in the fields of machine learning astronomy and magnetic cataclysmic variables. The first section will discuss machine learning algorithms, whilst magnetic cataclysmic variables will be discussed in the final section.

5.1 Machine Learning in Astronomy

This thesis has shown, using ISINA as an example, how machine learning algorithms can help and aid scientific discoveries in astronomy. Emphasis has been given mainly to ISINA however many identification/classification algorithms exist, all trying to find recurrent patterns within large datasets. Particularly identification algorithms have been brought forward in order to try and identify *XMM* sources [Pineau et al., 2008] using probabilistic frameworks. Richards et al. [2009] has used similar techniques on much larger datasets in

order to identify quasars within the SDSS dataset, whilst Scaringi et al. [2009] have used neural networks in order to identify broad absorption line quasars using SDSS spectra. Identification is not only restricted to individual sources but also relates to identifying structures with the sky. Gezari [2008] has shown how tidal disruption of stars by a supermassive black hole can be identified by monitoring light-curve shapes in a semi-automated fashion. Moreover galactic streams within our Galaxy can also be identified [Grillmair, 2008], however for now mainly relying on visual inspection, but algorithms are also being developed for this task in order to automate the process for large datasets [Cole et al., 2008].

Not only do machine learning algorithms help identify astronomical sources but also help to classify the objects in question. Brett et al. [2004] have shown, using an unsupervised neural network (no training) how various binary light curves (RR Lyrae, δ Scuti, cepheid variables, eclipsing close binaries) can be automatically clustered based solely on folded light curve shape. On the other hand Elting et al. [2008] have demonstrated the feasibility of photometric-based classification of stars using multi-dimensional clustering, whilst Andrae and Melchior [2008] have shown, using shapelets, how morphological galaxy classification is possible using automated algorithms.

All of the above are examples taken from the ever growing resource of algorithms being created in order to tackle some of the tasks facing astronomy in the coming century. With the advent of better observational resources, the need for autonomous algorithms will become inevitable and will complement the available data more and more as the data growth will keep increasing. Currently all-sky automated sky surveys such as the *WASP* project, using a dedicated telescope, is producing >30Tb of data per year, monitoring 100 million objects, all of which needs to be searched for transients and exo-planets.

The next generation of radio telescopes such as the *SKA* and *LOFAR* will be able to produce all-sky images of the radio sky with unprecedented timing and angular resolutions, yielding incredibly large dataflows of the order of a few hundred Pb per year, surveying over 100 billion objects. Similarly the *GAIA* mission will obtain photometry for over 1 billion objects, which will all need to be identified and classified accordingly.

The production of automated algorithms is thus a necessity in order to tackle astronomy in the coming century, a necessity which might one day change the way we analyse and interpret data and thus is named the *fourth paradigm* [Ball and Brunner, 2009].

5.2 Magnetic Cataclysmic Variables

This thesis has also examined the properties of the magnetic cataclysmic variable population, with particular emphasis on hard X-ray selected systems. We have shown how contemporary hard X-ray observatories, such as *INTEGRAL*, are able to detect more of these systems, and shown how future observatories will have the potential to increase the observed numbers even further.

In the process of studying the hard X-ray selected systems we have also examined the properties arising from the P_{orb}-P_{spin} plane for the whole mCV population. The analysis has shown how some of the observations are consistent with the theory arising from magnetic accretion, however some questions have still been left open. For example, it is not certain yet why very few systems, detected in any band, are found with $P_{spin}/P_{orb} > 0.3$ and $P_{spin}/P_{orb} = 1$, named as the synchronicity gap. One possible explanation however lies within the fact that IPs are driven towards their equilibrium at $P_{spin}/P_{orb} \approx 0.1$ by spinning up if they spin too slow by stream accretion and spinning down

if they spin too fast by propelling material away. Further evidence for this could also possibly come from the fact that no hard X-ray selected IP has been found above $P_{spin}/P_{orb} = 0.1$. This seems to be the dividing line (even though blurred) between two different kinds of IPs accreting through different mechanisms (stream and propeller), which would also point to different emission mechanisms for hard X-ray photons seen from these systems. All of the above also helps to explain the nature of the observed hardness correlations presented in the previous chapter in Figures 4.9, 4.10 and 4.11 where synchronicity (and thus spin period) correlates with the hardness ratio defined as 30-60keV/17-30keV count rate. Fast spinning systems display hard spectra whilst slowly spinning systems display softer spectra. This seems to be best explained by the footprint geometry on the WD poles, where a tall shock produced by the slowly spinning systems yields a softer spectrum, whilst fast spinning systems produce a short and wide shock yielding a hard spectrum. We believe the reason for the apparent hardness correlations results from how fast the shock can cool. In tall shocks (slowly spinning systems) the plasma has a greater distance too cool before it reaches the WD surface and hence will display a broad range of bremsstrahlung temperatures. Conversely short shocks do not have much to travel before they cool on the WD surface, making the resulting bremsstrahlung temperature gradient steeper (and thus harder spectra). This would also help explain why not many systems have been observed above $P_{spin}/P_{orb} = 0.1$, since their shocks would be even higher, and thus softer in the hard X-ray range, making these systems difficult to detect. Moreover it would help explain the low detection numbers of polars, since these systems are also believed to possess very high shocks.

Asynchronous polars however still pose a potential problem to the explanations above, since we would not necessarily expect to see these systems in the

hard X-ray domain given their close proximity to the polars in the P_{orb}-P_{spin} plane. However these systems are very mysterious with only four detections in any band, making them hard to compare against other known mCVs.

We conclude by mentioning the further need to study mCVs in the hard X-ray domain to better understand the origin of the hardness correlation and to better understand the properties arising from the P_{orb}-P_{spin} plane. Moreover we point out that if the shock height interpretation is correct, we would expect to use hardness as an indicator of both shock height, and as a consequence infer at what stage of it's evolution a particular mCV lies, hence inferring their spin and orbital period. This would help better understand the evolutionary process of these exotic magnetic systems.

Bibliography

R. Andrae and P. Melchior. Morphological Galaxy Classification with Shapelets. In C. A. L. Bailer-Jones, editor, *American Institute of Physics Conference Series*, volume 1082 of *American Institute of Physics Conference Series*, pages 129–133, December 2008. doi: 10.1063/1.3059023.

S. Araujo-Betancor, B. T. Gänsicke, H.-J. Hagen, and et al. 1RXS J062518.2+733433: A new intermediate polar. *A&A*, 406:213–219, July 2003. doi: 10.1051/0004-6361:20030787.

T. Augusteijn, M. H. M. Heemskerk, G. A. A. Zwarthoed, and et al. Periodicities in the optical brightness variations of the intermediate polar TV Columbae. *A&AS*, 107:219–233, October 1994.

N. M. Ball and R. J. Brunner. Data Mining and Machine Learning in Astronomy. *ArXiv e-prints*, June 2009.

E. J. Barlow, C. Knigge, A. J. Bird, and et al. 20-100 keV properties of cataclysmic variables detected in the INTEGRAL/IBIS survey. *MNRAS*, 372:224–232, October 2006. doi: 10.1111/j.1365-2966.2006.10836.x.

E. Bertin and S. Arnouts. SExtractor: Software for source extraction. *A&AS*, 117:393–404, June 1996.

A. J. Bird, E. J. Barlow, L. Bassani, and et al. The First IBIS/ISGRI Soft Gamma-Ray Galactic Plane Survey Catalog. *ApJL*, 607:L33–L37, May 2004. doi: 10.1086/421772.

A. J. Bird, E. J. Barlow, L. Bassani, and et al. The Second IBIS/ISGRI Soft Gamma-Ray Survey Catalog. *ApJ*, 636:765–776, January 2006. doi: 10.1086/498090.

A. J. Bird, A. Malizia, A. Bazzano, and et al. The Third IBIS/ISGRI Soft Gamma-Ray Survey Catalog. *ApJS*, 170:175–186, May 2007. doi: 10.1086/513148.

A. J. Bird, A. Bazzano, L. Bassani, F. Capitanio, M. Fiocchi, A. B. Hill, A. Malizia, V. A. McBride, S. Scaringi, V. Sguera, J. B. Stephen, P. Ubertini, A. J. Dean, F. Lebrun, R. Terrier, M. Renaud, F. Mattana, D. Gotz, J. Rodriguez, G. Belanger, R. Walter, and C. Winkler. The 4th IBIS/ISGRI soft gamma-ray survey catalog. *ArXiv e-prints*, October 2009.

J. M. Bonnet-Bidaud, M. Mouchet, D. de Martino, and et al. The white dwarf revealed in the intermediate polar V709 Cassiopeiae. *A&A*, 374:1003–1008, August 2001. doi: 10.1051/0004-6361:20010756.

J. M. Bonnet-Bidaud, M. Mouchet, D. de Martino, and et al. RX J2133.7+5107: identification of a new long period Intermediate Polar. *A&A*, 445:1037–1040, January 2006. doi: 10.1051/0004-6361:20053303.

J. M. Bonnet-Bidaud, D. de Martino, M. Falanga, and et al. IGR J00234+6141: a new INTEGRAL source identified as an intermediate polar. *A&A*, 473:185–189, October 2007. doi: 10.1051/0004-6361:20077877.

D. R. Brett, R. G. West, and P. J. Wheatley. The automated classification

of astronomical light curves using Kohonen self-organizing maps. *MNRAS*, 353:369–376, September 2004. doi: 10.1111/j.1365-2966.2004.08093.x.

L. Brieman, J. Friedman, C. J. Stone, and et al. *Classification and Regression Trees*. Wadsworth International Group, Belmont, California, 1984.

J. Brunschweiger, J. Greiner, M. Ajello, and et al. Intermediate polars in the Swift/BAT survey: spectra and white dwarf masses. *A&A*, 496:121–127, March 2009. doi: 10.1051/0004-6361/200811285.

D. A. H. Buckley and I. R. Tuohy. H0534 - 581: A new intermediate polar? *ApJ*, 349:296–312, January 1990. doi: 10.1086/168314.

O. W. Butters, E. J. Barlow, A. J. Norton, and et al. RXTE confirmation of the intermediate polar status of Swift J0732.5-1331. *A&A*, 475:L29–L32, November 2007. doi: 10.1051/0004-6361:20078700.

O. W. Butters, A. J. Norton, P. Hakala, and et al. RXTE determination of the intermediate polar status of XSS J00564+4548, IGR J17195-4100, and XSS J12270-4859. *A&A*, 487:271–276, August 2008. doi: 10.1051/0004-6361:200809942.

A. Chen, D. O'Donoghue, R. S. Stobie, and et al. Cataclysmic variables in the Edinburgh-Cape Blue Object SurveyQ3. *MNRAS*, 325:89–110, July 2001. doi: 10.1046/j.1365-8711.2001.04322.x.

N. Cole, H. Jo Newberg, M. Magdon-Ismail, and et al. Tracing the Sagittarius Tidal Stream with Maximum Likelihood. In C. A. L. Bailer-Jones, editor, *American Institute of Physics Conference Series*, volume 1082 of *American Institute of Physics Conference Series*, pages 216–220, December 2008. doi: 10.1063/1.3059049.

M. Cropper. The Polars. *Space Science Reviews*, 54:195–295, December 1990.

M. Cropper, K. Wu, G. Ramsay, and et al. Effects of gravity on the structure of post-shock accretion flows in magnetic cataclysmic variables. *MNRAS*, 306:684–690, July 1999.

R. M. Cutri, M. F. Skrutskie, S. van Dyk, and et al. *2MASS All Sky Catalog of point sources.* The IRSA 2MASS All-Sky Point Source Catalog, NASA/IPAC Infrared Science Archive. http://irsa.ipac.caltech.edu/applications/Gator/, June 2003.

D. de Martino, J.-M. Bonnet-Bidaud, M. Mouchet, and et al. The long period intermediate polar 1RXS J154814.5-452845. *A&A*, 449:1151–1160, April 2006. doi: 10.1051/0004-6361:20053877.

D. de Martino, J.M. Bonnet-Bidaud, M. Falanga, and et al. XMM-Newton discovery of pulsations from IGR J21237+4218=V2069 Cyg. *The Astronomer's Telegram*, 2089:1–+, June 2009.

A. J. Dean, D. J. Clark, J. B. Stephen, and et al. Polarized Gamma-Ray Emission from the Crab. *Science*, 321:1183–, August 2008. doi: 10.1126/science.1149056.

C. Done and P. Magdziarz. The X-Ray Spectrum of the Polar by Cam. In L. Scarsi, H. Bradt, P. Giommi, and F. Fiore, editors, *The Active X-ray Sky: Results from BeppoSAX and RXTE*, pages 376–+, 1998.

R. A. Downes, R. F. Webbink, M. M. Shara, and et al. A Catalog and Atlas of Cataclysmic Variables: The Final Edition. *Journal of Astronomical Data*, 11:2–+, December 2005.

C. Elting, C. A. L. Bailer-Jones, and K. W. Smith. Photometric Classification of Stars, Galaxies and Quasars in the Sloan Digital Sky Survey DR6 Using Support Vector Machines. In C. A. L. Bailer-Jones, editor, *American Institute of Physics Conference Series*, volume 1082 of *American Institute of Physics Conference Series*, pages 9–14, December 2008. doi: 10.1063/1.3059095.

P. A. Evans, A. P. Beardmore, and J. P. Osborne. Swift-XRT identification of IGR J19267+1325 as an Intermediate Polar. *The Astronomer's Telegram*, 1669:1–+, August 2008.

J. Faulkner. Ultrashort-Period Binaries, Gravitational Radiation, and Mass Transfer. I. The Standard Model, with Applications to WZ Sagittae and Z Camelopardalis. *ApJL*, 170:L99+, December 1971. doi: 10.1086/180848.

C. Ferguson, E. J. Barlow, A. J. Bird, A. J. Dean, A. B. Hill, S. E. Shaw, J. B. Stephen, S. Sturner, T. V. Tikkanen, G. Weidenspointner, and D. R. Willis. The INTEGRAL Mass Model - TIMM. *A&A*, 411:L19–L23, November 2003. doi: 10.1051/0004-6361:20031403.

M. T. Friend, J. S. Martin, R. Connon-Smith, and et al. The 8190-A Sodium Doublet in Cataclysmic Variables - Part Three - too Cool for Credibility. *MNRAS*, 246:654–+, October 1990.

B. T. Gänsicke, T. R. Marsh, A. Edge, and et al. Cataclysmic variables from a ROSAT/2MASS selection - I. Four new intermediate polars. *MNRAS*, 361: 141–154, July 2005. doi: 10.1111/j.1365-2966.2005.09138.x.

B. T. Gänsicke, M. Dillon, J. Southworth, J. R. Thorstensen, P. Rodríguez-Gil, A. Aungwerojwit, T. R. Marsh, P. Szkody, S. C. C. Barros, J. Casares, D. de Martino, P. J. Groot, P. Hakala, U. Kolb, S. P. Littlefair, I. G. Martínez-Pais,

G. Nelemans, and M. R. Schreiber. SDSS unveils a population of intrinsically faint cataclysmic variables at the minimum orbital period. *MNRAS*, 397: 2170–2188, August 2009. doi: 10.1111/j.1365-2966.2009.15126.x.

R. D. Geckeler and R. Staubert. Periodic changes of the accretion geometry in the nearly-synchronous polar RX J1940.1-1025. *A&A*, 325:1070–1076, September 1997.

S. Gezari. Identification of Tidal Disruption Events by Pan-STARRS1. In C. A. L. Bailer-Jones, editor, *American Institute of Physics Conference Series*, volume 1082 of *American Institute of Physics Conference Series*, pages 268–274, December 2008. doi: 10.1063/1.3059061.

A. Goldwurm, P. David, L. Foschini, and et al. The INTEGRAL/IBIS scientific data analysis. *A&A*, 411:L223–L229, November 2003. doi: 10.1051/0004-6361:20031395.

C. J. Grillmair. Finding Stellar Streams in Photometric Surveys. In C. A. L. Bailer-Jones, editor, *American Institute of Physics Conference Series*, volume 1082 of *American Institute of Physics Conference Series*, pages 226–232, December 2008. doi: 10.1063/1.3059052.

A. Gros, A. Goldwurm, M. Cadolle-Bel, and et al. The INTEGRAL IBIS/ISGRI System Point Spread Function and Source Location Accuracy. *A&A*, 411:L179–L183, November 2003.

C. Hellier. The Four Periodicities of the Cataclysmic Variable Tv-Columbae. *MNRAS*, 264:132–+, September 1993.

C. Hellier and A. P. Beardmore. The accretion flow in the discless intermediate polar V2400 Ophiuchi. *MNRAS*, 331:407–416, March 2002. doi: 10.1046/j.1365-8711.2002.05199.x.

M. Hernanz and G. Sala. A Classical Nova, V2487 Oph 1998, Seen in X-rays Before and After Its Explosion. *Science*, 298:393–395, October 2002.

J. Kaluzny and I. Semeniuk. Photometric Observations of the Intermediate Polar AO Piscium. *Information Bulletin on Variable Stars*, 3145:1–+, February 1988.

A. R. King and G. A. Wynn. The spin period of EX Hydrae. *MNRAS*, 310: 203–209, November 1999. doi: 10.1046/j.1365-8711.1999.02974.x.

A. R. King, R. Whitehurst, and J. Frank. Synchronous rotation in AM Herculis systems. I - Equilibrium configurations. *MNRAS*, 244:731–737, June 1990.

A. Kniazev, M. Revnivtsev, S. Sazonov, and et al. IGR J08390-4833 is a new cataclysmic variable from the INTEGRAL all- sky survey. *The Astronomer's Telegram*, 1488:1–+, April 2008.

C. Knigge. The donor stars of cataclysmic variables. *MNRAS*, 373:484–502, December 2006. doi: 10.1111/j.1365-2966.2006.11096.x.

J. Kuijpers and J. E. Pringle. Comments on radial white dwarf accretion. *A&A*, 114:L4–L6, October 1982.

D. Q. Lamb. Recent developments in the theory of AM HER and DQ HER stars. In D. Q. Lamb and J. Patterson, editors, *Cataclysmic Variables and Low-Mass X-ray Binaries*, volume 113 of *Astrophysics and Space Science Library*, pages 179–218, 1985.

C. M. Lance, M. L. McCall, and A. K. Uomoto. Portrait of a Novel Nova: V1500 Cygni. *ApJS*, 66:151–+, February 1988. doi: 10.1086/191251.

R. Landi, L. Bassani, A. J. Dean, and et al. INTEGRAL/IBIS and Swift/XRT

observations of hard cataclysmic variables. *MNRAS*, 392:630–640, January 2009. doi: 10.1111/j.1365-2966.2008.14086.x.

F. Lebrun, J. P. Leray, P. Lavocat, and et al. ISGRI: The INTEGRAL Soft Gamma-Ray Imager. *A&A*, 411:L141–L148, November 2003. doi: 10.1051/0004-6361:20031367.

T. R. Marsh and S. R. Duck. Stroboscopic Doppler tomography of FO AQR. *New Astronomy*, 1:97–119, October 1996. doi: 10.1016/S1384-1076(96)00008-5.

N. Masetti, L. Bassani, A. Bazzano, and et al. Unveiling the nature of INTEGRAL objects through optical spectroscopy. IV. A study of six new hard X-ray sources. *A&A*, 455:11–19, August 2006a. doi: 10.1051/0004-6361:20065111.

N. Masetti, L. Morelli, E. Palazzi, and et al. Unveiling the nature of INTEGRAL objects through optical spectroscopy. V. Identification and properties of 21 southern hard X-ray sources. *A&A*, 459:21–30, November 2006b. doi: 10.1051/0004-6361:20066055.

N. Masetti, E. Mason, L. Morelli, and et al. Unveiling the nature of INTEGRAL objects through optical spectroscopy. VI. A multi-observatory identification campaign. *A&A*, 482:113–132, April 2008a. doi: 10.1051/0004-6361:20079332.

N. Masetti, P. Parisi, E. Palazzi, and et al. Unveiling the nature of INTEGRAL objects through optical spectroscopy. VII. Identification of 20 Galactic and extragalactic hard X-ray sources. *ArXiv e-prints*, November 2008b.

C. W. Mauche. Chandra High Energy Transmission Grating Spectrum of AE Aquarii. *ArXiv e-prints*, October 2009.

I. M. McHardy, E. Koerding, C. Knigge, and et al. Active galactic nuclei as scaled-up Galactic black holes. *Nature*, 444:730–732, December 2006. doi: 10.1038/nature05389.

M. P. Muno, J. S. Arabadjis, F. K. Baganoff, and et al. The Spectra and Variability of X-Ray Sources in a Deep Chandra Observation of the Galactic Center. *ApJ*, 613:1179–1201, October 2004. doi: 10.1086/423164.

A. J. Norton and K. Mukai. A precessing accretion disc in the intermediate polar XY Arietis? *A&A*, 472:225–232, September 2007. doi: 10.1051/0004-6361:20077761.

A. J. Norton, A. P. Beardmore, A. Allan, and et al. YY Draconis and V709 Cassiopeiae: two intermediate polars with weak magnetic fields. *A&A*, 347: 203–211, July 1999.

A. J. Norton, A. P. Beardmore, A. Retter, and et al. The nature of TW Pictoris. *MNRAS*, 312:362–370, February 2000.

A. J. Norton, G. A. Wynn, and R. V. Somerscales. The Spin Periods and Magnetic Moments of White Dwarfs in Magnetic Cataclysmic Variables. *ApJ*, 614:349–357, October 2004. doi: 10.1086/423333.

A. J. Norton, E. J. Barlow, O. W. Butters, and et al. Are the INTEGRAL Intermediate Polars Different? In R. M. Bandyopadhyay, S. Wachter, D. Gelino, and C. R. Gelino, editors, *A Population Explosion: The Nature & Evolution of X-ray Binaries in Diverse Environments*, volume 1010 of *American Institute of Physics Conference Series*, pages 230–234, May 2008a. doi: 10.1063/1.2945047.

A. J. Norton, O. W. Butters, T. L. Parker, and et al. The Accretion Flows and

Evolution of Magnetic Cataclysmic Variables. *ApJ*, 672:524–530, January 2008b. doi: 10.1086/523932.

B. Paczynski and R. Sienkiewicz. Gravitational radiation and the evolution of cataclysmic binaries. *ApJL*, 248:L27–L30, August 1981. doi: 10.1086/183616.

P. Parisi, M. Masetti, E. Jimenez, and et al. Optical identification of IGR J18308-1232 as a Cataclysmic Variable. *The Astronomer's Telegram*, 1710: 1–+, September 2008.

J. Patterson. The DQ Herculis Stars. In *Cataclysmic Variables and Related Physics*, volume 10 of *Annals of the Israel Physical Society*, pages 193–+, 1993.

F.-X. Pineau, S. Derriere, L. Michel, and et al. Statistical identification of 2XMMi sources. In C. A. L. Bailer-Jones, editor, *American Institute of Physics Conference Series*, volume 1082 of *American Institute of Physics Conference Series*, pages 15–21, December 2008. doi: 10.1063/1.3059033.

M. L. Pretorius. Time-resolved optical observations of five cataclysmic variables detected by INTEGRAL. *ArXiv e-prints*, January 2009.

S. Rappaport, F. Verbunt, and P. C. Joss. A new technique for calculations of binary stellar evolution, with application to magnetic braking. *ApJ*, 275: 713–731, December 1983. doi: 10.1086/161569.

G. T. Richards, A. D. Myers, A. G. Gray, and et al. Efficient Photometric Selection of Quasars from the Sloan Digital Sky Survey. II. ˜1,000,000 Quasars from Data Release 6. *ApJS*, 180:67–83, January 2009. doi: 10.1088/0067-0049/180/1/67.

H. Ritter and U. Kolb. Catalogue of cataclysmic binaries, low-mass X-ray binaries and related objects (Seventh edition). *A&A*, 404:301–303, June 2003. doi: 10.1051/0004-6361:20030330.

K. Saitou, M. Tsujimoto, K. Ebisawa, and et al. Suzaku X-Ray Study of an Anomalous Source XSS J12270-4859. *ArXiv e-prints*, April 2009.

S. Scaringi, A. J. Bird, D. J. Clark, and et al. ISINA: INTEGRAL Source Identification Network Algorithm. *MNRAS*, 390:1339–1348, November 2008. doi: 10.1111/j.1365-2966.2008.13765.x.

S. Scaringi, C. E. Cottis, C. Knigge, and M. R. Goad. Classifying Broad Absorption Line Quasars: Metrics, Issues and a New Catalogue Constructed from SDSS DR5. *ArXiv e-prints*, July 2009.

R. Schwarz, A. D. Schwope, A. Staude, and et al. Doppler tomography of the asynchronous polar BY Camelopardalis. *A&A*, 444:213–220, December 2005. doi: 10.1051/0004-6361:20053711.

A. D. Schwope, D. A. H. Buckley, D. O'Donoghue, and et al. RX J2115.7-5840: a short-period, asynchronous polar. *A&A*, 326:195–202, October 1997.

D. Steeghs, C. Knigge, J. Drew, and et al. IPHAS identification of IGR J19267+1325 as a Cataclysmic Variable. *The Astronomer's Telegram*, 1669: 1–+, August 2008.

J. B. Stephen, L. Bassani, A. Malizia, and et al. Using the ROSAT catalogues to find counterparts for the second IBIS/ISGRI survey sources. *A&A*, 445: 869–873, January 2006. doi: 10.1051/0004-6361:20053958.

M. Sugizaki, K. Kinugasa, K. Matsuzaki, and et al. Discovery of a New Pul-

sating X-Ray Source with a 1549.1 Second Period, AX J183220-0840. *ApJL*, 534:L181–L184, May 2000. doi: 10.1086/312676.

Y. Terada, T. Hayashi, M. Ishida, and et al. Suzaku Observation of a White Dwarf as a new Candidate of Cosmic-ray Origin. In F. A. Aharonian, W. Hofmann, and F. Rieger, editors, *American Institute of Physics Conference Series*, volume 1085 of *American Institute of Physics Conference Series*, pages 689–692, December 2008. doi: 10.1063/1.3076769.

J. R. Thorstensen and C. J. Taylor. Spectroscopy and orbital periods of four cataclysmic variable stars. *MNRAS*, 326:1235–1242, October 2001. doi: 10.1046/j.1365-8711.2001.04680.x.

S. Tremaine, K. Gebhardt, R. Bender, and et al. The Slope of the Black Hole Mass versus Velocity Dispersion Correlation. *ApJ*, 574:740–753, August 2002. doi: 10.1086/341002.

J. Tueller, W. H. Baumgartner, C. B. Markwardt, and et al. The 22-Month Swift-BAT All-Sky Hard X-ray Survey. *ArXiv e-prints*, March 2009.

P. Ubertini, F. Lebrun, G. Di Cocco, and et al. IBIS: The Imager on-board INTEGRAL. *A&A*, 411:L131–L139, November 2003. doi: 10.1051/0004-6361:20031224.

S. van Amerongen, H. Kraakman, E. Damen, and et al. Spin-Up of the White Dwarf in the Intermediate Polar Ao-Piscium = H2252-035. *MNRAS*, 215: 45P–+, July 1985.

F. Verbunt and C. Zwaan. Magnetic braking in low-mass X-ray binaries. *A&A*, 100:L7–L9, July 1981.

B. Warner. *Cataclysmic Variable Stars*. Cataclysmic Variable Stars, by Brian Warner, pp. 592. ISBN 052154209X. Cambridge, UK: Cambridge University Press, September 2003., September 2003.

M. G. Watson, S. R. Rosen, D. O'Donoghue, and et al. Optical properties of the new polar RXJ1940.2-1025. *MNRAS*, 273:681–698, April 1995.