

White dwarf-red dwarf binaries in the Sloan Digital Sky Survey

I. Sample definition[★]

T. Augusteijn¹, R. Greimel^{2,3,4}, E. J. M. van den Besselaar⁵, P. J. Groot⁵, and L. Morales-Rueda⁵

¹ Nordic Optical Telescope, Apartado 474, 38700 Santa Cruz de La Palma, Spain
e-mail: tau@not.iac.es

² Isaac Newton Group of Telescopes, Apartado de correos 321, 38700 Santa Cruz de la Palma, Spain

³ Institute of Physics, University of Graz, Universitätsplatz 5, 8010 Graz, Austria
e-mail: rgreimel@gmail.com

⁴ Visiting scientist, Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

⁵ Department of Astrophysics, Radboud University Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands
e-mail: [besselaar;pgroot;lmr]@astro.ru.nl

Received 31 August 2007 / Accepted 16 October 2007

ABSTRACT

Context. A significant fraction of binary stars consisting of a white dwarf and a low-mass main-sequence star (red dwarf) are expected to be close binaries that are the end products of common-envelope (CE) evolution.

Aims. To gain a better understanding of CE evolution, we want to study white dwarf-red dwarf binaries. For this it is fundamental to establish a well-defined sample.

Methods. To reduce contamination by more distant sources, such as quasars, we have selected candidate white dwarf-red dwarf binaries from the catalogue of proper motion stars drawn from the intersection of the Sloan Digital Sky Survey (SDSS) and the USNO-B1.0 catalogue. To separate single from binary sources and to cut spurious sources out, we define selection criteria based on a combination of the $(u - g)$, $(g - r)$, and $(r - i)$ colours.

Results. We evaluate and discuss the selection criteria on the basis of both the publicly available SDSS spectra and the predicted colours of white dwarf-red dwarf binaries. We define a sample of 651 binary candidates. However, we find that for r magnitudes brighter than ~ 16.5 the proper motion catalogue is heavily contaminated with sources that have incorrect colours due to the r magnitudes being relatively too faint, most likely due to saturation effects.

Conclusions. We show that the level of contamination can be reduced by either excluding sources brighter than $r = 16.5$ mag or by excluding sources bluer than $(u - g) = -0.5$ or $(g - r) = -0.6$ and brighter than $r = 16.0$ mag. We expect $\sim 85\%$ of the remaining sources to be genuine white dwarf-red dwarf binaries. We estimate that we exclude $\sim 5\text{--}10\%$ of the white dwarf-red dwarf binaries in the proper motion catalogue.

Key words. binaries: spectroscopic – stars: late-type – white dwarfs

1. Introduction

A large fraction of the stars in the Galaxy are part of a binary or a multiple system (Abt 1983; Lada 2006). In many binaries the orbital separation is sufficiently small ($P_{\text{orb}} < \text{few} \times 100$ days), such that the expanding envelope of an evolving star will engulf its companion, creating a common-envelope (CE). Due to friction the binary components will lose angular momentum, and the core of the giant and the companion will spiral in toward each other. Eventually, the envelope is ejected, and the system will reappear as a binary system consisting of the original secondary and the core of the primary (e.g. Paczyński 1976). The removal of its envelope will affect the evolution of the primary, limiting the mass of its core. In most cases the core will appear as a hot sub-dwarf that will subsequently cool to a white dwarf. However, the detailed physics of the CE phase (e.g., the efficiency of energy transport through the envelope, the

distribution of mass ejection (e.g., spherical or in the plane), possible additional energy sources (e.g., from accretion or recombination), etc.; see e.g., Iben & Livio 1993) is not well understood, and there are large uncertainties in the precise outcome of the CE phase.

In general, one would expect that there are many binaries that contain a white dwarf and a low-mass main-sequence star (red dwarf), a significant fraction of which are expected to have evolved through a CE phase and should be close binaries (see, e.g., de Kool & Ritter 1993). The white dwarf and red dwarf in these binaries are not expected to have changed much since the CE phase, and the components have well defined properties and are well understood. Therefore, the population of white dwarf-red dwarf binaries provides an ideal sample for testing models for CE evolution through the study of the properties of the population (e.g., number density, orbital period distribution) and the individual components (e.g., masses, surface temperatures).

Currently there are several hundred detached white dwarf-red dwarf binaries known (Silvestri et al. 2006, and references therein), and for ~ 100 sources the binary period has been determined (Ritter & Kolb 2003). The sample of white dwarf-red

[★] A list of the complete sample is only available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via <http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/486/843>

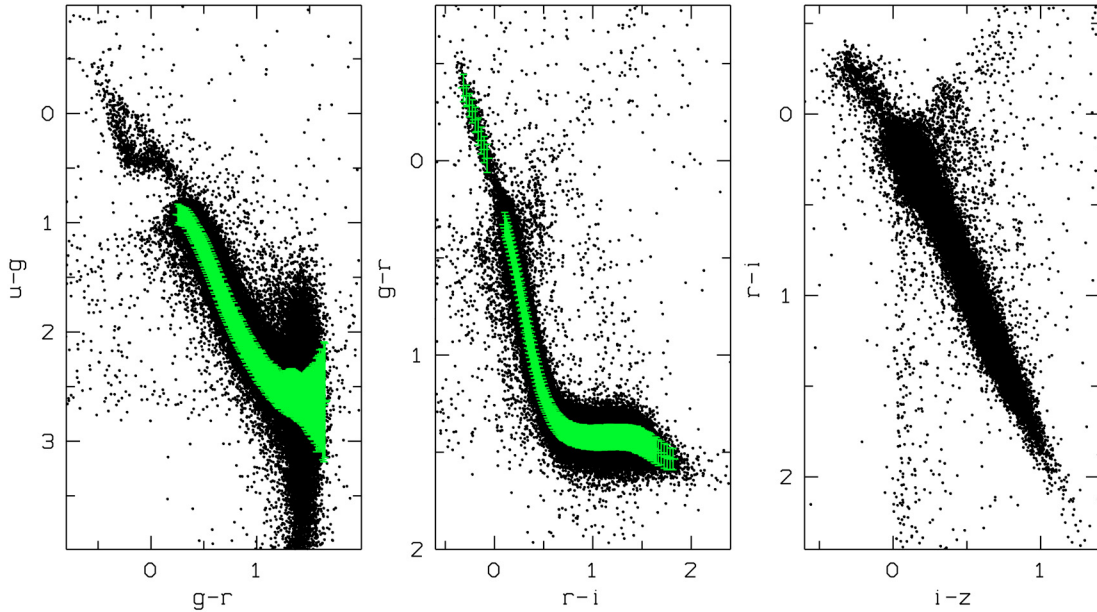


Fig. 1. The $(u - g)$ vs. $(g - r)$, the $(g - r)$ vs. $(r - i)$ and the $(r - i)$ vs. $(i - z)$ colour–colour diagrams for every 3rd source in the SDSS \cap USNO-B proper motion catalogue [Gould & Kollmeier \(2004\)](#). In the $(u - g)$ vs. $(g - r)$ and $(g - r)$ vs. $(r - i)$ colour–colour diagrams the average position and width of the main locus of stars is indicated (see Sect. 4).

dwarf binaries is very inhomogeneous, but most commonly white dwarfs selected on their (E)UV brightness were searched for a red (or infrared) excess ([Wachter et al. 2003](#); [Green et al. 2000](#)). This leads to a strong selection towards young, hot white dwarfs and relatively early type secondaries. E.g., among the binaries consisting of either a white dwarf or a white dwarf precursor primary and a red dwarf which have a known orbital period, more than half contain a (hot) sub-dwarf ([Ritter & Kolb 2003](#)), while one would expect the intrinsic population to be dominated by cooler white dwarfs.

White dwarf-red dwarf binaries are expected to show a large range in (photometric) properties, depending mostly on the temperature of the white dwarf and the spectral type of the red dwarf. In general, the white dwarf will dominate in the blue and the red dwarf in the red but the overall spectral energy distribution can vary strongly depending on the characteristics of the components. To be able to both detect white dwarf-red dwarf binaries with a hot white dwarf and a late-type red dwarf, and with a cool white dwarf and an early-type red dwarf one needs a broad coverage in wavelength. This is complicated by the fact that white dwarf-red dwarf binaries are intrinsically faint objects which makes it hard to distinguish both components if there is a large difference in luminosity between them.

The Sloan Digital Sky Survey (SDSS; [York et al. 2000](#)) is potentially a powerful tool to detect large numbers of white dwarf-red dwarf binaries as it provides deep ($r < 22.3$ mag), accurate photometry in five pass-bands (u, g, r, i and z ; [Fukugita et al. 1996](#)), covering a large range in wavelength. However, quasars cover a similar range in colours as white dwarf-red dwarf binaries (see, e.g., [Fan 1999](#)). It is not simple to separate these populations on colours alone, and any sample of white dwarf-red dwarf binaries will be highly contaminated. Here we present a sample of white dwarf-red dwarf binary candidates selected from the catalogue of proper motion stars ([Gould & Kollmeier 2004](#)) drawn from the intersection of the SDSS and the USNO-B1.0 catalogue ([Monet et al. 2003](#)). Using this catalogue, the contamination by more distant sources (quasars, binary giants) should be small. In addition, we have defined a

set of refined colour limits to include as much as possible all types of white dwarf-red dwarf binaries that are covered by the SDSS. The aim of this paper is to provide a well defined sample of white dwarf-red dwarf binaries drawn from the SDSS. Such a sample is intended as, and is essential for, a proper study of common-envelope evolution.

Below we discuss and evaluate the selection criteria on the basis of SDSS spectra of the sources, and on the basis of the predicted colours of white dwarf-red dwarf binaries. We will show that the contamination of the final sample is minimal while we exclude very few genuine sources. In an accompanying paper we present a more detailed analysis of the sources in our sample with existing SDSS spectra and of the whole photometric sample ([van den Besselaar et al. 2008](#), hereafter Paper II).

2. The SDSS \cap USNO-B proper motion catalogue

The SDSS \cap USNO-B proper motion catalogue ([Gould & Kollmeier 2004](#)) consists of 390 000 objects with a proper motion $\mu \geq 20$ mas yr⁻¹ and magnitudes $r \leq 20$. In Fig. 1 we present the $(u - g)$ vs. $(g - r)$, the $(g - r)$ vs. $(r - i)$ and the $(r - i)$ vs. $(i - z)$ colour–colour diagrams (CCDs) of the sources in the catalogue. The main stellar locus in all the diagrams is the main-sequence. In the $(u - g)$ vs. $(g - r)$ CCD the separate group of sources in the top-left are mostly single white dwarfs. In the $(g - r)$ vs. $(r - i)$ and the $(r - i)$ vs. $(i - z)$ CCDs these single white dwarfs form the extension of the main-sequence in the top-left, while the sources below the main-sequence in the $(g - r)$ vs. $(r - i)$ CCD are most likely dwarf carbon stars ([Margon et al. 2002](#)). The sources above and to the right of the main stellar locus are expected to be white dwarf-red dwarf binaries. The precise position in these diagrams depends on the combination of the surface gravity, temperature and atmospheric composition of the white dwarf and the spectral type of the red dwarf.

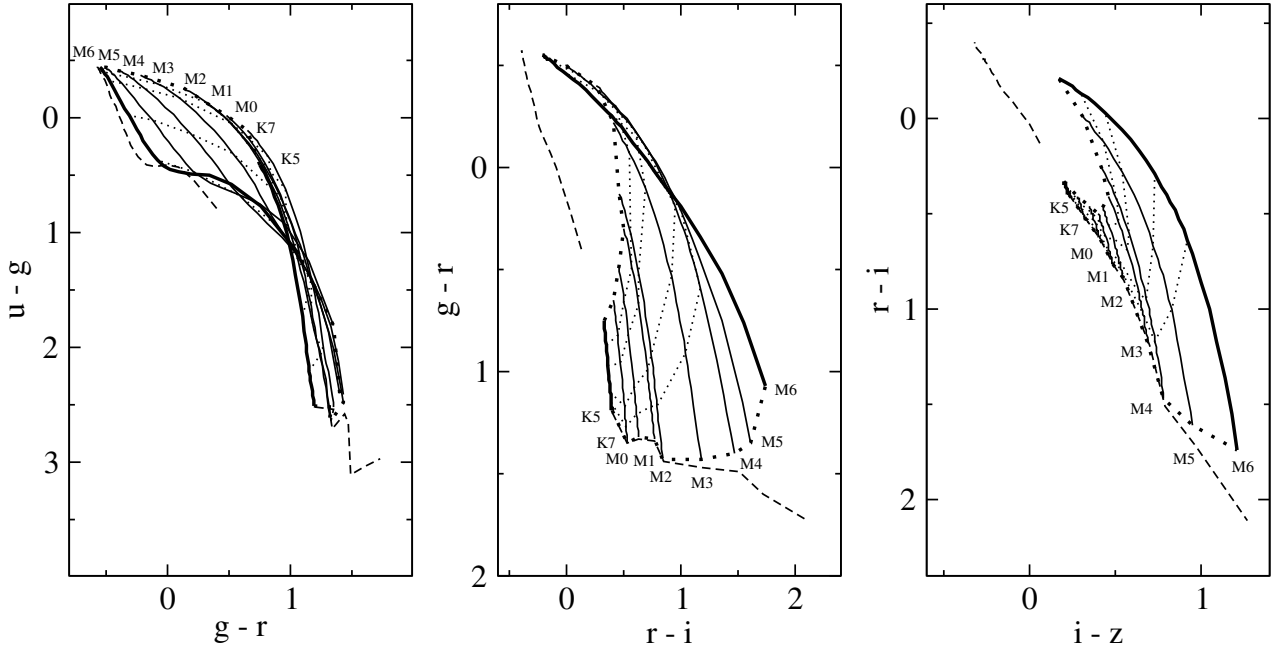


Fig. 2. The expected position of white dwarf-red dwarf binaries with a DA white dwarf in the different CCDs. The different solid lines correspond to binaries with a specific spectral type of the secondary as indicated in the figure. The surface gravity of the white dwarfs is fixed at $\log g = 8.0$. The dotted lines correspond to binaries with the same temperature white dwarfs, where we show (from top to bottom) effective temperatures of 80 000, 50 000, 34 000, 22 000, 14 000 and 6000 K, respectively. The dashed line in the upper left part of each panel indicates the position of single white dwarfs and the dashed line in the lower right part of each panel indicates the position of single red dwarfs.

3. White dwarf-red dwarf binaries

To explore the distribution of white dwarf-red dwarf binaries across the CCDs we have investigated the expected position of such binaries. For this we need the absolute magnitudes of the components in the binary in the different SDSS passbands to be able to combine their contributions and derive the expected colours. Here we have used values for single white dwarfs and single red dwarfs. This assumes that both components in the binary have the same brightness and colour as a single star of identical type, which should give a good representation for non-interacting binaries. The objective of this is to determine the expected brightness and colour of a white dwarf-red dwarf binary as a function of temperature and surface gravity (and chemical composition of the atmosphere) of the white dwarf, and the spectral type of the red dwarf.

3.1. Binary components

To derive the absolute magnitudes for white dwarfs as a function of their properties, specifically their effective temperature and surface gravity, we have used a model grid of H-rich (DA) and He-rich (DB) white dwarfs kindly made available to us by Detlev Koester (see for details Paper II). To derive the absolute magnitudes for the red dwarfs we have used the flux calibrated spectra provided by Pickles (1998).

We specifically use these spectra as they were also used in the analysis of the spectra in our sample (Paper II). To convert the Eddington fluxes of the WD models to observed fluxes (in $\text{erg s}^{-1} \text{cm}^{-1} \text{\AA}^{-1}$) we have used the mass-gravity relations of (Napiwotzki et al. 2005). For the scaling of the red dwarf spectra we used the filter response curve and the absolute V magnitude provided by Pickles (1998), and the zero point flux for the V filter provided by Bessell et al. (1998). Absolute magnitudes for the SDSS filters were obtained using the filter response curves

provided on the SDSS web page¹ where we applied corrections of +0.04 and -0.02 to the magnitudes in the u and z filter, respectively, to convert the resulting “true” AB magnitudes to SDSS catalogue magnitudes².

3.2. Binary colours

From the absolute magnitudes defined above for white dwarfs and red dwarfs it is straightforward to derive the expected colours of a white dwarf-red dwarf binary which can be compared to the CCDs for the sources from the proper motion catalogue. In Fig. 2 we present the resulting CCDs for white dwarf-red dwarf binaries with a DA white dwarf for a range in spectral type of red dwarfs (K5V to M6V), and a range in white dwarf temperatures (6000–80 000 K) for a fixed surface gravity of $\log g = 8.0$. Also indicated in the figure are the colours we derived for single white dwarfs and single red dwarfs.

Comparing Fig. 2 to the CCDs of the sources in the catalogue presented in Fig. 1 it is clear that, except for binaries with the latest types of red dwarfs ($\gtrsim M3$), it is not possible to separate white dwarf-red dwarf binaries from the main locus of stars in the $(r-i)$ vs. $(i-z)$ CCD. However, except for the earliest type of red dwarfs ($\lesssim K7$), the $(g-r)$ vs. $(r-i)$ CCD does provide a clear separation of the white dwarf-red dwarf binaries from the main locus of stars. White dwarf-red dwarf binaries are also well separated from single red dwarfs (even for the earliest type of red dwarfs if the temperature of the white dwarf is $\gtrsim 15\,000$ K) in the $(u-g)$ vs. $(g-r)$ CCD, but for binaries with late-type red dwarfs ($\gtrsim M5$) the colours overlap with those of single white dwarfs. The tracks with the range in properties for the white dwarfs and red dwarfs presented in Fig. 2 provide distinct areas in the CCDs

¹ <http://www.sdss.org/dr5/instruments/imager/index.html#filters>

² <http://www.sdss.org/dr5/algorithms/fluxcal.html>

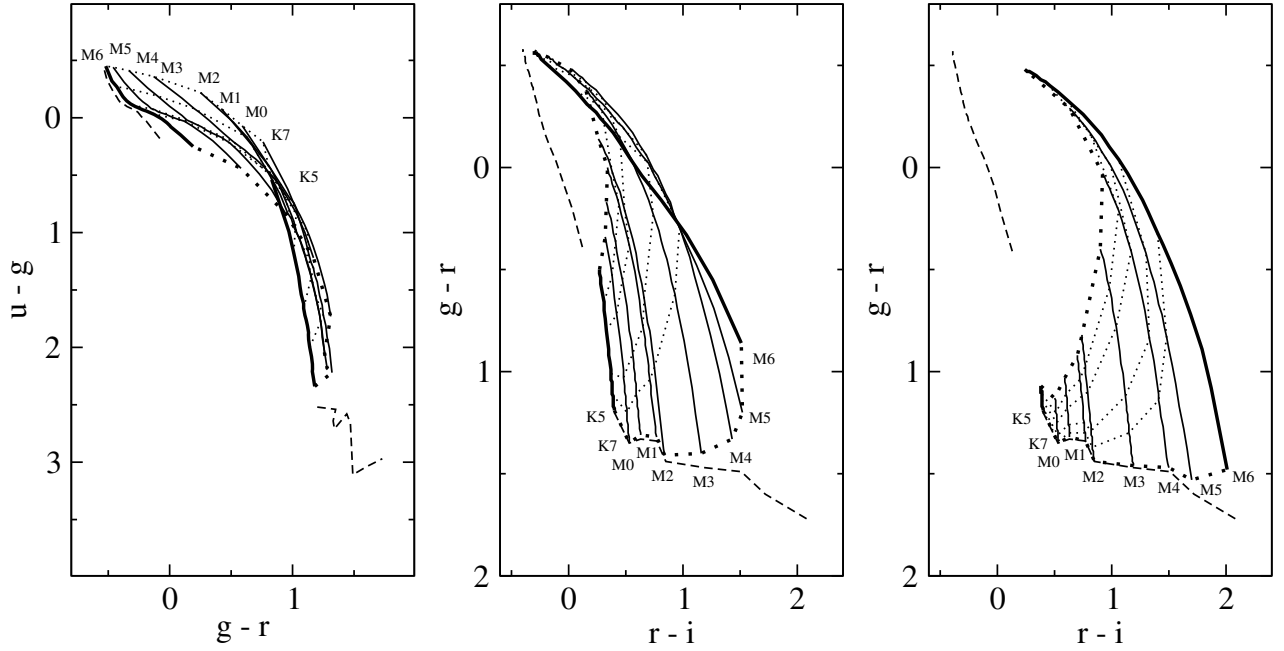


Fig. 3. The middle and right panel show the $(g-r)$ vs. $(r-i)$ CCDs in the same way as in Fig. 2, but now for white dwarf-red dwarf binaries with a DA white dwarf with $\log g = 7.0$ (middle panel), and $\log g = 9.0$ (right panel). The left panel shows the $(u-g)$ vs. $(g-r)$ CCD with the expected position of white dwarf-red dwarf binaries with a DB white dwarf with $\log g = 8.0$. The dotted lines correspond to binaries with the same temperature white dwarfs, where in this case we show (from top to bottom) effective temperatures of 50 000, 34 000, 22 000, 15 000 and 10 000 K, respectively.

where we expect to find white dwarf-red dwarf binaries. One can actually discern the population of white dwarf-red dwarf binaries in both the $(u-g)$ vs. $(g-r)$ and $(g-r)$ vs. $(r-i)$ CCDs in Fig. 1 (see also Smolčić et al. 2004). However, looking at Fig. 1 it is also clear that there is a lot of contamination of likely spurious sources in the data of the catalogue. Specifically, we note the groups of sources that appear to radiate away or are parallel to the main locus of stars in the different CCDs.

We have also investigated the effect of varying the surface gravity of the white dwarf on the expected colours of a white dwarf-red dwarf binary. In Fig. 3 we present the $(g-r)$ vs. $(r-i)$ CCDs for white dwarf-red dwarf binaries with a DA white dwarf and a surface gravity of $\log g = 7.0$ and $\log g = 9.0$. As expected, the tracks for a given spectral type of the red dwarf will move to the upper left for $\log g = 7.0$, i.e., a bigger white dwarf, as the contribution of the white dwarf to the overall emission from the binary will be larger for the same effective temperature, while for $\log g = 9.0$ the tracks will move to the lower right. In general, one would expect white dwarf-red dwarf binaries which are bluest in $(r-i)$ to contain a low surface gravity white dwarf, while the reddest binaries in $(r-i)$ are expected to contain a high surface gravity white dwarf.

With the exception of the $(u-g)$ colour, due to the fact that there is no Balmer jump in DB white dwarfs, the magnitude and colours for DA and DB white dwarfs are very similar. In the left panel of Fig. 3 we present the $(u-g)$ vs. $(g-r)$ CCD for white dwarf-red dwarf binaries with a DB white dwarf for a range in white dwarf temperatures (10 000–50 000 K) and a fixed surface gravity of $\log g = 8.0$. Although the tracks have a somewhat different shape than those of white dwarf-red dwarf binaries with a DA white dwarf there is not much qualitative difference. The tracks in the other CCDs for white dwarf-red dwarf binaries with a DB white dwarf are virtually the same as the tracks with DA white dwarfs (see Fig. 2).

In general, we can conclude from Figs. 2 and 3 that for a large range of surface gravities and temperatures for both DA and DB white dwarfs, and a large range of spectral types for the red dwarf, white dwarf-red dwarf binaries are well separated from the main locus of stars in both the $(u-g)$ vs. $(g-r)$ and $(g-r)$ vs. $(r-i)$ CCDs.

4. The main stellar locus

The main problem in defining the sample of high proper motion white dwarf-red dwarf binaries is to separate the candidate binaries from single stars and to filter out spurious sources in a well-defined, objective way. From the results presented in the previous section we can conclude that white dwarf-red dwarf binaries can in principle be separated from single main-sequence stars and single white dwarfs in a $(g-r)$ vs. $(r-i)$ CCD. On the other hand, these binaries could also be separated from single main-sequence stars in a $(u-g)$ vs. $(g-r)$ CCD, but there is a significant overlap between the expected position of white dwarf-red dwarf binaries and that of single white dwarfs. The $(r-i)$ vs. $(i-z)$ CCD can not be used to find white dwarf-red dwarf binaries as there is a large overlap between their expected position in this CCD and the main stellar locus. As a first step we will define the position of the main stellar locus in each CCD by calculating the average value of $(u-g)$ as a function of $(g-r)$, and of $(g-r)$ as a function of $(r-i)$, respectively. To do this, the data were binned as a function of $(g-r)$ in the $(u-g)$ vs. $(g-r)$ CCD, and as a function of $(r-i)$ in the $(g-r)$ vs. $(r-i)$ CCD.

In the $(u-g)$ vs. $(g-r)$ CCD the white dwarfs are well separated from the main-sequence but show a more complicated structure as DA and DB white dwarfs have a different distribution. Here we only define the position of the main locus of main-sequence stars limited over the colour range $(g-r) = 0.25$ to 1.65 and using a bin-size of 0.01 mag. The main stellar locus in the $(g-r)$ vs. $(r-i)$ CCD consists mainly of white dwarfs for

$(g-r) \leq 0.2$, and main-sequence stars for redder colours. To define the main stellar locus, we have used for the main-sequence stars a bin-size of 0.01 mag over the range $(r-i) = 0.10$ to 1.65 and, as the number of sources becomes smaller, a bin size of 0.03 mag over the range $(r-i) = 1.65$ to 1.83. As the number of objects in the part of the main locus dominated by white dwarfs is small we have used a bin size of 0.03 mag over the range $(r-i) = -0.33$ to -0.06 . To remove any extreme outliers we have only considered sources that have $-1 < (u-g) < 6$, $-0.7 < (g-r) < 1.8$, $(r-i) < 2.4$ and $(g-r) < \frac{20}{9} \times (r-i) + 0.5$ (the latter to exclude the relatively many sources that are just below the top part of the main locus in the $(g-r)$ vs. $(r-i)$ CCD).

To define the location of the main stellar locus and exclude as much as possible the contribution from stars that are not on the locus, we have tried various iterative schemes to determine the average value in each bin along the abscissa in each CCD. As the catalogue is dominated by stars in the main stellar locus, all schemes converge very rapidly, and the simplest scheme giving consistent results consists of doing a single step 2-sigma clipping before deriving the final position and width of the locus in each bin. Varying the bin size does not have any significant effect on the measured averages and widths (σ) along the main stellar locus. In the transition region between the white dwarfs and main-sequence stars in the $(g-r)$ vs. $(r-i)$ CCD (i.e. from $(r-i) = -0.06$ to 0.10) it is difficult to determine the average and width of the main stellar locus in a well-defined way. One might construct more sophisticated ways to calculate the averages and widths in this region, but we decided to just simply interpolate over it. We will later use the location and width of the main stellar locus of stars to select white dwarf-red dwarf binaries. The specific way we define these values in this region does not have any effect on the sample we select.

The result of these procedures is indicated in Fig. 1 as a set of points at the centre of each bin along the abscissa. These points show the average and sigma of the sources included in each bin, for each bin of the ordinate values³.

5. Selection criteria

A major issue in deriving the sample of white dwarf-red dwarf binaries is to define the separation between the main stellar locus of stars and the binary candidates. The basic idea is to use the position and width of the main stellar locus determined above to derive a well-defined limit. One would expect to have a wide range of colours for white dwarf-red dwarf binaries depending on the effective temperature and size of the white dwarf, and the spectral type of the secondary. As there will be cases in which either the white dwarf or the red dwarf can completely dominate the (optical) emission, the main issue is to limit the contamination by single stars and spurious sources. To explore this in detail we have used the available SDSS spectra (Adelman-McCarthy et al. 2007, Data Release 5) for sources from the proper motion catalogue to study the effect of varying the separation between the main locus and the binary candidates.

As a first step we concentrated on the $(g-r)$ vs. $(r-i)$ CCD because we expect this to best separate the white dwarf-red dwarf binaries from single stars. To limit the number of objects, we have only studied those sources that are $1-\sigma$ above the main locus in the $(g-r)$ vs. $(r-i)$ CCD. The limiting value in $(g-r)$ as a function of $(r-i)$ is calculated by taking, for each bin in

$(r-i)$, the average plus $1-\sigma$ in $(g-r)$ and linearly interpolating between these values. At the extremes, we extrapolated the limits as a function of $(r-i)$ at a constant level in $(g-r)$, corresponding to the value of the average plus $1-\sigma$ in the last bin in $(r-i)$. In this way we have selected a total of 63 602 sources from the catalogue. Out of these, a total of 554 already have a spectrum taken as part of the SDSS. Given the very conservative limits, we expect this sample to still contain a large fraction of main-sequence stars from the main stellar locus but at the same time not exclude any likely white dwarf-red dwarf binaries.

As a next step three of the authors (TA, RG, EvdB) have independently classified the spectra visually in a number of broad groups being: white dwarfs, early-type ($\sim A-G$) dwarfs, late-type ($\sim K-M$) dwarfs, white dwarf-red dwarf binaries (WD+RD), cataclysmic variables (CVs) and quasars (QSOs). We also noted separately if a source showed $H\alpha$ emission. There was a very large overlap between the classification of the sources, and a consensus was easily reached for all. The identification of $H\alpha$ emission is slightly more arbitrary as it was not always straightforward to decide whether there is truly emission. This is especially the case when there might be a combination of $H\alpha$ emission and ($H\alpha$ or other nearby) absorption and/or when the signal-to-noise of the spectrum is relatively low. One important thing to note is that here we only looked at the features in the spectrum to classify sources, and we did not specifically take their photometric colours into account.

The resulting distribution in the CCDs of the sources classified in this way is shown in Fig. 4 where the different symbols indicate the different classifications and a plus sign indicates the (likely) presence of emission. As expected, the plots still show the contours of the main stellar locus with the white dwarf-red dwarf binaries above it (where the relatively sharp lower-envelope in the $(g-r)$ vs. $(r-i)$ CCD is the result of the $1-\sigma$ selection we have made). However, the $(g-r)$ vs. $(r-i)$ CCD also shows sources in the top-right with colours well beyond what is expected for white dwarf-red dwarf binaries, while the $(u-g)$ vs. $(g-r)$ CCD shows sources to the left of the main stellar locus where one does not expect sources at all. Looking at the classification of these sources and their position in the CCDs, it appears that these are likely regular main-sequence sources with an incorrect r -band magnitude.

To improve on this further we will use the information provided by the $(u-g)$ vs. $(g-r)$ CCD (see Fig. 4). As the white dwarf, even for moderate temperatures, is in many cases expected to dominate in the blue and there is a significant overlap between the expected position in the CCD of white dwarf-red dwarf binaries and that of single white dwarfs (see Fig. 2), we simply have included all sources bluer than $(u-g) = +0.85$ (the blue cut-off of the main-sequence). In the $(u-g)$ vs. $(g-r)$ CCD we only take into account the separation between white dwarf-red dwarf binaries and main-sequence stars using the position and width of the main locus as defined in Sect. 4.

Given the position and width of the main stellar locus in both CCDs we can determine how distant any object is from the main locus in each case. We define these values in units of the width (σ) of the main locus, where we interpolate the values of the average and width in the ordinate as a function of the value of the abscissa in each case. This results in two numbers for each object that defines its position with respect to the main stellar locus.

The results for all the sources that are more than $1-\sigma$ above the main locus in the $(g-r)$ vs. $(r-i)$ CCD are shown in Fig. 5 as dots. The distance of a source in $(u-g)$ as a function of $(g-r)$ is shown along the abscissa, and the distance in $(g-r)$ as a

³ The detailed results of the determination of the position of the main stellar locus of stars can be obtained from the authors in tabular form, listing the average and width for each combination of colours.

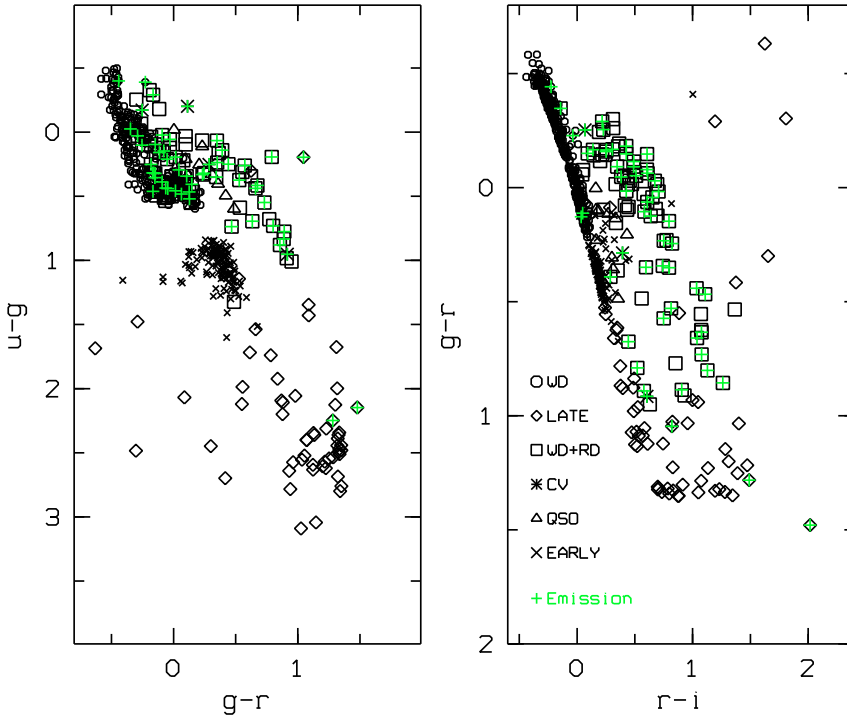


Fig. 4. The $(u-g)$ vs. $(g-r)$ and $(g-r)$ vs. $(r-i)$ CCDs of the sources that are $1-\sigma$ above the main locus in the $(g-r)$ vs. $(r-i)$ CCD and have a spectrum taken as part of the SDSS. The different symbols indicate the visual classification that was made.

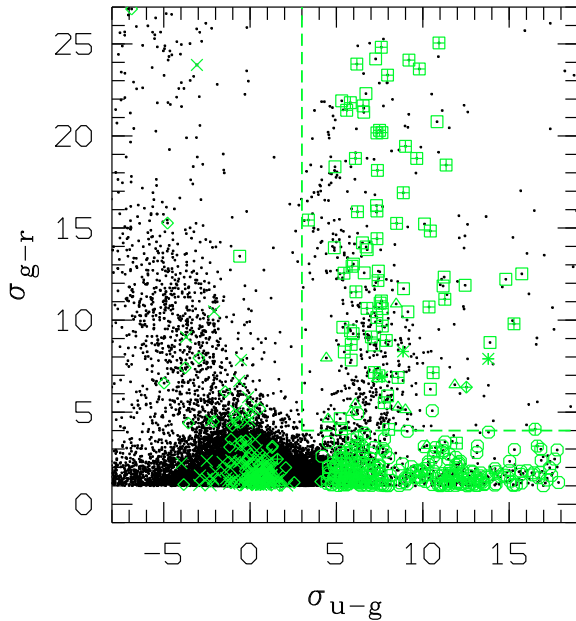


Fig. 5. The distance to the main stellar locus in units of σ for the sources that are more than $1-\sigma$ above the main locus in the $(g-r)$ vs. $(r-i)$ CCD. The small symbols correspond to all the objects that are $1-\sigma$ above the main locus in the $(g-r)$ vs. $(r-i)$ CCD. The large symbols correspond to those objects that have a spectrum taken as part of the SDSS where the different symbols indicate the classification as in Fig. 4.

function of $(r-i)$ is shown along the ordinate. Note that the distance in $(g-r)$ is limited to values greater than 1.0 as this was imposed to limit the number of objects to investigate. Also shown in Fig. 5 are the sources we have classified from their existing SDSS spectra, where the different symbols indicate the classification as in Fig. 4. As expected, there is a strong concentration of single (“early” \sim A–G and “late” \sim K–M) stars at low values in both colours. Single white dwarfs are concentrated at low values in σ_{g-r} and relatively high values in σ_{u-g} , while the

sources classified as white dwarf-red dwarf binaries are mostly at higher, positive values. CVs and QSOs are in between the white dwarf-red dwarf binaries, though they do not reach very high distances away from the main loci.

5.1. Selection in the $(g-r)$ vs. $(r-i)$ colour–colour diagram

The sources at high values in σ_{g-r} (≥ 5) and negative values in σ_{u-g} are probably wrong measurements. As these are mostly classified as normal stars it is likely that the r magnitude is incorrect (too faint), which moves them in the $(g-r)$ vs. $(r-i)$ CCD to a position above the main locus, into the area where we also expect white dwarf-red dwarf binaries. In the $(u-g)$ vs. $(g-r)$ CCD the incorrect r magnitude moves them from the main stellar locus to the blue, and well below the location of the main locus. This clearly shows that limiting the selection to only the $(g-r)$ vs. $(r-i)$ CCD would still cause a significant level of contamination in the sample (see Fig. 4).

Investigating Fig. 5 in detail one can see that to define a proper sample from the whole proper motion catalogue, a cut in σ_{g-r} is needed to separate the white dwarf-red dwarf binaries from the single stars, and a cut in σ_{u-g} is needed to remove contamination with sources that have incorrect colours. The sources we have classified as white dwarf-red dwarf binaries from their SDSS spectra do show a well defined distribution with a relatively abrupt boundary in σ_{u-g} and a more gradual variation as a function of σ_{g-r} . To explore the latter in a quantitative way we have taken all the sources with $\sigma_{u-g} > 3$ (the precise limit does not affect the following, see below) and looked at the number of sources that we have classified as white dwarf-red dwarf binaries compared to the total number of sources that were classified as a function of σ_{g-r} .

One potential problem is that any white dwarf-red dwarf binary which is dominated in brightness by one of the two components will have colours close to the main stellar locus, and might not be recognised as a binary from our visual classification. This could under-estimate the relative number of white

Table 1. The number of white dwarf-red dwarf binaries (WD+RD) and sources with emission for $\sigma_{u-g} > 3$. In parenthesis the fraction (%) of the total is given.

Type	σ_{g-r}				
	1.0	2.0	3.0	4.0	5.0
All	373	178	117	95	88
WD+RD	81(22)	81(46)	80(68)	79(83)	76(86)
Emission	49(13)	48(27)	48(41)	46(48)	44(50)

dwarf-red dwarf binaries close to the main locus. In principle, any source classified as a single white dwarf or red dwarf might be an unrecognised white dwarf-red dwarf binary, but in itself that does not provide a way to check if we are specifically including or excluding them. For that reason we have taken all those sources that were classified as white dwarfs, red dwarfs, or white dwarf-red dwarf binaries with emission. The reasoning behind this is that for any of these sources the H α emission might indicate that the source is actually a binary, where the emission is due to heating of part of the atmosphere of a red dwarf by a relatively hot white dwarf. The relative size and temperature of the components determine if the spectrum is dominated by either component, but this sample should not be biased against white dwarf-red dwarf binaries which are dominated in brightness by one of the two components. If any white dwarf or red dwarf in this sample is just a single star with emission, you would expect them to be concentrated towards the main locus, so this would tend to over-estimate the fraction of sources at low values of σ_{g-r} .

The resulting fractions for both the sources classified as white dwarf-red dwarf binaries and the sources with H α emission are given in Table 1. The changes in percentage as a function of σ_{g-r} is very similar for both groups of sources which indicates that we are unlikely to exclude any large fraction of unrecognised white dwarf-red dwarf binaries. In both cases the fraction steadily increases and then levels off. Based on these results, and as a balance between including white dwarf-red dwarf binaries and excluding any contamination, we decided to take $\sigma_{g-r} = 4$ as the limit.

5.2. Selection in the $(u-g)$ vs. $(g-r)$ colour-colour diagram

The sources we have classified from their existing SDSS spectra do not provide a limit as well defined for σ_{u-g} as for σ_{g-r} as there is only one source in the range $\sigma_{u-g} = 0-4$. However, the overall distribution of all sources presented in Fig. 5 follows a similar pattern to that of the sources that have been classified, where the group of sources starting off at $\sigma_{u-g} \sim 7$, $\sigma_{g-r} \sim 4$ and going up in σ_{g-r} are all likely to be white dwarf-red dwarf binaries. The main complication when defining a limit as a function of σ_{u-g} is that this group of sources bends toward low values of σ_{u-g} for higher values of σ_{g-r} . This is due to the location of white dwarf-red dwarf binaries in the $(u-g)$ vs. $(g-r)$ CCD getting close to that of early-type main-sequence stars; see Figs. 1 and 2.

The sources at low σ_{g-r} are likely to be contaminating single early-type stars or single white dwarfs from the main stellar locus (depending on the precise choice of σ_{u-g}). Assuming that in the range of $\sigma_{u-g} \approx 0-4$ the sources with $\sigma_{g-r} > 10$ are likely white dwarf-red dwarf binaries and sources with $\sigma_{g-r} < 10$ are likely contamination, we can study the fraction of white dwarf-red dwarf binary candidates to be included as a function of σ_{u-g} . For $\sigma_{g-r} > 4$ the ratio of likely white dwarf-red dwarf

binaries to contaminating sources (basically those above and below $\sigma_{g-r} = 10$) is 4/33 for $1 < \sigma_{u-g} < 2$, 7/19 for $2 < \sigma_{u-g} < 3$ and 22/6 for $3 < \sigma_{u-g} < 4$. There is a marked change in this ratio going from $2 < \sigma_{u-g} < 3$, which is dominated by contaminating sources, to $3 < \sigma_{u-g} < 4$, which is dominated by likely white dwarf-red dwarf binaries, thus we selected the best limit to be $\sigma_{u-g} = 3$.

6. The sample

The best balance between including white dwarf-red dwarf binaries and excluding single stars and spurious measurements is obtained by combining the selection in both CCDs. In the $(u-g)$ vs. $(g-r)$ CCD we include the sources that are more than 3σ in $(u-g)$ above the main locus of stars, and in the $(g-r)$ vs. $(r-i)$ CCD those that are more than 4σ in $(g-r)$ above the main locus of stars (indicated by the dashed lines in Fig. 5). By applying these limits we arrive to a sample of 651 sources. The CCDs of the final sample are shown in Fig. 6. In the $(u-g)$ vs. $(g-r)$ CCD the sample is limited on the red side in $(g-r)$ as given in the $(g-r)$ vs. $(r-i)$ CCD (where we show the limits up to the highest value of $(g-r)$ allowed). In the $(g-r)$ vs. $(r-i)$ CCD the sample is limited at $(r-i) = 2.4$ as defined at the start of Sect. 4.

6.1. Magnitude distribution

As mentioned before, the distribution of points in Figs. 1, 4 and 5 seems to indicate that there are many sources that have incorrect r magnitudes. To investigate this in more detail we looked at the distribution of the sources in r magnitude. The distribution for the sample, basically the sources that are above the main locus of stars in both the $(u-g)$ vs. $(g-r)$ and $(g-r)$ vs. $(r-i)$ CCDs (i.e. with $\sigma_{u-g} > 3$ and $\sigma_{g-r} > 4$; see Fig. 5), is shown as a solid line in Fig. 7. The distribution for sources that are above the main locus in the $(g-r)$ vs. $(r-i)$ CCD, but in, or below the main locus in the $(u-g)$ vs. $(g-r)$ CCD (i.e., with $\sigma_{u-g} < 3$ and $\sigma_{g-r} > 4$) is shown as a dashed line in Fig. 7).

The distribution of the sample (solid line in Fig. 7) shows the monotonic increase towards fainter magnitudes that one would expect, with a turnover at $r \sim 19.5$ mag indicating the completeness limit of the sample. However, there is also an unexpected increase in the distribution towards the bright end, starting at $r \sim 16.5$ mag, which indicates some contamination. The sources that were excluded from the sample (dashed line in Fig. 7) have a very different distribution from that of the sample, being dominated by sources with very bright ($r \sim 14-15$) mag, and a more or less flat distribution down to the limiting magnitude of the catalogue at $r = 20$. This is not what one would expect of a magnitude limited sample, but is consistent with the idea that most of these sources have incorrect colours (r magnitudes that are relatively too faint). These incorrect r magnitudes move the sources from the main locus, up into the area covered by white dwarf-red dwarf binaries in the $(g-r)$ vs. $(r-i)$ CCD, and to the left of the main locus in the $(u-g)$ vs. $(g-r)$ CCD, away from the area covered by white dwarf-red dwarf binaries.

The increase in the distribution of the sample towards bright magnitudes coincides more or less with that seen in the distribution of the sources that were excluded, indicating that the sample is affected for magnitudes brighter than $r \sim 16.5$. If these sources are similar to the majority of sources that were excluded, then the main reason why these sources are not excluded by the selection procedure is likely to be because we include all sources with $(u-g) < 0.85$ in the $(u-g)$ vs. $(g-r)$ CCD. While sources

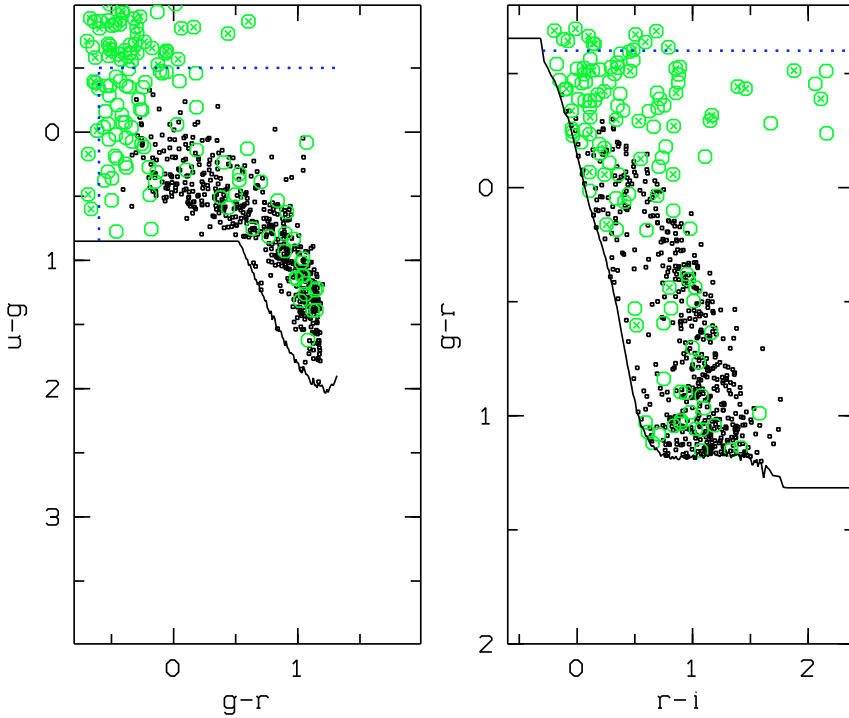


Fig. 6. The $(u-g)$ vs. $(g-r)$ and $(g-r)$ vs. $(r-i)$ CCDs of the final sample. The solid line in the CCDs shows the applied selection criteria used to define our sample. The sources brighter than $r = 16.5$ are indicated with a circle. The dashed line shows the additional colour selection (see text), where the points excluded by this selection are indicated with crosses.

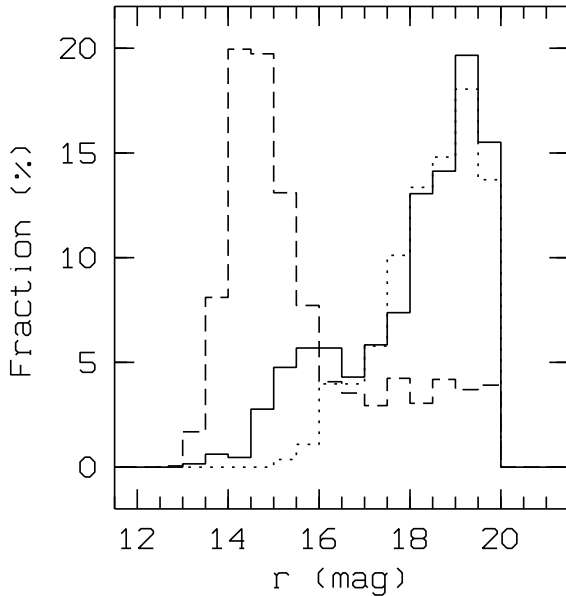


Fig. 7. A histogram of the r magnitudes of the sources that are included in the $(g-r)$ vs. $(r-i)$ CCD but excluded in the $(u-g)$ vs. $(g-r)$ CCD ($\sigma_{u-g} < 3$ and $\sigma_{g-r} > 4$; dashed line), of the sources that are included in both CCDs ($\sigma_{u-g} > 3$ and $\sigma_{g-r} > 4$; solid line), and the sources that are included in both CCDs, but have $(u-g) > 0.85$ (dotted line).

with redder $(u-g)$ colours and (relatively) too faint r magnitudes will be excluded as they move to the left, below the main locus of stars, this is not the case for the bluer sources. To verify this we have also looked at the magnitude distribution of only those sources in our sample which have $(u-g) > 0.85$. This is shown as the dotted line in Fig. 7. For comparison, this curve was scaled such that the area for $r > 16.5$ mag is the same as that for the complete sample. The resulting distribution shows, as one would expect, a monotonic rise from the bright end towards fainter magnitudes. This indicates that the peak at brighter magnitudes ($r \sim 16$) in the distribution of the whole sample is

limited to sources with $(u-g) < 0.85$, which are likely contaminating sources that have an r magnitude that is (relatively) too faint. In Fig. 6 the sources with magnitudes $r < 16.5$ are indicated with a circle.

One remaining question to address is the cause for the contaminating sources. One option is that these are completely spurious sources, e.g. an artifact of the data or the reduction. However, at least some of the sources with incorrect colours are real sources that are just displaced in colour (see specifically Fig. 4). The sources with $r < 16.5$ mag in the different filters turn out to be relatively bright in all bands, and it seems likely that the incorrect magnitudes are due to some saturation effect. The spread in magnitude of the sources that are affected can be understood as due to variations in sky transparency and seeing during the observations, which will cause the source magnitude at which the detector will show saturation effects to vary.

For a bright source one might expect the magnitude to be effected in any of the passbands. However, saturation of a source in the u or i band is very unlikely and if the magnitude of a source would be fainter than its true value in either of these bands it would move a source down from the main stellar locus and the source would actually be excluded. The effect of saturation in the g band resulting in a too faint magnitude is precisely opposite to that in the r band. In this case the source would move from the main locus up into the area covered by white dwarf-red dwarf binaries in the $(u-g)$ vs. $(g-r)$ CCD, but it would be down below the main locus in the $(r-i)$ vs. $(g-r)$ CCD, away from the area covered by white dwarf-red dwarf binaries. The main difference is that, in contrast to the $(u-g)$ vs. $(g-r)$ CCD, the area excluded in the $(r-i)$ vs. $(g-r)$ CCD also includes the region covered by white dwarfs, and sources with too faint g magnitudes that belong to the main locus of stars will always be excluded.

As the contaminating sources with bright r magnitudes seem to be spread-out also in the $(g-r)$ colour (see Fig. 6) it is not possible to exclude them. However, as white dwarf-red dwarf binaries are not expected to have $(u-g)$ and $(g-r)$ colours

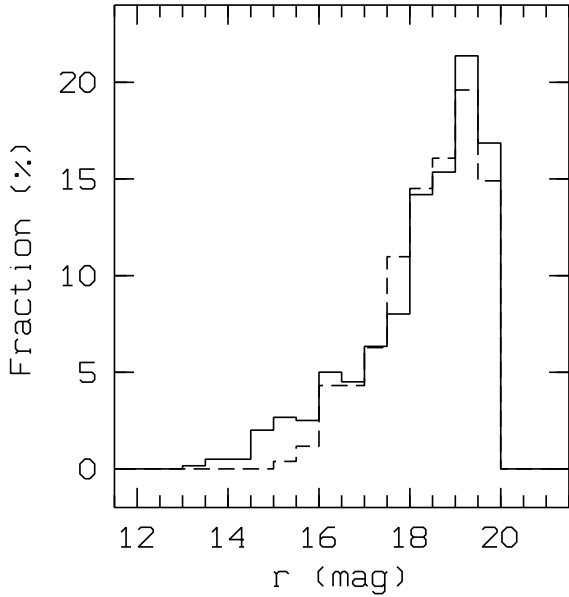


Fig. 8. A histogram of the r magnitudes of the sources in the final sample excluding sources which are bluer than $(u - g) = -0.5$ or $(g - r) = -0.6$ (solid line), and the sources in the sample that also have $(u - g) > 0.85$ (dashed line).

bluer than any single white dwarf, their number can be limited by excluding sources which are bluer than $(u - g) = -0.5$ or $(g - r) = -0.6$ (which corresponds to the blue cut-off for single white dwarfs identified in the SDSS; Kleinman et al. 2004).

The distribution in r magnitudes of the sources in the sample excluding sources which are bluer than $(u - g) = -0.5$ or $(g - r) = -0.6$ is shown in Fig. 8 with a solid line. Also shown, with the dotted line, is the distribution for these sources with $(u - g) > 0.85$, where this curve was scaled such that the area for $r > 16.5$ mag is the same as that for the complete sample. The distributions are now more similar than before (see Fig. 7), but there still is a systematic difference starting at $r \sim 16$ mag. In Fig. 6 the sources with colours bluer than $(u - g) = -0.5$ or $(g - r) = -0.6$ or magnitudes $r < 16.0$ are indicated with a crosses. It can be seen that these sources largely overlap with the sources that have $r < 16.5$. On the other hand, sources that are relatively red and have colours more in line with what is expected for white dwarf-red dwarf binaries are less affected. Increasing the brightness limit to exclude contamination should also reduce the fraction of true white dwarf-red dwarf binaries that are excluded (see Sect. 7.3).

7. Discussion

7.1. Sample properties

From Fig. 6 one can see that the distribution of our sample of white dwarf-red dwarf binaries, drawn from the proper motion catalogue (Gould & Kollmeier 2004), is well defined in both CCDs. It is very different from the distribution of all the sources in the catalogue (see Fig. 1) which, away from the main stellar locus of stars, is dominated by contaminating sources. Also, the shape of the distribution of the sample is as expected (compare Figs. 6 to 2), and we should detect white dwarf-red dwarf binaries with a large range in white dwarf temperatures and secondary spectral type. In particular, our sample does not suffer from the strong bias toward high temperature young white dwarfs, nor from the strong limitations posed by

“contamination” from distant sources (QSOs, Compact Emission line Galaxies and binary giants).

The distribution of sources with existing SDSS spectra in Fig. 4 actually does show a strong bias towards white dwarf-red dwarf binaries with hot white dwarfs (see also the discussion below), which is not apparent in our sample (see Fig. 6). However, although we specifically attempted to limit any biases in our sample some selection biases will also exist. One can get some idea of these selection biases by comparing the predicted colours of white dwarf-red dwarf binaries shown in Figs. 2 and 3 with those in Fig. 6. E.g., we specifically select against white dwarf-red dwarf binaries with relatively early-type red dwarfs and low temperature ($T_{\text{eff}} \leq 10000$ K) white dwarfs because they overlap with the main stellar locus, which is excluded.

The predicted colours of white dwarf-red dwarf binaries, which at least should give a qualitative description, also provide some predictions of the general properties of the sample. E.g., the sample seems to be limited to white dwarf-red dwarf binaries with red dwarfs of spectral type no later than $\sim M6$ (as defined by Pickles 1998). This is reasonable as it is the same as the cut-off spectral type for single red dwarfs present in the whole catalogue (see Fig. 1). The colour tracks do show that white dwarf-red dwarf binaries with the latest type red dwarfs should stand-out in a $(r - i)$ vs. $(i - z)$ CCD. As also shown by the tracks one would actually not find white dwarf-red dwarf binaries with very late-type red dwarfs in the bottom-right of a $(g - r)$ and $(r - i)$ CCD, but in the top-left, just to the right of the main locus for single white dwarfs.

Figure 2 indicates that for white dwarf-red dwarf binaries with a relatively hot white dwarf and late-type red dwarf, the $(u - g)$ vs. $(g - r)$ CCD should provide a direct way to differentiate white dwarf-red dwarf binaries as a function of white dwarf temperature and red dwarf spectral type. The same is true for white dwarf-red dwarf binaries with a relatively cool white dwarf and early-type red dwarf in the $(g - r)$ vs. $(r - i)$ CCD. In fact, individual tracks of the predicted colours of white dwarf-red dwarf binaries with a certain spectral type red dwarf and a varying white dwarf temperature define 3-dimensional tracks in $(u - g)$, $(g - r)$ and $(r - i)$, which are used in an accompanying paper (Paper II) to characterise white dwarf-red dwarf binaries based on photometry. Figure 3 also indicates that white dwarf-red dwarf binaries that are on the far left in the $(g - r)$ vs. $(r - i)$ CCD are likely to have a white dwarf with a low surface gravity, and white dwarf-red dwarf binaries that are on the far right are likely to have a white dwarf with a high surface gravity.

A particularly interesting thing to note of the sources in our sample that already have a spectrum is the high incidence of sources with $H\alpha$ emission lines. These sources are spread-out over the CCDs (see Fig. 4) which indicates that they are systems with a wide range in white dwarf temperatures and red dwarf spectral types. There are two possible explanations for this emission: either it is due to heating of the red dwarf by the white dwarf; or it is due to chromospheric activity of the red dwarf. In the former case, this directly implies that the system must be a close binary. As for the latter case, chromospheric activity is correlated with fast rotation (Schrijver & Zwaan 2000, and references therein) and in a close binary the red dwarf will rotate fast due to tidal interaction, i.e., also in this case the emission might point to the system being a close binary.

7.2. Comparison with other samples

The SDSS itself also includes a search for white dwarf-red dwarf binaries (Raymond et al. 2003; Silvestri et al. 2006). The original

criteria for target selection as part of the spectroscopic follow-up within the SDSS were $(u-g) < 0.45$, $(g-r) < 0.7$, $(r-i) > 0.30$, and $(i-z) < 0.40$ (Raymond et al. 2003), which were later modified to $[(u-g) - 1.314(g-r)] < 0.61$, $(r-i) > 0.91$, and $(i-z) < 0.49$ (Silvestri et al. 2006), with a magnitude cut-off of $g < 20$. The applied colour selection criteria limit the level of “contamination”, and a large fraction of the selected sources are indeed white dwarf-red dwarf binaries, but these colour criteria cover only part of the region in colours where white dwarf-red dwarf binaries are expected. It should be noted that these limits are only used to select targets for spectroscopy, and that these targets have a low priority for spectroscopic follow-up. However, the colours of white dwarf-red dwarf binaries overlap with many other objects targeted by the SDSS. This heavily affects the final spectroscopic sample as “the majority of systems in this sample were serendipitously targeted for spectroscopy by other higher priority pipelines (e.g. quasars, white dwarfs). Therefore, our spectroscopic compilation of close binary systems is not well defined photometrically nor is it statistically complete” (Silvestri et al. 2006).

These ill defined and variable selection criteria make it hard to draw any firm conclusion from this sample. E.g., comparing the colour limits used for spectroscopic targeting of white dwarf-red dwarf binaries to the expected position of white dwarf-red dwarf binaries as shown in Fig. 2, one can see that these criteria select white dwarf-red dwarf binaries with relatively late-type secondaries and relatively cool white dwarfs. This could explain the low number of binaries with early-type red dwarfs compared to single red dwarfs found by Silvestri et al. (2006). However, these authors also noted a relatively high number of white dwarf-red dwarf binaries with hot white dwarfs in their sample compared to single white dwarfs, which can not be explained by these selection criteria. On the other hand, targeting selection criteria for single white dwarfs and QSOs (see Fan 1999), which were the primary spectroscopic targets, include mostly only white dwarf-red dwarf binaries with relatively late-type red dwarfs and relatively hot white dwarfs. The reason for this is basically that the colours of white dwarf-red dwarf binaries with an early-type red dwarf, or with a low temperature white dwarf and a late-type red dwarf, will be too red to be included in either sample. The resulting effect of these, very different, selection biases might explain the distribution of white dwarf temperatures and red dwarf spectral types found for the white dwarf-red dwarf binaries in the SDSS sample presented by Raymond et al. (2003) and Silvestri et al. (2006). For a comparison of the properties of the sample from Silvestri et al. (2006) with our sample we refer to Paper II.

A separate study of white dwarf-red dwarf binaries in the SDSS was presented by Smolčić et al. (2004). This was based on the discovery of a “bridge”, containing these binaries, connecting the white dwarfs and red dwarfs in a $(u-g)$ vs. $(g-r)$ CCD of the complete SDSS Data Release 1 catalogue. This “bridge” can also be seen in Fig. 1 and is directly predicted by the expected colours of white dwarf-red dwarf binaries shown in Fig. 2. In principle this does provide a better defined sample than that presented by Raymond et al. (2003) and Silvestri et al. (2006), as it also includes white dwarf-red dwarf binaries with relatively cool white dwarfs. However, the primary limitation remains the presence of “contamination” from distant sources (mostly QSOs), and significant areas of the CCDs covered by white dwarf-red dwarf binaries are excluded to limit this. The specific colour criteria used in the paper by Smolčić et al. (2004) also include a large part of the area covered by single M stars which increases the relative amount of contamination.

A different search for (white dwarf-red dwarf) binaries in the SDSS was made by (Pourbaix et al. 2004) using “Colour-Induced Displacement”. This method relies on the variation in centroid position as a function of wavelength for binaries with different colour components. This method is specifically biased towards binaries with a significant separation and with components of similar brightness, which covers only a fraction of the intrinsic population of white dwarf-red dwarf binaries and as such does not provide a proper sample. However, this method can be used as a tool to separate close and wide binaries in any sample.

7.3. Limitations of the sample

One specific problem discussed in Sect. 6.1 is the contamination of bright sources with incorrect colours. A rough estimate of the degree of contamination can be made by comparing the distributions in Fig. 7, where we assume that the distribution for sources that have $(u-g) > 0.85$ (dotted line) is not contaminated by these bright sources (see Sect. 6.1). In the whole sample 20% of the sources have magnitudes brighter than $r = 16.5$, for which we estimate that $\sim 75\%$ (nearly 15% of the whole sample) are contaminating sources. If we exclude all sources brighter than magnitude $r = 16.5$ we estimate that we exclude $\sim 5\%$ of the proper candidates in the sample. When we exclude the sources bluer than $(u-g) = -0.5$ or $(g-r) = -0.6$ and compare the distributions in Fig. 8, we find that 8% of the sources have magnitudes brighter than $r = 16.0$, for which we estimate that $\sim 80\%$ (nearly 7% of all the sources included) are contaminating sources. If we exclude the bluer sources and all sources brighter than magnitude $r = 16$ we estimate that we exclude less than $\sim 2\%$ of the proper candidates in the sample.

A different kind of problem is that there are still QSOs present in our sample as is shown by the fact that we have classified 9 out of the 95 sources with an existing SDSS spectrum as QSOs. Clearly there is a problem with the determination of the proper motion for these sources. As noted in the proper motion catalogue (Gould & Kollmeier 2004), some contamination is still expected, especially towards fainter magnitudes. Indeed, 7 out of the 9 QSOs that were included have magnitudes fainter than $r = 19$, while 8 out of the 9 have a measured proper motion below 30 mas yr^{-1} , i.e., relatively close to the lower limit of 20 mas yr^{-1} of the sample. We also note that the position of the QSOs in Fig. 4 indicates that it might be possible to exclude a relatively large fraction of these sources by applying more complicated/arbitrary colour selections in the $(g-r)$ vs. $(r-i)$ CCD. It should also be noted that QSOs are targeted in particular for spectroscopic follow-up. This is likely to increase the fraction of these sources among those with an existing SDSS spectrum. Therefore, we conclude that the contamination by QSOs in the whole sample is likely to be well below 10%.

If there are serious problems with the proper motion measurements, one might also expect binary giants (which are expected to have similar colours to that of white dwarf-red dwarf binaries) in the sample, and among the sources with existing spectra. However, the surface density of binary giants down to the magnitude limit of the catalogue compared to that of QSOs is probably low and we do not expect any significant contamination by these sources.

CVs are interacting white dwarf-red dwarf binaries and we do not specifically select against them, but they are effectively a kind of contamination. Of the 4 CVs among the sources with existing spectra, 3 are included in the final sample. This corresponds to 3% of all sources in the sample with spectra. This

fraction goes down to 2% if we either only take the sources fainter than $r = 16.5$, or when we exclude the sources bluer than $(u - g) = -0.5$ or $(g - r) = -0.6$ and take the sources fainter than $r = 16.0$, as in both these cases there is one additional CV excluded.

The last remaining potential sources of contamination are single white dwarfs and single red dwarfs. Among the sources with existing spectra that are included in the sample there are 3 that were classified as white dwarfs, one of which shows $H\alpha$ emission, and there is one red dwarf with $H\alpha$ emission. These objects are only a minor source of contamination (at a level of $\sim 1-3\%$), which shows that our selection criteria are effective at separating the white dwarf-red dwarf binary candidates from single stars. This is especially true if we consider that both the sources with $H\alpha$ emission are strong candidates to be true white dwarf-red dwarf binaries. However, it should be noted that none of the sources with an existing spectrum in our sample is close to the boundary with the area where single red dwarfs are located in the CCDs (see Fig. 6).

A separate question is what the fraction of true white dwarf-red dwarf binaries is in the sample. Taking the sources that have been classified and given the discussion in Sect. 5, specifically the results presented in Table 1, one can estimate that at least 83% (85% if we include the source classified as a single white dwarf with $H\alpha$ in emission and the one classified as a single red dwarf with $H\alpha$ in emission as white dwarf-red dwarf binaries) of the final sample are genuine white dwarf-red dwarf binaries. The fraction is 85% (86%) if we only take the sources fainter than $r = 16.5$, and it is 84% (86%) if we exclude sources which are bluer than $(u - g) = -0.5$ or $(g - r) = -0.6$ and only take the sources fainter than $r = 16.0$.

One specific thing to note here is that the selection effects on the sample of sources with a spectrum is probably different from our sample, so it is not entirely clear if the above estimates can be generalised to the whole sample. However, except for the lowest magnitude bin, which has relatively few sources, the magnitude distribution of the sources with spectra is similar to the distribution for the whole sample for magnitudes fainter than $r = 16.5$, and similar to the distribution for the whole sample excluding sources which are bluer than $(u - g) = -0.5$ or $(g - r) = -0.6$ and for magnitudes fainter than $r = 16.0$ (see Sect. 6). From this we conclude that for the sample either excluding sources brighter than $r = 16.5$, or excluding sources brighter than $r = 16.0$ and bluer than $(u - g) = -0.5$ or $(g - r) = -0.6$, we estimate that $\sim 85\%$ of the remaining sources are true white dwarf-red dwarf binaries.

We also need to know how many genuine white dwarf-red dwarf binaries are *excluded*. Of the 82 sources with an existing spectra that were classified as white dwarf-red dwarf binaries, 79 were included in the final sample. Of the sources that were excluded, 2 are in the region of single white dwarfs in the CCDs, and one is among the white dwarf-red dwarf binaries in the $(g - r)$ vs. $(r - i)$ CCD, but in the region of early type single stars in the $(u - g)$ vs. $(g - r)$ CCD (see Figs. 4 and 5). The latter source might have some incorrect magnitude measurement (at $r = 19.83$ and $u = 21.64$ mag it is one of the faintest in the whole sample), while the other two are probably in their proper place, but have colours that makes them impossible to include without increasing the contamination of the sample by single white dwarfs. By simply taking the numbers of white dwarf-red dwarf binaries we can estimate that our selection criteria exclude $\sim 4\%$ of all genuine white dwarf-red dwarf binaries. If we also include single white dwarfs and red dwarfs with $H\alpha$ emission as likely white dwarf-red dwarf binaries we estimate that we

exclude $\sim 8\%$. Again, it should be noted that none of the sources with an existing spectrum in our sample is close to the boundary with the single red dwarfs region in the CCDs, but we do not expect that to alter our estimates very much.

If the sample is limited to sources fainter than $r = 16.5$ mag, or to sources redder than $(u - g) = -0.5$ or $(g - r) = -0.6$ and fainter than $r = 16.0$, the total number of white dwarf-red dwarf binaries that are excluded will be slightly higher (see above) and we estimate that we miss $\sim 10\%$ of the white dwarf-red dwarf binaries.

Apart from the limitations of the sample as a result of our selection criteria there is also a specific selection effect due to the limit on the proper motion of the sources in the proper motion catalogue (Gould & Kollmeier 2004). In principle this just puts a lower limit on the transverse velocity which varies linearly with distance. However, the effect on a sample of observed sources is more complicated and depends on a combination of the magnitude range covered by the sample, and the brightness and velocity distribution of the intrinsic population (see Paper II).

In a following paper (Paper II) we will investigate the properties of the sample as defined in Sect. 6 through the analysis of the sources with existing SDSS spectra. The results from this analysis are then used to determine the space density of the (thick disc, thin disc and halo) population of white dwarf-red dwarf binaries.

Acknowledgements. E.v.d.B. and P.J.G. are supported by NWO-VIDI grant 639.042.201 to P.J. Groot. We thank Detlev Koester for making his white dwarf model atmospheres available. Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is <http://www.sdss.org/>. The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Korean Scientist Group, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

References

- Abt, H. A. 1983, *ARA&A*, 21, 343
 Adelman-McCarthy, J. K., et al. 2007, *VizieR Online Data Catalog*
 Bessell, M. S., Castelli, F., & Plez, B. 1998, *A&A*, 333, 231
 de Kool, M., & Ritter, H. 1993, *A&A*, 267, 397
 Fan, X. 1999, *AJ*, 117, 2528
 Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, *AJ*, 111, 1748
 Gould, A., & Kollmeier, J. A. 2004, *ApJS*, 152, 103
 Green, P. J., Ali, B., & Napiwotzki, R. 2000, *ApJ*, 540, 992
 Iben, I., & Livio, M. 1993, *PASP*, 105, 1373
 Kleinman, S., Harris, H., Eisenstein, D., et al. 2004, *ApJ*, 607, 426
 Lada, C. J. 2006, *ApJ*, 640, L63
 Margon, B., Anderson, S. F., Harris, H. C., et al. 2002, *AJ*, 124, 1651
 Monet, D. G., Levine, S. E., Canzian, B., et al. 2003, *AJ*, 125, 984
 Napiwotzki, R., Karl, C. A., Nelemans, G., et al. 2005, in 14th European Workshop on White Dwarfs, ed. D. Koester, & S. Moehler, *ASP Conf. Ser.*, 334, 375
 Paczyński, B. 1976, in *Structure and Evolution of Close Binary Systems*, IAU Symp., 73, 75
 Pickles, A. 1998, *PASP*, 110, 863
 Pourbaix, D., Ivezić, Ž., Knapp, G. R., Gunn, J. E., & Lupton, R. H. 2004, *A&A*, 423, 755
 Raymond, S., Szkody, P., Hawley, S., et al. 2003, *AJ*, 125, 2621
 Ritter, H., & Kolb, U. 2003, *A&A*, 404, 301
 Schrijver, C. J., & Zwaan, C. 2000, *Solar and Stellar Magnetic Activity*, Cambridge Astrophys. Ser.
 Silvestri, N. M., Hawley, S. L., West, A. A., et al. 2006, *AJ*, 131, 1674
 Smolčić, V., Ivezić, Ž., Knapp, G. R., et al. 2004, *ApJ*, 615, L141
 van den Besselaar, E. J. M., Augusteijn, T., Greimel, R., et al. 2008, *A&A*, submitted (Paper II)
 Wachter, S., Hoard, D. W., Hansen, K. H., et al. 2003, *ApJ*, 586, 1356
 York, D., Adelman, J., Anderson, J., et al. 2000, *AJ*, 120, 1579