



---

# DSC validation report for the ApsisVal3.3a run

---

prepared by: M. Süveges  
reference: GAIA-C8-TN-MPIA-MSU-005-1  
issue: 1  
revision: 0  
date: 2019-03-12  
status: Draft

## Abstract

## Contents

<b>1 Overview</b>	<b>3</b>
<b>2 Validation data set</b>	<b>3</b>
2.1 Class composition . . . . .	3
2.2 Sky maps of validation objects . . . . .	4
2.3 Apparent magnitudes . . . . .	6
2.4 Proper motions . . . . .	7
2.5 Colour-magnitude diagrams . . . . .	9
<b>3 DSC performance on “regular” classes</b>	<b>10</b>
3.1 ExtraTree classifier based on BP-RP spectra . . . . .	10
3.1.1 Confusion matrix . . . . .	10
3.1.2 Classification probability vectors . . . . .	12
3.1.3 Class probabilities vs. $G$ . . . . .	14
3.1.4 Class probabilities vs. $G_{BP} - G_{RP}$ . . . . .	16
3.1.5 Class probabilities vs. astrometry . . . . .	18
3.1.6 Colour-magnitude diagrams of predicted classes . . . . .	20
3.1.7 Sky maps of correct/incorrect classifications . . . . .	22
3.1.8 Purity and completeness . . . . .	23
3.2 Astrometric classifier . . . . .	25
3.2.1 Confusion matrix . . . . .	25

3.2.2	Classification probability vectors	26
3.2.3	Class probabilities vs. $G$	27
3.2.4	Class probabilities vs. $G_{BP} - G_{BP}$	28
3.2.5	Class probabilities vs. astrometry	29
3.3	Position–parallax–G-magnitude classifier	30
3.3.1	Confusion matrix	30
3.3.2	Classification probability vectors	31
3.3.3	Class probabilities vs. $G$	32
3.3.4	Class probabilities vs. $G_{BP} - G_{BP}$	33
3.3.5	Class probabilities vs. astrometry	34
3.4	Photometric SVM classifier	35
3.5	Bayesian combination	35

# 1 Overview

This report was prepared using the validation data of DSC from the apsis\_val3.3a on February 1, 2019.

## 2 Validation data set

### 2.1 Class composition

	DSC not.	Total	Processed	Loss	Training	Test
Catacl.	(-4)	3350	2404	946	0	2404
CU7 QSO cand.	(-3)	0	0	0	0	0
WD-M dwarf bin.	(-2)	251	232	19	0	232
WD-MS bin.	(-1)	3160	3014	146	0	3014
Star	0	315886	306045	9841	59039	247006
WD	1	31571	30928	643	9697	21231
Binary	2	123737	112550	11187	38624	73926
QSO	4	431360	428061	3299	19902	408159
Galaxy	5	189396	184473	4923	19382	165091

TABLE 1: Class distribution of the DSC validation data set. In the leftmost column, the (bracketed) negative integers represent classes into which DSC does not classify.

The class composition of the complete validation data set and its subsets used as training and test sets are summarized in Table 1. The classes DSC classifies into are stars, white dwarfs (WDs), physical binaries, quasars and galaxies, henceforth termed ‘regular classes’ in this document. The enumeration values corresponding to them in the Java codes are (as shown in the second column of Table 1) 0, 1, 2, 4, 5, respectively. Of these objects, empirical training of the ExtraTree classifier used a subset of approximately 150000. The remaining more than 900000 objects were used to test and validate the performance of DSC.

Apart from these regular classes, the validation set contains also classes which do not have an unambiguous true class, such as cataclysmic variables and two different white dwarf-star binary types. For all these, at least two predicted class labels could be accepted as true (either of star, WD or binary). In order to enable MSC to deal with them, in the past they were included in the binary class. However, this caused considerable confusion in the classification of not only stars, binaries and white dwarfs, but also that of quasars. For this run, these objects were therefore not used to train ExtraTrees (cf. column ‘Training’ in Table 1).

For a detailed description of the present DSC validation set, see the technical note MSU-

001.

## 2.2 Sky maps of validation objects

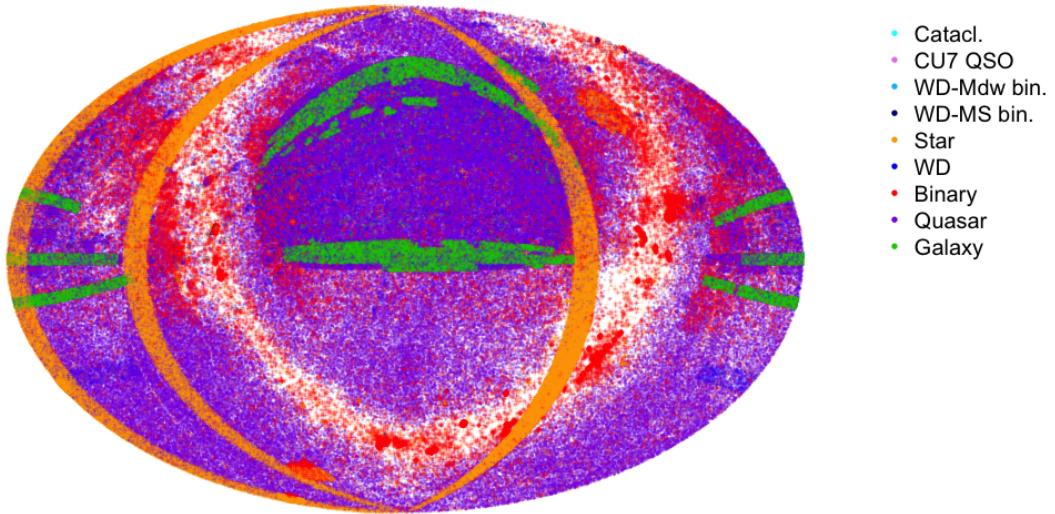


FIGURE 1: Sky map of the sources. The colour of the different classes is consistent throughout this document: orange for stars, blue for white dwarfs, red for physical binaries, purple for quasars, green for galaxies, cyan for cataclysmic variables, light blue for white dwarf-M dwarf binaries, and dark blue for white dwarf-main sequence star binaries.

Figures 1 and 2 show the distribution of the different classes of validation objects over the sky. The images mostly reflect the catalogs from which the objects were collected. For example, among the stars (top left of Fig. 2), the three densely populated bands consist of all Simbad objects in  $10^\circ$  wide stripes along three meridians for which all classification provided on Simbad corresponded to single stars. The white dwarfs (top right) and the white dwarf-star binaries (bottom) follow in majority the footprint of the Sloan Digital Sky Survey. A large part of the quasar set too was drawn from the SDSS Data Release 14, but since it was combined with other, all-sky catalogs of quasars (e.g. AllWise), the dominant structure in it is the band of the Milky Way, devoid of extragalactic objects.

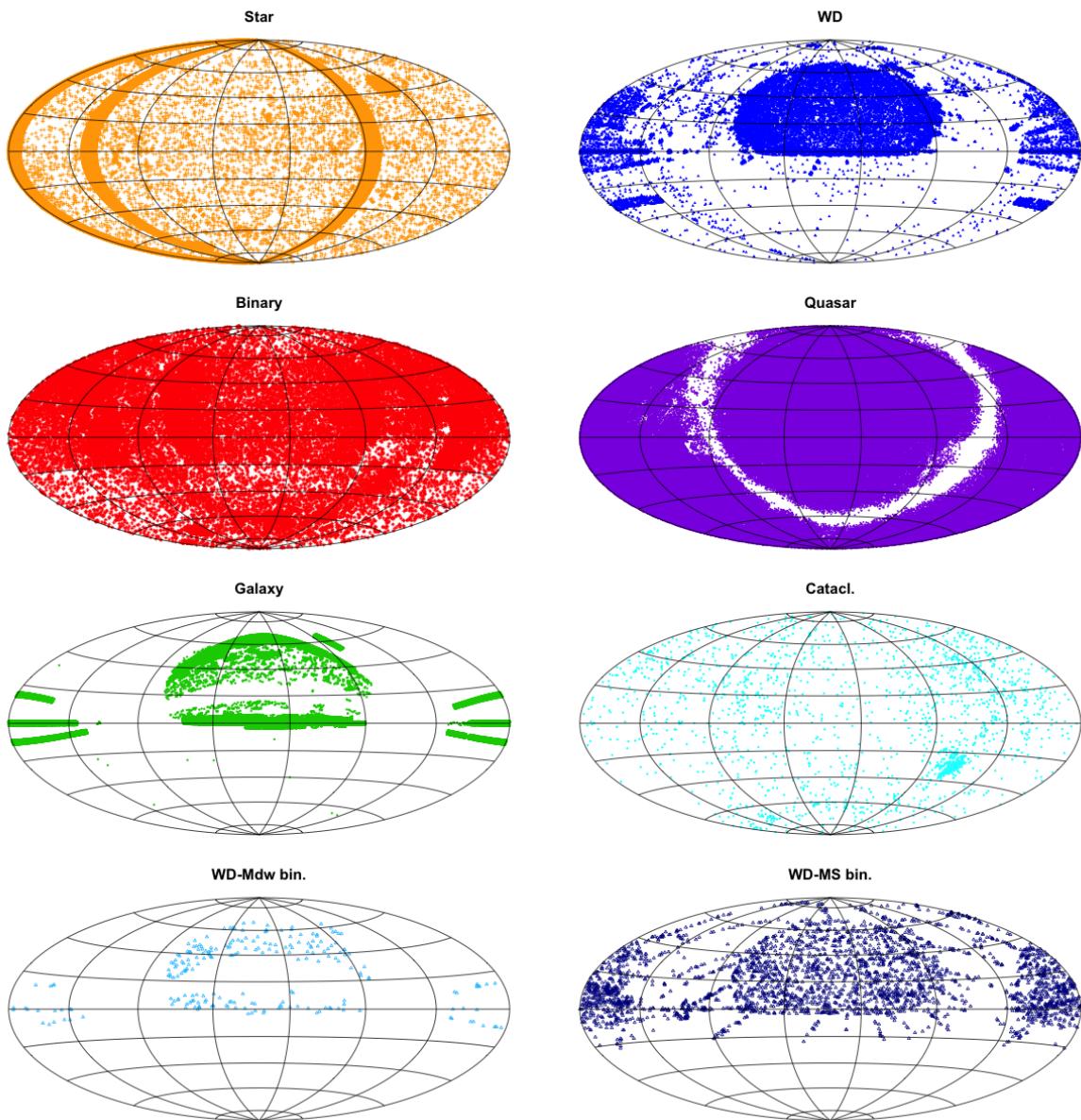


FIGURE 2: Sky maps of the sources by class.

## 2.3 Apparent magnitudes

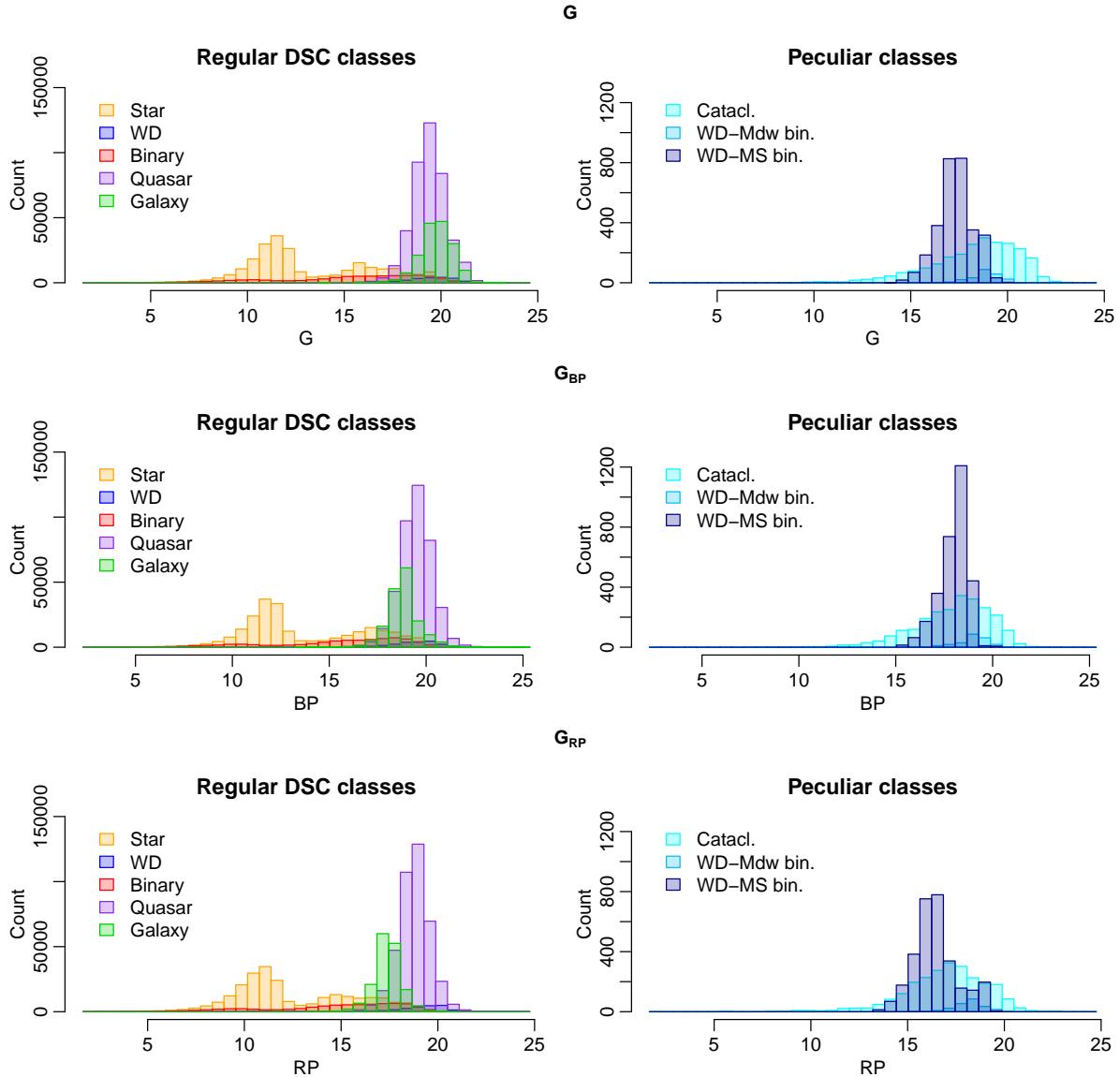


FIGURE 3: Distribution of the magnitude  $G$  (top row),  $G_{BP}$  (middle row) and  $G_{RP}$  (bottom row). The left column shows the regular DSC classes, the right column, those with multiple true labels.

## 2.4 Proper motions

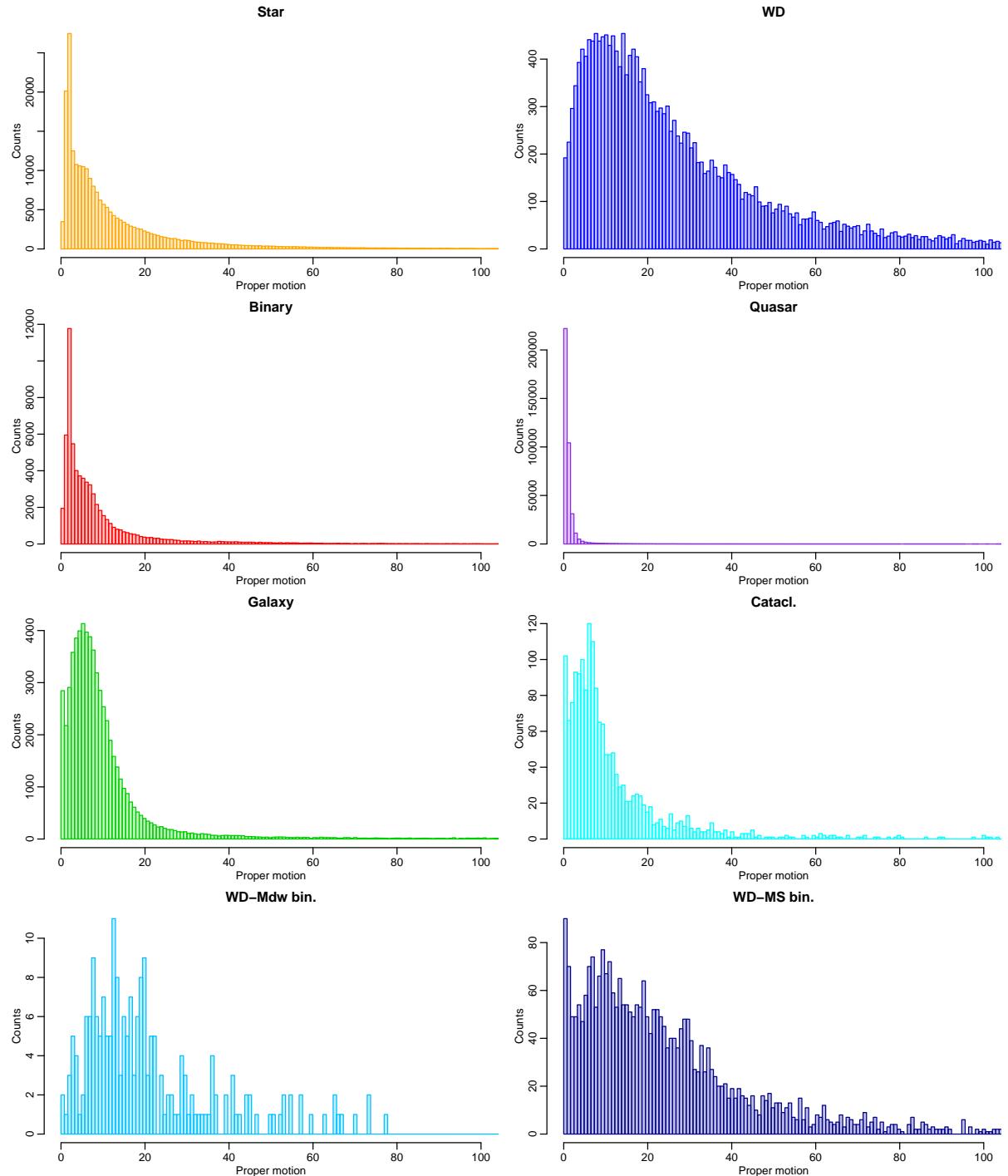


FIGURE 4: Class-wise distribution of the proper motions in the validation set.

Figure 4 shows the distribution of the proper motions  $\sqrt{\mu_\alpha^2 \cos^2 \delta + \mu_\delta^2}$  of the validation objects by class.

- The distribution of the stars and binaries, as expected, are very similar. However, both appears to be composed of (at least) two unimodal populations, one narrowly concentrated at around 3 mas/year and another one, more spread-out, around 7 mas/year. These roughly correspond to two important sources of validation objects, the former to the variable star catalog of D. W. Kim () containing Magellanic Cloud objects, the latter to the broad general sample from Simbad.
- White dwarfs (and the related classes WD–M dwarf and WD–main sequence star binaries) have on average high proper motion and a broad proper motion distribution, due to the fact that these mostly faint objects are observable only relatively close to Earth.
- The narrower distribution of the cataclysmic variables might be due to an observing bias: their luminous state makes them detectable at larger distances.
- Quasars have on average low proper motion (distribution peaking below 2 mas/year).
- Galaxies, although in theory they should have a distribution similar to quasars, have on average higher proper motion and a broader distribution than quasars. Their extended morphology causes difficulties in their astrometric processing.
- In the distribution of galaxies, cataclysmic variables and WD–main sequence star binaries, there is an excess of objects with near-zero proper motion. In majority, they seem to have exactly 0 values in the DR2 tables.

## 2.5 Colour-magnitude diagrams

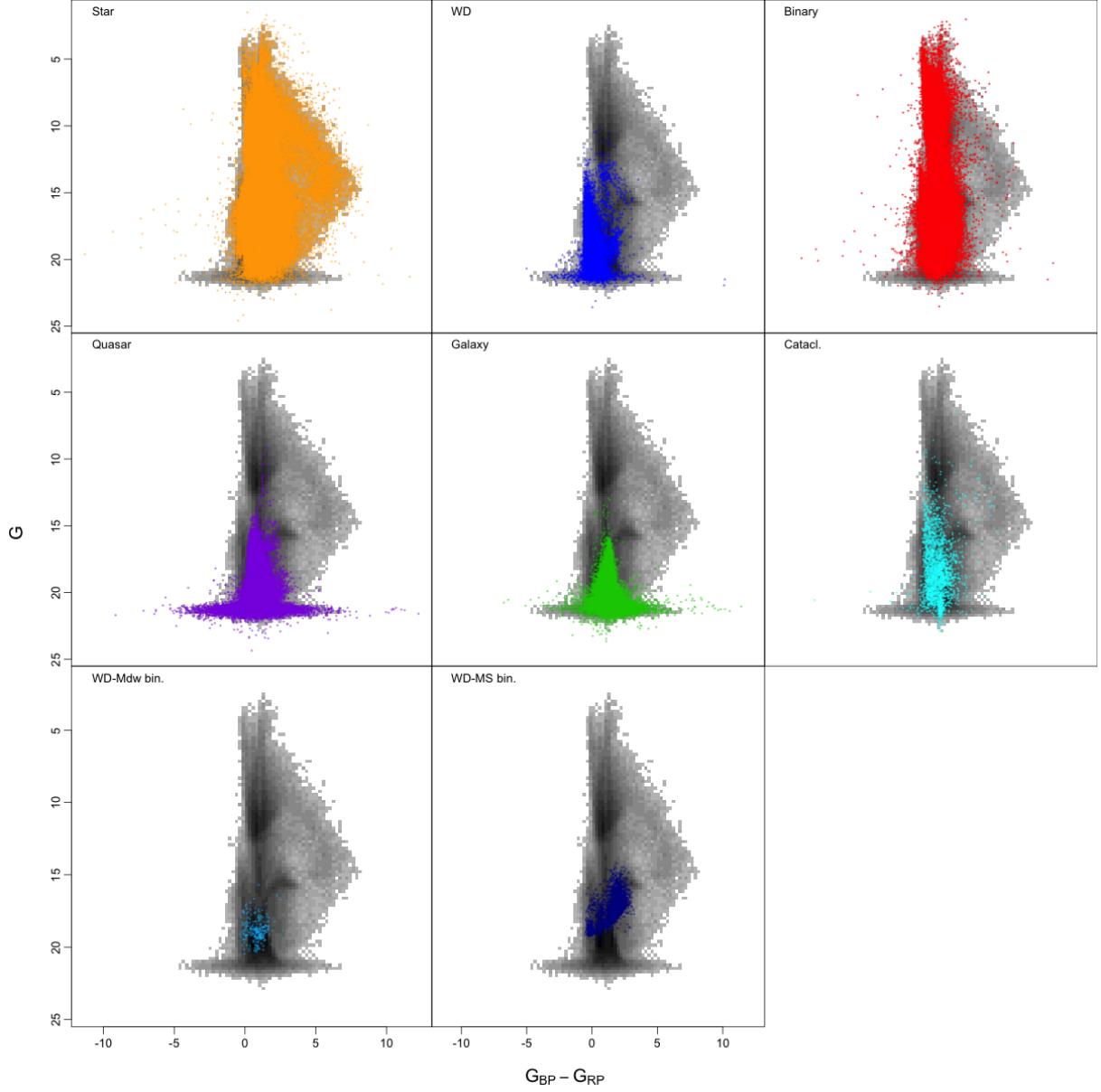


FIGURE 5: The class-wise colour-magnitude diagram of the validation data set, using the  $G$  magnitude versus the  $G_{BP} - G_{RP}$  colour. The grey background 2D histogram shows the complete validation data set where the greyscale corresponds to counts on logarithmic scale. The classes are highlighted in colour.

Figure 5 shows the validation set on the  $(G_{BP} - G_{RP}, G)$  colour-magnitude diagram. This, too, reflects the choice of catalogs used for validation set construction.

The extended horizontal stripe at  $G$  magnitudes fainter than 20.5 is in majority composed of 39711 quasars (1684 from AllWise and 38027 from SDSS DR14), and of 37113 galaxies from SDSS.

Figure 5 also shows predominantly class 0 (single star) objects with colours redder than  $G_{BP} - G_{RP} = 5$ . Other types of objects are rare with this colour. Upon examination, the single stars in this region turn out to come from the general selection from Simbad (from the two stripes visible in the top left panel of Fig. 2). About two-third of them have no detailed type information there (other than they are stellar objects), some have detected variability without further specification, and some are candidate or confirmed long-period variables. We find only very few binaries with such colours, although all binary systems with no contradictory classifications in Simbad were included in the validation set.

### 3 DSC performance on “regular” classes

This section presents the results of DSC on those objects that can be labeled unambiguously as one of DSC’s standard classes: star, white dwarf, physical binary, quasar or galaxy.

#### 3.1 ExtraTree classifier based on BP-RP spectra

The Extremely Randomised Trees algorithm (ET or ExtraTrees), based on the middle wavelength ranges of the mean BP and RP spectra, predicts class labels using an ensemble of 1000 random trees. The 24 edge pixels from both ends of both the BP and RP spectrum were discarded, and then the remaining twice 72 pixels were normalised to have unit sum. No other information has been used in this classifier than this normalised spectrum. Neither integrated magnitudes, nor parallax or proper motion have been used; two other classifiers are based on these features, whose output can hence be regarded as independent from that of the photometry-based ET.

##### 3.1.1 Confusion matrix

The performance of the ET is shown in Fig. 6. The misclassification rate is the lowest in the galaxy class (about 4%), and 10% among quasars, which implies that DSC classification output can be used to select these two object types by downstream modules. Further aid to select samples of galaxies and quasars with the desired purity/contamination balance is given in Section 3.1.8.

There is an extensive confusion between physical binaries and single stars. This is at least partly as expected, since it is in part due to a genuine similarity between some binaries

	Star	WD	Binary	Quasar	Galaxy	Error	NA
Star	98118	2850	127923	8820	3855	<b>0.59</b>	5440
WD	234	16543	1515	2639	204	<b>0.22</b>	96
Binary	5783	1790	55546	4169	3844	<b>0.22</b>	2794
Quasar	1876	6961	20270	365514	12367	<b>0.1</b>	1171
Galaxy	354	120	2668	2695	159080	<b>0.04</b>	174
Contam.	<b>0.08</b>	<b>0.41</b>	<b>0.73</b>	<b>0.05</b>	<b>0.11</b>		

FIGURE 6: The confusion matrix, contamination and error rates and account of NAs from the ExtraTree classifier based on BP/RP spectra. Rows correspond to true class, columns to predicted class. In the upper left  $5 \times 5$  block, as well as in the account of NAs, counts are given. The column corresponding to errors and the row corresponding to contamination (both orange-shaded according to quality) shows fractions.

and the single stars, which limits the potential of DSC to distinguish between single and multiple star systems. The correct identification is possible for those binaries that consist of two sufficiently different spectral types with not too great a difference in luminosity (so that the shape of the BP and RP spectra of the system is noticeably different from that of a single star), or where the spectral shapes of the components are similar, but the system is detectably brighter than would correspond to its spectral type. Only the former is feasible for the present-state ET classifier, since the latter would require the inclusion of a good estimation of the absolute luminosity of the system, which in turn needs reliable parallaxes for all objects (including all object types, and faint or extended objects too).

Further work is necessary to investigate precisely what subset of binaries can be reliably distinguished based on the shape of their BP/RP spectra. For this, I see two principal ways: (i) to simulate binaries; (ii) to use the present results to select those binaries that have been successfully classified as binaries, and from these, create iteratively a self-consistent subset that is reliably distinguished from single stars. The present validation set contains a very broad variety of binaries, ranging from spectroscopic or double-lined binaries to eclipsing binaries with components of very different luminosity, but most often without sufficient individual information of the components. A manual selection of a large enough sample of binaries with well-known physical parameters requires a lot of time, and does not seem to be the most fruitful way to go.

A contribution to the low performance of ET on binaries can come from the fact that training labels, supposed to be known, can easily be erroneous. For instance, the general star sample from Simbad can contain many objects that have not been detected as bina-

ries simply because not yet investigated spectroscopically. This, while obviously affecting the classification quality through the incorrect training labels, raises the interesting question: could some of these assumed stars, apparently misclassified as binaries, be in truth undetected binaries?

### 3.1.2 Classification probability vectors

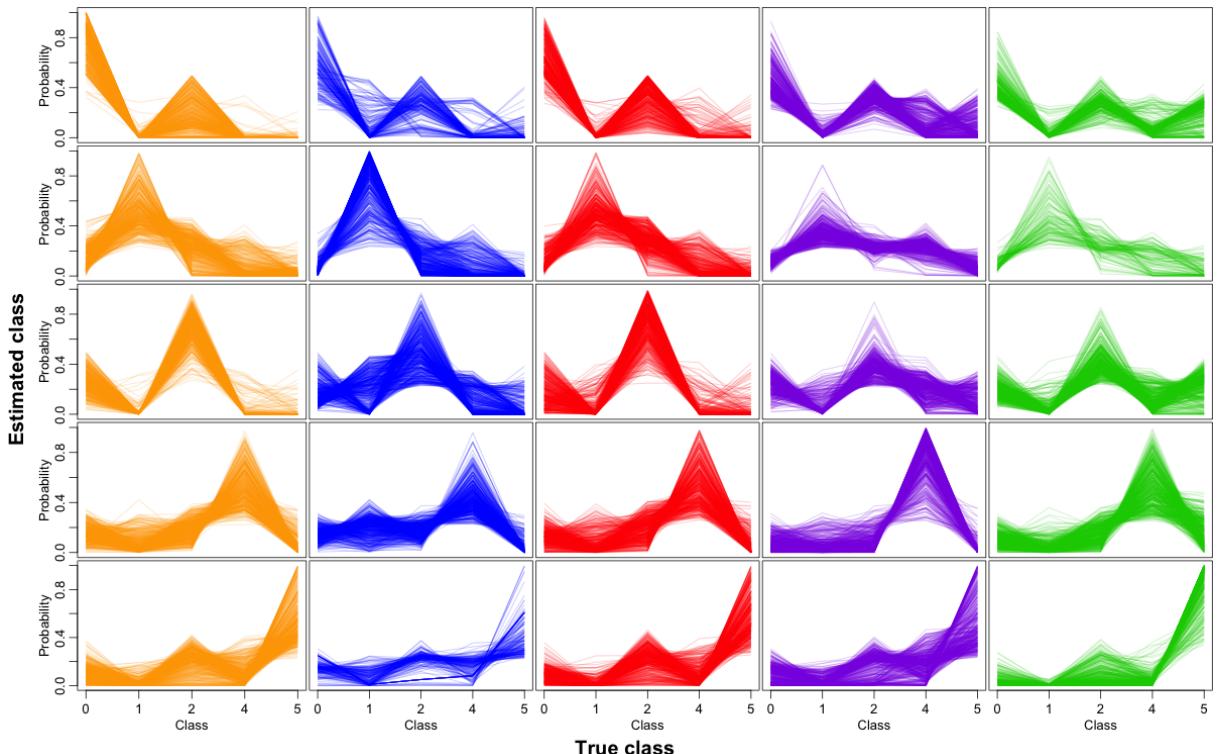


FIGURE 7: The probability vectors produced by the ET classifier, for a sample of at most 500 objects of each combination of true and estimated class. The true class is colour-coded (the same as in Figure 2), and is arranged by column (leftmost column: stars, second from left: white dwarfs, middle: binaries, fourth: quasars, rightmost: galaxies). Rows correspond to estimated classes (top: predicted star, second from top: predicted white dwarf, middle: predicted binary, fourth: predicted quasar, fifth: predicted galaxy). Correctly classified objects thus are shown in the diagonal.

Figure 7 shows the output probability vectors, arranged in a grid of rows corresponding to true classes, columns to predicted classes.

Its left and middle panels in the top and middle rows suggest a similarity between the probability vectors of single and binary stars. Objects in these panels, namely, binaries and stars that were classified either to be stars or binaries (not necessarily correctly),

most often have a general shape with two high probability values at the star and the binary class and near-zero values everywhere else. This seems to confirm that there is a continuous transition between systems detectably binary and both single stars and undetectable binaries, but all these objects can be quite reliably distinguished from white dwarfs, quasars or galaxies.

Correctly classified stars (top left panel) appear to have higher probability values for the class star than quasars or galaxies misclassified as stars (top row, two rightmost panels). Similarly, the probability of being a quasar for a real quasar is higher than the probability of being a quasar for a star misclassified as quasar. The same is true for white dwarfs or galaxies. In other words, generally, the probability of the modal class is on average higher when the modal class is the correct one than when the modal class is wrong, at least for stars, white dwarfs, quasars and galaxies. This implies that by a cut put on classification probability, we can remove part of the contamination from these predicted classes. This appears to be much less so for binaries: for both true binaries and stars misclassified as binaries, the probability for class binary seems to be high, though at the highest levels, there is some difference. This emphasizes further the difficulties of discriminating binaries and single stars. Section 3.1.8 will provide a further look into this.

### 3.1.3 Class probabilities vs. $G$

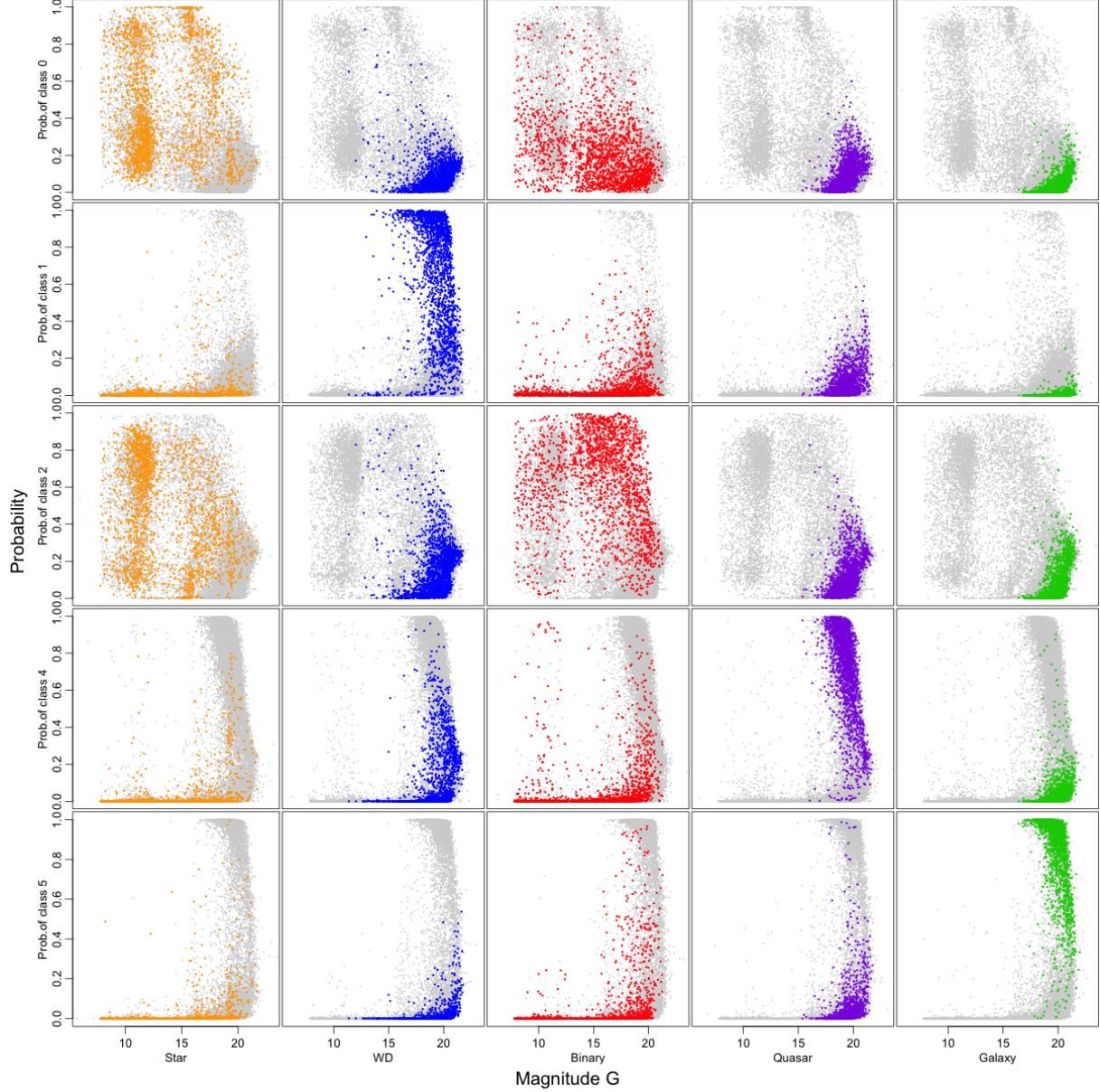


FIGURE 8: Probabilities produced by the ExtraTree classifier based on BP/RP spectra, for a sample of at most 2000 objects of each (true) class, as a function of apparent  $G$  magnitude. Each column contains results for a (true) DSC class (leftmost: stars, second: white dwarfs, middle: physical binaries, fourth: quasars, fifth: galaxies). The true class is also indicated by colour, according to the colour code used throughout this document. The uppermost row shows the star probability by ET, the second row from top, white dwarf probability, the third, physical binary probability, the fourth, quasar probability, and the bottom row, galaxy probability. The grey dots show the full test set, for comparison.

Figure 8 confirms that for stars, white dwarfs, galaxies and quasars, class probabilities are the highest when correct: in panels in the diagonal, which show the probability values for the correct class, the coloured dots are at much higher levels than in the other panels showing the probabilities for incorrect classes. The exception is the probabilities for class binary when stars are misidentified as binaries: in the leftmost panel of the middle row, showing this case, we find a cloud of bright stars at high binary probability (left upper quadrant of the panel). These objects represent the majority of the contamination in the estimated binary class, although other stars (fainter and with lower modal probability) also contribute.

Bright objects, which are almost exclusively stars and binaries (galactic sources), appear to have high probabilities only on the classes binary and star. It appears the classifier learnt the simple rule that “if it is bright, then it is a star or a binary”, even though there is no direct information about the absolute brightness of the object among the inputs to the ET classifier. This information can come through indirectly: brighter objects have smoother spectra with higher SNR, which can be detectable by the classifier. Distinction between stars and binaries among these bright objects is, however, not good, as mentioned above and illustrated in the leftmost panel of the middle row of Fig. 8.

### 3.1.4 Class probabilities vs. $G_{BP} - G_{RP}$

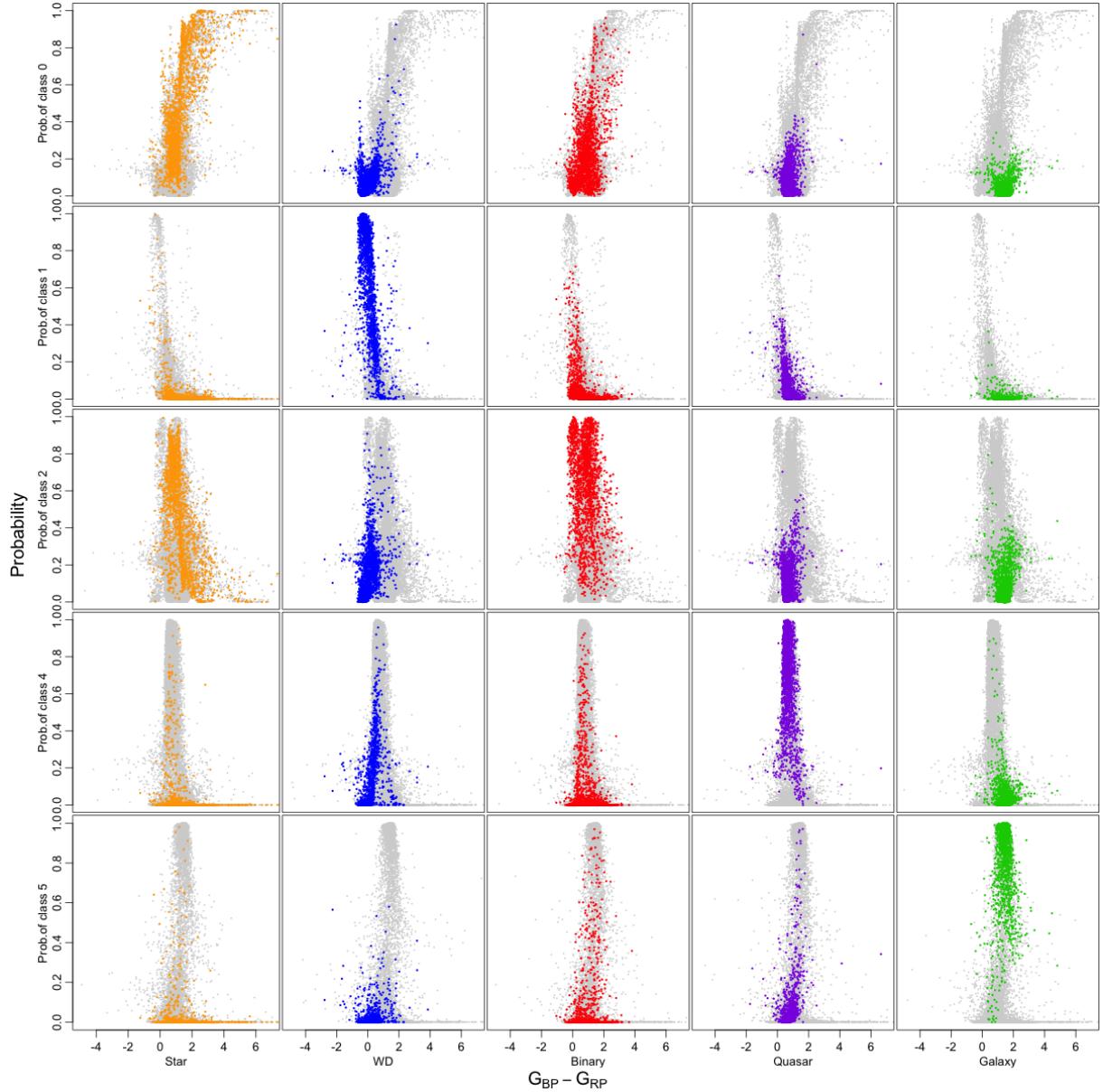


FIGURE 9: Probabilities produced by the ExtraTree classifier based on BP/RP spectra, for a sample of at most 2000 objects of each (true) class, as a function of  $G_{BP} - G_{RP}$ . Each column contains results for a regular class (leftmost: stars, second: white dwarfs, middle: physical binaries, fourth: quasars, fifth: galaxies). The true class is also colour-coded. The uppermost row shows the ETS probability for class star for the objects; the second row from top, for class white dwarf; the third, for class physical binary; the fourth, for class quasar; and the bottom row, for class galaxy. The grey dots show the full test set, for comparison.

The colour dependence of the classification probabilities is shown in Fig. 9. As was concluded from Fig. 5 in Section 2.5, the reddest objects in the DSC validation set are all stars. Fig. 9 suggests that the classifier learnt this (see its leftmost column), which implies that when applied to real data possibly containing red binaries, these might be missed using this model.

For white dwarfs, it is interesting to observe that the probabilities of the (correct) white dwarf class are increasing with bluer colour (second diagonal panel from top), which makes sense: the bluer the object, the more likely it is to be a white dwarf according to the classifier. Still among the white dwarfs (second row from top), the probabilities for other (wrong) classes behave the opposite way: the redder the white dwarf, the higher the probability is for an incorrect classification, suggesting that the ET classifier can learn to use correctly important features in the data.

### 3.1.5 Class probabilities vs. astrometry

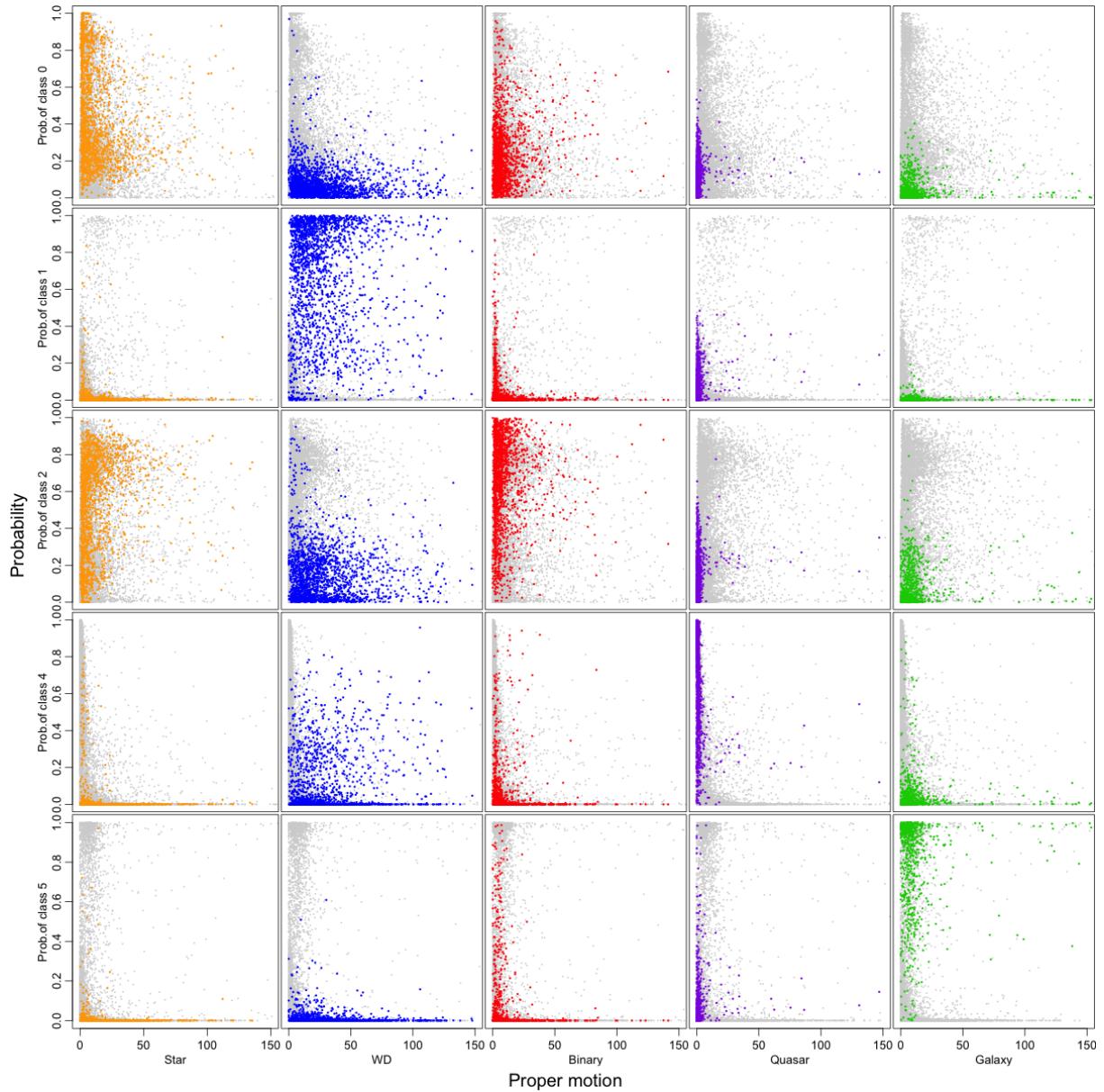


FIGURE 10: Probabilities produced by the ExtraTree classifier based on BP/RP spectra, for a sample of at most 2000 objects of each (true) class, as a function of proper motion. Each column contains results for a regular class (leftmost: stars, second: white dwarfs, middle: physical binaries, fourth: quasars, fifth: galaxies). The true class is also colour-coded. The uppermost row shows the ET probability for class star for the objects, the second row from top, for class white dwarf, the third, for class physical binary, the fourth, for class quasar, and the bottom row, for class galaxy. The grey dots show a mixture of all classes, for comparison.

Figure 10 shows the probability values on the different classes as a function of proper motion. Apart from reflecting the fact that galactic objects can have on average the highest proper motions, galaxies stay at lower proper motion values, and the pointlike quasars have the smallest values, no interesting further structure can be discerned in the plots. We expect the photometry-based results to be largely independent from proper motions. All clustering in these plots would indicate conditional dependence, and possibly a training set bias: the dominance of a group of specific objects with specific proper motions in the training set.

### 3.1.6 Colour-magnitude diagrams of predicted classes

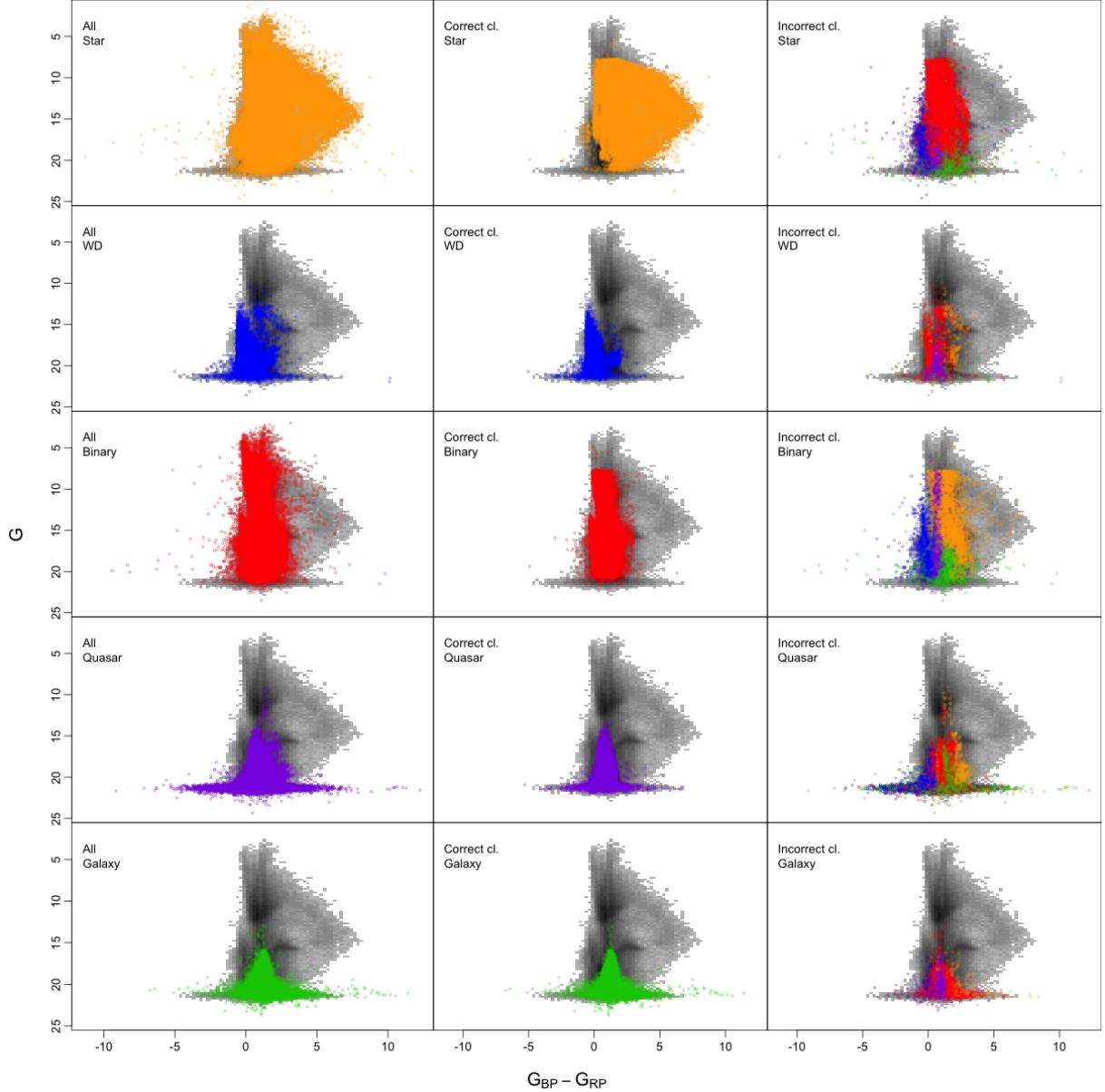


FIGURE 11: The class-wise ( $G, G_{BP} - G_{RP}$ ) colour-magnitude diagrams of the regular DSC classes in the validation data set, as a function of classification quality. The different regular classes of DSC occupy different rows (stars: top, white dwarfs: second from top, binaries: middle row, quasars: fourth row, galaxies: bottom row). The grey background 2D histogram shows the complete validation data set and is the same on all panels. The left column shows all objects from the class, regardless of estimated label. Middle column: correctly classified objects from the class. The colours in these two columns indicate the true class. Right column: misclassified objects from the class, the colour here showing the (mis-)estimated class label from the ET classifier (orange: star, blue: white dwarf, red: binary, purple: quasar, green: galaxy).

The colour-magnitude diagrams, shown in Fig. 11, also suggest a reasonable behaviour of the ET classifier. The distribution of the correctly labelled objects (middle column) follow well that of the true labels (although colour and magnitude was not directly used in the classifier). The mislabelled cases are in majority on the outskirts of the distribution of their class, such as, for example, the binaries misclassified as stars (third row, rightmost panel, orange dots): they are on the red edge of the binary distribution, where stars dominate in the validation set. The pattern of misclassifications is similar in every class. The position of objects getting a wrong label occupy a position which is compatible with their wrongly assigned class (again, despite that colour and magnitude were not directly used in the classifier).

### 3.1.7 Sky maps of correct/incorrect classifications

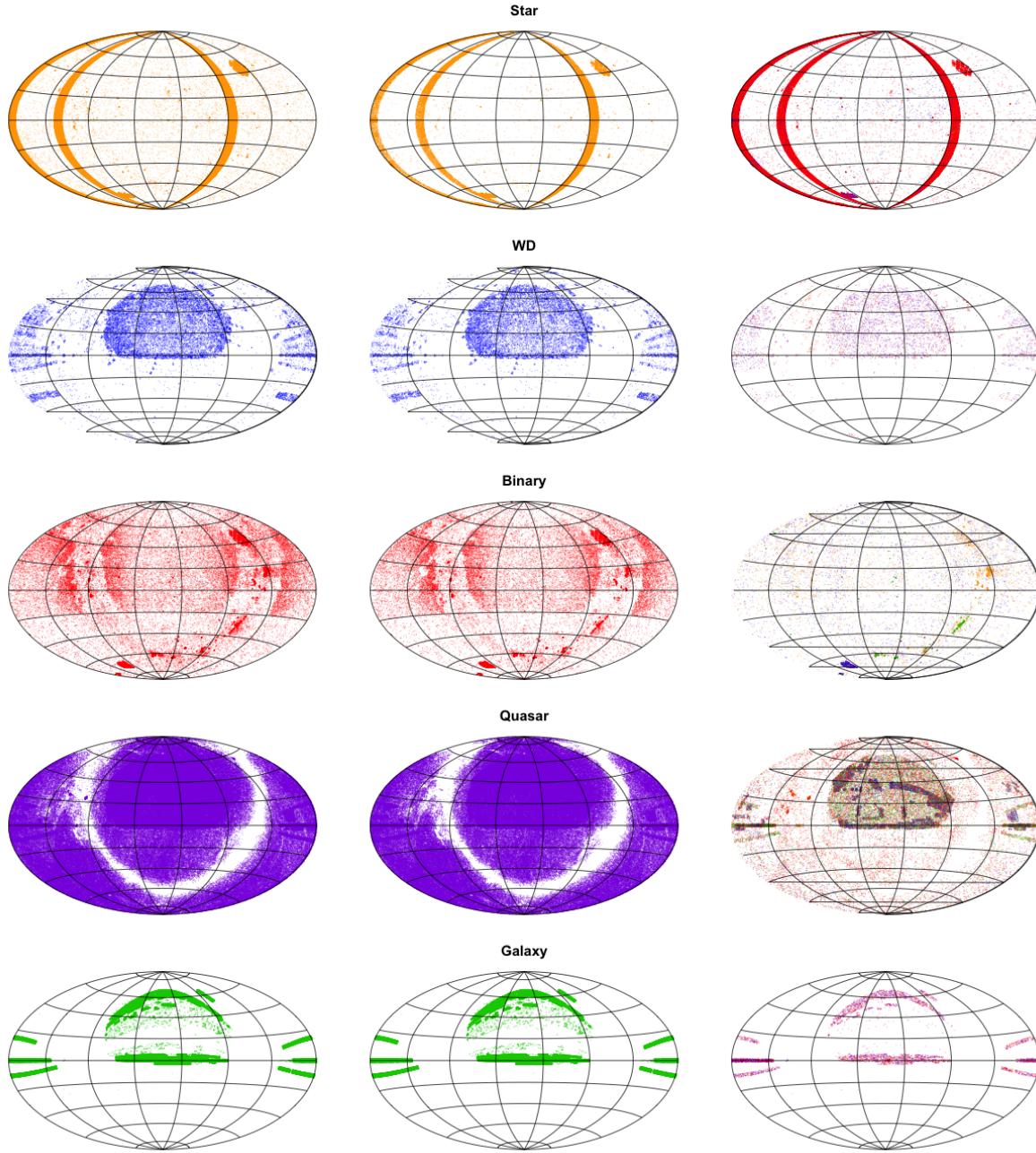


FIGURE 12: Sky maps of the sources by true class (left column, from top: stars, white dwarfs, binaries, quasars, galaxies), the correctly classified members of the class (middle column), and the incorrectly classified members of the class (right column). In the left and middle columns, colours indicate the true class. In the right column, the colours indicate which class was (mistakenly) given to the objects by the ET (orange: predicted star, blue: predicted white dwarf, red: predicted binary, purple: predicted quasar, green: predicted galaxy).

Figure 12 shows the sky distribution of all validation sources by true class, the distribution of the correctly classified cases and that of the incorrectly classified ones. No particular dependence on position can be discerned.

### 3.1.8 Purity and completeness

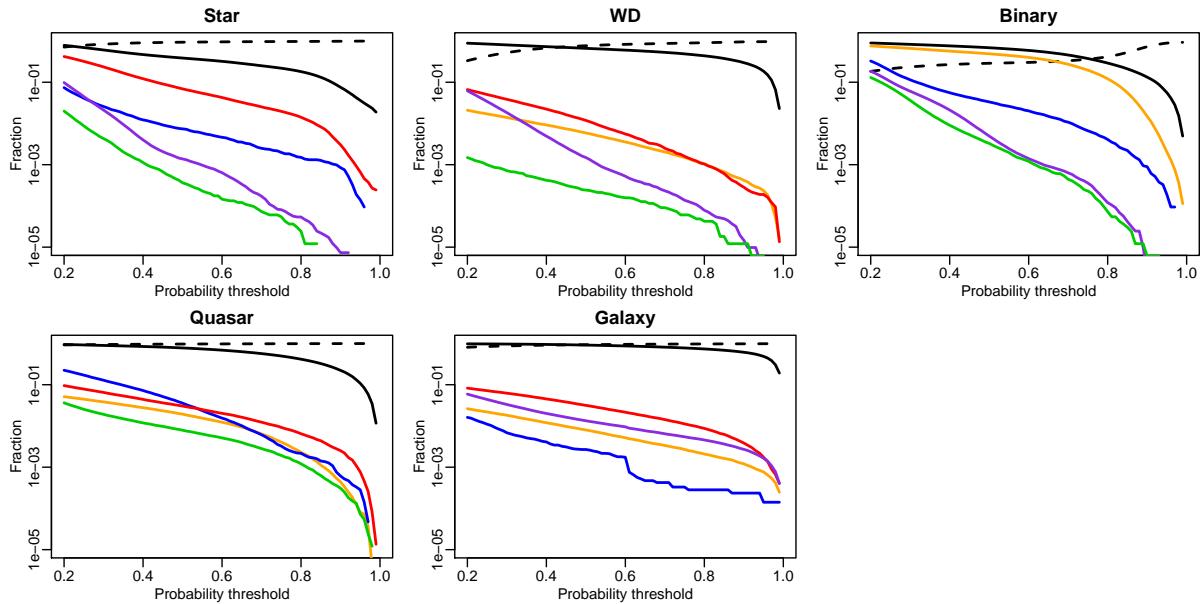


FIGURE 13: Characterisation of the purity, the completeness and the contamination in samples based on the ET classification results. The panels present the content of selected samples from each of the five predicted DSC classes, as a function of probability threshold. In panel showing selection from (say) class  $C$ , the solid black line shows what fraction of the *true* class  $C$  is correctly selected depending on the probability threshold. The broken black line shows the fraction of true positives in the *selected* class  $C$ . The coloured lines show what fraction of the *other true* classes gets mistakenly selected into class  $C$ ; the colours indicate the true class of the contamination. Note the logarithmic scale of the  $y$  axis.

Selection of certain types of objects (galaxies, quasars, binaries for downstream modules) can be required to satisfy certain completeness or purity requirements. These requirements can sometimes be opposing for different modules using the same object type. A custom-tailored selection can be most easily achieved by setting a probability threshold which must be exceeded by the classification probability for an object to be selected. A low threshold will result in a near-complete sample, which has a large fraction of contaminating objects (false negatives), while a high threshold produces a purer sample, which nevertheless loses a large fraction of the true objects.

Figure 13 provides basic information aiming to aid the choice of a threshold. It shows the fractions of true positives, correct identifications and rates of contaminating objects as a function of selection threshold. Taking binaries as an example, the solid black curve shows the selected fraction of *true* binaries (that is,  $N_{selected}/N_{truebinaries}$ ) versus the selection threshold. Using a threshold of 0.4 (that is,  $p_{binary} > 0.4$  in order to get into the sample), a relatively large fraction of the real binaries will be identified (approximately 80%). With a threshold of 0.9, this drops to barely more than 10% of detected real binaries.

Checking the broken black line, we can estimate the fraction of true positives in the samples, although this estimate directly depends on the composition of the DSC test set, and cannot be used to assess contamination fraction for a real, unknown Gaia sample. Using the low threshold, only a minority of the selected sample, 15% will be true binary, although these indeed represent about 80% of all true binaries. This sample is huge, and the price to pay for a relatively complete binary sample is a very large contamination. Using the higher threshold, the purity of the sample increases to about 90%, but at the same time, completeness decreases to 10%.

To get a more objective idea about the contamination we can expect in our sample with a specific threshold, it is better to consult the coloured lines. These show what fraction of other *true* classes make it into the selected sample using a given threshold. Continuing with the example of binary sample selection, we can read off the orange line (which shows the fraction of *true* stars that gets selected as a function of threshold) that the threshold  $p_{binary} > 0.4$  selects nearly as high a fraction of stars as that of binaries. This means that supposing 700 million true single and 700 million true binary star systems in a new Gaia data set, such a threshold would provide a sample containing half single stars, half binaries. Using the threshold of 0.9 on the probability of the binary class, binaries dominate the selected sample (about ten times more binaries than single stars).

The bottom middle panel suggests that galaxies can be selected most advantageously: completeness remains high in a wide range of thresholds, and the fraction of other classes contaminating the galaxy sample is relatively low. However, considering that galactic objects are more numerous than galaxies by a factor of about 1000, this still means more contamination than true galaxies in a sample aimed at completeness. For quasars, completeness decreases with increasing threshold faster than for galaxies, so for a complete sample, a high fraction of contamination is still expected.

## 3.2 Astrometric classifier

The astrometric classifier provides a vector of class probabilities using a Gaussian mixture model defined in the two-dimensional space of proper motions in  $\alpha \cos \delta$  and  $\delta$ .

### 3.2.1 Confusion matrix

	Star	WD	Binary	Quasar	Galaxy	Error	NA
Star	108216	20335	93299	0	0	<b>0.51</b>	25156
WD	9721	2561	5076	0	0	<b>0.85</b>	3873
Binary	32385	4969	24113	0	0	<b>0.61</b>	12459
Quasar	8255	529	1425	0	0	<b>1</b>	397950
Galaxy	428	113	267	0	0	<b>1</b>	164283
Contam.	<b>0.32</b>	<b>0.91</b>	<b>0.81</b>	NaN	NaN		

FIGURE 14: The confusion matrix of the astrometric classifier. Rows correspond to true class, columns to predicted class. In the upper left  $5 \times 5$  block, as well as in the account of NAs, counts are given. The column corresponding to errors and the row corresponding to contamination (both orange-shaded according to quality) show fractions.

### 3.2.2 Classification probability vectors

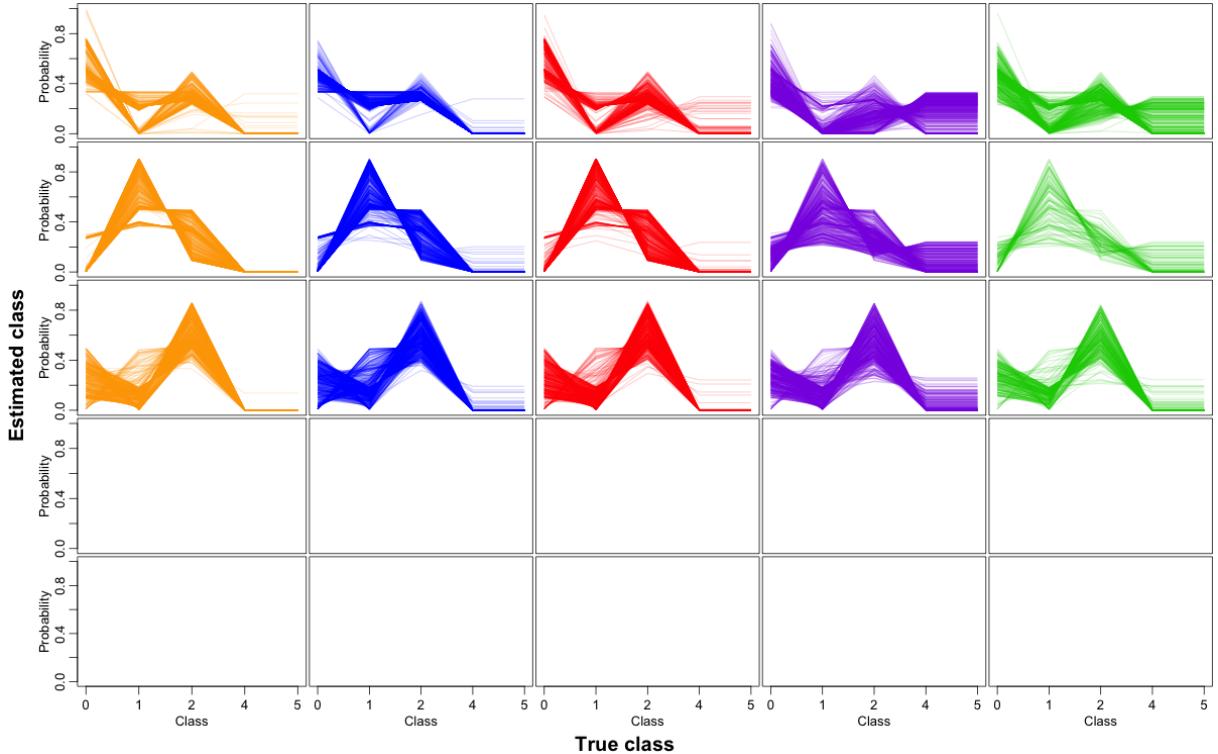


FIGURE 15: The probability vectors produced by the astrometric classifier, for a sample of at most 500 objects of each combination of true and estimated class. The true class is colour-coded (the same as in Figure 2), and are arranged by column (leftmost column: stars, second from left: white dwarfs, middle: binaries, fourth: quasars, rightmost: galaxies). Rows correspond to estimated classes (top: labelled by ET as star, second from top: labelled as white dwarf, middle: labelled as binary, fourth: labelled as quasar, fifth: labelled as galaxy).

### 3.2.3 Class probabilities vs. $G$

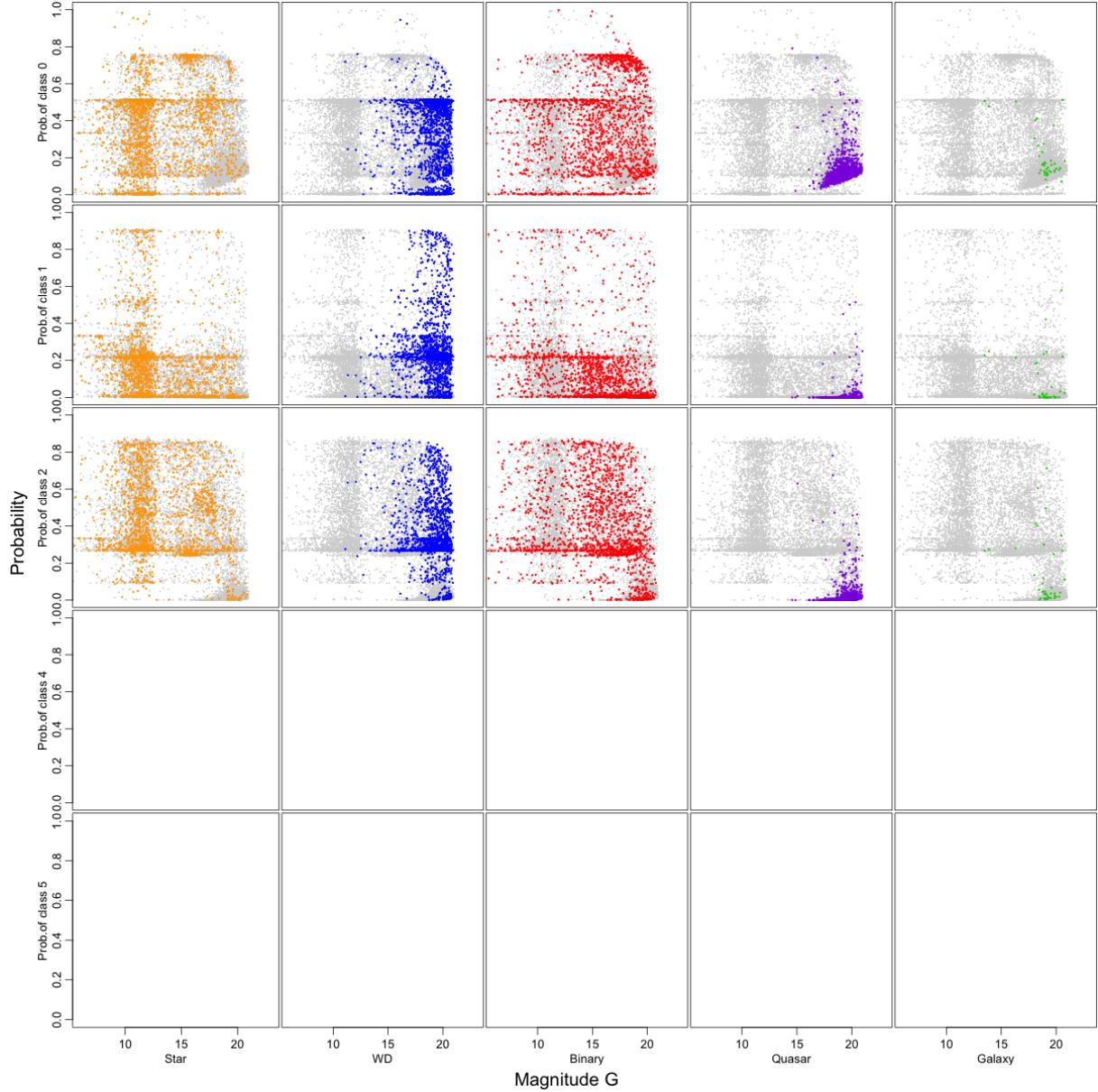


FIGURE 16: Probabilities produced by the astrometric classifier, for a sample of at most 20000 objects of each (true) class, as a function of apparent  $G$  magnitude. Each column contains results for a regular class (leftmost: stars, second: white dwarfs, middle: physical binaries, fourth: quasars, fifth: galaxies). The true class is also colour-coded (the colour key is the same as in Figure 15). The uppermost row shows the astrometric probability for class star for the objects, the second row from top, for class white dwarf, the third, for class physical binary, the fourth, for class quasar, and the bottom row, for class galaxy. The grey dots show a mixture of all classes, for comparison.

### 3.2.4 Class probabilities vs. $G_{BP} - G_{RP}$

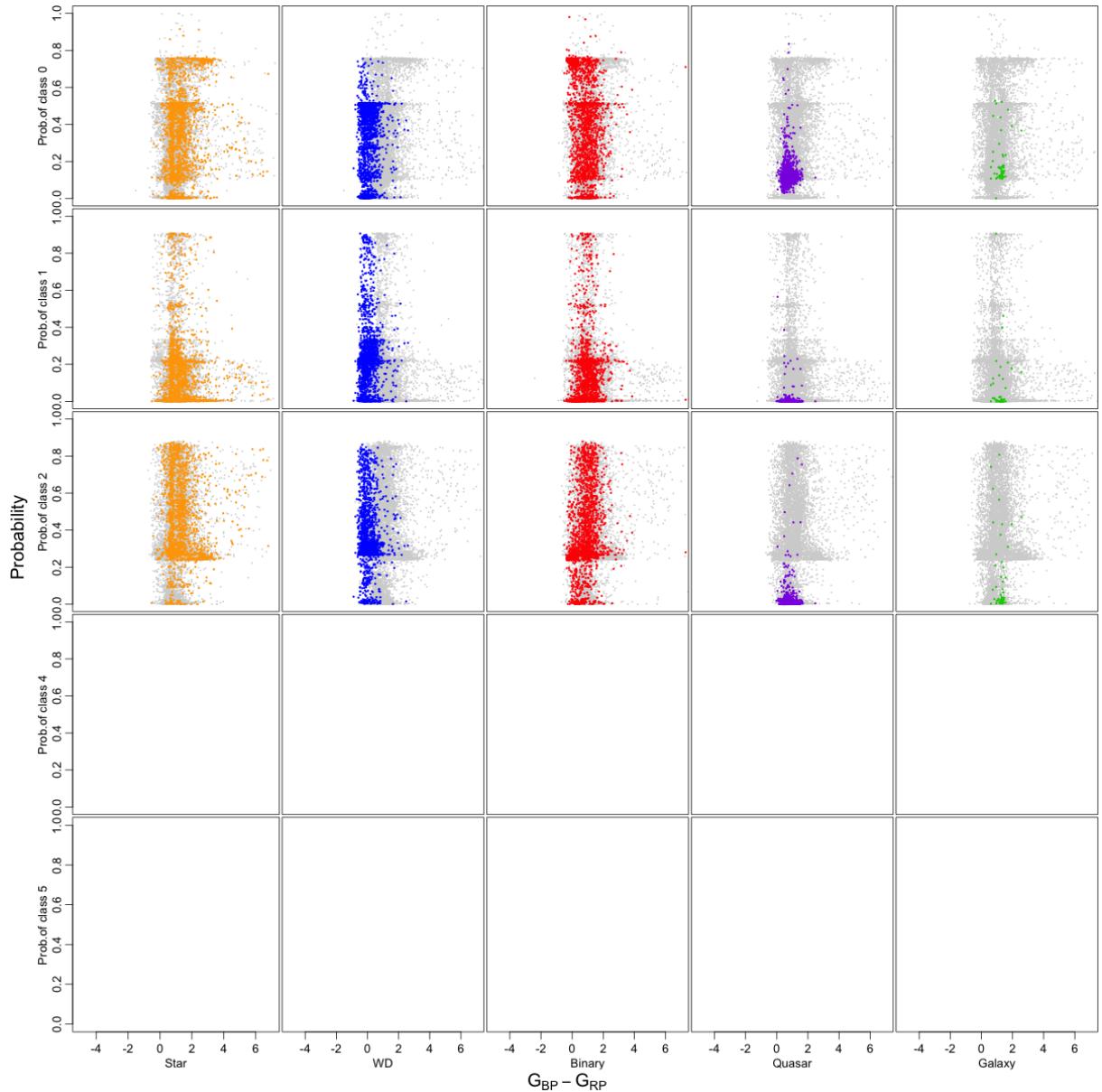


FIGURE 17: Probabilities produced by the astrometric classifier, for a sample of at most 20000 objects of each (true) class, as a function of  $G_{BP} - G_{RP}$ . Each column contains results for a regular class (leftmost: stars, second: white dwarfs, middle: physical binaries, fourth: quasars, fifth: galaxies). The true class is also colour-coded (the colour key is the same as in Figure 15). The uppermost row shows the astrometric probability for class star for the objects, the second row from top, for class white dwarf, the third, for class physical binary, the fourth, for class quasar, and the bottom row, for class galaxy. The grey dots show a mixture of all classes, for comparison.

### 3.2.5 Class probabilities vs. astrometry

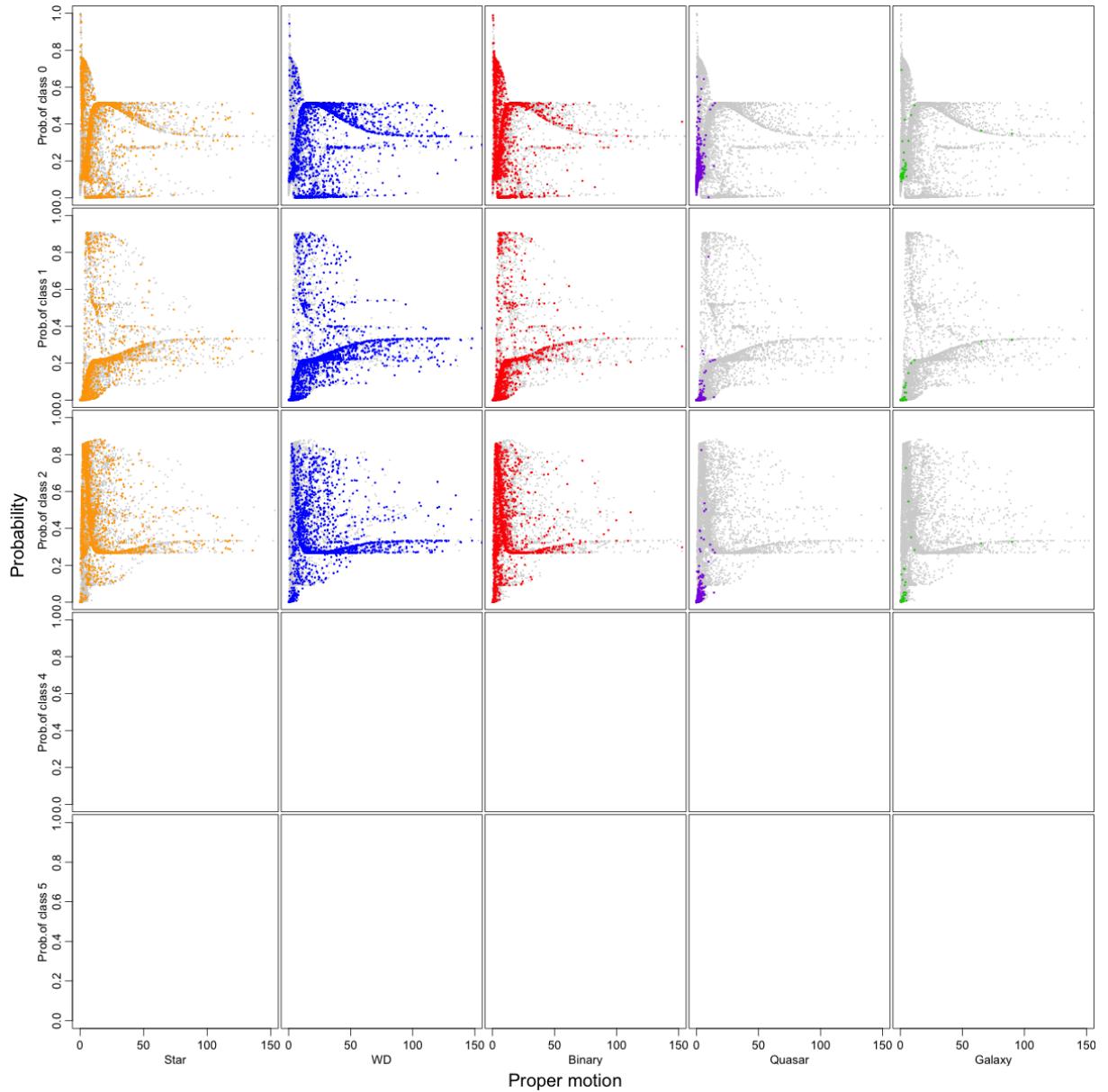


FIGURE 18: Probabilities produced by the astrometric classifier, for a sample of at most 2000 objects of each (true) class, as a function of proper motion. Each column contains results for a regular class (leftmost: stars, second: white dwarfs, middle: physical binaries, fourth: quasars, fifth: galaxies). The true class is also colour-coded (the colour key is the same as in Figure 15). The uppermost row shows the astrometric probability for class star for the objects, the second row from top, for class white dwarf, the third, for class physical binary, the fourth, for class quasar, and the bottom row, for class galaxy. The grey dots show a mixture of all classes, for comparison.

### 3.3 Position–parallax–G-magnitude classifier

This classifier combines the likelihood for various classes based on the G-magnitude and the parallax with the density of the class as a function of galactic latitude.

#### 3.3.1 Confusion matrix

	Star	WD	Binary	Quasar	Galaxy	Error	NA
Star	28525	2897	194580	0	0	<b>0.87</b>	21004
WD	4537	12957	776	0	0	<b>0.29</b>	2961
Binary	17304	664	48097	0	0	<b>0.27</b>	7861
Quasar	245515	4949	121373	0	0	<b>1</b>	36322
Galaxy	3161	245	421	0	0	<b>1</b>	161264
Contam.	<b>0.9</b>	<b>0.4</b>	<b>0.87</b>	NaN	NaN		

FIGURE 19: The confusion matrix of the position–parallax–G-magnitude classifier. Rows correspond to true class, columns to predicted class. In the upper left  $5 \times 5$  block, as well as in the account of NAs, counts are given. The column corresponding to errors and the row corresponding to contamination (both orange-shaded according to quality) shows fractions.

### 3.3.2 Classification probability vectors

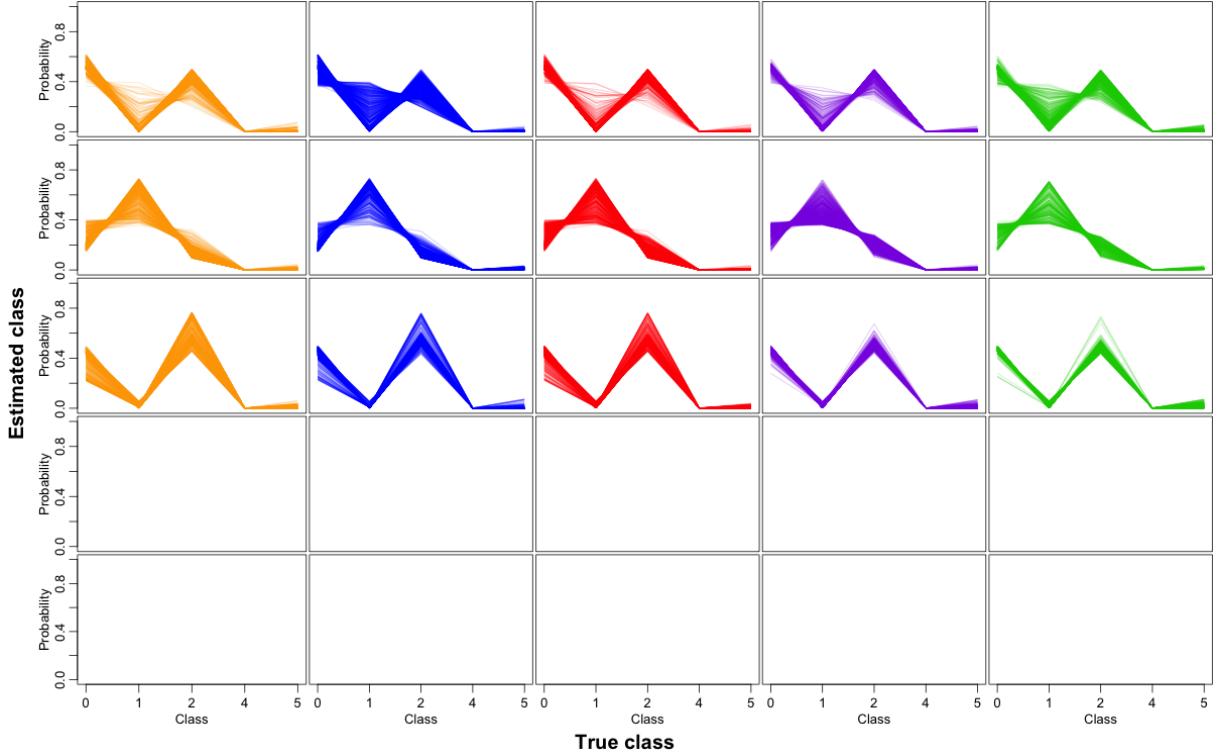


FIGURE 20: The probability vectors produced by the position–parallax–G-magnitude classifier, for a sample of at most 500 objects of each combination of true and estimated class. The true class is colour-coded (the same as in Figure 2), and are arranged by column (leftmost column: stars, second from left: white dwarfs, middle: binaries, fourth: quasars, rightmost: galaxies). Rows correspond to estimated classes (top: labelled by ET as star, second from top: labelled as white dwarf, middle: labelled as binary, fourth: labelled as quasar, fifth: labelled as galaxy).

### 3.3.3 Class probabilities vs. $G$

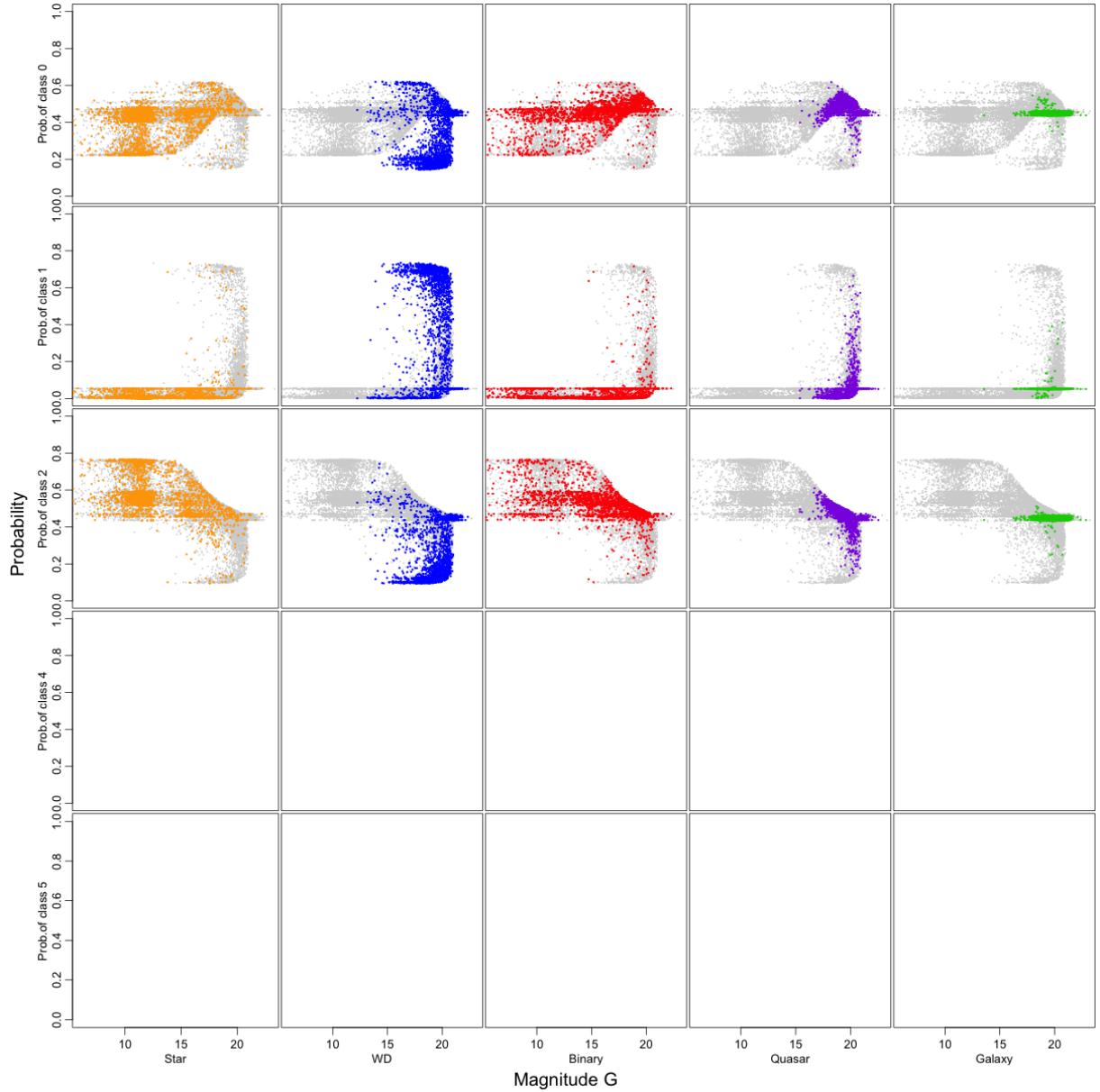


FIGURE 21: Probabilities produced by the position–parallax–G-magnitude classifier, for a sample of at most 20000 objects of each (true) class, as a function of apparent  $G$  magnitude. Each column contains results for a regular class (leftmost: stars, second: white dwarfs, middle: physical binaries, fourth: quasars, fifth: galaxies). The true class is also colour-coded (the colour key is the same as in Figure 20). The uppermost row shows the astrometric probability for class star for the objects, the second row from top, for class white dwarf, the third, for class physical binary, the fourth, for class quasar, and the bottom row, for class galaxy. The grey dots show a mixture of all classes, for comparison.

### 3.3.4 Class probabilities vs. $G_{BP} - G_{RP}$

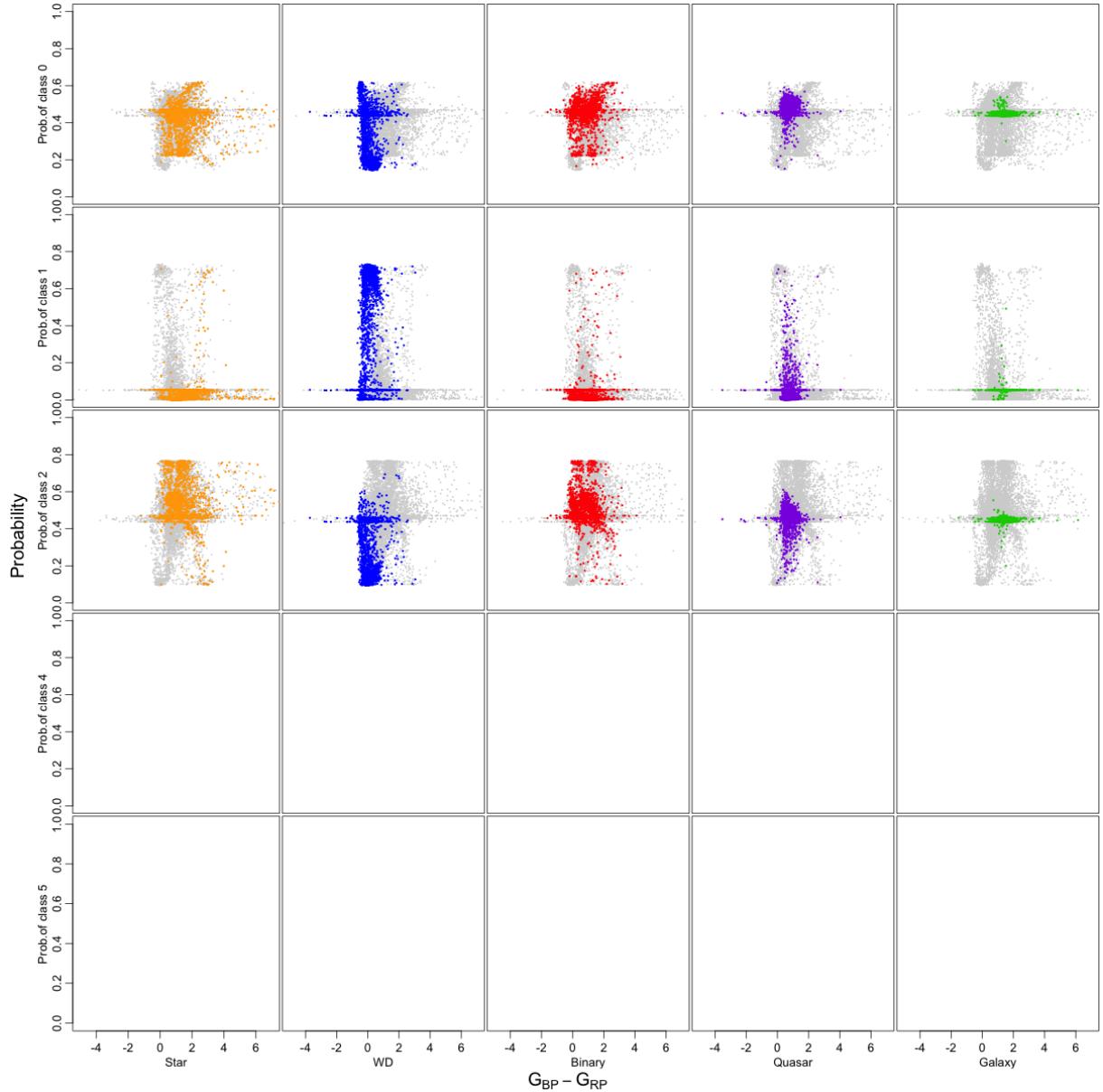


FIGURE 22: Probabilities produced by the position–parallax–G-magnitude classifier, for a sample of at most 20000 objects of each (true) class, as a function of  $G_{BP} - G_{RP}$ . Each column contains results for a regular class (leftmost: stars, second: white dwarfs, middle: physical binaries, fourth: quasars, fifth: galaxies). The true class is also colour-coded (the colour key is the same as in Figure 20). The uppermost row shows the astrometric probability for class star for the objects, the second row from top, for class white dwarf, the third, for class physical binary, the fourth, for class quasar, and the bottom row, for class galaxy. The grey dots show a mixture of all classes, for comparison.

### 3.3.5 Class probabilities vs. astrometry

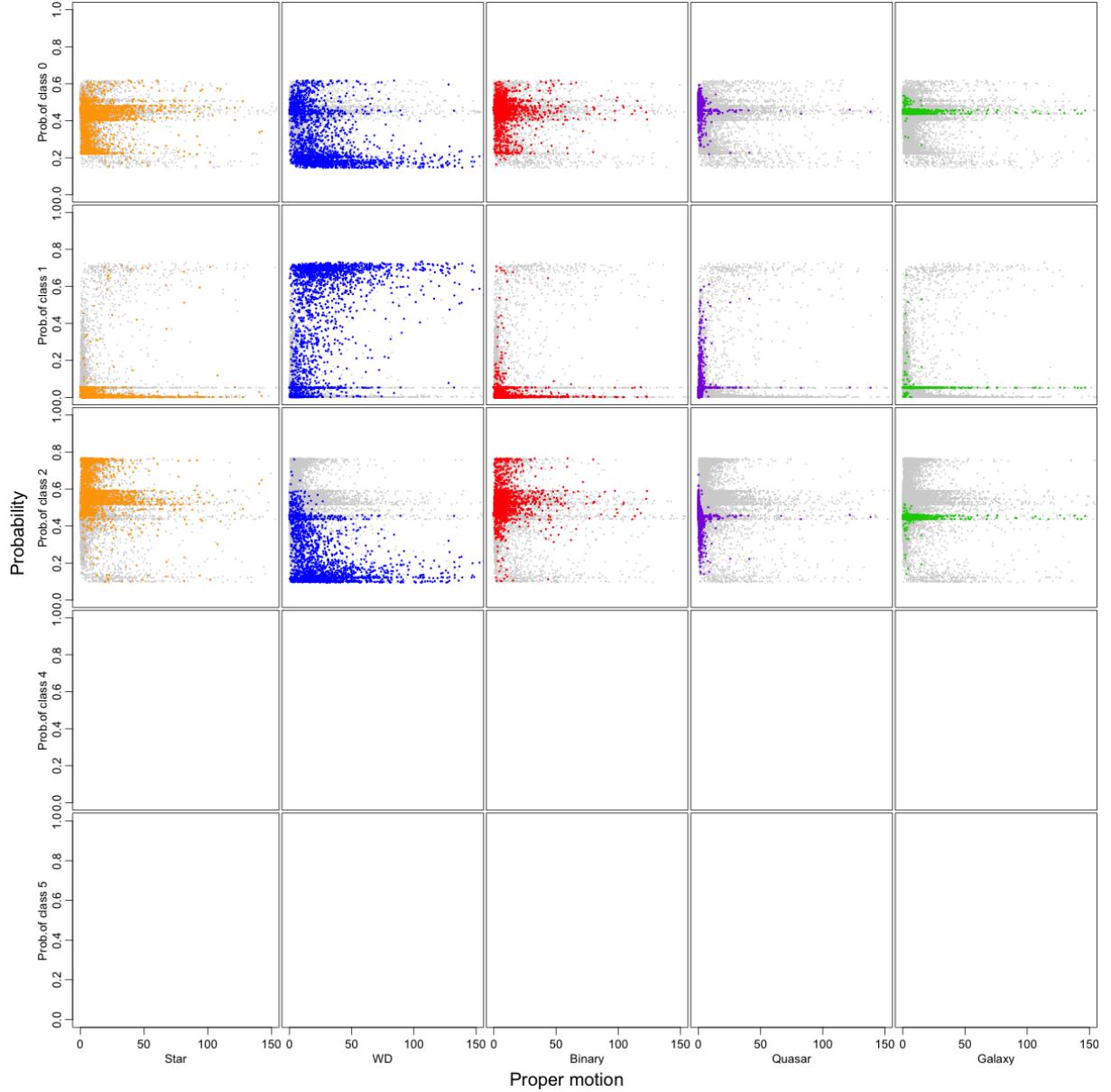


FIGURE 23: Probabilities produced by the position–parallax–G-magnitude classifier, for a sample of at most 20000 objects of each (true) class, as a function of proper motion. Each column contains results for a regular class (leftmost: stars, second: white dwarfs, middle: physical binaries, fourth: quasars, fifth: galaxies). The true class is also colour-coded (the colour key is the same as in Figure 20). The uppermost row shows the astrometric probability for class star for the objects, the second row from top, for class white dwarf, the third, for class physical binary, the fourth, for class quasar, and the bottom row, for class galaxy. The grey dots show a mixture of all classes, for comparison.

### 3.4 Photometric SVM classifier

This classifier applies a family of Support Vector Machine classifiers, each trained in a given magnitude range on a particular set of synthetic spectra, then combines them into one output probability vector.

### 3.5 Bayesian combination

The output probability vectors from the above classifiers (Sections 3.2, 3.3, and 3.4) are combined into a single probability vector, synthetizing prior knowledge about the relative frequency of the classes and the information about its position, magnitude, proper motion and spectrum.