Gaia DR2 and the Washington Double Star Catalog: A Tale of Two Databases

Richard Harshaw
Brilliant Sky Observatory, Cave Creek, AZ
rharshaw51@cox.net

Abstract: A large-data extraction of the Washington Double Star Catalog (WDS) from Gaia's Data Release 2 (DR2) was done in early 2018. This data was read into an Excel spreadsheet and calculations on distance, relative motion through space, R2 fit to trend lines in the data, and relative radial velocities vis a vis system escape velocity were done. The process generates "weighting factors" for each of these four areas. Actual binaries will score high in the overall weighting of these factors, while optical pairs will score low. The result reveals that less than half of the WDS (42%) is actually physical with a very large group (33%) of unknown status (due to a lack of good data to perform the necessary calculations).

1. Introduction

The release of Gaia DR2 in April of 2018 bears seeds that will revolutionize double star astronomy forever. From the mission's home page, we find this statement of vision: "Gaia is an ambitious mission to chart a three-dimensional map of our Galaxy, the Milky Way, in the process revealing the composition, formation and evolution of the Galaxy." [ESA,2018] And we can define "ambitious" as details on 1,692,919,135 stars! Launched on Dec. 19, 2013, the mission is planned to run to late 2019. To date, there have been two releases of data (DR1, released Sept. 14, 2016; and DR2, released Apr. 25, 2018). Two more are planned (DR3 slated for release late 2020, and DR4 slated for release at the end of 2022. The content of DR2 was about 3 orders of magnitude denser than DR1, while DR3 and DR4 will be refinements to the already incredible data in DR2.

2. The Gaia Instrumentation

Gaia was placed in a Lissajous orbit around point L2. Figure 1 gives a rough idea of where Gaia has been parked.

Missions parked at any of the Lagrangian points (the "L" points in Figure 1) will reside in places that require a minimum of expenditure of propellant to maintain mission orientation and stability and represent areas of space relatively clear of space debris by the gravitational balancing points of the earth and sun. (The James Webb Space Telescope is also planned for insertion at L2.)

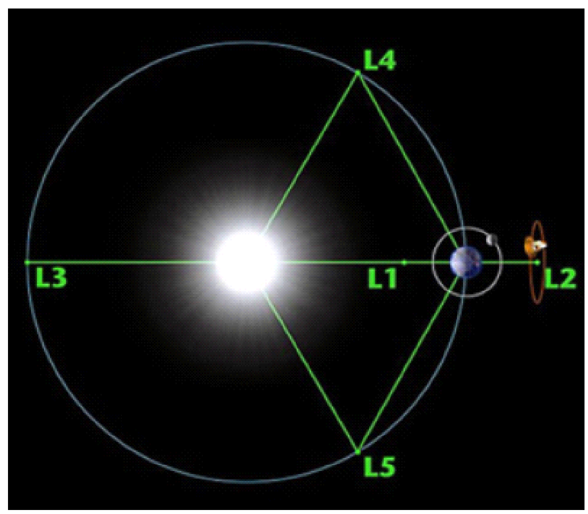

Figure 1. Gaia's Parking Station at L2

Instruments aboard the Gaia spacecraft include two star imaging systems (employing rectangular mirrors of large size), two photometers (red and blue), and a radial velocity spectrometer. These instruments are of exquisite precision and have given us literally an astronomical cache of data points about the stars surveyed.

3. Contents of DR2

One may access DR2 through the VizieR portal of the CDS service maintained by the Université de Strasbourg in France. I use this URL as my launching point for VizieR:

http://vizier.u-strasbg.fr/viz-bin/VizieR

Gaia DR2 and the Washington Double Star Catalog: A Tale of Two Databases

Currently, there is an icon at the top of the VizieR opening page that directs you to DR2 with a single mouse click (see Figure 2), but should the page design change after this paper is published, you can still retrieve DR2 by typing I/345/gaia2 into the catalog call window at the top of the page.

Once you click the Gaia DR2 hyperlink (or type the catalog call code), a menu of options fills the page. Besides the main DR2 catalog, there are 19 extractions that can be called for special research projects, but none of these falls under the domain of double star astronomy.

I therefore normally use the first entry on the selection form, "I/345/gaia2" by clicking the box in front of its listing and then clicking "Query Selected Tables".

That action opens the selection panel for what data you wish to view in DR2—and the list is very extensive! (The list extends over 5 screens down on my oversized monitor!)

You should then check the boxes for all the data you wish to retrieve. After that, you are ready to call up data on any star by using the Target Position input window at the top of the page. Here is where having the *precise* position of a double star is critical (and thankfully, the WDS does offer that datum on any star it lists). I usually open a downloaded version of the WDS and select the Precise Position for the star I want to research and paste it into the Target window.

Before you press the "Submit" button, I suggest you change the search radius from its default (2 arc minutes) to 1 arc minute, as you may well get data returned on hundreds of stars around the position you designate (depending, of course, on where your target lies with respect to galactic longitude and latitude). Clicking "Submit" will then return all the data for the stars within 1 arc minute of the precise position you designate.

As a safety check (to be sure I have the right stars in my query), I also click the "Start Aladin Lite" hyperlink above the results table to view a digitized image of the sky at this position.

DR2 can return data to you on a large number of parameters, including (but not limited to) the DR2 identifying number, precise position in the sky, parallax (and its error), proper motion vectors (and their errors), G magnitude (roughly equivalent to the visual magnitude; this number is *derived* from the R and B photometer data), B (blue) magnitude, R (red) magnitude, radial velocity and its error (in km/sec), and the effective surface temperature (T_{eff}, derived from the color photometers and estimates of distance based on the parallax). Data that is also reported (but derived from parallax and magnitudes) includes estimates of the star's radius and

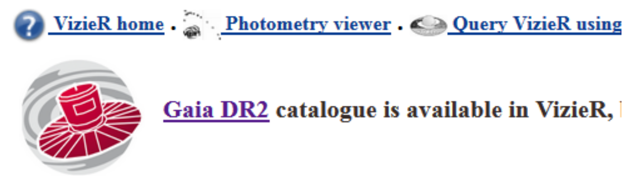

Figure 2: Link to DR2 at the top of the VizieR landing page.

luminosity (both in solar equivalents). There is a lot more available than this, but these are the main items of interest to most double star researchers.

The inclusion of estimates of stellar radii and luminosity based on parallax, magnitudes and temperature are also helpful as a second check on what the data might suggest. For instance, if the data show that the parallaxes of two stars are too different to permit the conclusion that the stars are binary, then a check of the luminosity for each star may help confirm that. If the companion star (usually fainter than the primary) is, in fact, quite a lot more luminous than the primary, this would confirm the suggestion that the stars are at vastly different distances.

4. The Power of Querying Large Databases

David Rowe, chief technical officer of PlainWave Instruments, is an expert on large databases. Soon after its release, he downloaded the entire DR2 database (it runs almost 8 TB!). After informing a team of us who use his Speckle Tool Box to do data reductions on our speckle measurements, we asked Dave if he could extract the Washington Double Star Catalog (WDS) from DR2? Within 24 hours, we had our catalog!

Rowe's extraction of the WDS from DR2 gives us extremely powerful tools to use in our research and saves us untold hours of time using the standard VizieR search process to get data on the stars we select for research.

Within a day or two of downloading a copy of Dave's WDS extraction, I had written routines that allowed my Excel version to compute distances to the stars (given their parallaxes and error estimates, when available, using the weighted parallax method I describe in Harshaw 2018. I had also developed subroutines that analyzed the proper motion data to compare how much the stars should have moved over a given time due to PM alone (and hence know if the PM vectors were accurate or not), and even derive estimates of the minimum distance between the members of a pair (the distance in parsecs times the separation in arc seconds) and stellar mass using the mass/luminosity relationship and DR2's B-R color indexes.

5. Factors That May Indicate Physicality

Let us define "physicality" as the likelihood that a given pair is traveling together through space, may have a common origin, and may, in fact, be in orbit around a common center of gravity. What things must be true for a pair to be physical?

First, the stars must have the same (or nearly the same) **parallax**. Rarely do both stars of a pair have the exact same parallax. So that is why it is necessary to estimate the likelihood that two stars are close enough to be gravitationally bound together by computing the total distance each parallax and its error projects and seeing if the distance "windows" overlap.

For example, suppose that two stars of a pair have parallaxes of 5.68 mas ± 0.06 mas for the primary star and 5.81 mas ± 0.05 mas for the companion. The parallaxes are not exactly equal, but what are the distance windows suggested by the parallaxes and the errors? For the primary star, the parallax could be anywhere from 5.68 + 0.06 = 5.74 mas to 5.68 – 0.06 = 5.62 mas. The distance in parsecs is simply the reciprocal of the parallax in arc seconds (not mas, so to do the calculation we must divide the parallax by 1000), giving us, for the primary star, 174 parsecs to 178 parsecs. For the companion, the parallax limits are 5.86 mas and 5.76 mas, resulting in distances of 171 parsecs to 174 parsecs. Since each star's distance window shares a common bound (174 parsecs), there is a significant chance (although not 100%) that the stars are within a parsec of each other. Is this close enough for gravitational binding?

On the surface, probably not. A survey of the 2,500+ orbits and their parameters from the 6th Orbit Catalog shows that very few known binaries have separations that exceed 3,000 AU, and most are closer than 1,000 AU. A parsec is approximately 192,000 AU, so two stars one parsec apart are probably too far apart to be gravitationally bound, even if they are very massive stars.

In a case like this, I use the weighted parallax method (cited earlier) to determine a weighted parallax (which yields 174 parsecs). If we know the separation (rho), we can easily estimate the *minimum* separation by multiplying the distance in parsecs by the separation in arc seconds. For example, if our sample pair had a value of rho of 1.26", the minimum separation would be 219 AU, well within the binding range of a physical pair. (We say *minimum separation* since we do not know the exact orientation of the system to our line of sight. The rho value we measure is a *projected* value on the plane of the sky. More than likely, the stars are not oriented orthogonal to our line of sight.)

In the analysis of the WDS extraction from DR2, I used a simple formula to assess how strong the parallax data might be in establishing physicality. That expression is given in Equation 1:

$$P_{px} = 1 - \left| \frac{P_d - C_d}{\frac{1}{2}\left(P_d + C_d\right)} \right| \qquad [1]$$

where P_d is the distance to the primary (in parsecs) and C_d is the distance to the companion. Thus, two stars with the same parallax will evaluate P to 1.00, while two stars with different parallaxes will evaluate P to something less than 1.00.

Once the parallax factor has been computed, we can move on to the analysis of the **proper motion vectors**. Two stars that are in orbit around one another should have identical, or very nearly identical, proper motions. Large differences in proper motion would suggest the stars are not gravitationally related. I evaluate the proper motion vectors with this rather complex expression (Equation 2):

$$P_{pm} = \left| 1 - \frac{\sqrt{\left(P_{pmra} - C_{pmra}\right)^2 + \left(P_{pmdec} - C_{pmdec}\right)^2}}{\sqrt{P_{pmra}^2 + P_{pmdec}^2} + \sqrt{C_{pmra}^2 + C_{pmdec}^2}} \right| * 0.15 \qquad [2]$$

where P_{pmra} is the primary star's PM vector in RA, P_{pmdec} is the primary's PM vector in DEC, and C_{pmra} and C_{pmdec} are the corresponding vectors for the companion. The 0.15 factor at the end is a weighting factor that I will explain in more detail shortly.

Thus it can be seen that two stars with exactly the same proper motion evaluate P_{pm} to 0.15 while stars with different proper motions will evaluate to something less than 0.15.

A third factor to consider is the line of best fit **R^2 value** for what appear to be linear and short arc traces on a plot of the data returned by a datarequest submitted to the USNO. I import all datarequest text files into an Excel spreadsheet I created that evaluates a number of items and plots the measurements on Cartesian coordinates, corrected for precession of the equinoxes. Most of the time, a plot of the data produces a scattered grouping of data like the one in Figure 3.

However, sometimes a data plot resembles Figure 4.

Here, the data clearly lies along a line. (The orange line is the proper motion resultant for the period between the first and last observations.) If we then click

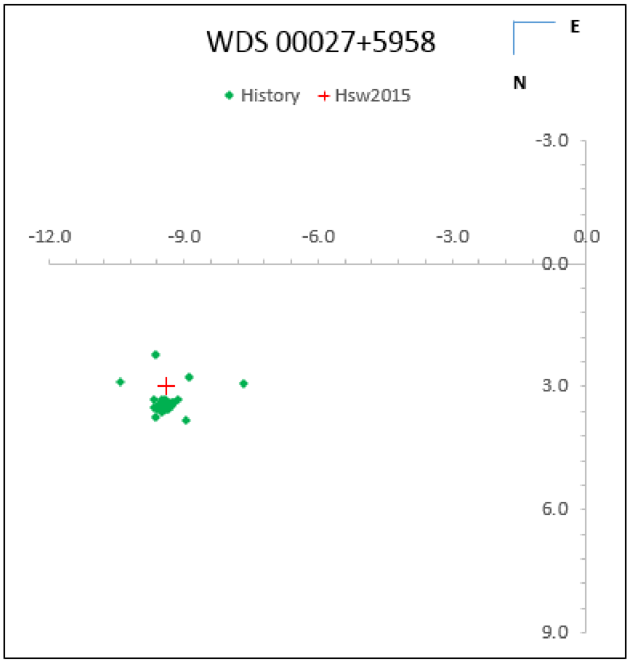

Figure 3. Plot of scattered measurements

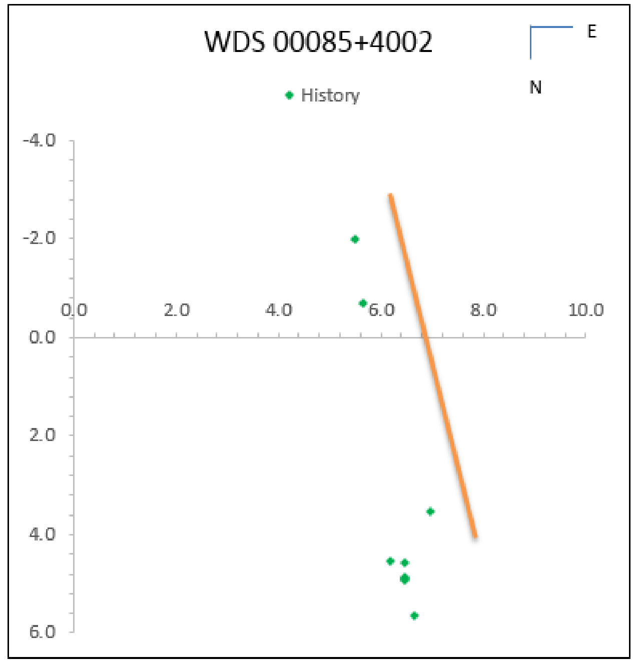

Figure 4. Plot of a linear nature

on any of the data points and ask Excel to generate a linear trend line, the graph looks like the one in Figure 5:

The closer the R^2 value is to 1.00, the better the fit of the data. 0.7135 is not a bad fit, but I have plots with values as high as 0.9989.

Similarly, an *exponential* trend line can be generated for cases that show an arc developing, like Figure 6.

When generating short arc trend lines, it is important that the projected curve enclose the (0,0) point of the graph. If the curve does not “wrap around” (0,0), the stars are not likely bound in an orbit.

However, extreme caution must be used when using the Excel Insert Trendline function with the request to see the R^2 value. By default, Excel assigns *equal weight to every measurement* while in practice, we need to assign varying weights to the measurements based on criteria covered already by the USNO at the 6th Orbit Catalog web page [USNO 2018].

The factor I use for R^2 values is only computed for those cases where I have generated graphs and have an R^2 value. If the plot shows a linear pattern, 10% of the R^2 value is *subtracted* from the final weight; if the pattern is of a short arc, 10% of the R^2 value is *added* to the final weight.

A fourth criteria to consider when determining physicality is the relative **radial velocities** of the stars compared to the escape velocity of the system. If the stars are close enough by parallax to be physical, we can assess the probable masses of the stars using the mass/luminosity equations and determine a rough estimate of the masses of each star knowing their tempera-

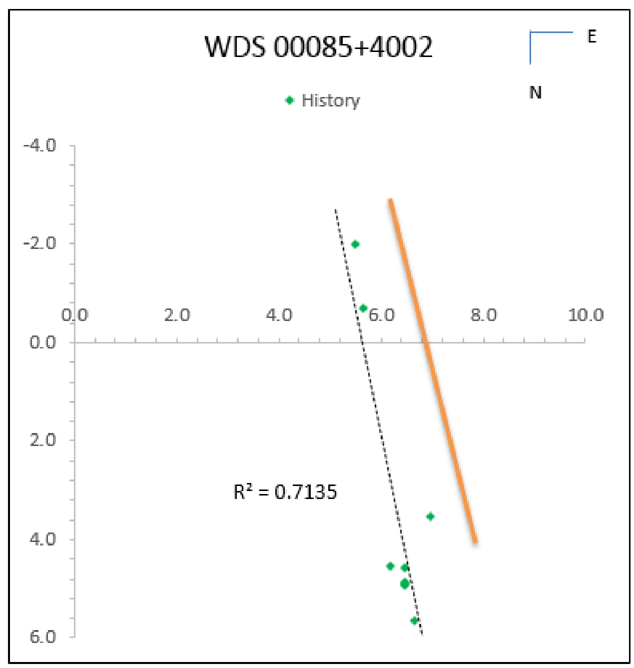

Figure 5. Plot with a linear trend line and R^2 value shown

Gaia DR2 and the Washington Double Star Catalog: A Tale of Two Databases

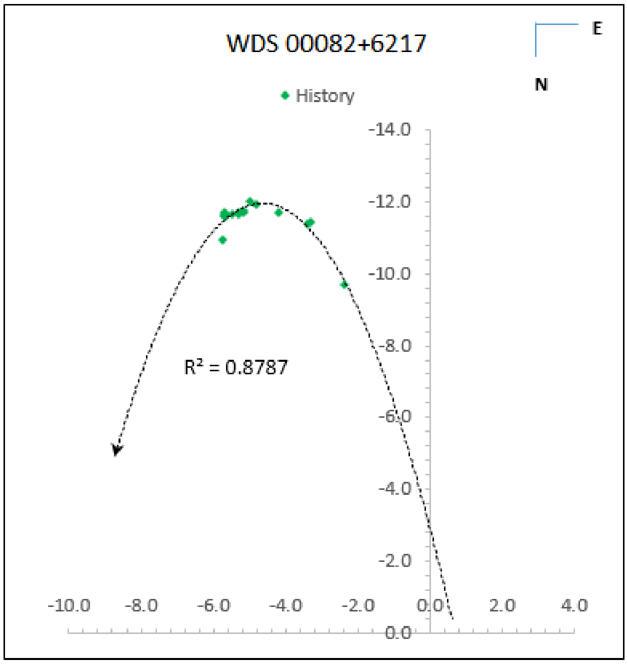

Figure 6. Plot of a short arc with high R^2 value

ture and luminosity. (The radius is also a helpful factor, but for DR2, the radius is a *derived* factor, not a directly -measured one like T_{eff}. Or course, DR2's luminosity figures are also derived from the measured magnitudes and distances as suggested by the parallaxes.)

To determine the escape velocity we use Equation 3:

$$V_{esc} = \frac{\sqrt{2GM_{Tot}}}{r} \qquad [3]$$

where G is the gravitational constant, M_{Tot} is the total mass of the binary system, and r is the distance between the two stars in parsecs [Wiley and Rica, 2015].

Meanwhile, the orbital velocity can be approximated by Equation 4:

$$V^2 = GM_{Tot}\left(\frac{2}{R} - \frac{1}{a}\right) \qquad [4]$$

Where G is the gravitational constant, M_{Tot} is the total system mass, R is the separation in AU, and a is the semi-minor axis in AU. To be on the safe side, I assume the system is not orthogonal to our line of sight and assume that the result of the equation is the absolute *maximum* possible for the system. Clearly if the difference in the radial velocities measured by Gaia exceeds this value, the pair will not be physical for long.

Estimating the total mass is tricky for cases where we do not have an orbital solution. We can estimate the masses based on the assumed values for luminosity, T_{eff}, and radius. Luminosity is related to mass by Equation 5:

$$\frac{L}{L_\odot} = \left(\frac{M}{M_\odot}\right)^n \qquad [5]$$

where L is the luminosity of the star in question, $L_\odot$ is the luminosity of the sun, M is the mass of the star in question, $M_\odot$ is the mass of the sun, and n is an exponent that varies with the luminosity class of the star. For bright massive stars, n averages 3; for sun-like stars, it runs closer to 4; and is about 2.5 for dim red dwarfs of low mass. But this holds for the Main Sequence only.

A second handle on the mass, possible with DR2's T_{eff} data, is given by Equation 6:

$$L \propto M^{3.5} \qquad [5]$$

where L is luminosity and M is mass, both relative to the sun.

A final approach to the mass problem can be derived by using the new Gaia DR2 H-R Diagram, shown in tiny form in Figure 7, and available on-line at https://physics.stackexchange.com/questions/402299/what-is-this-clump-of-hot-and-dim-stars-gaia-h-r-diagram and other URLs.

Here, one would need to enter the bottom of the chart using the B-R color index and then move up to the luminosity function to get an estimate on the star's spectral and luminosity classes. This is tedious work and the rough scaling on the vertical axis makes accuracy a challenge.

6. Combining Weight Factors

So we now have four weight factors that can be combined to give us an indication of the physicality of a pair:

1. Parallax analysis
2. Proper motion analysis
3. R^2 Fit analysis
4. Relative Radial Velocity analysis

Using parallax as the chief arbiter of physicality, I assign a relative weight of 75% to the parallax factor. I assign 10% weight each to the proper motion and R^2 factors, and 5% to the radial velocity factor, for a total of 100%.

Building the factor equations and weighting process into my Excel version of the WDS extraction from DR2 quickly yields estimates of physicality for all the stars in the WDS catalog. The closer to 1.00 the total weights

Gaia DR2 and the Washington Double Star Catalog: A Tale of Two Databases

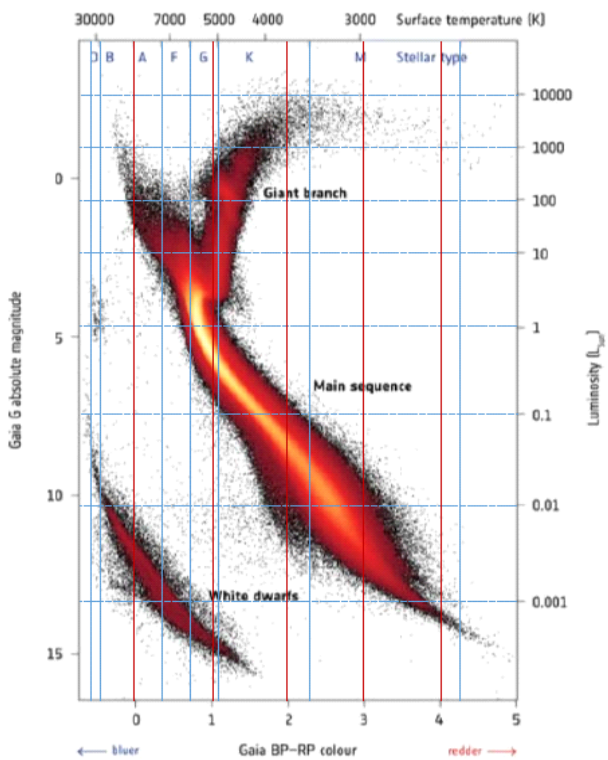

Figure 7. The Gaia DR2 H-R Diagram

are, the more likely the pair is to be physical. Conversely, as values drop towards 0.00, the pair is less and less likely to be physical.

I am currently working now on obtaining mass estimates for the 6,500 pairs I have datarequest graphs on, but have weights (less the radial velocity factor) for all the pairs in the WDS.

In Excel, one can format any cell to draw a bar graph of the value of another cell. Figure 8 shows what a section of those bar graphs looks like.

In Figure 8, stars 1, 2, 4, 5 and 6 all show fairly high values indicating physicality. Star 3 is somewhere in the middle of the range, while the last two stars lack enough critical data from DR2 to enable us to make a calculation.

7. Results

Once probability estimates had been made for all the stars in the WDS, it was a simple matter to create a data filter and view how many pairs had high probability (over 85%, class "Y") of being physical, how many had medium-high probability (65% to 85%, class "Y?"), how many might be physical (50% to 65%, class "Maybe"), how many were questionable (35% to 50%,

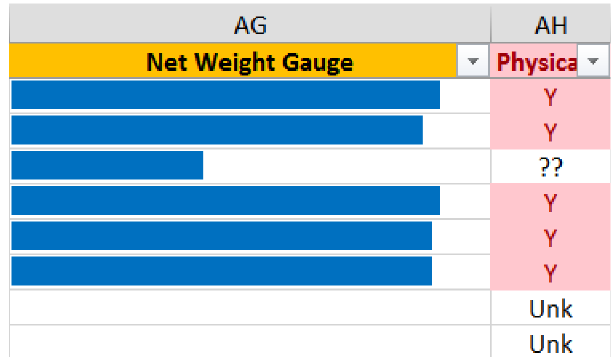

Figure 8. New Weight Average "Gauge"

class "??"), and not physical (< 35%, class "No"), with a class of "Unknown" when two or more of the weight factors could not be calculated. The numbers could then be used to create a pie chart showing the overall results. Here is what was returned (Figure 9).

Because the chart is small, I will place the results

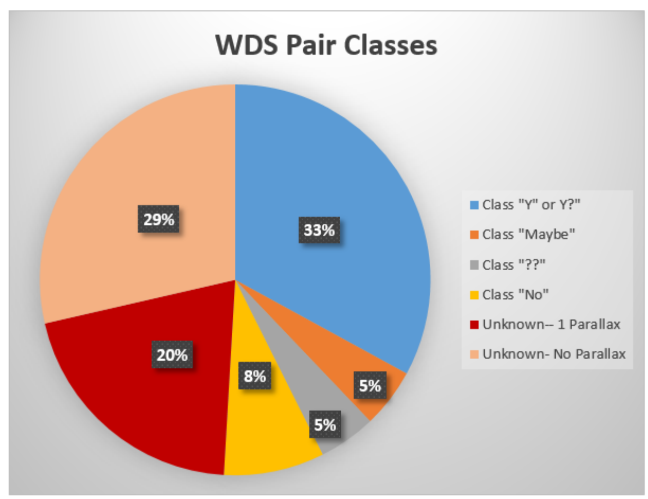

Figure 9. Pie Chart showing the makeup of the WDS

Table 1. Results of WDS / DR2 Extraction

Type of Pair	Number	Percentage
Definitely physical (Y, Y?)	46,061	33%
Maybe	6,871	4.5%
??	6,613	4.5%
No	11,642	8%
Unknown	68,595	49%

into a table (Table 1).

To my surprise, only 33% of the 139,781 WDS records extracted from DR2 show definite signs of physicality while a solid factor (8%) do not, with a further 49% of unknown status due to a lack of good data.

8. Discussion

The Excel spreadsheet is large (38 MB) and will be available to readers of this journal by clicking this link: http://www.jdso.org/volume14/number4/Harshaw_vol14_pg734/WDSGaiaDR2_Ver3.xlsx.

It is recommended that before a program of observing and measuring a double star is begun, the database should be consulted to see if the selected star is physical or not. Since it is the goal of the research to advance us towards *orbital solutions*, then non-physical pairs should not be high on the priority list.

The author will continue working on mass estimates for cases where DR2 shows radial velocities for both stars. There will also be future reports on suspected linear cases that as yet have no solutions as well as strong short arc candidates that may be ready for an orbital solution.

9. Conclusion

Gaia DR2 is a game changer for double star astronomy. Rather than make observational astrometry obsolete, on the contrary, DR2 gives us solid information to help us plan and direct our research. Look how far we have come in 250 years with hit-or-miss observation programs. Think what we can do going forward knowing now exactly where to look for new data!

10. Acknowledgements

The author wishes to thank Dave Rowe for his rapid response to team inquiries for a WDS extraction from Gaia DR2. Of course, the Washington Double Star Catalog was used to build the DR2 extraction, so special thanks to the United States Naval Observatory. And a tremendous debt of gratitude is owed the European Space Agency (ESA) for the Gaia mission and the treasure trove of data being supplied to us about our stellar neighbors in the galaxy.

11. References

ESA 2018, http://sci.esa.int/gaia/

Harshaw, Richard, 2018, “Measurements of 427 Double Stars With Speckle Interferometry: The Winter/Spring 2017 Observing Program at Brilliant Sky Observatory, Part 1”, *JDSO,* **14**-2, 284-330.

USNO 2018: http://ad.usno.navy.mil/wds/orb6/orb6text.html#grading

Wiley, E. O. and Rica, F. M., 2015, “Dynamic Studies of Struve Double Stars: STF4 and STF 236AB Appear Gravitationally Bound”, *JDSO*, **11**-1, 2-8.