Publication Year	2023
Acceptance in OA	2025-02-25T15:26:01Z
Title	Gaia Data Release 3. Apsis. III. Non-stellar content and source classification
Authors	Delchambre, L., Bailer-Jones, C. A. L., Bellas-Velidis, I., DRIMMEL, Ronald, Garabato, D., Carballo, R., Hatzidimitriou, D., Marshall, D. J., Andrae, R., Dafonte, C., Livanou, E., Fouesneau, M., Licata, E. L., Lindstrøm, H. E. P., Manteiga, M., Robin, C., Silvelo, A., Abreu Aramburu, A., Álvarez, M. A., Bakker, J., Bijaoui, A., Brouillet, N., BRUGALETTA, Elisa, Burlacu, A., Casamiquela, L., Chaoul, L., Chiavassa, A., Contursi, G., Cooper, W. J., Creevey, O. L., Dapergolas, A., de Laverny, P., Demouchy, C., Dharmawardena, T. E., Edvardsson, B., Frémat, Y., García-Lario, P., García-Torres, M., Gavel, A., Gomez, A., González-Santamaría, I., Heiter, U., Jean-Antoine Piccolo, A., Kontizas, M., Kordopatis, G., Korn, A. J., LANZAFAME, Alessandro, Lebreton, Y., Lobel, A., Lorca, A., Magdaleno Romeo, A., Marocco, F., Mary, N., Nicolas, C., Ordenovic, C., Pailler, F., Palicio, P. A., Pallas-Quintela, L., Panem, C., Pichon, B., POGGIO, Eloisa, Recio-Blanco, A., Riclet, F., Rybizki, J., Santoveña, R., Sarro, L. M., Schultheis, M. S., Segol, M., Slezak, I., SMART, Richard Laurence, SORDO, Rosanna, Soubiran, C., Süveges, M., Thévenin, F., Torralba Elipe, G., Ulla, A., Utrilla, E., VALLENARI, Antonella, van Dillen, E., Zhao, H., Zorec, J.
Publisher's version (DOI)	10.1051/0004-6361/202243423
Handle	http://hdl.handle.net/20.500.12386/36226
Journal	ASTRONOMY & ASTROPHYSICS
Volume	674

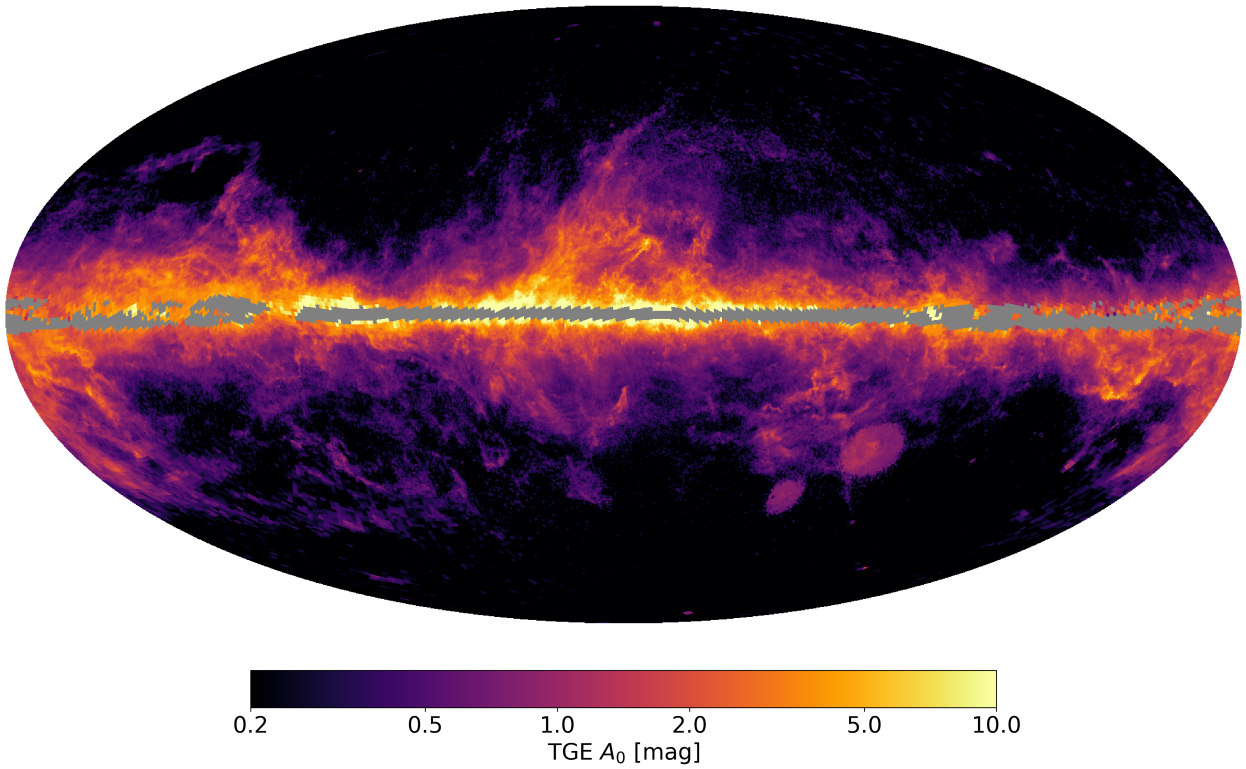

Fig. 24. HEALPix map of the total Galactic extinction, built from HEALPixes between levels 6 and 9 (0.839 to 0.013 deg^2), which are identified as being at the optimum resolution over their field of view.

the given uncertainty by the square root of the number of tracers used (`num_tracers_used`). The full range of A_0 extinction measures of the tracers (`a0_min`, `a0_max`) is also provided.

The first table, `total_galactic_extinction_map`, contains HEALPix maps at four different HEALPix levels, from level 6 (49 152 HEALPixes with an area of 0.84 deg^2) to level 9 (3 145 728 HEALPixes with an area of 0.013 deg^2), with the HEALPix level indicated with the parameter `healpix_level`. This range of HEALPix levels ensures that a minimum number of tracers per HEALPix will be found at high Galactic latitudes, where the sky density of tracers is low, while allowing a higher resolution in areas of the sky where the density of tracers is high. (At level 9 only 1% of the sky has more than 40 tracers per HEALPix.)

For any given direction we determine the optimum HEALPix level, that is, the set of the smallest HEALPixes with at least ten tracers to ensure a reliable estimate of the extinction and its uncertainties. However, as the base resolution is HEALPix level 6, all HEALPixes with fewer than ten tracers at this level are tagged as 'optimum'. As in the level 6 map, the optimum map has full sky coverage at $|b| > 5°$ (i.e. all HEALPixes at $|b| > 5°$ have at least three tracers, so an A_0 value is reported for each of them). In the HEALPix scheme, each HEALPix at level n contains four sub-HEALPixes at level $n + 1$, meaning that each of the four sub-HEALPixes must have at least ten tracers to allow all four to be tagged as optimum. This algorithm is repeated iteratively over each level, starting at the base level 6, until the lack of tracers in a sub-HEALPix prevents further subdivision, or until level 9 is reached. In the table `total_galactic_extinction_map`, the optimum HEALPixes are flagged as such with the boolean flag `optimum_hpx_flag`. This algorithm ensures that the subset of optimum HEALPixes do not overlap with one another, yet cover the entire sky.

The second table, `total_galactic_extinction_map_opt`, is a single optimum HEALPix map at level 9 provided for convenience, where each HEALPix adopts the extinction value of the optimum HEALPix `total_galactic_extinction_map` coincident with or containing the HEALPix. That is, if a HEALPix at level 6 is tagged as optimum in `total_galactic_extinction_map`, then all 64 of its level-9 sub-HEALPixes in the `total_galactic_extinction_map_opt` map will be assigned the `a0` value of the level 6 HEALPix. The parameter `optimum_hpx_level` in this table indicates, for each HEALPix, the HEALPix level of the optimum HEALPix from which its `a0` value is based.

7.3. Performance

At the base level 6, only 2.8% of the sky (1379 out of 49152 HEALPixes) close to the Galactic plane (with $|b| < 5°$) has no `a0` values because of an insufficient number of tracers. The fraction of HEALPixes with an insufficient number of tracers increases at the higher HEALPix levels as the HEALPixes become smaller: 5.2% at level 7, 30.4% at level 8, and 66.3% at level 9. The average number of tracers for the HEALPixes with A_0 estimates is 268.3 at level 6, but only 10.7 at level 9, while the average number of tracers for the optimum HEALPix map is 30.3. The optimum HEALPix map, `total_galactic_extinction_map_opt`, shown in Fig. 24, has the same sky coverage as the level 6 map, but is of higher resolution when a sufficient number of tracers are available. To better demonstrate this, we show a zoom into the Rho Ophiuchi region in Fig. 25. Over the whole sky, only about 1% of the HEALPixes at level 9 have more than 40 tracers, and thus the potential to be mapped at higher resolution. Figures

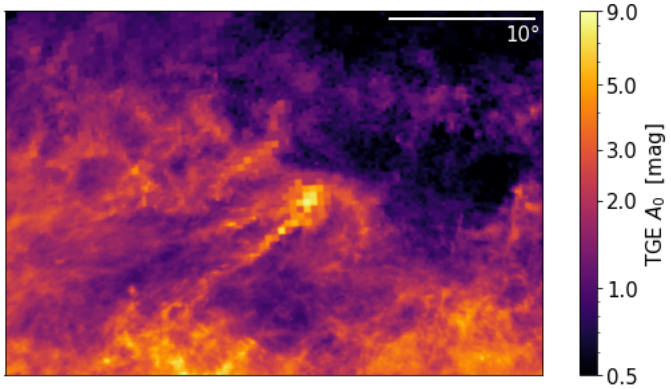

Fig. 25. A_0 towards Rho Ophiuchi from the TGE optimum HEALPix map (Fig. 24) centred at $(l, b) = (-5°, 18°)$. The solid white line in the upper right corner provides the angular scale of the image. The variable resolution of the optimum HEALPix map is particularly obvious towards the middle of the figure.

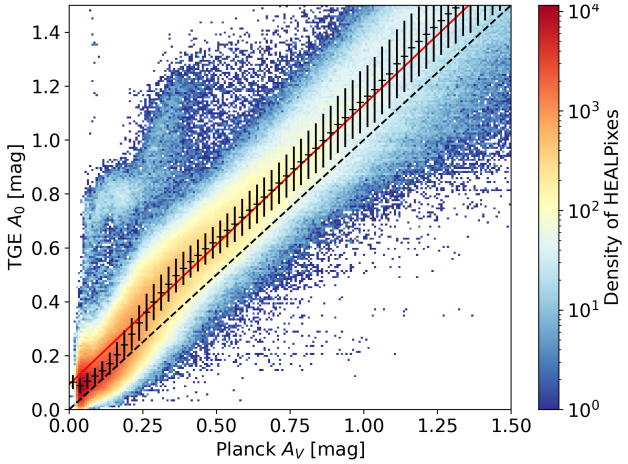

Fig. 26. Extinction comparison between the TGE A_0 optimum HEALPix map and the Planck A_V HEALPix level 9 map at small extinction values. The colour scale shows the density of HEALPixes, the red dashed line represents unity, and the points with error bars are the median A_0 and average absolute deviation computed in A_V bins of width 0.025 mag. The red line is the result of a linear fit to the points.

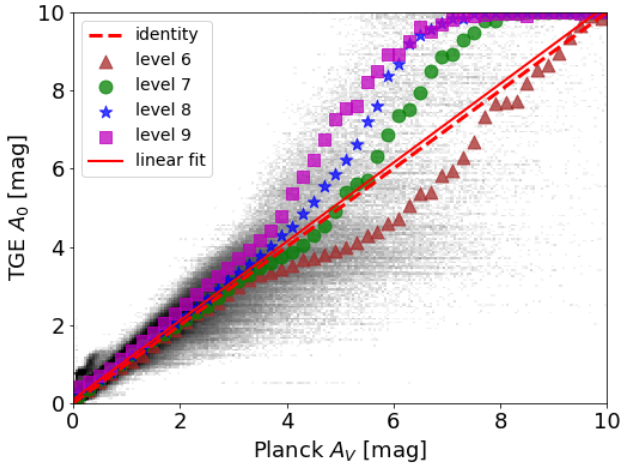

Fig. 27. Comparison of the extinction between the TGE A_0 optimum HEALPix map and the Planck A_V HEALPix level 9 map for extinctions up to 10 mag. The background grey scale is a density plot of the entire optimal HEALPix TGE map (comprising the optimal HEALPixes at several HEALPix levels). The dashed red line represents unity and the solid red line is a linear fit of the medians of all HEALPixes in the optimum HEALPix map with $0.5 \leq A_V \leq 3$. Coloured symbols refer to the median A_0 computed in A_V bins of width 0.2 mag for various HEALPix levels that are used to assign the A_0 value.

showing the individual all-sky maps at levels 6 through 9 can be found in the online documentation, along with maps of the `a0_uncertainty`. We note that the `a0_uncertainty` is smallest in HEALPix level 6 with a mean value of 0.03 mag; this is due to the larger number of tracers contained in the HEALPixes at this level, whereas the mean `a0_uncertainty` of the HEALPixes in `total_galactic_extinction_map` tagged as optimum (`optimum_hpx_flag` = 1) is of 0.06 mag, as they cover various HEALPix levels.

In Fig. 26, the TGE A_0 estimate at the optimum HEALPix level 9 is plotted against the dust optical depth expressed as A_V from Planck Collaboration Int. XLVIII (2016)[16], once re-binned at the same HEALPix level. We see good agreement, as a linear fit using the median points with $0.2 \leq A_V \leq 3$ results in a slope of 1.04 ± 0.05, albeit with an offset of 0.09 ± 0.05. It should be noted that the ratio of A_V/A_0 for giants (stars with effective temperature $3000 < T_{\rm eff} < 5700$ K) is ~0.98 (see the online documentation), meaning that the slope of TGE (converted to A_V) over Planck(A_V) is $1.04 \times 0.98 = 1.02$. Also worth bearing in mind is that there are a number of Planck maps of the dust distribution available on the Planck Legacy Archive; for example, using the map described in Planck Collaboration XXIX (2016) we find a slope of 0.90 ± 0.04 and an offset of 0.05 ± 0.04.

Performing a linear fit in the same extinction range between TGE A_0 and Schlegel et al. (1998) A_V results in a slope of 0.98 ± 0.04 (offset: 0.10 ± 0.04, in agreement with the 1.04 ± 0.05 obtained using Planck. However, the same linear fit performed between TGE and the Bayestar's map (Green et al. 2019) results in a slope of 1.20 ± 0.04 (offset: 0.01 ± 0.04), suggesting that the Bayestar map is systematically underestimating the extinction with respect to other extinction maps; see discussion in Andrae et al. (2022).

Towards the limit where the extinction measured by Planck tends to zero, the TGE A_0 tends to a non-zero value. This offset is found empirically by fitting a third-order polynomial to the median points for $A_0 < 0.4$ and obtaining the TGE A_0 value at Planck $A_V = 0$. The resulting offset is 0.10 ± 0.03 mag and starts to become evident at $A_V < 0.1$ mag. The existence of this offset is likely due to the fact that the GSP-Phot extinction prior forces its extinction estimate to be non-negative, which creates a statistical bias at very low extinction values. Indeed, this A_0 offset is of the order expected if the true uncertainty of the A_0 estimates per source were 0.1 magnitude. See Andrae et al. (2022) for further discussion.

Comparing TGE A_0 to Planck A_V over a larger interval highlights a possible bias at extinctions $A_V \geq 4$ mag. In Fig. 27, TGE is plotted versus Planck over an interval of ten magnitudes. A large dispersion in A_0 is observed for the optimal map for $A_V > 4$ mag, and it can be seen that the different HEALPix levels do not behave in the same way. The coarser resolutions (levels 6 and 7) initially predict less extinction than Planck (for $4 \leq A_0 \leq 5$ mag) whereas the finer resolutions either agree or predict higher extinction. Above an A_V of 5 mag, only level 6 predicts less extinction than Planck, while the others predict more. Even for $A_V < 4$ mag, where TGE and Planck are in very good agreement, a difference can be seen where the lower resolutions predict lower extinction. This is likely due to a selection

[16] The Planck collaboration reports $E(B-V)$ that we convert to A_V via $A_V = R_V E(B-V)$ and $R_V = 3.1$. See the Planck Legacy Archive (`http://pla.esac.esa.int`) for details.

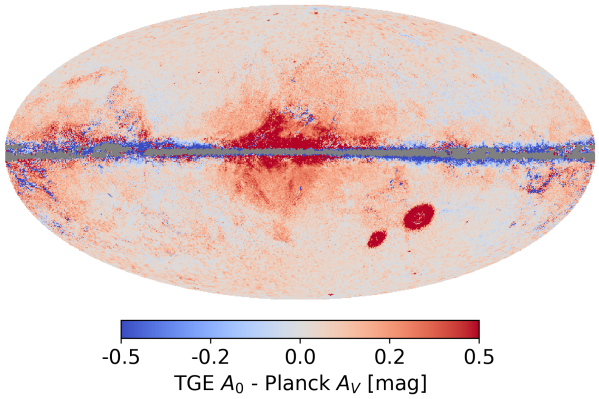

Fig. 28. Residual sky map of TGE A_0 minus Planck A_V, using the optimum HEALPix level 9 map. Red values show regions where TGE predicts more extinction than Planck, whereas blue values show the opposite.

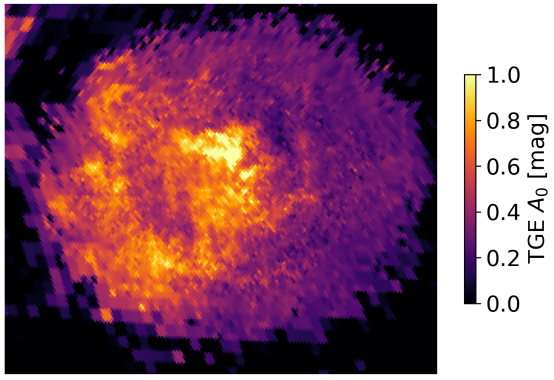

Fig. 29. A_0 towards the LMC from the TGE Optimum HEALPix map (Fig. 24), centred at $(l, b) = (280.0°, -33.0°)$. The estimated offset of $A_0 = 0.6$ mag has been subtracted. The solid white line in the bottom left corner provides the angular scale of the image.

effect where in a given HEALPix with variable extinction, more stars will be observed where the extinction is smaller. This will bias the extinction estimate for the HEALPix to lower values, and will be more obvious for larger HEALPixes.

Finally in Fig. 28 the residual map of TGE A_0 minus Planck A_V is shown. TGE underestimates extinction with respect to Planck toward molecular clouds, where dust emission remains optically thin but where TGE estimates may be biased toward smaller values as unresolved areas with below average extinction are oversampled, as mentioned above; see further discussion regarding high-extinction regions in the following section. Meanwhile, within about 30° towards the Galactic centre, TGE shows more extinction than Planck, apart from the foreground molecular complexes we just mentioned.

7.4. Use of TGE results

The TGE extinction maps estimate the total Galactic extinction A_0 from the Milky Way ISM toward extragalactic sources, where A_0 is the monochromatic extinction at 541.4 nm. As mentioned above, A_V/A_0 is approximately equal to 0.98 for cool stars at $A_0 < 3$ mag. However, in general, the effective extinction in a passband depends on the SED of the source; see the online documentation for a discussion on how to derive the extinction from A_0 for any passband.

As the selected extinction tracers were required to be beyond a certain minimum distance to ensure that they were outside the ISM layer of the Milky Way's disc, sources in nearby galaxies may also be selected as tracers. This means that the extinction towards the LMC and SMC will be a combination of Galactic extinction, inter-galactic extinction, and extinction in the Magellanic clouds (although the latter will be the dominant contribution). Another factor that will influence the amount of reported extinction in these directions stems from the distance prior used in GSP-Phot, which assumes that the sources are Galactic. As such, the extinction will be overestimated. An evaluation of this overestimation can be obtained via a comparison with an external data set. Indeed, in Fig. 26, there is a cloud of points with a locus stretching from around $A_V = 0.2$, $A_0 = 0.8$ to $A_V = 0.4$, $A_0 = 1.2$ that consists entirely of lines of sight towards the Magellanic clouds. Comparing the median TGE A_0 (1.0 mag) to the median Planck A_V (0.4 mag) towards the LMC reveals a difference of 0.6 mag. These values are both higher than the extinction found using near-infrared observations ($A_V = 0.3$ mag; Imara & Blitz 2007) and in the visible ($A_V = 0.24$ mag; Wagner-Kaiser & Sarajedini 2013). This difference is likely not only due to the GSP-Phot distance prior, but also to variations in dust properties in the LMC/SMC. Although the absolute level of extinction in these Galactic satellites needs to be interpreted with caution, the relative variations evidencing structured patterns are most certainly real (see Fig. 29).

Because extinction tracers are required to be outside the dust layer of the Milky Way, they must be at greater distances at lower Galactic latitudes. This, together with the effect of increasing extinction and *Gaia*'s magnitude limit, means that at very low latitudes it is not possible to find a sufficient number of tracers outside the ISM layer of the Milky Way with which to make a reliable estimate of the total Galactic extinction. This explains the band of HEALPixes at $b \approx 0$ with no extinction values. Indeed we recommend that the map should not be used for latitudes $|b| < 5°$. Also, GSP-Phot sets an upper limit of ten magnitudes on its estimate of A_0 per source, and so any HEALPixes with an extinction near this value should be interpreted as a lower bound. However, as suggested by Fig. 27, our maps may instead be over-estimating extinction toward these lines of sight with respect to Planck, though we point out that HEALPixes with $A_0 > 4$mag are at low Galactic latitude and make up only 2% of the sky. Furthermore, Planck estimates towards the Galactic plane may be underestimated as a consequence of assuming a single mean dust temperature for the whole line of sight. Further details of the TGE data products are documented in the online documentation.

8. Beyond *Gaia* DR3

We present the non-stellar and classification modules from CU8 in their present status, as for *Gaia* DR3. However, they are in constant evolution and changes are already planned for *Gaia* DR4 and later, which we summarise for each module in this section.

Although the intrinsic performance of DSC is very good, once we take into account class prior – as we do for all results shown in this paper – the purities of the classified samples are modest. In preparation for *Gaia* DR4, we will aim to improve this, for example by optimising the feature set in Allosmod and how this is used. We will also reconsider the class definitions and the training data, in particular for white dwarfs and physical binaries. As Specmod uses the entire BP/RP spectrum, we expected better performance (compared to Allosmod), and so

we will investigate improving the classifier. We may also introduce filters to remove the classifications of the lowest quality data (which are the main determinant of the low purities).

OA will be upgraded by implementing its own outlier detector, which will be mostly based on unsupervised clustering algorithms. Additionally, we will improve the statistical description and the templates that were used for *Gaia* DR3. The functionality offered by the GUASOM visualisation tool will be extended in order to allow the user to perform and explore their own clustering analysis.

QSOC will use epoch BP/RP spectra re-sampled into logarithmic wavelength bins in order to overcome the issues we encountered while using the Hermite spline polynomials associated with the internal representation of the BP/RP spectra. This internal representation effectively tends to produce wiggles whose strength can be comparable to those of quasar emission lines in faint $G \geq 19$ mag spectra (Creevey et al. 2023). This solution will concurrently allow us to use sampled BP/RP spectra with uncorrelated noise on their flux, as the algorithm described in Delchambre (2016) is not optimised to deal with full covariance matrices.

The performance of the UGC redshift estimator strongly depends on the training set used. As more epochs are incorporated in the BP/RP spectra, we expect to have more (and generally fainter) sources with redshifts above 0.4 available for inclusion in the training set, thus improving the performance especially for higher redshifts. We will also investigate optimisation of the SVM model parameters in order to reduce the large variability in the performance with redshift and to minimise the positive bias for bright, low-redshift objects.

In future data releases, we can expect the TGE maps to improve with future improvements of GSP-Phot (Andrae et al. 2022). In particular, we expect that the number of sources with stellar parameters will increase, which will improve the reliability of the TGE maps, and possibly allow for maps at a resolution higher than HEALPix level 9.

Acknowledgements. This work presents results from the European Space Agency (ESA) space mission *Gaia*. *Gaia* data are being processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the *Gaia* MultiLateral Agreement (MLA). The *Gaia* mission website is https://www.cosmos.esa.int/gaia. The *Gaia* archive website is https://archives.esac.esa.int/gaia. Some of the results in this work have been derived using the HEALPix package (Górski et al. 2005; healpix.sourceforge.net). The full ackowledgements are available in Appendix A.

References

Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, ApJS, 240, 23
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Álvarez, M. A., Dafonte, C., Manteiga, M., Garabato, D., & Santoveña, R. 2021, Neural Computing and Applications
Andrae, R., Fouesneau, M., Sordo, R., et al. 2022, A&A, 674, A27 (*Gaia* DR3 SI)
Bailer-Jones, C. A. L. 2021, Gaia Data Processing and Analysis Consortium (DPAC) Technical Note GAIA-C8-TN-MPIA-CBJ-094, http://www.cosmos.esa.int/web/gaia/public-dpac-documents
Bailer-Jones, C. A. L., Smith, K. W., Tiede, C., Sordo, R., & Vallenari, A. 2008, MNRAS, 391, 1838
Bailer-Jones, C. A. L., Andrae, R., Arcay, B., et al. 2013, A&A, 559, A74
Bailer-Jones, C. A. L., Fouesneau, M., & Andrae, R. 2019, MNRAS, 490, 5615
Bastian, U., & Portell, J. 2020, Gaia Data Processing and Analysis Consortium (DPAC) technical note GAIA-C3-TN-ARI-BAS-020, http://www.cosmos.esa.int/web/gaia/public-dpac-documents
Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, A&A, 652, A86
Chang, C.-C., & Lin, C.-J. 2011, ACM Transactions on Intelligent Systems and Technology, 2, 27:1, Software Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
Cortes, C., & Vapnik, V. 1995, Mach. Learn., 20, 273
Creevey, O., Sordo, R., Pailler, F., et al. 2023, A&A, 674, A26 (*Gaia* DR3 SI)
Delchambre, L. 2015, MNRAS, 446, 3545
Delchambre, L. 2016, MNRAS, 460, 2811
Delchambre, L. 2018, MNRAS, 473, 1785
Fitzpatrick, E. L. 1999, PASP, 111, 63
Flesch, E. W. 2021, ArXiv e-prints [arXiv:2105.12985]
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2023, A&A, 674, A28 (*Gaia* DR3 SI)
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration (Bailer-Jones, C., et al.) 2023, A&A, 674, A41 (*Gaia* DR3 SI)
Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 887, 93
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Imara, N., & Blitz, L. 2007, ApJ, 662, 969
Kohonen, T. 1982, Biol. Cybern., 43, 59
Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, A&A, 649, A2
Montegriffo, P., De Angeli, F., Andrae, R., et al. 2023, A&A, 674, A3 (*Gaia* DR3 SI)
Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51
Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, A&A, 597, A79
Planck Collaboration XXIX. 2016, A&A, 586, A132
Planck Collaboration Int. XLVIII. 2016, A&A, 596, A109
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Wagner-Kaiser, R., & Sarajedini, A. 2013, MNRAS, 431, 1565

[1] Institut d'Astrophysique et de Géophysique, Université de Liège, 19c, Allée du 6 Août, 4000 Liège, Belgium
[2] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[3] National Observatory of Athens, I. Metaxa and Vas. Pavlou, Palaia Penteli 15236, Athens, Greece
[4] INAF – Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
[5] CIGUS CITIC – Department of Computer Science and Information Technologies, University of A Coruña, Campus de Elviña s/n, A Coruña 15071, Spain
[6] Dpto. de Matemática Aplicada y Ciencias de la Computación, Univ. de Cantabria, ETS Ingenieros de Caminos, Canales y Puertos, Avda. de los Castros s/n, 39005 Santander, Spain
[7] Department of Astrophysics, Astronomy and Mechanics, National and Kapodistrian University of Athens, Panepistimiopolis, Zografos 15783, Athens, Greece
[8] IRAP, Université de Toulouse, CNRS, UPS, CNES, 9 Av. colonel Roche, BP 44346, 31028 Toulouse Cedex 4, France
[9] Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen Ø, Denmark
[10] DXC Technology, Retortvej 8, 2500 Valby, Denmark
[11] CIGUS CITIC, Department of Nautical Sciences and Marine Engineering, University of A Coruña, Paseo de Ronda 51, 15071 A Coruña, Spain
[12] Thales Services for CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[13] ATG Europe for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[14] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
[15] Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[16] INAF – Osservatorio Astrofisico di Catania, Via S. Sofia 78, 95123 Catania, Italy
[17] Telespazio for CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[18] GEPI, Observatoire de Paris, Université PSL, CNRS, 5 place Jules Janssen, 92190 Meudon, France

[19] CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[20] Centre for Astrophysics Research, University of Hertfordshire, College Lane, AL10 9AB Hatfield, UK
[21] APAVE SUDEUROPE SAS for CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[22] Theoretical Astrophysics, Division of Astronomy and Space Physics, Department of Physics and Astronomy, Uppsala University, Box 516, 751 20 Uppsala, Sweden
[23] Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium
[24] European Space Agency (ESA), European Space Astronomy Centre (ESAC), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[25] Data Science and Big Data Lab, Pablo de Olavide University, 41013 Seville, Spain
[26] Observational Astrophysics, Division of Astronomy and Space Physics, Department of Physics and Astronomy, Uppsala University, Box 516, 751 20 Uppsala, Sweden
[27] Dipartimento di Fisica e Astronomia "Ettore Majorana", Università di Catania, Via S. Sofia 64, 95123 Catania, Italy
[28] LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 Place Jules Janssen, 92190 Meudon, France
[29] Université Rennes, CNRS, IPR (Institut de Physique de Rennes) – UMR 6251, 35000 Rennes, France
[30] Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[31] IPAC, Mail Code 100-22, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
[32] Dpto. de Inteligencia Artificial, UNED, c/ Juan del Rosal 16, 28040 Madrid, Spain
[33] INAF – Osservatorio astronomico di Padova, Vicolo Osservatorio 5, 35122 Padova, Italy
[34] Institute of Global Health, University of Geneva, Geneva, Switzerland
[35] Applied Physics Department, Universidade de Vigo, 36310 Vigo, Spain
[36] Sorbonne Université, CNRS, UMR7095, Institut d'Astrophysique de Paris, 98bis bd. Arago, 75014 Paris, France

Appendix A: Full acknowledgements

The *Gaia* mission and data processing have financially been supported by, in alphabetical order by country:

- the Algerian Centre de Recherche en Astronomie, Astrophysique et Géophysique of Bouzareah Observatory;
- the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF) Hertha Firnberg Programme through grants T359, P20046, and P23737;
- the BELgian federal Science Policy Office (BELSPO) through various PROgramme de Développement d'Expériences scientifiques (PRODEX) grants and the Polish Academy of Sciences - Fonds Wetenschappelijk Onderzoek through grant VS.091.16N, and the Fonds de la Recherche Scientifique (FNRS), and the Research Council of Katholieke Universiteit (KU) Leuven through grant C16/18/005 (Pushing AsteRoseismology to the next level with TESS, GaiA, and the Sloan DIgital Sky SurvEy – PARADISE);
- the Brazil-France exchange programmes Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Coordenação de Aperfeicoamento de Pessoal de Nível Superior (CAPES) - Comité Français d'Evaluation de la Coopération Universitaire et Scientifique avec le Brésil (COFECUB);
- the Chilean Agencia Nacional de Investigación y Desarrollo (ANID) through Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) Regular Project 1210992 (L. Chemin);
- the National Natural Science Foundation of China (NSFC) through grants 11573054, 11703065, and 12173069, the China Scholarship Council through grant 201806040200, and the Natural Science Foundation of Shanghai through grant 21ZR1474100;
- the Tenure Track Pilot Programme of the Croatian Science Foundation and the École Polytechnique Fédérale de Lausanne and the project TTP-2018-07-1171 'Mining the Variable Sky', with the funds of the Croatian-Swiss Research Programme;
- the Czech-Republic Ministry of Education, Youth, and Sports through grant LG 15010 and INTER-EXCELLENCE grant LTAUSA18093, and the Czech Space Office through ESA PECS contract 98058;
- the Danish Ministry of Science;
- the Estonian Ministry of Education and Research through grant IUT40-1;
- the European Commission's Sixth Framework Programme through the European Leadership in Space Astrometry (ELSA) Marie Curie Research Training Network (MRTN-CT-2006-033481), through Marie Curie project PIOF-GA-2009-255267 (Space AsteroSeismology & RR Lyrae stars, SAS-RRL), and through a Marie Curie Transfer-of-Knowledge (ToK) fellowship (MTKD-CT-2004-014188); the European Commission's Seventh Framework Programme through grant FP7-606740 (FP7-SPACE-2013-1) for the *Gaia* European Network for Improved data User Services (GENIUS) and through grant 264895 for the *Gaia* Research for European Astronomy Training (GREAT-ITN) network;
- the European Cooperation in Science and Technology (COST) through COST Action CA18104 'Revealing the Milky Way with *Gaia* (MW-Gaia)';
- the European Research Council (ERC) through grants 320360, 647208, and 834148 and through the European Union's Horizon 2020 research and innovation and excellent science programmes through Marie Skłodowska-Curie grant 745617 (Our Galaxy at full HD – Gal-HD) and 895174 (The build-up and fate of self-gravitating systems in the Universe) as well as grants 687378 (Small Bodies: Near and Far), 682115 (Using the Magellanic Clouds to Understand the Interaction of Galaxies), 695099 (A sub-percent distance scale from binaries and Cepheids – CepBin), 716155 (Structured ACCREtion Disks – SACCRED), 951549 (Sub-percent calibration of the extragalactic distance scale in the era of big surveys – UniverScale), and 101004214 (Innovative Scientific Data Exploration and Exploitation Applications for Space Sciences – EXPLORE);
- the European Science Foundation (ESF), in the framework of the *Gaia* Research for European Astronomy Training Research Network Programme (GREAT-ESF);
- the European Space Agency (ESA) in the framework of the *Gaia* project, through the Plan for European Cooperating States (PECS) programme through contracts C98090 and 4000106398/12/NL/KML for Hungary, through contract 4000115263/15/NL/IB for Germany, and through PROgramme de Développement d'Expériences scientifiques (PRODEX) grant 4000127986 for Slovenia;
- the Academy of Finland through grants 299543, 307157, 325805, 328654, 336546, and 345115 and the Magnus Ehrnrooth Foundation;
- the French Centre National d'Études Spatiales (CNES), the Agence Nationale de la Recherche (ANR) through grant ANR-10-IDEX-0001-02 for the 'Investissements d'avenir' programme, through grant ANR-15-CE31-0007 for project 'Modelling the Milky Way in the *Gaia* era' (MOD4Gaia), through grant ANR-14-CE33-0014-01 for project 'The Milky Way disc formation in the *Gaia* era' (ARCHEOGAL), through grant ANR-15-CE31-0012-01 for project 'Unlocking the potential of Cepheids as primary distance calibrators' (UnlockCepheids), through grant ANR-19-CE31-0017 for project 'Secular evolution of galxies' (SEGAL), and through grant ANR-18-CE31-0006 for project 'Galactic Dark Matter' (GaDaMa), the Centre National de la Recherche Scientifique (CNRS) and its SNO *Gaia* of the Institut des Sciences de l'Univers (INSU), its Programmes Nationaux: Cosmologie et Galaxies (PNCG), Gravitation Références Astronomie Métrologie (PNGRAM), Planétologie (PNP), Physique et Chimie du Milieu Interstellaire (PCMI), and Physique Stellaire (PNPS), the 'Action Fédératrice *Gaia*' of the Observatoire de Paris, the Région de Franche-Comté, the Institut National Polytechnique (INP) and the Institut National de Physique nucléaire et de Physique des Particules (IN2P3) co-funded by CNES;
- the German Aerospace Agency (Deutsches Zentrum für Luft- und Raumfahrt e.V., DLR) through grants 50QG0501, 50QG0601, 50QG0602, 50QG 0701, 50QG0901, 50QG1001, 50QG1101, 50QG1401, 50QG1402, 50QG1403, 50QG1404, 50QG1904, 50QG2101, 50QG2102, and 50QG2202, and the Centre for Information Services and High Performance Computing (ZIH) at the Technische Universität Dresden for generous allocations of computer time;
- the Hungarian Academy of Sciences through the Lendület Programme grants LP2014-17 and LP2018-7 and the Hungarian National Research, Development, and Innovation Office (NKFIH) through grant KKP-137523 ('SeismoLab');
- the Science Foundation Ireland (SFI) through a Royal Society - SFI University Research Fellowship (M. Fraser);

– the Israel Ministry of Science and Technology through grant 3-18143 and the Tel Aviv University Center for Artificial Intelligence and Data Science (TAD) through a grant;
– the Agenzia Spaziale Italiana (ASI) through contracts I/037/08/0, I/058/10/0, 2014-025-R.0, 2014-025-R.1.2015, and 2018-24-HH.0 to the Italian Istituto Nazionale di Astrofisica (INAF), contract 2014-049-R.0/1/2 to INAF for the Space Science Data Centre (SSDC, formerly known as the ASI Science Data Center, ASDC), contracts I/008/10/0, 2013/030/I.0, 2013-030-I.0.1-2015, and 2016-17-I.0 to the Aerospace Logistics Technology Engineering Company (ALTEC S.p.A.), INAF, and the Italian Ministry of Education, University, and Research (Ministero dell'Istruzione, dell'Università e della Ricerca) through the Premiale project 'MIning The Cosmos Big Data and Innovative Italian Technology for Frontier Astrophysics and Cosmology' (MITiC);
– the Netherlands Organisation for Scientific Research (NWO) through grant NWO-M-614.061.414, through a VICI grant (A. Helmi), and through a Spinoza prize (A. Helmi), and the Netherlands Research School for Astronomy (NOVA);
– the Polish National Science Centre through HARMONIA grant 2018/30/M/ST9/00311 and DAINA grant 2017/27/L/ST9/03221 and the Ministry of Science and Higher Education (MNiSW) through grant DIR/WK/2018/12;
– the Portuguese Fundação para a Ciência e a Tecnologia (FCT) through national funds, grants SFRH/BD/128840/2017 and PTDC/FIS-AST/30389/2017, and work contract DL 57/2016/CP1364/CT0006, the Fundo Europeu de Desenvolvimento Regional (FEDER) through grant POCI-01-0145-FEDER-030389 and its Programa Operacional Competitividade e Internacionalização (COMPETE2020) through grants UIDB/04434/2020 and UIDP/04434/2020, and the Strategic Programme UIDB/-00099/2020 for the Centro de Astrofísica e Gravitação (CENTRA);
– the Slovenian Research Agency through grant P1-0188;
– the Spanish Ministry of Economy (MINECO/FEDER, UE), the Spanish Ministry of Science and Innovation (MICIN), the Spanish Ministry of Education, Culture, and Sports, and the Spanish Government through grants BES-2016-078499, BES-2017-083126, BES-C-2017-0085, ESP2016-80079-C2-1-R, ESP2016-80079-C2-2-R, FPU16/03827, PDC2021-121059-C22, RTI2018-095076-B-C22, and TIN2015-65316-P ('Computación de Altas Prestaciones VII'), the Juan de la Cierva Incorporación Programme (FJCI-2015-2671 and IJC2019-04862-I for F. Anders), the Severo Ochoa Centre of Excellence Programme (SEV2015-0493), and MICIN/AEI/10.13039/501100011033 (and the European Union through European Regional Development Fund 'A way of making Europe') through grant RTI2018-095076-B-C21, the Institute of Cosmos Sciences University of Barcelona (ICCUB, Unidad de Excelencia 'María de Maeztu') through grant CEX2019-000918-M, the University of Barcelona's official doctoral programme for the development of an R+D+i project through an Ajuts de Personal Investigador en Formació (APIF) grant, the Spanish Virtual Observatory through project AyA2017-84089, the Galician Regional Government, Xunta de Galicia, through grants ED431B-2021/36, ED481A-2019/155, and ED481A-2021/296, the Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC), funded by the Xunta de Galicia and the European Union (European Regional Development Fund – Galicia 2014-2020 Programme), through grant ED431G-2019/01, the Red Española de Supercomputación (RES) computer resources at MareNostrum, the Barcelona Supercomputing Centre - Centro Nacional de Supercomputación (BSC-CNS) through activities AECT-2017-2-0002, AECT-2017-3-0006, AECT-2018-1-0017, AECT-2018-2-0013, AECT-2018-3-0011, AECT-2019-1-0010, AECT-2019-2-0014, AECT-2019-3-0003, AECT-2020-1-0004, and DATA-2020-1-0010, the Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya through grant 2014-SGR-1051 for project 'Models de Programació i Entorns d'Execució Parallels' (MPEXPAR), and Ramon y Cajal Fellowship RYC2018-025968-I funded by MICIN/AEI/10.13039/501100011033 and the European Science Foundation ('Investing in your future');
– the Swedish National Space Agency (SNSA/Rymdstyrelsen);
– the Swiss State Secretariat for Education, Research, and Innovation through the Swiss Activités Nationales Complémentaires and the Swiss National Science Foundation through an Eccellenza Professorial Fellowship (award PCEFP2_194638 for R. Anderson);
– the United Kingdom Particle Physics and Astronomy Research Council (PPARC), the United Kingdom Science and Technology Facilities Council (STFC), and the United Kingdom Space Agency (UKSA) through the following grants to the University of Bristol, the University of Cambridge, the University of Edinburgh, the University of Leicester, the Mullard Space Sciences Laboratory of University College London, and the United Kingdom Rutherford Appleton Laboratory (RAL): PP/D006511/1, PP/D006546/1, PP/D006570/1, ST/I000852/1, ST/J005045/1, ST/K00056X/1, ST/K000209/1, ST/K000756/1, ST/L006561/1, ST/N000595/1, ST/N000641/1, ST/N000978/1, ST/N001117/1, ST/S000089/1, ST/S000976/1, ST/S000984/1, ST/S001123/1, ST/S001948/1, ST/S001980/1, ST/S002103/1, ST/V000969/1, ST/W002469/1, ST/W002493/1, ST/W002671/1, ST/W002809/1, and EP/V520342/1.

The GBOT programme uses observations collected at (i) the European Organisation for Astronomical Research in the Southern Hemisphere (ESO) with the VLT Survey Telescope (VST), under ESO programmes 092.B-0165, 093.B-0236, 094.B-0181, 095.B-0046, 096.B-0162, 097.B-0304, 098.B-0030, 099.B-0034, 0100.B-0131, 0101.B-0156, 0102.B-0174, and 0103.B-0165; and (ii) the Liverpool Telescope, which is operated on the island of La Palma by Liverpool John Moores University in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias with financial support from the United Kingdom Science and Technology Facilities Council, and (iii) telescopes of the Las Cumbres Observatory Global Telescope Network.

Appendix B: Combining probabilities for DSC-Combmod

Combmod in DSC combines the posterior probabilities from Specmod and Allosmod into a new posterior probability, taking care to ensure that the global prior is only counted once. If Specmod and Allosmod used the same classes, and operated on independent data, then combining their probabilities would be simple. However, Specmod has three classes (star, white dwarf,

physical binary star) that correspond to the single star class in Allosmod. It is also possible that Specmod or Allosmod provides no result. The combination method is therefore a bit more complicated. The basic idea is that a fraction of the Allosmod probability for the single 'superclass' is taken to correspond to each subclass in Specmod, with that fraction equal to the prior. We assume that Specmod and Allosmod are independent, which is not quite true as the colours in Allosmod are derived from the BP/RP spectra used by Specmod.

- Let P_k^m be the posterior probability from classifier m for class k.
- Let π_k^m be the prior probability used in classifier m for class k.
- For Specmod, $m = s$ and $k = 1 \ldots 5$ corresponding to quasar, galaxy, star, white dwarf, physical binary star respectively.
- For Allosmod, $m = a$ and $k = 1 \ldots 3$ corresponding to quasar, galaxy, star, respectively.
- For each classifier, classes are disjoint and exhaustive, so the probabilities sum to one.
- The priors for the two classifiers are consistent, so $\pi_1^a = \pi_1^s$, $\pi_2^a = \pi_2^s$, and $\pi_3^a = \sum_{k=3}^5 \pi_k^s$.

For the classes that correspond one-to-one, the combined posterior probability is obtained by multiplying the likelihoods (the posterior divided by the prior, to within a normalisation factor) and then multiplying by the prior. This is

$$P_k^c = a \frac{P_k^s}{\pi_k^s} \frac{P_k^a}{\pi_k^a} \pi_k^a = a P_k^s P_k^a \frac{1}{\pi_k^s} \quad k \in \{1, 2\}, \tag{B.1}$$

where a is a data-dependent but class-independent normalisation factor. For each of the three stellar classes in Specmod, we assume that a fraction π_k^s/π_3^a for $k \in \{3, 4, 5\}$ of the posterior probability P_3^a is the Allosmod posterior probability for that class. Thus the combined probability for each of these three classes is

$$P_k^c = a \frac{P_k^s}{\pi_k^s} \frac{P_3^a}{\pi_3^a} \frac{\pi_k^s}{\pi_3^a} \pi_k^s = a P_k^s P_3^a \frac{\pi_k^s}{(\pi_3^a)^2} \quad k \in \{3, 4, 5\}\,. \tag{B.2}$$

If Specmod probabilities are not available (missing), the combined posterior probability for the classes that correspond one-to-one is equal to the Allosmod probabilities:

$$P_k^c = P_k^a \quad k \in \{1, 2\} \quad \text{(no Specmod results)}\,. \tag{B.3}$$

For the three stellar classes, we distribute the corresponding Allosmod probability to these classes in proportion to the priors, i.e.

$$P_k^c = P_3^a \frac{\pi_k^s}{\sum_{k=3}^{k=5} \pi_k^s} \quad k \in \{3, 4, 5\} \quad \text{(no Specmod probabilities)}\,. \tag{B.4}$$

If Allosmod probabilities are not available, we simply copy the Specmod probabilities:

$$P_k^c = P_k^s \quad k \in \{1, 2, 3, 4, 5\} \quad \text{(no Allosmod probabilities)}\,. \tag{B.5}$$

If neither the Specmod nor the Allosmod probabilities are available, the Combmod probabilities will be empty.

The above equations run the risk of divide by zero if probabilities are exactly zero. To avoid this we 'soften' the Specmod and Allosmod probabilities prior to combination by adding 10^{-8}. This is only done in the combination: the Specmod and Allosmod probabilities written to the catalogue are not modified.

The above probability combination is not complicated conceptually, but it can lead to counter-intuitive results. Bailer-Jones (2021) works through various examples to demonstrate and explain this.

Appendix C: Adjusting the DSC probabilities to accommodate a new prior

All DSC probabilities are posterior probabilities that have taken into account the class priors listed in Table 2. Posteriors are equal to the product of a likelihood and a prior that has then been normalized. It is therefore simple to adjust the DSC probabilities to reflect a different prior probability: we simply divide each output by the prior used (to strip this off), multiply by the new prior, and then normalise the resulting probability vector. That is, if P_k^d is the DSC probability in the catalogue (for any of its classifiers) for class k, and if π_k^d is the corresponding catalogue prior (Table 2), then the new posterior probabilities corresponding to a new prior π_k^{new} are

$$\frac{P_k^d}{\pi_k^d} \pi_k^{\text{new}} \Big/ \sum_{k'} \left(\frac{P_{k'}^d}{\pi_{k'}^d} \pi_{k'}^{\text{new}} \right). \tag{C.1}$$