A&A 674, A33 (2023)
https://doi.org/10.1051/0004-6361/202243709


Astronomy & Astrophysics
Special issue

*Gaia Data Release 3*

# *Gaia* Data Release 3

## The Galaxy in your preferred colours: Synthetic photometry from *Gaia* low-resolution spectra

Gaia Collaboration: P. Montegriffo[1], M. Bellazzini[1,⋆], F. De Angeli[2], R. Andrae[3], M. A. Barstow[4], D. Bossini[5], A. Bragaglia[1], P. W. Burgess[2], C. Cacciari[1], J. M. Carrasco[6], N. Chornay[2], L. Delchambre[7], D. W. Evans[2], M. Fouesneau[3], Y. Frémat[8], D. Garabato[9], C. Jordi[6], M. Manteiga[10], D. Massari[1], L. Palaversa[11,2], E. Pancino[12,13], M. Riello[2], D. Ruz Mieres[2], N. Sanna[12], R. Santoveña[9], R. Sordo[14], A. Vallenari[14], N. A. Walton[2], A. G. A. Brown[15], T. Prusti[16], J. H. J. de Bruijne[16], F. Arenou[17], C. Babusiaux[18,17], M. Biermann[19], O. L. Creevey[20], C. Ducourant[21], L. Eyer[22], R. Guerra[23], A. Hutton[24], S. A. Klioner[25], U. L. Lammers[23], L. Lindegren[26], X. Luri[6], F. Mignard[20], C. Panem[27], D. Pourbaix[28,29,†], S. Randich[12], P. Sartoretti[17], C. Soubiran[21], P. Tanga[20], C. A. L. Bailer-Jones[3], U. Bastian[19], R. Drimmel[30], F. Jansen[31,⋆⋆], D. Katz[17], M. G. Lattanzi[30,32], F. van Leeuwen[2], J. Bakker[23], J. Castañeda[33], C. Fabricius[6], L. Galluccio[20], A. Guerrier[27], U. Heiter[34], E. Masana[6], R. Messineo[35], N. Mowlavi[22], C. Nicolas[27], K. Nienartowicz[36,37], F. Pailler[27], P. Panuzzo[17], F. Riclet[27], W. Roux[27], G. M. Seabroke[38], F. Thévenin[20], G. Gracia-Abril[39,19], J. Portell[6], D. Teyssier[40], M. Altmann[19,41], M. Audard[22,37], I. Bellas-Velidis[42], K. Benson[38], J. Berthier[43], R. Blomme[8], D. Busonero[30], G. Busso[2], H. Cánovas[40], B. Carry[20], A. Cellino[30], N. Cheek[44], G. Clementini[1], Y. Damerdji[7,45], M. Davidson[46], P. de Teodoro[23], M. Nuñez Campos[24], A. Dell'Oro[12], P. Esquej[47], J. Fernández-Hernández[48], E. Fraile[47], P. García-Lario[23], E. Gosset[7,29], R. Haigron[17], J.-L. Halbwachs[49], N. C. Hambly[46], D. L. Harrison[2,50], J. Hernández[23], D. Hestroffer[43], S. T. Hodgkin[2], B. Holl[22,37], K. Janßen[51], G. Jevardat de Fombelle[22], S. Jordan[19], A. Krone-Martins[52,53], A. C. Lanzafame[54,55], W. Löffler[19], O. Marchal[49], P. M. Marrese[56,13], A. Moitinho[52], K. Muinonen[57,58], P. Osborne[2], T. Pauwels[8], A. Recio-Blanco[20], C. Reylé[59], L. Rimoldini[37], T. Roegiers[60], J. Rybizki[3], L. M. Sarro[61], C. Siopis[28], M. Smith[38], A. Sozzetti[30], E. Utrilla[24], M. van Leeuwen[2], U. Abbas[30], P. Ábrahám[62,63], A. Abreu Aramburu[48], C. Aerts[64,65,3], J. J. Aguado[61], M. Ajaj[17], F. Aldea-Montero[23], G. Altavilla[56,13], M. A. Álvarez[9], J. Alves[66], R. I. Anderson[67], E. Anglada Varela[48], T. Antoja[6], D. Baines[40], S. G. Baker[38], L. Balaguer-Núñez[6], E. Balbinot[68], Z. Balog[19,3], C. Barache[41], D. Barbato[22,30], M. Barros[52], S. Bartolomé[6], J.-L. Bassilana[69], N. Bauchet[17], U. Becciani[54], A. Berihuete[70], M. Bernet[6], S. Bertone[71,72,30], L. Bianchi[73], A. Binnenfeld[74], S. Blanco-Cuaresma[75], T. Boch[49], A. Bombrun[76], S. Bouquillon[41,77], L. Bramante[35], E. Breedt[2], A. Bressan[78], N. Brouillet[21], E. Brugaletta[54], B. Bucciarelli[30,32], A. Burlacu[79], A. G. Butkevich[30], R. Buzzi[30], E. Caffau[17], R. Cancelliere[80], T. Cantat-Gaudin[6,3], R. Carballo[81], T. Carlucci[41], M. I. Carnerero[30], L. Casamiquela[21,17], M. Castellani[56], A. Castro-Ginard[15], L. Chaoul[27], P. Charlot[21], L. Chemin[82], V. Chiaramida[35], A. Chiavassa[20], G. Comoretto[40,83], G. Contursi[20], W. J. Cooper[84,30], T. Cornez[69], S. Cowell[2], F. Crifo[17], M. Cropper[38], M. Crosta[30,85], C. Crowley[76], C. Dafonte[9], A. Dapergolas[42], P. David[43], P. de Laverny[20], F. De Luise[86], R. De March[35], J. De Ridder[64], R. de Souza[87], A. de Torres[76], E. F. del Peloso[19], E. del Pozo[24], M. Delbo[20], A. Delgado[47], J.-B. Delisle[22], C. Demouchy[88], T. E. Dharmawardena[3], S. Diakite[89], C. Diener[2], E. Distefano[54], C. Dolding[38], H. Enke[51], C. Fabre[90], M. Fabrizio[56,13], S. Faigler[91], G. Fedorets[57,92], P. Fernique[49,93], F. Figueras[6], Y. Fournier[51], C. Fouron[79], F. Fragkoudi[94,95,96], M. Gai[30], A. Garcia-Gutierrez[6], M. Garcia-Reinaldos[23], M. García-Torres[97], A. Garofalo[1], A. Gavel[34], P. Gavras[47], E. Gerlach[25], R. Geyer[25], P. Giacobbe[30], G. Gilmore[2], S. Girona[98], G. Giuffrida[56], R. Gomel[91], A. Gomez[9], J. González-Núñez[44,99], I. González-Santamaría[9], J. J. González-Vidal[6], M. Granvik[57,100], P. Guillout[49], J. Guiraud[27], R. Gutiérrez-Sánchez[40], L. P. Guy[37,101], D. Hatzidimitriou[102,42], M. Hauser[3,103], M. Haywood[17], A. Helmer[69], A. Helmi[68], M. H. Sarmiento[24], S. L. Hidalgo[104,105], N. Hładczuk[23,106], D. Hobbs[26], G. Holland[2], H. E. Huckle[38], K. Jardine[107], G. Jasniewicz[108], A. Jean-Antoine Piccolo[27], Ó. Jiménez-Arranz[6], J. Juaristi Campillo[19], F. Julbe[6], L. Karbevska[37,109], P. Kervella[110], S. Khanna[68,30], G. Kordopatis[20], A. J. Korn[34], Á Kóspál[62,3,63], Z. Kostrzewa-Rutkowska[15,111], K. Kruszyńska[112], M. Kun[62], P. Laizeau[113], S. Lambert[41], A. F. Lanza[54], Y. Lasne[69], J.-F. Le Campion[21], Y. Lebreton[110,114], T. Lebzelter[66], S. Leccia[115], N. Leclerc[17], I. Lecoeur-Taibi[37], S. Liao[116,30,117], E. L. Licata[30], H. E. P. Lindström[30,118,119], T. A. Lister[120], E. Livanou[102], A. Lobel[8], A. Lorca[24], C. Loup[49], P. Madrero Pardo[6], A. Magdaleno Romeo[79], S. Managau[69],

⋆ Corresponding author: Michele Bellazzini, e-mail: michele.bellazzini@inaf.it.
⋆⋆ Retired.

R. G. Mann[46], J. M. Marchant[121], M. Marconi[115], J. Marcos[40], M. M. S. Marcos Santos[44], D. Marín Pina[6], S. Marinoni[56,13], F. Marocco[122], D. J. Marshall[123], L. Martin Polo[44], J. M. Martín-Fleitas[24], G. Marton[62], N. Mary[69], A. Masip[6], A. Mastrobuono-Battisti[17], T. Mazeh[91], P. J. McMillan[26], S. Messina[54], D. Michalik[16], N. R. Millar[2], A. Mints[51], D. Molina[6], R. Molinaro[115], L. Molnár[62,63,124], G. Monari[49], M. Monguió[6], A. Montero[24], R. Mor[6], A. Mora[24], R. Morbidelli[30], T. Morel[7], D. Morris[46], T. Muraveva[1], C. P. Murphy[23], I. Musella[115], Z. Nagy[62], L. Noval[69], F. Ocaña[40,125], A. Ogden[2], C. Ordenovic[20], J. O. Osinde[47], C. Pagani[4], I. Pagano[54], P. A. Palicio[20], L. Pallas-Quintela[9], A. Panahi[91], S. Payne-Wardenaar[19], X. Peñalosa Esteller[6], A. Penttilä[57], B. Pichon[20], A. M. Piersimoni[86], F.-X. Pineau[49], E. Plachy[62,63,124], G. Plum[17], E. Poggio[20,30], A. Prša[126], L. Pulone[56], E. Racero[44,125], S. Ragaini[1], M. Rainer[12,127], C. M. Raiteri[30], P. Ramos[6,49], M. Ramos-Lerate[40], P. Re Fiorentin[30], S. Regibo[64], P. J. Richards[128], C. Rios Diaz[47], V. Ripepi[115], A. Riva[30], H.-W. Rix[3], G. Rixon[2], N. Robichon[17], A. C. Robin[59], C. Robin[69], M. Roelens[22], H. R. O. Rogues[88], L. Rohrbasser[37], M. Romero-Gómez[6], N. Rowell[46], F. Royer[17], K. A. Rybicki[112], G. Sadowski[28], A. Sáez Núñez[6], A. Sagristà Sellés[19], J. Sahlmann[47], E. Salguero[48], N. Samaras[8,129], V. Sanchez Gimenez[6], M. Sarasso[30], M. S. Schultheis[20], E. Sciacca[54], M. Segol[88], J. C. Segovia[44], D. Ségransan[22], D. Semeux[90], S. Shahaf[130], H. I. Siddiqui[131], A. Siebert[49,93], L. Siltala[57], A. Silvelo[9], E. Slezak[20], I. Slezak[20], R. L. Smart[30], O. N. Snaith[17], E. Solano[132], F. Solitro[35], D. Souami[110,133], J. Souchay[41], A. Spagna[30], L. Spina[14], F. Spoto[75], I. A. Steele[121], H. Steidelmüller[25], C. A. Stephenson[40,134], M. Süveges[135], J. Surdej[7,136], L. Szabados[62], E. Szegedi-Elek[62], F. Taris[41], M. B. Taylor[137], R. Teixeira[87], L. Tolomei[35], N. Tonello[98], F. Torra[33], J. Torra[6,†], G. Torralba Elipe[9], M. Trabucchi[138,22], A. T. Tsounis[139], C. Turon[17], A. Ulla[140], N. Unger[22], M. V. Vaillant[69], E. van Dillen[88], W. van Reeven[141], O. Vanel[17], A. Vecchiato[30], Y. Viala[17], D. Vicente[98], S. Voutsinas[46], T. Wevers[2,142], Ł. Wyrzykowski[112], A. Yoldas[2], P. Yvard[88], H. Zhao[20], J. Zorec[143], S. Zucker[74] and T. Zwitter[144]

*(Affiliations can be found after the references)*

Received 4 April 2022 / Accepted 19 May 2022

**ABSTRACT**

*Gaia* Data Release 3 provides novel flux-calibrated low-resolution spectrophotometry for ≃220 million sources in the wavelength range 330 nm ≤ $\lambda$ ≤ 1050 nm (XP spectra). Synthetic photometry directly tied to a flux in physical units can be obtained from these spectra for any passband fully enclosed in this wavelength range. We describe how synthetic photometry can be obtained from XP spectra, illustrating the performance that can be achieved under a range of different conditions – for example passband width and wavelength range – as well as the limits and the problems affecting it. Existing top-quality photometry can be reproduced within a few per cent over a wide range of magnitudes and colour, for wide and medium bands, and with up to millimag accuracy when synthetic photometry is standardised with respect to these external sources. Some examples of potential scientific application are presented, including the detection of multiple populations in globular clusters, the estimation of metallicity extended to the very metal-poor regime, and the classification of white dwarfs. A catalogue providing standardised photometry for ≃$2.2\times10^8$ sources in several wide bands of widely used photometric systems is provided (*Gaia* Synthetic Photometry Catalogue; GSPC) as well as a catalogue of ≃$10^5$ white dwarfs with DA/non-DA classification obtained with a Random Forest algorithm (*Gaia* Synthetic Photometry Catalogue for White Dwarfs; GSPC-WD).

**Key words.** catalogs – surveys – techniques: photometric – techniques: spectroscopic – stars: general – Galaxy: general

## 1. Introduction

Photometry, together with astrometry, are probably the most ancient and fundamental techniques upon which our knowledge of the Universe and of astrophysical phenomena relies, and can be traced back to the ancient Greeks (see e.g. Sterken et al. 2011, for a short historical introduction and references; SMY11 hereafter). Photometry consists in sampling the spectra of astronomical sources by measuring their incoming flux passing through a transmission curve (TC) that allows the user to detect only the light within a defined wavelength range (spectral window). For a pure black-body spectrum, photometric measurements in two different spectral windows are sufficient to estimate the derivative of the black-body curve, unequivocally establishing its temperature. As the spectra of stars and other kinds of celestial sources deviate from black body, more than two spectral windows must be sampled to properly infer the most relevant astrophysical properties and/or to obtain an adequate classification (Young 1992a).

The earliest TC used for photometry was the sensitivity curve of the human eye, the details of which depend on the physiology of the observer. Today, actual TCs, which in the following we also refer to as passbands, are defined by the combination of the TC of an optical filter – which is designed to select the desired spectral window –, the sensitivity curve of a photon-counting detector (typically a CCD for observations in the optical spectral range), and the TC of the optical elements that collect the light from a source and properly convey it to the detector (telescope and camera), plus a contribution from the terrestrial atmosphere if observations are performed on the Earth's surface.

A photometric system is defined by a set of passbands and a set of standard stars observed in these passbands with an instrumental setup and a data-reduction procedure that is as controlled and homogeneous as possible (Bessell 2005; Sterken et al. 2011). The magnitude and colour differences between the standard stars define a relative photometric scale.

Following SMY11, a 'closed' photometric system is established by taking all the relevant measurements with the same observing site and instrumental setup (or the best possible approximation of this condition; these latter authors report the Walraven (Pel & Lub 2007) and the Geneva (Rufener 1971) systems as examples of closed systems). This approach should

† Deceased.

maximise the homogeneity and consequently the precision of the measurements gathered. On the other hand, we refer to 'open' systems as those with a sufficient number of standard stars distributed over the sky to allow broad accessibility, such that any observer can attempt to obtain photometry in that scale using their own instrumental setup, chosen to match the original one as closely as possible. This, in principle, allows a general use of the system and fruitful comparison between observations obtained at different places and in different epochs[1]. However, as the exact reproduction of the original observing conditions – in particular of the actual TCs – is virtually impossible to achieve, colour-dependent transformations are required to convert instrumental magnitudes into the desired scale. Transformations require repeated observations of standard stars during an observing run, and, in general, they are prone to subtle but sizable systematic errors (Young 1992a; SMY11). In general, a transformation should imply non-linear colour terms that may be hard to constrain and that are often neglected (see Young 1992a,b, 1994a, and references therein). In any case, ground-based photometric measures must be corrected for time- and (slightly) colour-dependent atmospheric extinction, a complex process in itself that may significantly contribute to the systematic error budget.

Finally, to convert magnitudes into physical fluxes, spectrophotometry is required, which comes in the form of synthetic photometry through the system TCs on the flux-calibrated spectrum of (at least) one standard star (see, e.g. Fukugita et al. 1996). According to Landolt (2011), spectrophotometry is the only kind of photometry that can be considered absolute, as it is directly linked to fluxes in physical units[2].

The synthetic description, the set of definitions, and the nomenclature above, as well as many general concepts that are used throughout this paper, are largely based on the reviews collected in Milone & Sterken (2011), in particular SMY11, Landolt (2011), but also in Bessell (2005), Sirianni et al. (2005), Sterken (2007a,b), Young (1992a,b, 1994a), and Manfroid (1992). We refer the interested reader to these papers and the references therein, as well as to Magnier et al. (2020a) and Thanjavur et al. (2021), for two examples of very recent, state-of-the-art applications to wide-area surveys.

Here we limit our discussion to photometry in the optical wavelength range, that is approximately between 300 nm and 1100 nm. In this context, it is worth reiterating the definition of photometric precision and accuracy provided by (Young 1994b, as reported by SMY11), as a reference: *precision* refers to the repeatability of a measurement, while *accuracy* means the absence of error, as measured against some external reference, such as a set of standard stars. The typical precision and accuracy of ground-based photometry in the past century is of the order of $\gtrsim$1% (Stubbs & Tonry 2006). Such a limit is sufficient for many applications but is somewhat lacking when compared with other physical quantities that are known with an accuracy of better than one part in a million (Young 1992a).

A significant step forward in the precision of ground-based photometry was obtained by modern digital panoramic surveys, such as the Sloan Digital Sky Survey (SDSS, York et al. 2000) or Pan-STARRS1 (PS1, Chambers et al. 2016). The acquisition of multi-colour photometry for many millions of stars over huge areas of the sky, with strictly the same setup and innovative techniques of photometric calibration, has allowed for the first time to achieve precision of <0.01 mag on an industrial scale. This achievement converted the de facto closed systems associated to these surveys into open systems, providing abundant standard stars with which to transform suitable observations taken outside the survey into the standard system that they define (see Huang & Yuan 2022, for a synthetic review and references on modern surveys and calibration techniques).

However, it is widely recognised (see e.g. Huang et al. 2021; Magnier et al. 2020a) that the all-sky, space-based, three-band photometry provided by the ESA space mission *Gaia* (Gaia Collaboration 2016) presents high-quality photometric measurements with photometric precision rivaling the best available, especially for wide sky coverage. In its Early Data Release 3 (EDR3) realisation (Riello et al. 2021), it effectively reaches submillimag precision in the range $10.0 \leq G \leq 17.0$ mag. Indeed, this exquisite degree of internal homogeneity has been used to significantly reduce residual systematic errors in the best set of SDSS standard stars (see e.g. Thanjavur et al. 2021; Huang & Yuan 2022).

The new *Gaia* Data Release 3 (*Gaia* DR3; Gaia Collaboration 2023a) provides – for the first time – internally (Carrasco et al. 2021; De Angeli et al. 2023) and externally calibrated (i.e. flux and wavelength calibrated; Montegriffo et al. 2023) very low resolution ($\lambda/\Delta\lambda \simeq 25-100$) spectra from the *BP* and *RP* spectrophotometers for about 220 million sources, mostly with $G < 17.65$ mag (see Fouesneau et al. 2023; De Angeli et al. 2023, for a complete list of sources with released *BP*/*RP* spectra). These spectra were used to infer astrophysical parameters, which are also released as part of *Gaia* DR3 (Creevey et al. 2023; Fouesneau et al. 2023; Andrae et al. 2023).

Another interesting product that can be obtained from externally calibrated[3] (EC) *BP* and *RP* (hereafter XP, for brevity) spectra is synthetic photometry. In principle, synthetic photometry can be obtained from EC XP spectra in any photometric system and for any passband enclosed in the spectral range covered by XP spectra (330−1050 nm) and whose characteristic width is larger than the line spread function (LSF) of XP spectra at the relevant wavelength[4]. In principle and in perspective, this may constitute a true revolution in optical photometry.

For the passbands of a given photometric system fulfilling the above conditions, we can get all-sky space-based photometry for all the sources for which XP spectra are available, in terms of magnitudes and flux in physical units. This is limited to $\simeq$220 million sources in *Gaia* DR3 but will amount to the entire *Gaia* data set in future releases ($\sim$2 billion sources down to $G \simeq 20.5$ mag). The relative flux scale relies on the precision of the EC XP spectra, while the absolute flux scale is based on the *Gaia* grid of SpectroPhotometric Standard Stars (SPSS, Pancino et al. 2021, and references therein).

[1] We can also refer to an open system as a standard system, as it offers the opportunity to standardise a given magnitude scale, transforming instrumental magnitudes into magnitudes in the desired open system.

[2] Transformation of instrumental magnitudes into a standard system is often referred to as absolute photometric calibration. This is not completely unjustified as, in principle, it is a process transforming magnitudes in an arbitrary scale into magnitudes in a standard system that in turn can be (and in most cases is) tied to physical fluxes by spectrophotometry of some of its standard stars.

[3] Meaning, in this context, "flux-calibrated using spectrophotometric data external to the direct data product of the *Gaia* satellite", namely the Spectro Photometric Standard Stars by Pancino et al. (2021), see Montegriffo et al. (2023)

[4] Parametrised here as the ratio $Rf$ between the full width at half maximum (FWHM) of the passband and of the XP Line Spread Function. It is useful to anticipate here the (conservative) criterion found in Appendix B: flux-conserving SP from XP spectra in a given photometric band can be achieved (also in presence of a strong spectral feature) if $Rf \geq 1.4$. See Appendix B for additional details and discussion.

Therefore, in principle, synthetic photometry from XP spectra (XP Synthetic Photometry, XPSP, hereafter) can supply absolute optical photometry for hundreds of millions of stars in any suitable system over the entire sky, thus for example transforming any closed system into an open system (albeit limited by the exact knowledge of the TCs). This in turn can provide, among the various possibilities: (a) the basis for the validation and/or re-calibration of existing photometric surveys; (b) the basis for validation and/or calibration of *future* photometric surveys from the ground or from space; and (c) the opportunity to experiment with the performance of a photometric system on a huge data set of real data on real sources before its actual realisation. As we show in the following, the potentiality of the method and of the product have not yet been fully realised because of systematic errors depending on spectral type that still affect EC XP spectra. The present contribution should be considered as one step in a process that is designed to maximise our exploitation of *Gaia* spectrophotometric data and will be continued in future data releases.

This paper is intended to illustrate how to get synthetic photometry from *Gaia* DR3 data. We showcase the performance of the synthetic photometry that can be currently obtained from XP spectra and outline its limitations. We also show a few examples of possible applications, and provide a few general-use photometric catalogues from XPSP, which are publicly accessible through the *Gaia* mission archive or other public repositories.

An obvious internal application made possible by XPSP is to provide additional means for validation of the EC XP spectra by comparison with huge external sets of high-quality photometry (SDSS, PS1, etc., see below and Montegriffo et al. 2023). For example, Montegriffo et al. (2023) demonstrated that $H_{\rm p}$, $B_{\rm T}$, and $V_{\rm T}$ HIPPARCOS photometry (van Leeuwen et al. 1997), which is considered a benchmark of precision (Bessell 2005), is reproduced by XPSP with typical accuracy of better than 2.5 millimag over the entire sky (see their Fig. 44). Similarly, we can provide some cross-validation with the results of DPAC/CU8[5] (Creevey et al. 2023; Fouesneau et al. 2023), by treating the same observational material in a completely different way.

The paper is organised as follows. In Sect. 2 we illustrate our formalism, starting from the representation of XP spectra in the *Gaia* context (De Angeli et al. 2023). We also introduce the concept of standardisation within the XPSP context. In Sect. 3 we show the performance of XPSP for widely used wide-band photometric systems, including the effects of standardisation. We deal separately with TCs including the spectral range $\lambda \leq 400$ nm, as this is particularly critical for XP spectra and requires special treatment. In Sect. 4 we show some examples of XPSP using medium-width and narrow passbands, including emission line photometry. We also illustrate the case of a photometric system brought into life for the first time by means of XPSP, the *Gaia* C1 system (Jordi et al. 2006). In Sect. 5 we present some example of performance verification in a scientific context and in Sect. 6 we illustrate the XPSP products offered to the general user in *Gaia* DR3, namely tools to get XPSP in the preferred system of the user and value-added catalogues. In Sect. 7 some caveats and recommendations for best use are reported. Finally, in Sect. 8 we summarise our results and discuss perspectives and developments of XPSP for future *Gaia* data releases. For increased readability, we collect some figures, tables, and discussions relevant to the quantitative understanding of the performance of XPSP and for its actual use and provide these in a series of appendices at the end of the paper. A list of the principal *Gaia*-related acronyms used throughout the paper is presented in Table I.1. Finally, as synthetic photometry can only be obtained from externally calibrated spectra, in the following we often drop the EC label for brevity, referring to the EC XP spectra used to get synthetic magnitudes simply as XP spectra.

[5] The Data Processing and Analysis Consortium (DPAC) is the consortium responsible of the processing of data from the *Gaia* mission (Gaia Collaboration 2016). It is structured in Coordination Units (CUs), each dealing with a specific subsystem of the processing system. The core mission of CU8 is to provide astrophysical parameters (AP) of the sources in the *Gaia* catalogue.

## 2. Methods

Synthetic photometry is based on the computation of a properly normalised mean flux (as defined in Bessell & Murphy 2012) obtained by integrating the product of a transmission curve $S(\lambda)$ and a spectral energy distribution (SED) over a given wavelength or frequency interval (depending on the photometric system definition). Following Bessell & Murphy (2012), for the photometric systems considered in this work, the mean flux can be expressed as

$$\langle f_\lambda \rangle = \frac{\int f_\lambda(\lambda)\, S(\lambda)\, \lambda\, \mathrm{d}\lambda}{\int S(\lambda)\, \lambda\, \mathrm{d}\lambda} \tag{1}$$

in VEGAMAG and Johnson-Kron-Cousins systems, and

$$\langle f_\nu \rangle = \frac{\int f_\lambda(\lambda)\, S(\lambda)\, \lambda\, \mathrm{d}\lambda}{\int S(\lambda)\, (c/\lambda)\, \mathrm{d}\lambda} \tag{2}$$

in the AB system (see Fukugita et al. 1996; Bessell 2005; Sirianni et al. 2005, and references therein).

In this work we express wavelengths $\lambda$ in units of nanometres (nm), energy flux per wavelength units $f_\lambda$ in units of W m$^{-2}$ nm$^{-1}$, and energy flux per frequency units $f_\nu$ in units of W m$^{-2}$ Hz$^{-1}$. $S(\lambda)$ designs a photonic response curve (i.e. it includes the quantum efficiency curve of the CCD).

The synthetic flux can be converted into a magnitude by

$$\mathrm{mag} = -2.5 \log\langle f_{\lambda|\nu} \rangle + \mathrm{ZP}, \tag{3}$$

where the zero point (ZP) in VEGAMAG is computed with respect to a reference SED:

$$\mathrm{ZP} = +2.5 \log\langle f_\lambda^{\mathrm{ref}} \rangle + V, \tag{4}$$

and in the AB case is

$$\mathrm{ZP} = -56.10. \tag{5}$$

VEGAMAG and Johnson-Kron-Cousins only differ in the choice of the reference SED: in the first case, we adopt the same reference as Riello et al. (2021), that is, an unreddened A0V star with $V = 0.0$, while in the other case we assume the Alpha Lyrae SED provided by Bohlin (2014) and $V = 0.03$ mag as reference.

In this context, *Gaia* EC XP spectra are no exception, and synthetic fluxes and magnitudes can be derived as described above. However, in the *Gaia* Archive, the XP spectra are stored as the projection on a set of basis functions, that is, as coefficients and corresponding covariance matrix. The SEDs (*BP* and *RP* separately) can then be reconstructed by linear combination of the bases, given the coefficients, as described in De Angeli et al. (2023) and Montegriffo et al. (2023):

$$f_\lambda^{\mathrm{XP}}(\lambda) = \sum_{i=1}^{N} b_i^{\mathrm{XP}} \phi_i^{\mathrm{XP}}(\lambda). \tag{6}$$

The two partially overlapping SEDs can be combined into a single distribution by computing a weighted mean with the weight for *BP* and *RP* given by:

$$w_{\rm BP}(\lambda) = \begin{cases} 1 & \text{if } \lambda < \lambda_{\rm lo} \\ 1 - \frac{\lambda-\lambda_{\rm lo}}{\lambda_{\rm hi}-\lambda_{\rm lo}} & \text{if } \lambda_{\rm lo} < \lambda < \lambda_{\rm hi} \\ 0 & \text{if } \lambda > \lambda_{\rm hi}, \end{cases} \tag{7}$$

$$w_{\rm RP}(\lambda) = 1 - w_{\rm BP}(\lambda), \tag{8}$$

where $[\lambda_{\rm lo}, \lambda_{\rm hi}]$ is the overlapping region (see Montegriffo et al. 2023, for further details). Combining Eq. (1) with Eq. (6), we obtain a very efficient algorithm to compute synthetic fluxes in a VEGAMAG system by means of the quantities:

$$s_i^{\rm XP} = \frac{\int w_{\rm XP}(\lambda)\, \phi_i^{\rm XP}(\lambda)\, S(\lambda)\, \lambda\, {\rm d}\lambda}{\int S(\lambda)\, \lambda\, {\rm d}\lambda}, \tag{9}$$

so that the mean synthetic flux of the source is simply given by

$$\langle f_\lambda \rangle = \sum_{i=1}^{N} b_i^{\rm BP} s_i^{\rm RP} + \sum_{i=1}^{N} b_i^{\rm RP} s_i^{\rm RP}. \tag{10}$$

The computation of Eq. (9) coefficients for an AB system is straightforward. In practice, a photometric system containing $K$ passbands is reduced to a set of *TC bases* composed of two $K \times N$ matrices $\mathbb{S}^{\rm BP}$ and $\mathbb{S}^{\rm RP}$; given a *Gaia* source with spectral coefficients $\boldsymbol{b}^{\rm BP}$ and $\boldsymbol{b}^{\rm RP}$, and covariances $K_{\rm bb}{}^{\rm BP}$ and $K_{\rm bb}{}^{\rm BP}$, the $K$ synthetic fluxes $\boldsymbol{f}$ in the photometric system are readily given by

$$\boldsymbol{f} = \mathbb{S}^{\rm BP} \cdot \boldsymbol{b}^{\rm BP} + \mathbb{S}^{\rm RP} \cdot \boldsymbol{b}^{\rm RP}. \tag{11}$$

A covariance matrix can be computed for fluxes $\boldsymbol{f}$ as:

$$K_{\rm ff} = \mathbb{S}^{\rm BP} \cdot K_{\rm bb}{}^{\rm BP} \cdot \mathbb{S}^{\rm BP^T} + \mathbb{S}^{\rm RP} \cdot K_{\rm bb}{}^{\rm RP} \cdot \mathbb{S}^{\rm RP^T}. \tag{12}$$

The nominal uncertainties on $\boldsymbol{f}$ fluxes are given by the square root of diagonal elements of $K_{\rm ff}$. In practice, for issues related to the uncertainties in the XP spectra (De Angeli et al. 2023; Montegriffo et al. 2023), these can be significantly underestimated. In Sect. 2.1 we derive empirical corrections to properly trace the uncertainty in synthetic fluxes and magnitudes. We computed the filter bases for a number of commonly used photometric systems, many of them discussed and validated below. XPSP in these and other systems can be obtained from the *Gaia* Archive as described in Sect. 6.1.

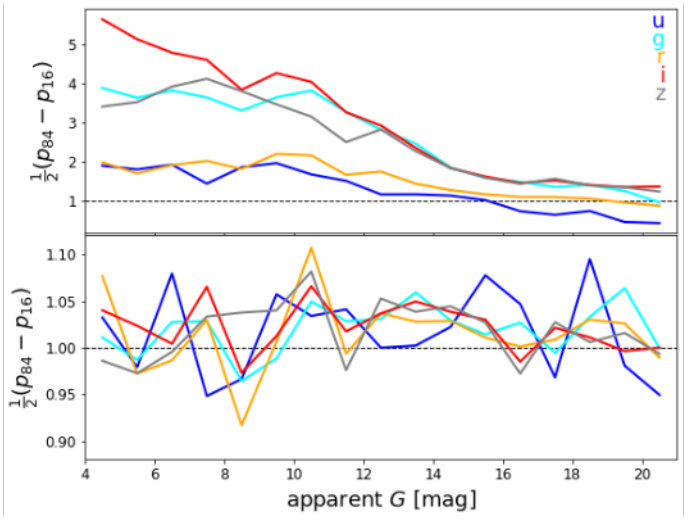


**Fig. 1.** Illustration of underestimated uncertainties for the standardised SDSS system. We summarise the underestimation as half the difference between the 84th and the 16th percentiles of flux differences (normalised by their combined nominal uncertainties) of randomly split sources falling into this apparent $G$ magnitude bin. If uncertainties are correctly estimated, this quantity should be 1, as indicated by the horizontal dashed line. *Top panel*: nominal uncertainties. *Bottom panel*: calibrated uncertainties (we highlight the very different $y$-axis range).

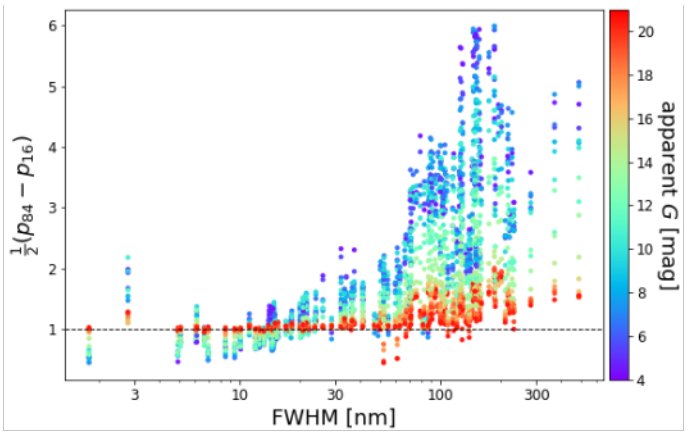


**Fig. 2.** Systematic underestimation of nominal uncertainties for synthetic fluxes as function of FWHM of each band in all photometric systems considered in this paper.

### 2.1. Empirical estimate of errors

In order to validate the uncertainty estimates for the passband fluxes, as derived from XPSP, we took a sample of 43 653 randomly selected sources covering a suitable range of colour and magnitude, and for each source we randomly split its epoch observations into two groups (hereafter '*BP*/*RP* split-epoch validation dataset'; for further details see De Angeli et al. 2023). We then compute two separate mean *BP* and *RP* spectra and their resulting synthetic fluxes for every pair. This procedure results in two statistically independent measurements for each source, which should be consistent within their respective uncertainty estimates. We emphasise that the randomised grouping of epoch observations is essential because it prevents any potential intrinsic time variability of a source from compromising the uncertainty validation.

As expected, this test revealed that the nominal uncertainty estimates of the synthetic fluxes are systematically underestimated for most photometric systems (see De Angeli et al. 2023, for a discussion on the underestimation of errors in the underlying XP spectra). In such cases, the distributions of flux differences within a pair of randomly split sources normalised by their combined uncertainties would be substantially broader than a unit Gaussian. In particular, we notice that this underestimation of uncertainties appears to depend on the apparent $G$ magnitude of a source. We illustrate this for the example of the standardised SDSS system (see Sect. 3.1) in the top panel of Fig. 1. Here, we clearly see that the distribution of normalised flux differences in the randomly split sources is broader than a unit Gaussian, because half the difference between the 84th and 16th percentiles is larger than 1. We also see that this underestimation of uncertainties has a different effect from one passband to another; the underestimation appears to be stronger for broader synthetic bands, as is evident from Fig. 2, yet we did not observe any dependence on the wavelength of the band.

In order to calibrate the uncertainty estimates for the synthetic fluxes, for each band in every photometric system

considered in this paper we tabulate the factors by which the distributions of normalised differences are too high, as a function of apparent $G$ magnitude (see top panel of Fig. 1). The calibrated uncertainties are then obtained by inflating the nominal uncertainties for every source according to the tabulated factors by which they are found to be too small. Again, this is illustrated in the bottom panel of Fig. 1 for the example of the standardised SDSS system. Evidently, the calibrated uncertainties now fully account for the flux differences in the pairs of randomly split sources.

We note that the Python software tool to deal with XP spectra, GaiaXPy (Sect. 6.1 and De Angeli et al. 2023), provides by default the nominal uncertainties for the synthetic fluxes, which are underestimated. However, it can optionally compute the calibrated uncertainties instead (by setting `error_correction=True`), for all the sets of passbands currently included in the GaiaXPy repository. Please refer to the GaiaXPy documentation (link included in Sect. 6.1) for instructions and for a full list of the systems for which this is available.

### 2.2. Standardisation

Externally calibrated XP spectra are known to suffer from systematic errors attributable to various factors (see Montegriffo et al. 2023). These issues manifest as systematic differences between XPSP magnitudes and the corresponding magnitudes of top-quality external sources that are taken as reference for a given photometric system (e.g. sets of primary and/or secondary standard stars). In general, for wide-band XPSP, these effects amount ZP differences within a few hundredths of a magnitude and/or to trends as a function of colour with a maximum amplitude of a few hundredths of a magnitude over wide colour ranges (with the exception of ultraviolet (UV) bands, $\lambda < 400$ nm, which are discussed separately in Sect. 2.2.2; see Appendix G; see also Montegriffo et al. 2023).

In future data releases, once we are able to keep these systematic errors on EC XP spectra to a minimum, XPSP will directly serve to re-define optical photometric systems based on exquisitely homogeneous space-based spectrophotometry. However, in the meantime, users might be interested in reproducing the existing photometric systems at best, with currently available XPSP. This can be achieved by a process that we call standardisation, following Bessell (2005). In our context, standardisation consists in (a) adopting an external photometric dataset as the reference set of standards for a given photometric system, such as SDSS Stripe 82 photometric standard stars (Ivezić et al. 2007; Thanjavur et al. 2021), (b) comparing the XPSP magnitudes for these standard stars ($\mathrm{mag_{synth}}$) with those from the reference source ($\mathrm{mag_{phot}}$) as a function of magnitude and colour, and (c) finding a correction that, when applied to XPSP magnitudes, minimises the differences ($\Delta\mathrm{mag} = \mathrm{mag_{phot}} - \mathrm{mag_{synth}}$), thus providing the best reproduction of the external system.

When dealing with pure magnitudes, that is, when the product of the measuring apparatus is an estimate of the integrated flux of the source through the considered TC, this kind of standardisation is typically achieved by means of polynomial transformation as a function of colour. As mentioned above, these may suffer from strong systematic effects, for example because a polynomial may not be adequate to model all the subtleties of the relations between the two systems. This kind of problem can be mitigated if the set of magnitudes to be transformed is from synthetic photometry. In this case, the safest and most widely adopted way to standardise magnitudes is to tweak the profile of the TC adopted for synthetic photometry in order to minimise $\Delta$mag and its trends with colour (Bessell 2005). This process is designed to remove the small differences between the TC of the reference system and the one to be transformed, possibly taking into account the effects that would require high-order terms in a polynomial transformation[6].

In our specific case, the tweaking is mainly used to minimise the effects of the residual systematic errors of EC XP spectra on synthetic photometry using the external standards as a kind of second-level calibrator. As we see in Sect. 3, for wide passbands in the range $\lambda \gtrsim 400$ nm, standardisation allows us to reproduce existing systems with typical accuracy from a few millimag to submillimag, depending on the specific passband, over broad ranges in colour and for the large majority of well-measured stars with published XP spectra in *Gaia* DR3. In Sect. 2.2.1, we describe the way in which we get standardisation by TC tweaking, and how we deal with passbands in the range $\lambda \lesssim 400$ nm (Sect. 2.2.2).

#### 2.2.1. Standardisation: general method

Figure 3 shows residuals between standard and synthetic $i_{\mathrm{SDSS}}$ magnitudes obtained with the Doi et al. (2010) TCs for the SDSS reference dataset presented in Sect. 3.1. The figure provides an example illustrating all the effects that need to be corrected within the standardisation process. Residuals are plotted as a function of $G$ magnitude (upper panel) and $G_{\mathrm{BP}}-G_{\mathrm{RP}}$ colour (lower panel). In both cases, the continuous red curve traces the median ($P_{50}$) of the residual distribution computed in bins of 0.4 mag in width, while the dashed curves are the loci of the 15.87% ($P_{16}$) and 84.13% ($P_{84}$) percentiles. There is a clear trend as a function of magnitude that is common to all photometric systems. This is interpreted as being (mainly) due to systematic overestimation of the background, which produces a negative offset in measured XP fluxes (see De Angeli et al. 2023, for a detailed discussion). In the following, we refer to this general magnitude-dependent trend as the 'hockey-stick' effect, described and discussed in Montegriffo et al. (2023)[7]. Independently of the actual nature of this effect, which will be further investigated in preparation for future *Gaia* data releases, we find that it can be effectively mitigated by applying a background-like correction and, consequently, we adopted this approach in the standardisation process.

The presence of additional offsets in the magnitude scale cannot be excluded, but the median of the residuals in the range where the hockey-stick effect is minimised ($G \lesssim 15.5$ mag) constrains their amplitude to <0.01 mag. A selection in magnitude ($G < 17.65$ mag)[8] has been applied to data plotted as a function of colour in order to minimise the disturbance due to the hockey-stick effect and to better appreciate the small colour term present in the data (linear trend with $G_{\mathrm{BP}}-G_{\mathrm{RP}}$ colour).

The standardisation process is composed of two phases that can be iterated a few times. For each passband: (1) the flux offset $f_{\mathrm{bg}}$ to be added to synthetic fluxes for the removal of the hockey-stick is evaluated; and (2) the TC shape is tweaked to remove the colour term.

[6] The underlying hypothesis is that a TC should exist that removes all the systematic differences between the two sets of magnitudes, assuming that both accurately trace the original SED of the observed sources.
[7] In the context of *Gaia* photometry, the hockey-stick effect is mentioned for the first time in Evans et al. (2018). A realisation of the same effect we are dealing with here is shown in the top panel of Fig. 23 of Riello et al. (2021), and is briefly discussed there.
[8] This is the general magnitude limit for XP spectra in *Gaia* DR3, see Sect. 3.1.

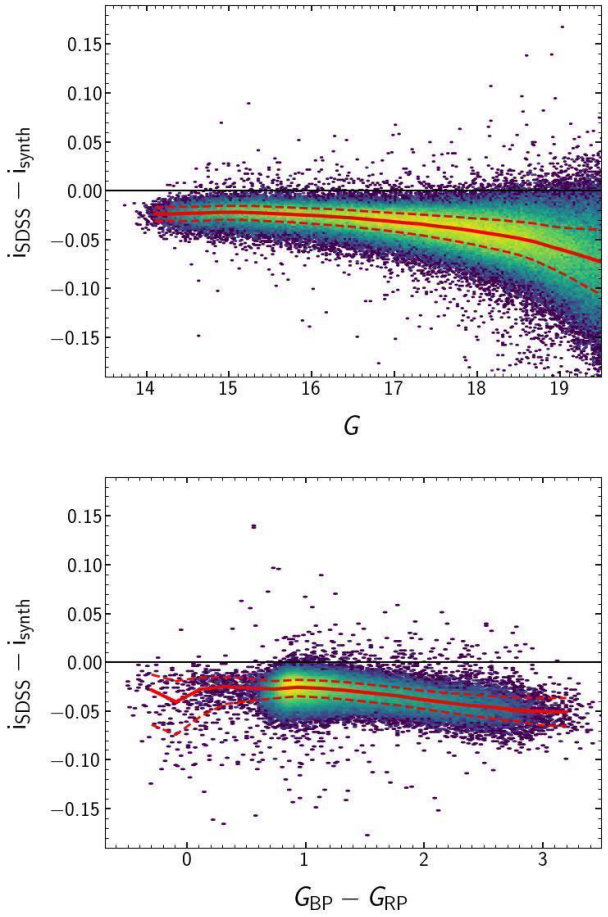


**Fig. 3.** Residuals between reference and synthetic magnitudes computed through a nominal filter transmission curve (Doi et al. 2010) for a set of standard stars plotted as a function of $G$ magnitudes (*top*) and $G_{BP}$−$G_{RP}$ colour (*bottom*). The red continuous curves represent a smoothed median line of the data, while the dashed curves trace the 16th and 84th percentile of the distribution of residuals.

To minimise the entanglement of the two effects, we perform process (1) on a subsample of available data by selecting a restricted $G_{BP}$−$G_{RP}$ colour range (to minimise disturbance due to the colour term) while process (2) is performed on a subsample selected in magnitude, avoiding fainter stars which are more affected by the background issue. Finally, we evaluate a correction factor for the zero point $ZP_{std}$ in order to mitigate any residual grey offset.

A standardised photometric system thus consists in a new set of basis functions $(\mathbb{S}^{BP}, \mathbb{S}^{RP})_{STD}$ computed with the tweaked TC, an array of flux offsets $\boldsymbol{f}_{bg}$ to be added to synthetic fluxes of Eq. (11),

$$\boldsymbol{f}_{STD} = \boldsymbol{f} + \boldsymbol{f}_{bg}, \tag{13}$$

and the array of ZP correction factors to be included in Eq. (3),

$$\mathrm{mag}_{Std} = -2.5\log\langle f_{Std}\rangle + \mathrm{ZP} + \mathrm{ZP}_{STD}, \tag{14}$$

where all the involved vectors have one component for each passband of the considered system. The evaluation of the background offset can be achieved only if available standards span a sufficiently wide range in magnitude (it must roughly cover from $G \simeq 13$ mag to $G \gtrsim 18$ mag). We typically select standards with colours within ≃0.5 mag of $G_{BP}$−$G_{RP} \simeq 1.0$ mag; data are then partitioned in magnitude bins of ≃0.5 mag. For each bin, we evaluate the median of the differences,

$$\mathrm{mag}_{phot} - \mathrm{mag}_{STD}, \tag{15}$$

which are arranged in the array $\boldsymbol{P}_{50}$. The background correction $f_{bg}$ is found as the value that minimises the cost function,

$$\rho = \sum_i (P_{50_i} - < \boldsymbol{P}_{50} >)^2 . \tag{16}$$

To implement the filter-tweaking algorithm, we model the shape of the standardised filter response by multiplying the nominal transmission $S(\lambda)$ with a linear combination of a (low) number of basis functions $\mathcal{S}_k$:

$$S^{\dagger}(\lambda) = S(\lambda) \cdot \sum_k \alpha_k \mathcal{S}_k(\lambda). \tag{17}$$

The basis functions used for the present work include mainly Legendre polynomials and Hermite functions. An important issue to keep in mind is that this method has an intrinsically low sensitivity: large variations in the shape of the filter may result in very small changes in the residuals, meaning that there is no unique solution to the problem. When several models give comparable results, we arbitrarily select TCs with shapes closer to the nominal one. The procedure for the optimisation of the model $S^{\dagger}(\lambda)$ is similar to that described for the correction of the hockey-stick effect: (1) We select calibrators with a $G$ magnitude brighter than a given value (depending on the specific data set); (2) we partition data in $G_{BP}$−$G_{RP}$ colour bins of ∼0.2 mag; (3) for each bin, we compute the median $P_{50}$ and the width $\sigma = 0.5(P_{84}-P_{16})$ of the distribution of the difference $(\mathrm{mag}_{phot}-\mathrm{mag}_{Std})$; and (4) the model is optimised by minimising the cost function

$$\rho = \sum_i \left(P_{50_i}^2 + \sigma_i^2\right), \tag{18}$$

where the $\sigma_i$ terms have been included as they were found to be effective in preventing odd solutions of the standardisation process that were sometimes found to arise. In all the cases considered here, the changes of the TC shapes induced by the standardisation are small; a typical example is shown in Fig. 4.

As a final remark, while the $f_{bg}$ values we derive are representative of the conditions of the adopted reference samples, which are typical uncrowded field stars, we cannot guarantee their universal validity, because we have not been able to test their possible variation as a function of position in the sky, local stellar density, and so on. However, in Sect. 3.5 we use a reference sample where the crowding conditions are significantly poorer than in the typical reference sample (as e.g. in those described in Sects. 3.1 and 3.2), and we verified that the $f_{bg}$ values estimated for broadly similar passbands in the different cases are similar, with typical differences of ≲20%.

### 2.2.2. The case of UV bands

As anticipated above and discussed in detail in Montegriffo et al. (2023), the strongest colour-dependent systematic errors affecting EC XP spectra occur in the spectral range $\lambda \lesssim 400$ nm, where the TC of the *BP* spectrophotometer is low and highly structured, with two very steep branches found at around 390 nm and at the blue cut-off at ≃330 nm, and two local maxima at $\lambda \simeq 338$ nm and ≃355 nm (see Fig. 5). In the following, for brevity, we refer to passbands whose predominant part of the spectral range is below 400 nm (and typically ≳300 nm) as UV bands.

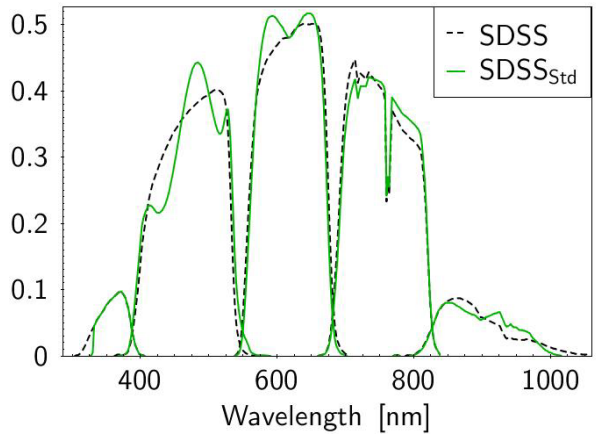


**Fig. 4.** Original SDSS transmission curves from Doi et al. (2010, black dashed line) are compared to their tweaked version obtained with the standardisation process (green continuous lines). It is important to remember that the shape of standardised TCs is designed to correct for the systematic errors that still affect EC XP spectra. No tweaking is applied to the *u* passband, as the standardisation of *u* magnitudes and fluxes is performed by means of polynomial transformation. The cut at 330 nm follows the TC of the *BP* spectrometer.

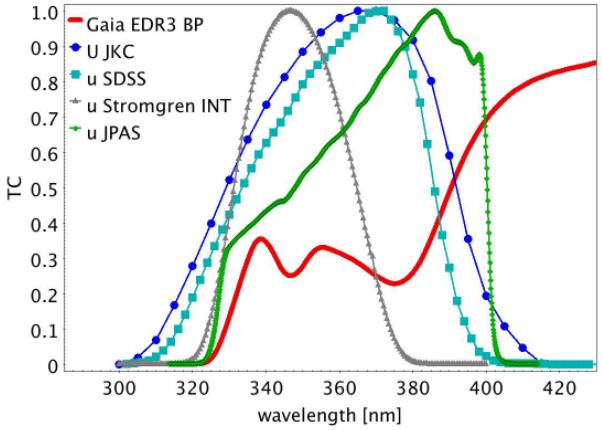


**Fig. 5.** Transmission curves of all the UV bands considered in this paper are compared with the transmission curve of the *BP* spectrometer. All the curves are normalised to their maximum.

The most widely used UV bands (a) span this spectral window, with most of the throughput in the region bluer than ≃375 nm, which is especially critical for XP spectra, and (b) have a blue cut-off exceeding the blue limit of *BP* (Fig. 5). Therefore, reproducing the photometry in these passbands with XPSP is quite challenging, with factor (b) effectively preventing the possibility of a full standardisation[9].

However, as that region of the stellar spectra is especially important and informative, we attempt a standardisation of SDSS *u* and Johnson-Kron-Cousin's (JKC, hereafter, as defined by Landolt 1992, standard stars) *U* bands. In these cases we were not able to obtain satisfactory standardisations by tweaking the TCs and we used high-degree colour-dependent polynomial transformations instead. Moreover, as the adopted solution does not provide satisfactory results over the whole *Gaia* DR3 sample of XP spectra, the use of standardised $u/U$ magnitudes is recommended only for a subset limited in signal-to-noise ratio (S/N; see Sects. 3.3 and 6.2).

As a first step we produced new passbands identical to the original ones (from Doi et al. 2010 for $u_{\rm SDSS}$ and from Bessell & Murphy 2012 for $U_{\rm JKC}$) but valued 0.0 for $\lambda <$ 330 nm. We then proceeded in a similar way as for the non-UV passbands. The hockey-stick correction was obtained, taking special care to minimise the effect of the large colour terms at work in this case. The median and $\sigma$ of the resulting residuals as a function of colour computed over bins were then fitted with high-order polynomials. We find that adopting colours from the same system as the considered UV bands provides simpler and more robust solutions, and therefore the polynomials are a function of (synthetic and non-standardised) $g-i$ and $B-V$ for $u_{\rm SDSS}$ and $U_{\rm JKC}$, respectively. The public tool to manage XP spectra (GaiaXPy, see Sect. 6.1) will allow the user to produce both raw and standardised XPSP (for the standardised systems), independently of the method adopted, that is, polynomial transformation for UV bands and TC tweaking for all the other cases.

The actual performance of the standardised version of the two UV bands considered in this section is discussed in Sect. 3.3 and Appendix G, while the recommendations for safe use are shown and discussed in Sects. 6.2 and 7.

## 3. Wide band synthetic photometry

In this section we illustrate the performance of XPSP in reproducing the photometry of existing and widely used wide-band photometric systems. We also show how residual inaccuracies are reduced below the 1% level by the process of standardisation (described in Sect. 2) with respect to selected sets of reliable photometric standard stars. To illustrate the process, we treat the cases of the SDSS and JKC systems more extensively, while for the other standardised systems, some of the relevant plots and tables are collected in Appendix G. Some experiments of validation using stellar models are also reported in Appendix C.

### 3.1. SDSS system and its standardisation

The Sloan Digital Sky Survey (SDSS; York et al. 2000) was the first modern digital survey producing precise photometry over a large portion of the Northern Sky. Its photometric system, defined in Fukugita et al. (1996), established a new standard, now widely used in Galactic and extra-galactic astronomy (see Ivezić et al. 2007; Thanjavur et al. 2021, and references therein).

As a reference set for the SDSS system, we used a selected subsample of the Stripe 82 standard stars recently presented and discussed by Thanjavur et al. (2021, T21 hereafter). Compared to the previous realisation of the same set (Ivezić et al. 2007), T21 has two to three times more epochs per source used in photometric averaging; systematic photometric zero-point errors as functions of RA and Dec are estimated and corrected for using *Gaia* EDR3 photometry[10]; and the same is used to correct *ugiz*

[9] The information in the spectral range $\lambda \lesssim 330$ nm is not present in the XP spectra and no correction can help to recover it. Hence, in cases where significant star-to-star differences in that wavelength range occur, the standardised $UV_{\rm synth}$ magnitudes cannot adequately reproduce $UV_{\rm phot}$ ones.

[10] It is important to note that, as a consequence, any spatial trend of the photometric zero-points in *Gaia* EDR3 should have been transferred to the T21 photometry. However, we also note that (a) when comparing standardised XPSP photometry with T21 we find residual trends of amplitude $\lesssim$10.0 mmag as a function of position, and (b) the comparison of XPSP photometry with HIPPARCOS photometry presented in Montegriffo et al. (2023) suggests that XPSP photometry should be spatially homogeneous to the level of a few mmag over most of the sky. This may suggest that spatial trends were not completely removed from T21 photometry.

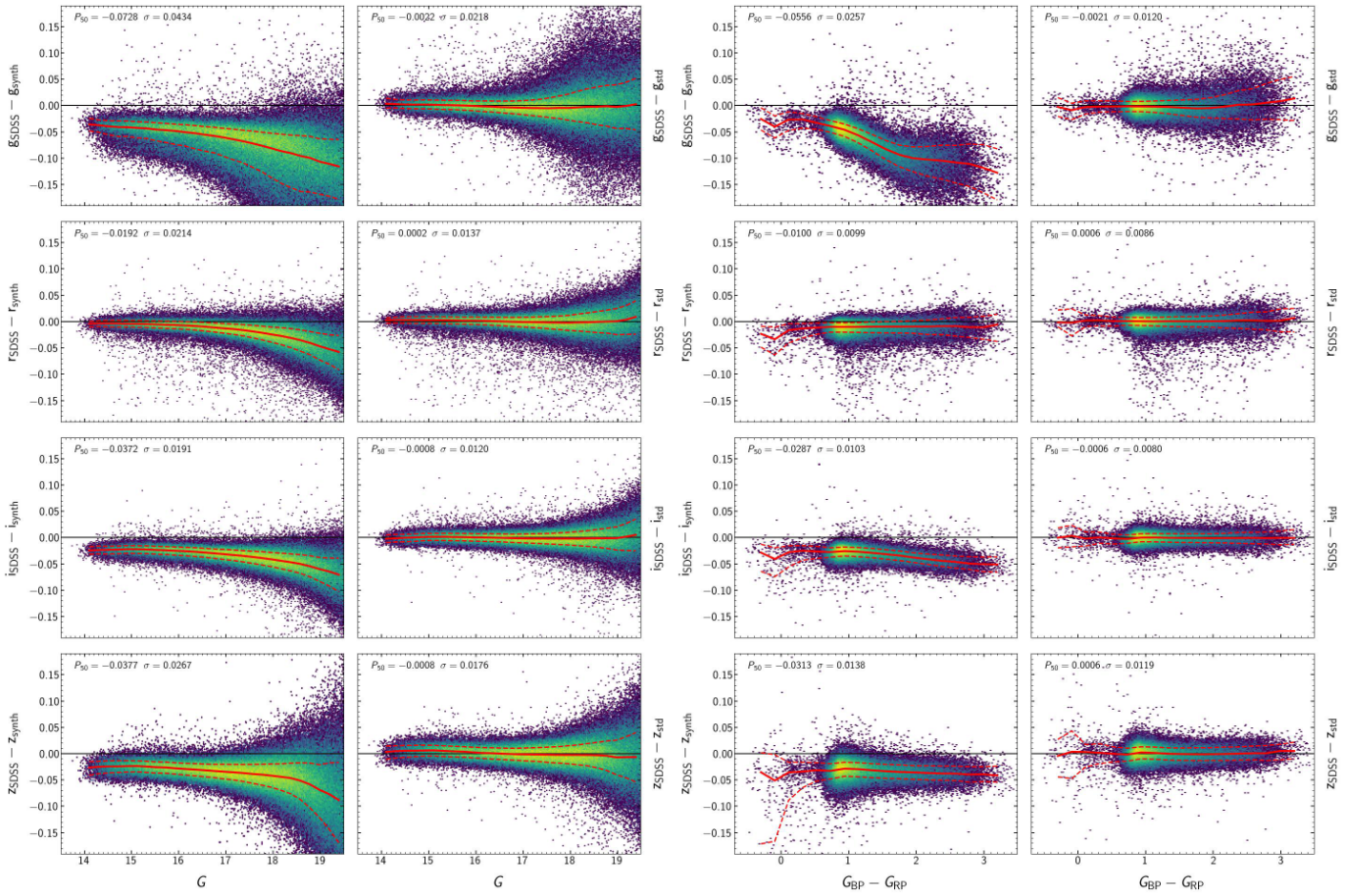


**Fig. 6.** Performance and standardisation of SDSS *ugriz* XP synthetic magnitudes using the selected subset of the Thanjavur et al. (2021) sample, which we adopted as reference. *Left set of panels*: Δmag as a function of $G$ magnitude for the entire sample using nominal XP synthetic magnitudes (*left panels*) and standardised XP synthetic magnitudes (*right panels*). In each panel, the continuous red line connects the median Δmag computed in 0.2 mag wide bins, and the dashed red lines connect the loci of the 15.87% ($P_{16}$) and the 84.13% ($P_{84}$) percentile computed in the same bins. The median ($P_{50}$) and the difference between $P_{84}$ and $P_{16}$ – here used as a proxy for the standard deviation $\sigma$ – for the entire sample are reported in the upper left corner of each panel. *Right set of panels*: Same for Δmag as a function of $G_{BP}-G_{RP}$ colour, limited to the subsample of reference stars with XP spectra released in *Gaia* DR3.

magnitudes relative to the $r$-band. This approach results in random photometric errors approximately 30% smaller than in the I07 catalogue and below ≈0.01 mag for stars brighter than 20.0, 21.0, 21.0, 20.5, and 19.0 mag in $u$, $g$, $r$, $i$, and $z$-bands, respectively.

To obtain our reference set to be used for comparison and standardisation of XP photometry, we cross-matched the EDR3 sources with XP spectra to the T21 sample and applied the following quality filters on *Gaia* data:

– `XP_num_of_transits` >= 15,
– `XP_num_of_contaminated_transits/XP_num_of_transits` < 0.1,
– `XP_num_of_blended_transits/XP_num_of_transits` < 0.1,
– `XP_number_of_neighbours`< 2,
– `XP_number_of_mates` < 2,
– `XP_number_of_visibility_periods_used` > 10,

where `XP` stands for *Gaia BP* and *RP*. A set of broad quality filters was applied on parameters from the T21 sample as well:

– `{u,g,r,i,z}Nobs` > 4,
– `{g, r, i}msig` $\cdot \sqrt{\texttt{\{g,r,i\}Nobs}} < 0.03$.

Detailed explanations of the used columns can be found in the Gaia DR3 documentation and the SDSS Data Model. The final reference sample includes approximately 280 879 sources. For the synthetic photometry, we adopt the official SDSS TCs from Doi et al. (2010).

Figure 6 is a good example of the typical plot with which we illustrate the performance of XPSP in reproducing the photometry of the external set adopted as a reference, for the various photometric systems. Two multi-panel sets of diagrams are presented, the set on the left showing Δ mag as a function of $G$ magnitude, and the set on the right showing Δ mag as a function of $G_{BP}-G_{RP}$ colour. Within each of the two sets, the left column displays the comparison with raw XPSP magnitudes before standardisation, while the right columns show the comparison after standardisation. In each panel, the continuous red line is the median ($P_{50}$) of the Δ mag distribution computed over independent bins of 0.4 mag in width, while the dashed red lines trace the 15.87% ($P_{16}$) and the 84.13% ($P_{84}$) percentiles computed in the same bins. It is important to recall that, in this figure, as well as in all other analogous figures for other systems shown below, if not otherwise stated, the plots as a function of magnitude refer to the entire reference sample, including stars fainter than $G$ = 17.65 mag, which in general do not have their XP spectra released. This is required to adequately constrain the hockey-stick effect in order to correct for it in the process of standardisation. On the other hand, the plots as a function of colour refer only to the subsample with $G$ < 17.65 mag in order to better

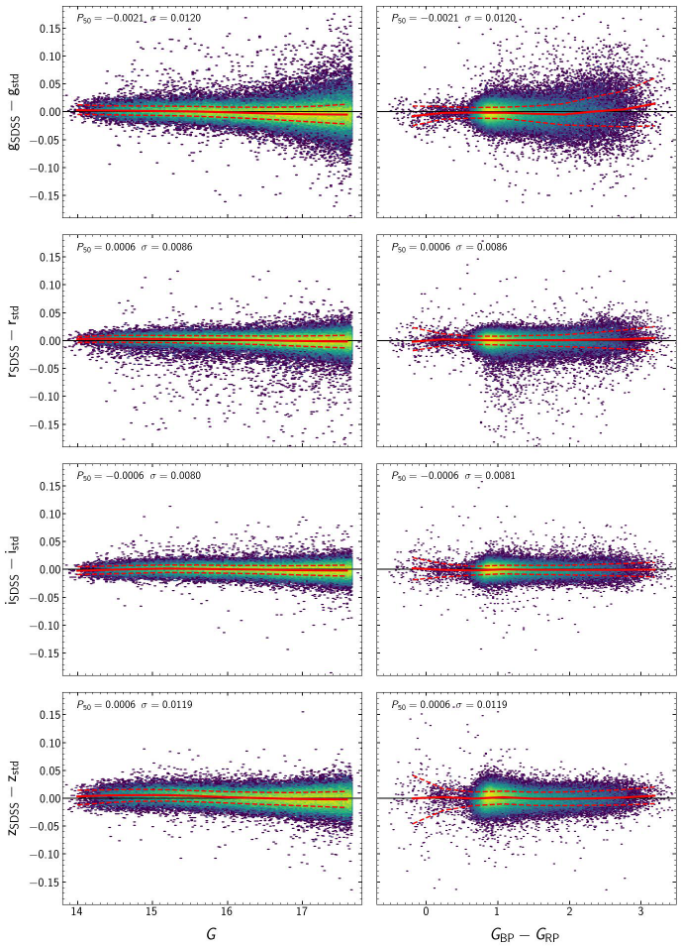


**Fig. 7.** Performances of standardised XPSP in the SDSS system (*griz*). We show Δmag as a function of $G$ magnitude (*left panels*) and $G_{BP}-G_{RP}$ colour (*right panels*) for the subsample of T21 stars whose XP spectra have been released in *Gaia* DR3. The arrangement and the meaning of the symbols is the same as in Fig. 6.

trace genuine colour terms, minimising the additional noise due to the hockey-stick effect.

The left rows of the two panels of Fig. 6 show the performances of raw XPSP in reproducing SDSS magnitudes. The deviation that is apparent for $G \gtrsim 16.0$ mag in the diagrams as a function of $G$ magnitude is due to the hockey-stick effect. Taking this factor into account, we conclude that *riz* photometry is reproduced remarkably well, with zero-point differences of <0.02 mag (as traced by $G \lesssim 15.5$ mag stars) and colour terms with amplitudes of ≲0.02 mag over the whole colour range covered by the reference sample. On the other hand, $\Delta g$ displays a colour term with an amplitude of ≃0.05 mag, and also produces a larger and asymmetric scatter about the median in the plot as a function of magnitude with respect to the other passbands. This reflects the coverage by $g$ band of regions of the XP spectra that suffer from colour-dependent systematic errors, including the first sudden drop of the *BP* TC around 390 nm. A fully analogous behaviour is observed for PanSTARRS $g$ (Appendix G; see also the case of JKC $B$ band discussed below), confirming that XP spectra are to be blamed for the mismatch. Standardisation significantly reduces all the discrepancies described above, as can be readily appreciated from the direct comparison between standardised and non-standardised $\Delta g$ distributions shown in Fig. 6. A larger scatter about the median remains in the $g$ band than in the redder passbands, and the scatter in $z$ is slightly greater than in $r$ and $i$, while in the latter passbands the performance of standardised SP appears to be excellent.

Figure 7 shows the final result of the standardisation process for the stars of the reference sample whose XP spectra are released in *Gaia* DR3, that is, those for which XPSP can be obtained. This figure shows the excellent quality of the final product[11]. The median difference over the entire subsample is <2.5 mmag for all the passbands considered here, and the standard deviation $\sigma$[12] is ≲12 mmag.

$P_{50}$, $P_{16}$, and $P_{84}$ values from Fig. 7 are listed as a function of $G$ magnitude in Table 1. The median Δmag are within ±6.0 mmag for all the considered passbands and for the entire magnitude range sampled, and are typically lower than ±3.0 mmag in wide ranges of magnitudes, especially in $r$ and $i$ bands. The typical scatter, as parametrised by $\sigma = 0.5(P_{84}-P_{16})$, amounts to ≲10 mmag down to $G = 16.5$ mag, for *riz*.

The adopted reference sample is dominated by dwarf stars, and almost completely lacks giants redder than $G_{BP}-G_{RP} = 1.5$. In Appendix G we test our standardised XPSP in the SDSS system against a selected sample of giant stars reaching $G_{BP}-G_{RP} = 3.5$, showing that it is accurate within ≃±10 mmag for these stars as well.

We verified that the bulk of the Δmag distributions are very similar to Gaussian curves. However, a few outliers, for example those with Δmag > 50 mmag at any $G$, can be noted in all the panels of Fig. 7. We explored whether or not some quality parameter included in the *Gaia* Archive correlates with these outliers. Our far-from-exhaustive exploration led to the conclusion illustrated in Fig. 8, shown as an example. Many of the outliers have $|C^\star| > 0.1$ (Riello et al. 2021). In this sample, there was no source with `phot_variable_flag = VARIABLE` from the `gaiaedr3.gaia_source` table, but in other cases we verified that sources classified as VARIABLE according to this flag account for several outliers in Δmag (see e.g. Fig. G.4).

This was found to be the case for all the passbands in all the photometric systems we tested in this way. We therefore conclude that the main reasons for anomalous individual inaccuracies in XPSP magnitudes are (a) contamination from nearby sources or, in any case, issues traced by $C^\star$, and (b) mean spectra obtained by combining epoch spectra of a variable source. It is worth noting that the majority of high-$C^\star$ outliers lie on the same side of the Δmag distribution, either preferentially positive or negative, as in Fig. 8, depending on the considered passband (see Appendix G, for an example).

[11] There is some redundancy between this figure and the right columns of panels of Fig. 6, as well as in analogous sets of figures produced for other photometric systems. Still we feel that it is worth showing both kinds of plots, as those as in Fig. 6 illustrate the comparison with raw XPSP and the effect of standardisation, while those as in Fig. 7 give a direct view of the XPSP perfomance for the material that is actually made available in *Gaia* DR3, from the *Gaia* archive.

[12] Estimated as half of the difference between the 84.13% and the 15.87% percentiles of the distribution of Δmag. In the following, we refer to this quantity as $\sigma$, if not otherwise stated, for brevity.

**Table 1.** SDSS system: median ($P_{50}$), 15.87% ($P_{16}$), and 84.13% ($P_{84}$) percentiles of the $\Delta$mag distributions of Fig. 7.

| $G$ (mag) | $P_{50}(\Delta g)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $P_{50}(\Delta r)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $P_{50}(\Delta i)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $P_{50}(\Delta z)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $n_\star$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 14.0 | 2.7 | −5.0 | 12.6 | 2.7 | −3.6 | 7.1 | −2.4 | −9.2 | 4.7 | 3.2 | −12.0 | 14.3 | 198 |
| 14.4 | 2.2 | −5.4 | 10.6 | 3.3 | −2.9 | 8.8 | −0.8 | −8.2 | 6.0 | 5.0 | −7.8 | 14.9 | 2208 |
| 14.8 | 1.6 | −6.3 | 10.3 | 3.0 | −3.5 | 9.3 | 0.5 | −6.6 | 7.2 | 5.5 | −5.6 | 15.9 | 4827 |
| 15.2 | 1.1 | −7.0 | 10.0 | 2.8 | −3.6 | 9.0 | 1.5 | −5.6 | 7.9 | 5.5 | −5.2 | 15.2 | 7550 |
| 15.6 | 0.3 | −8.2 | 9.5 | 2.1 | −4.7 | 8.6 | 0.9 | −6.1 | 7.8 | 4.0 | −6.6 | 13.4 | 10 406 |
| 16.0 | −1.0 | −10.3 | 8.3 | 1.6 | −5.6 | 8.4 | 0.3 | −7.0 | 7.1 | 1.9 | −8.5 | 11.7 | 11 724 |
| 16.4 | −2.6 | −12.7 | 8.1 | 0.7 | −7.1 | 8.2 | −0.2 | −7.8 | 7.1 | 0.4 | −10.7 | 10.7 | 14 678 |
| 16.8 | −3.9 | −15.5 | 8.2 | −0.5 | −9.1 | 7.9 | −1.2 | −9.3 | 6.5 | −1.0 | −12.6 | 10.0 | 18 968 |
| 17.2 | −4.7 | −18.8 | 10.0 | −1.1 | −11.2 | 8.5 | −1.8 | −10.8 | 6.9 | −1.8 | −14.7 | 10.8 | 22 809 |
| 17.6 | −5.1 | −22.1 | 13.4 | −1.6 | −13.3 | 9.4 | −2.1 | −12.0 | 7.7 | −2.2 | −16.7 | 12.1 | 16 586 |

**Notes.** $n_\star$ is the number of sources in the considered bin.

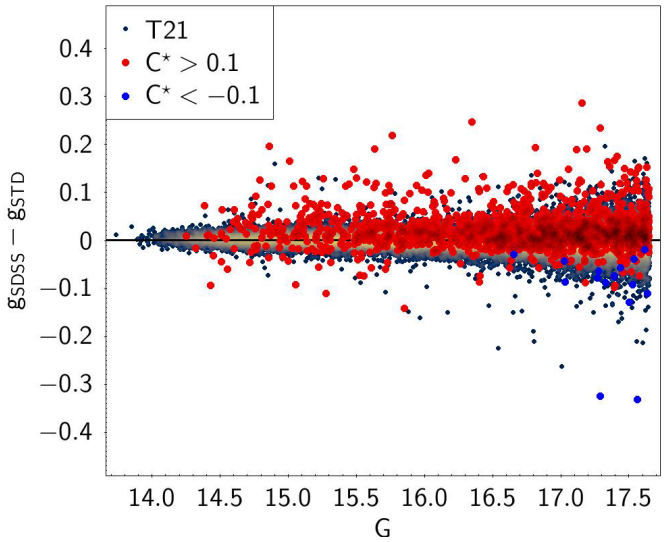


**Fig. 8.** Tracing outliers in the $\Delta g$ vs. $G$ plot for the subset of the T21 reference sample having XP spectra in *Gaia* DR3 and $G < 17.65$. Stars with a relatively large absolute value of $C^\star$ are highlighted in red, for $C^\star > 0.1$, and in blue, for $C^\star < -0.1$).

### 3.2. Johnson-Kron-Cousins system and its standardisation

Of the several photometric systems proposed since the advent of photoelectric and CCD (charge-coupled device) photometry, the Johnson-Kron-Cousins system (JKC hereafter) was – and still is– one of the most widely adopted. It was built starting from the Johnson *UBV* (Johnson & Morgan 1953; Johnson 1963), Kron *RI* (Kron et al. 1953), and Cousins *VRI* (Cousins 1973, 1983, 1984) systems. In 1992, Arlo U. Landolt published a catalogue of equatorial standard stars, which from then on became the fundamental defining set for the *UBVRI* JKC system, and has been used over the last three decades to calibrate the vast majority of all imaging observations in the *UBVRI* passbands. The original 1992 photoelectric set was later extended with observations far from the celestial equator and also with a large amount of CCD observations (hereafter Landolt collection, Landolt 1992, 2007, 2009, 2013; Landolt & Uomoto 2007; Clem & Landolt 2013, 2016). Moreover, from 1983, P. B. Stetson collected observations for approximately $10^5$ secondary *UBVRI* standards using about half a million proprietary and archival CCD images (hereafter Stetson collection) of various fields of astrophysical interest, including star clusters, supernova remnants, and dwarfs galaxies. We used the Landolt and Stetson collections to respectively standardise and validate (see Appendix G) the *UBVRI* synthetic photometry obtained from *Gaia* XP spectra presented here. The Landolt and Stetson collections are described in detail by Pancino et al. (2022); here we briefly describe the quality selections that were applied for the purpose of the present work.

First, we used *Gaia* and other literature catalogues to clean the collections from variables stars, binaries, blends, and stars with lower photometric quality. For the variables, we made use of the *Gaia* DR2 catalogue (Gaia Collaboration 2019), the ASAS-SN catalogue of variable stars (Shappee et al. 2014; Jayasinghe et al. 2018, 2019a,b), and the Zwicky Transient Facility catalogue of periodic variable stars (Chen et al. 2020). For binaries, we profited from the work done by the Survey of Surveys team (Tsantaki et al. 2022), who compiled all known spectroscopic binaries in large spectroscopic surveys and astroseismology missions (Price-Whelan et al. 2020; Kounkel et al. 2021; Traven et al. 2020; Merle et al. 2017; Birko et al. 2019; Qian et al. 2019; Tian et al. 2020; Deleuil et al. 2018; Kirk et al. 2016). In addition, we used the following cuts on parameters from the main `gaia_source` table to further remove possible contaminated and blended sources: `IPD_Frac_Odd_Win` and `IPD_Frac_Multi_Peak` above 7%, Renormalised Unit Weight Error (`RUWE`) above 1.4, and the recommended cut by Riello et al. (2021) on the renormalised *BP* and *RP* flux excess, $|C^*| > 2\sigma_{C^*}$, as well as a cut on the $\beta$[13] parameter by Riello et al. (2021) above 20%.

The original Landolt and Stetson collections agree very well with each other, with zero-point offsets of below 1% in all bands, and spreads of ±1–2%. However, some disagreement (3%–5%) was found for the reddest stars, which are less represented in the original Landolt (1992) set, with only half a dozen stars (Pancino et al. 2022). This is particularly evident for the *R* and *I* bands. For this reason, we decided to use only the Landolt collection for the standardisation and the Stetson one for the validation. This uncertainty for redder stars makes both the Landolt and Stetson collections less reliable for stars redder than $G_{BP}-G_{RP} \simeq 2$ mag, although both collections are rigorously calibrated on the original Landolt (1992) set. We used the Landolt sample to standardise the synthetic photometry in

[13] Defined as (`phot_bp_n_blended_transits` + `phot_rp_n_blended_transits`)/(`phot_bp_n_obs` + `phot_rp_n_obs`).

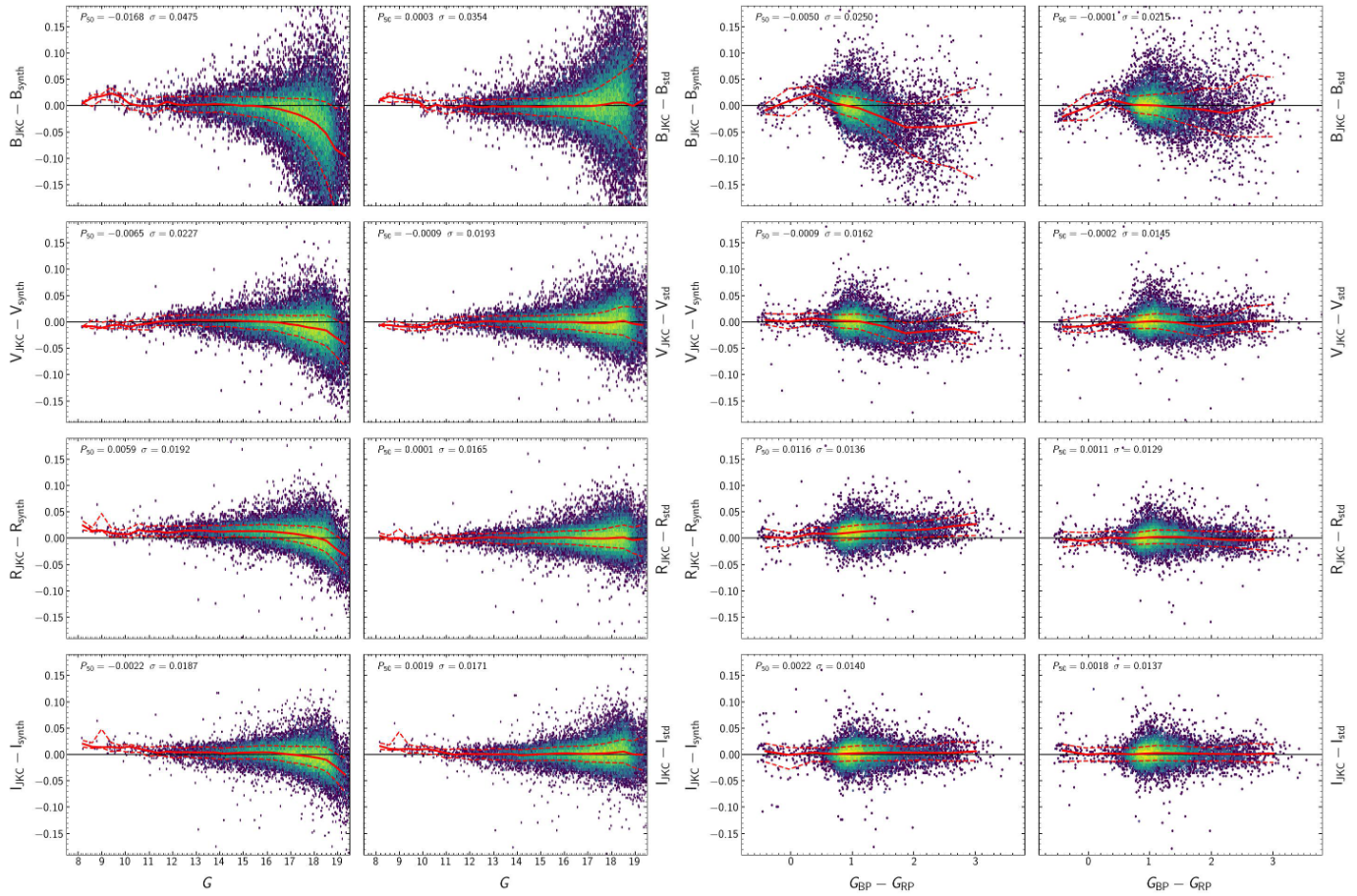


**Fig. 9.** Performance and standardisation of JKC *BVRI* XP synthetic magnitudes using the reference set of standard stars described in the text. The arrangement and symbols are the same as in Fig. 6.

the *UBVRI* system, which we obtained using the passbands by Bessell & Murphy (2012), as described in Sect. 2.2.

Figure 9, which is fully analogous to Fig. 6, shows the comparison between XPSP and reference *BVRI* magnitudes before and after standardisation. The overall ZP, for stars not seriously affected by the hockey-stick effect, are reproduced by raw XPSP to better than ≃0.02 mag in all the considered bands. A significant colour trend is observed for *B* that is very similar to the case of SDSS *g*. This analogy is not surprising because the two filters sample a similar range of the *BP* spectrum. Similarly, the performance of raw *V* is significantly poorer than raw *r*, with larger scatter and colour terms, which is likely due to *V* sampling more problematic regions of *BP* than *r* (≃490−660 nm versus ≃540−699 nm, respectively). We address the reader to Montegriffo et al. (2023) for an additional discussion on this specific system. Here we note that standardisation significantly mitigates the amplitude of the residual systematic errors displayed by raw XPSP.

The comparison between standardised XPSP magnitudes and those from the reference sample for $G < 17.65$ mag stars is presented in Fig. 10, as a function of magnitude and colour. Performances are very similar to the SDSS case described above. The remarkable differences are: (a) the loss of millimag accuracy for $G \lesssim 11.5$ mag in correspondence with a transition to different setups of the *BP* and *RP* spectrometers which affects the internal calibration of XP spectra in this bright magnitude range (onset of *gates*, change of window class, etc.; see e.g. De Angeli et al. 2023; Montegriffo et al. 2023), which is not sampled by SDSS stars[14]; and (b) the residual colour terms of order ≃10 mmag remaining in some colour range for the *B* and – to a lesser extent – *V* passbands. The median, and the $P_{16}$ and $P_{84}$ percentiles of the Δmag distributions for JKC *BVRI* magnitudes shown in Fig. 10 are listed in Table 2. The scatter about the median for $G < 16.5$ mag is ≲15 mmag in *VRI*, and ≲20 mmag in *B*.

Similarly to the case of the T21 sample, red giants are also relatively rare in the Landolt reference sample used here, with just a handful in the range $1.5 < G_{BP}-G_{RP} < 3.5$. We carefully verified that these red giants match the same locus of the bulk of the other stars in the sample in the $G_{BP}-G_{RP}$ versus Δmag diagrams, within <10.0 mmag.

### 3.3. Standardised ultraviolet bands

Figure 11 shows the performances of the JKC *U* band and SDSS *u* band magnitudes, which are standardised as described in Sect. 2.2.2. These are shown in comparison with the respective reference samples (see Sects. 3.1 and 3.2, respectively) for the subset of sources that will have XP spectra released in *Gaia* DR3 (see Appendix G for comparison of raw and standardised magnitudes). The median Δmag is a few millimag for $15.2 \leq G \leq 17.6$ mag in $u_{SDSS}$ and for $11.5 \leq G \leq 17.6$ mag in $U_{JKC}$, with the issue related to sources brighter than $G = 11.5$ mag discussed above decreasing the accuracy in this range (see Tables 3 and 4).

[14] Because the saturation limit of SDSS occurs around $G = 14.0$ mag.

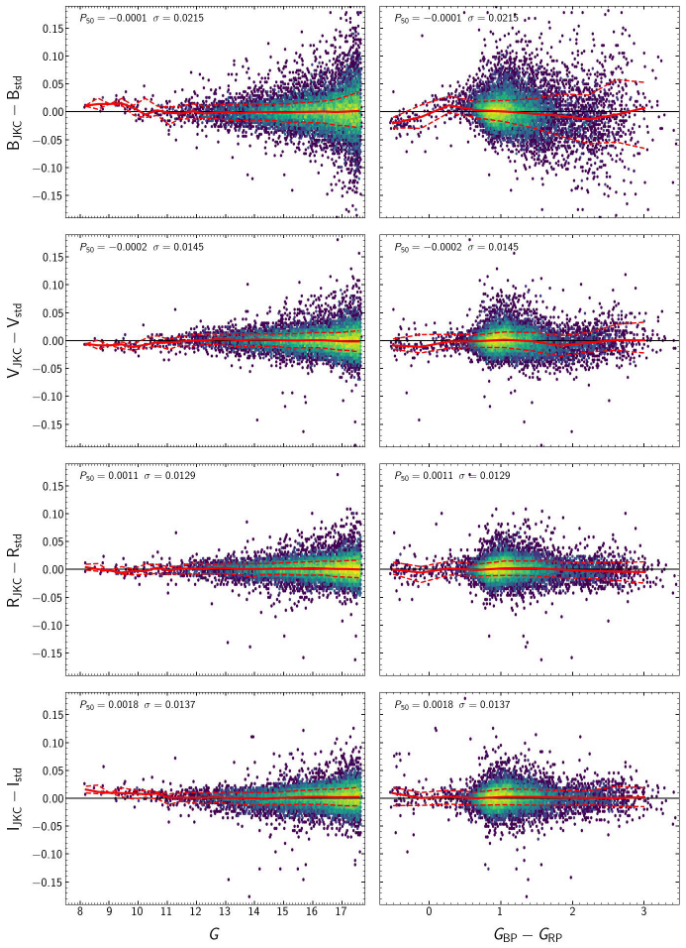


**Fig. 10.** Performances of standardised XPSP in the JKC system (*BVRI*). We show Δmag as a function of $G$ magnitude (*left panels*) and $G_{BP}-G_{RP}$ colour (*right panels*) for the subsample of reference stars whose XP spectra has been released in *Gaia* DR3. The arrangement and the meaning of the symbols is the same as in Fig. 6.

However, the scatter is significantly larger than in all redder wide-band magnitudes considered here, reaching 0.12−0.15 mag at $G = 16.4$ mag and >0.3 mag at $G = 17.6$ mag. In particular, the accuracy is generally poor for red sources, missing sufficient signal in the wavelength range covered by UV bands for reliable magnitudes to be provided.

Figure 12 shows the effect of adopting the selection in S/N, `flux_x/flux_x_error > 30`, for $x = u_{SDSS}$ and $U_{JKC}$, respectively, on the Δmag distribution of Fig. 11. This selection greatly reduces the scatter about the median, thus providing much more reliable individual magnitudes, but it implies a strong selection in magnitude and in colour in these samples, in practice removing all stars with $G_{BP}-G_{RP} \gtrsim 1.3$ mag and $G \gtrsim 16.5$. In Sect. 6.2 we show that when applied to larger samples, reliable UV magnitudes can be obtained for stars as red as $G_{BP}-G_{RP} \simeq 3.0$ mag, depending on their apparent magnitude, still maintaining a strong bias against red and faint stars.

In summary, as anticipated in Sect. 2.2.2, the performances for any band covering the XP range ≲400 nm are significantly poorer than in all the redder passbands. We do not discuss similar bands from other systems any further, as the results would be very similar to those shown in Figs. G.1 and G.2 for example, namely. strong colour-dependent trends with respect to reference external photometry. Given the high astrophysical relevance of these UV bands and the lack of all-sky sources for them, we made a concerted effort to provide standardised $u_{SDSS}$ and $U_{JKC}$ magnitudes and managed to obtain reasonably accurate and precise photometry for the subset of stars with sufficient signal in that region of the spectrum due to favourable combinations of magnitude and colour. While XPSP magnitudes in bands at ≲400 nm can be obtained for all the sources with XP spectra released in *Gaia* DR3, and sometimes a highly uncertain measurement can be better than no measure at all, we strongly recommend using these magnitudes only if `flux_x/flux_x_error > 30`, and, in particular, using preferentially the standardised $u_{SDSS}$ and $U_{JKC}$ provided in the *Gaia* Synthetic Photometry Catalogue (GSPC; Sect. 6.2). In any case, even standardised UV XPSP must be used with caution (Sect. 7 for further caveats).

### 3.4. PanSTARRS-1 system and its standardisation

PanSTARRS-1 (hereafter PS1, for brevity) is an ambitious multi-task project, the main aim of which is to survey all the sky above Dec = −30° in five passbands, *grizy* (3π survey, see Chambers et al. 2016, and references therein). The wide sky coverage and the high photometric precision reached qualify PS1 as one of the most widely used sources of stellar photometry (Magnier et al. 2020a; Xiao & Yuan 2022).

As a reference sample of standard stars, here we adopt two 15° × 15° patches located at the Galactic caps ($|b| > 60.0°$). Bona fide point sources with multi-epoch observations were selected using the difference between PSF magnitudes and Kron magnitudes as a diagnostic, following a kind suggestion by E. Magnier (priv. comm.)[15] and requiring the uncertainty on magnitudes to be <0.02 mag in all passbands. Once matched with the *Gaia* source catalogue, with a 1″ cone search, a reference sample of 76 491 stars was finally adopted. XPSP magnitudes are compared with PS1 PSF magnitudes based on the average of the chip measurements (`x_chp_psf`, where `x = grizy`), because these have the best corrections for systematic effects.

The performance before and after standardisation is similar to that obtained for SDSS magnitudes and is shown in Fig. G.3. Figure 13 shows the performance for the subset that will be included in *Gaia* DR3 for $G < 17.65$ mag. The accuracy of standardised XPSP is good in all passbands; see also Table G.1. The median Δmag amounts to a few millimag over the entire range of magnitudes considered, while the typical $\sigma$ ranges between 10 and 15 mmag for $G \leq 16.5$ mag. Slightly larger deviations are observed at the extremes of the colour range spanned by the reference sample in the $g$ band (red side) and in $y$ band (blue side). It is worth noting that the reference sample adopted has a limited coverage of colour and spectral type compared to those we use for SDSS and JKC systems for example. We therefore recommend special caution in using PS1 XPSP magnitudes outside the validated colour and magnitude ranges.

### 3.5. Standardised HST magnitudes

The *Hubble* Space Telescope (HST) is one of the most successful space missions ever, with a long-standing and huge impact on virtually all branches of astrophysics (see e.g. Macchetto 2010, and references therein). Unprecedented photometric precision is one of the many excellent achievements of the optical-IR

[15] In particular, in our catalogue, we only kept sources with `x_chp_psf-x_chp_kron > −0.3` and `x_chp_psf-x_chp_kron < 0.1`, and `x_chp_psf_nphot ≥ 10` for `x = grizy`.

**Table 2.** JKC system: median ($P_{50}$) and 15.87% ($P_{16}$) and 84.13% ($P_{84}$) percentiles of the Δmag distributions of Fig. 10.

| $G$ (mag) | $P_{50}(\Delta B)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $P_{50}(\Delta V)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $P_{50}(\Delta R)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $P_{50}(\Delta I)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $n_\star$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9.4 | 15.1 | 11.9 | 23.6 | −6.9 | −14.4 | −2.7 | −5.7 | −8.3 | 4.9 | 9.2 | 2.4 | 18.9 | 17 |
| 9.8 | 7.0 | −6.1 | 11.3 | −11.6 | −17.3 | −6.3 | −2.8 | −9.8 | 0.9 | 8.7 | −1.4 | 14.6 | 15 |
| 10.2 | −5.0 | −10.6 | 23.9 | −7.6 | −13.1 | 3.4 | −3.2 | −8.7 | 9.2 | 7.1 | 1.7 | 14.2 | 11 |
| 10.7 | 2.5 | −6.5 | 10.2 | −3.1 | −10.7 | 8.7 | 2.9 | −7.2 | 8.8 | 8.4 | −1.2 | 11.0 | 20 |
| 11.1 | −2.0 | −23.1 | 8.1 | −3.7 | −15.8 | 1.9 | −1.8 | −8.2 | 4.6 | 1.1 | −11.5 | 7.7 | 31 |
| 11.5 | −3.1 | −11.8 | 10.5 | −1.5 | −8.5 | 8.1 | 1.4 | −8.0 | 6.6 | 1.2 | −5.1 | 9.2 | 42 |
| 11.9 | 0.8 | −14.2 | 14.0 | −0.2 | −9.7 | 9.6 | 1.5 | −8.2 | 10.8 | 1.0 | −8.4 | 10.4 | 70 |
| 12.3 | −3.3 | −11.2 | 10.0 | 0.3 | −7.3 | 12.0 | 1.1 | −8.1 | 10.1 | 0.8 | −10.1 | 9.1 | 105 |
| 12.7 | −1.2 | −13.1 | 12.2 | 2.4 | −7.1 | 11.9 | 2.5 | −7.7 | 9.6 | 1.8 | −7.7 | 9.7 | 148 |
| 13.1 | −0.7 | −14.5 | 13.6 | 1.9 | −9.0 | 12.4 | 2.2 | −8.6 | 10.6 | 0.4 | −9.5 | 11.6 | 210 |
| 13.5 | −2.3 | −15.9 | 11.2 | 0.2 | −11.0 | 11.6 | 0.4 | −10.4 | 10.2 | 0.3 | −10.4 | 9.5 | 282 |
| 13.9 | −0.4 | −14.6 | 14.5 | 1.7 | −11.4 | 12.1 | 0.7 | −12.9 | 10.3 | 0.3 | −12.8 | 10.1 | 381 |
| 14.3 | −0.6 | −15.6 | 12.4 | 1.7 | −10.3 | 12.0 | 0.8 | −9.0 | 10.3 | −1.0 | −10.9 | 10.0 | 555 |
| 14.7 | −1.1 | −16.2 | 14.1 | 0.8 | −10.3 | 11.7 | 1.5 | −8.8 | 11.4 | 0.4 | −10.6 | 10.9 | 706 |
| 15.1 | −1.5 | −19.5 | 13.6 | 0.3 | −13.1 | 12.1 | 1.4 | −9.6 | 12.1 | 0.9 | −11.8 | 12.2 | 875 |
| 15.5 | −0.5 | −17.7 | 16.0 | 0.1 | −13.1 | 12.7 | 1.4 | −10.1 | 13.0 | 1.9 | −10.8 | 14.1 | 1149 |
| 16.0 | −0.3 | −19.6 | 17.1 | 0.6 | −12.8 | 13.9 | 2.3 | −10.5 | 13.7 | 2.5 | −11.0 | 14.8 | 1395 |
| 16.4 | 0.5 | −20.1 | 18.9 | 0.5 | −13.3 | 14.5 | 1.6 | −10.0 | 14.9 | 2.4 | −10.8 | 16.1 | 1692 |
| 16.8 | −0.6 | −24.0 | 21.7 | −0.6 | −16.4 | 13.8 | 0.8 | −13.0 | 14.1 | 1.8 | −12.5 | 16.0 | 2270 |
| 17.2 | 0.3 | −25.4 | 28.1 | −1.6 | −17.7 | 15.8 | 0.7 | −13.8 | 15.7 | 2.0 | −13.7 | 17.7 | 2973 |
| 17.6 | 1.5 | −28.8 | 35.6 | −1.6 | −20.0 | 17.6 | −0.0 | −14.9 | 16.7 | 2.3 | −13.4 | 19.6 | 2295 |

**Notes.** Here, $n_\star$ is the number of sources in the considered bin.

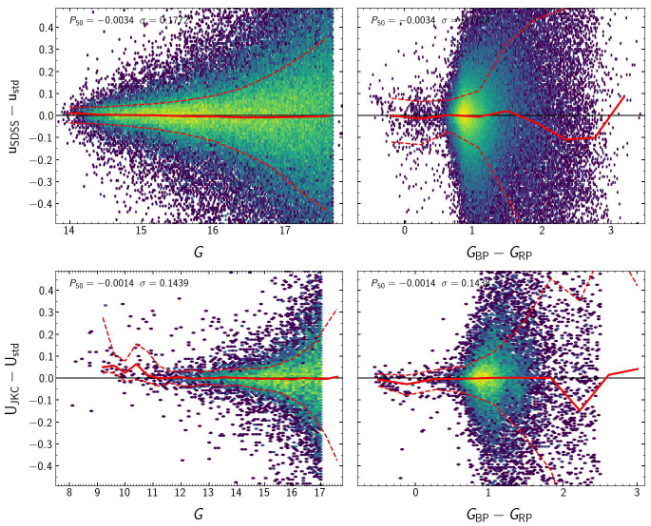


**Fig. 11.** *Upper panel*: performance of standardised XPSP in the SDSS $u$ band. The reference sample is the same as in Fig. 7. *Lower panel*: performance of standardised XPSP in the JKC $U$ band. The reference sample is the same as in Fig. 10. Please note that the bright limit of the two reference samples is very different.

**Table 3.** SDSS system: median ($P_{50}$) and 15.87% ($P_{16}$), and 84.13% ($P_{84}$) percentiles of the $\Delta u$ distribution of Fig. 11.

| $G$ (mag) | $P_{50}(\Delta u)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $n_{\rm star}$ |
|---|---|---|---|---|
| 14.0 | 11.3 | −34.3 | 37.2 | 198 |
| 14.4 | 4.2 | −38.2 | 38.1 | 2208 |
| 14.8 | 2.0 | −47.8 | 45.5 | 4827 |
| 15.2 | 0.7 | −61.7 | 53.8 | 7550 |
| 15.6 | −3.3 | −84.6 | 67.9 | 10 406 |
| 16.0 | −4.4 | −115.6 | 89.2 | 11 724 |
| 16.4 | −9.0 | −168.4 | 123.9 | 14 678 |
| 16.8 | −8.3 | −232.2 | 183.4 | 18 968 |
| 17.2 | −6.6 | −342.5 | 261.3 | 22 809 |
| 17.6 | −2.6 | −437.8 | 355.9 | 16 586 |

**Notes.** Here, $n_\star$ is the number of sources in the considered bin.

cameras on board this iconic space observatory (see Bedin et al. 2019, for a recent example).

HST cameras are equipped with large sets of narrow, medium, and wide filters, making for several very powerful and flexible photometric systems. XPSP for the subset of those that are enclosed within the XP spectral range may be very useful for many scientific applications. For instance, all-sky XPSP for the 220 M stars with XP spectra released in *Gaia* DR3 will hugely extend the realm reachable by photometry in HST systems to the entire sky and in a bright range of magnitudes ($4.0 \lesssim G \lesssim 17.65$ mag) not usually easily accessible to HST (and/or not convenient to be sampled with HST). It may be worth recalling an additional desirable feature, namely that these extensions come from space-based spectrophotometry.

Unfortunately, this high degree of complementarity between HST and *Gaia*, makes it extremely difficult to find proper samples for validation and standardisation of XPSP in the HST systems. Only a handful of well-measured *Gaia* sources can be found in the typical FoV of HST cameras ($\lesssim$4 arcmin$^2$) and HST is mainly used to measure very faint stars that would otherwise be unreachable for ground-based instruments, and therefore the typical overlap in magnitude between HST and *Gaia* is limited. Finally, even if there were samples with a significant number of stars common to both of them, HST observations would only be available in a limited number of passbands for any given camera.

**Table 4.** JKC system: median and 15.87% ($P_{16}$) and 84.13% ($P_{84}$) percentiles of the $\Delta U$ distributions of Fig. 11.

| $G$ (mag) | $P_{50}(\Delta U)$ (mmag) | $P_{16}$ (mmag) | $P_{84}$ (mmag) | $n_\star$ |
|---|---|---|---|---|
| 10.0 | 23.7 | −14.1 | 77.8 | 14 |
| 10.4 | 63.9 | −7.9 | 155.5 | 13 |
| 10.8 | 13.2 | −18.5 | 121.3 | 22 |
| 11.2 | −1.6 | −36.8 | 55.9 | 35 |
| 11.6 | 1.0 | −22.1 | 32.5 | 51 |
| 12.0 | 4.0 | −36.9 | 30.5 | 73 |
| 12.4 | −2.4 | −37.6 | 35.4 | 109 |
| 12.8 | −1.2 | −29.2 | 28.9 | 158 |
| 13.2 | −0.2 | −31.4 | 34.3 | 236 |
| 13.6 | 3.7 | −35.8 | 40.4 | 287 |
| 14.0 | 0.7 | −31.5 | 39.4 | 402 |
| 14.4 | −0.6 | −40.0 | 41.5 | 566 |
| 14.8 | −3.5 | −51.9 | 44.2 | 719 |
| 15.2 | −2.7 | −58.2 | 50.6 | 880 |
| 15.6 | −3.5 | −69.0 | 66.6 | 1192 |
| 16.0 | −6.7 | −96.8 | 88.0 | 1357 |
| 16.4 | 0.5 | −128.0 | 123.0 | 1689 |
| 16.8 | −3.2 | −180.5 | 161.3 | 2264 |
| 17.2 | −4.2 | −264.5 | 240.6 | 2923 |
| 17.6 | 5.7 | −376.0 | 316.2 | 2230 |

**Notes.** Here, $n_\star$ is the number of sources in the considered bin.

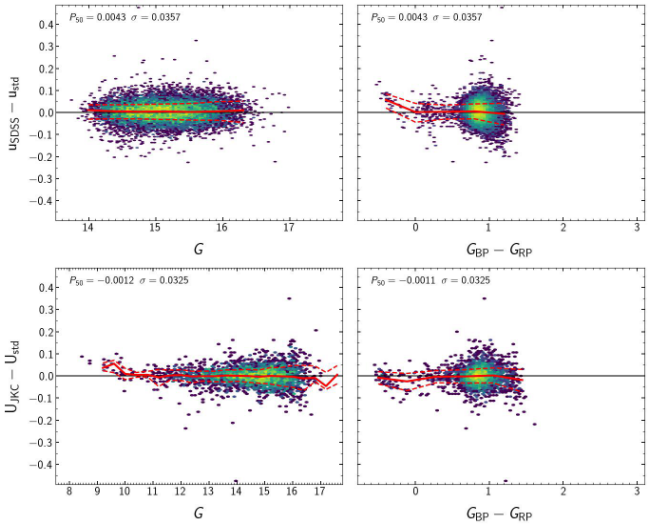


**Fig. 12.** Same as Fig. 11 but limited to stars with flux/flux$_{e}$rror > 30 in $u_{SDSS}$ and $U_{JKC}$, respectively. This is the S/N limit on individual magnitudes that we adopt for the GSPC (Sect. 6.2).

For these reasons, the sample we used for validation and standardisation of a few HST passbands, while absolutely excellent for the scientific application for which it was acquired, is clearly not ideal for our purpose. Still, it is adequate for testing the accuracy and precision of XPSP at the ≃1% level over a limited range in colour and magnitude.

We used the photometry in the WFC3/UVIS $F438W$ and ACS/WFC $F606W$, $F814W$ bands of a set of Galactic globular clusters (GCs) from the HUGS project (Nardiello et al. 2018) as a reference sample. These observations have the advantage of providing a large number of stars in the small FoV covered by the considered cameras, and remarkable overlap in magnitude

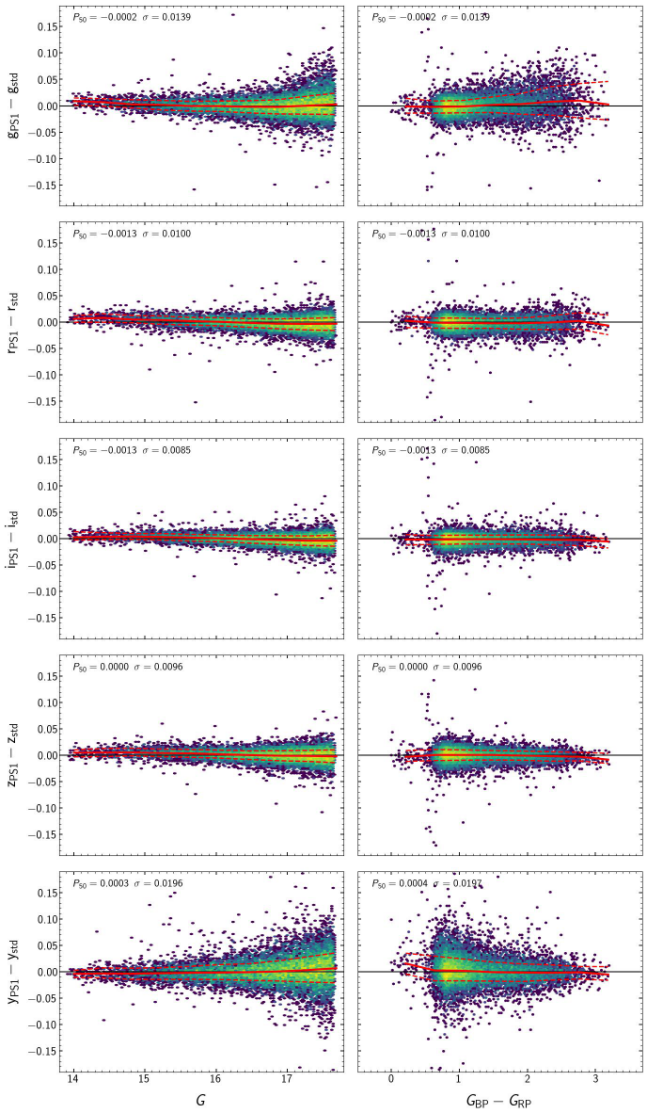


**Fig. 13.** Performance of standardised XPSP in the PanSTARRS-1 system (*grizy*). We show Δmag as a function of $G$ magnitude (*left panels*) and $G_{BP}$−$G_{RP}$ colour (*right panels*) for the subsample of reference stars whose XP spectra has been released in *Gaia* DR3. The arrangement and the meaning of the symbols is the same as in Fig. 6.

with the *Gaia* source catalogue. Unfortunately, for obvious scientific reasons, HUGS fields target crowded areas in the central region of the clusters, where blending and contamination of the relatively wide *BP* and *RP* apertures by nearby sources and/or high background may severely affect most of the XP spectra. For this reason, we applied very strong selections on the original HUGS samples, only keeping in the final reference sample used for standardisation stars (a) with |RADXS| < 0.1, QFIT > 0.9, and photometric uncertainty on individual HUGS magnitudes <0.1 mag[16]; and (b) with a number of *BP* and *RP* epoch spectra sufficient to ensure high S/N in XP spectra,

[16] RADXS and QFIT are quality parameters from the HUGS catalogs, see Nardiello et al. (2018) for details and discussion. The original comparison with the HUGS photometry included also the $F336W$ filter, that was later abandoned because its XPSP counterpart suffers from the strong systematics affecting $U$ bands. However the selections of RADXS and QFIT were imposed an all the considered bands including $F336W$. Analogously, a star was accepted for the final sample only if it had a valid magnitude in all the four passbands.

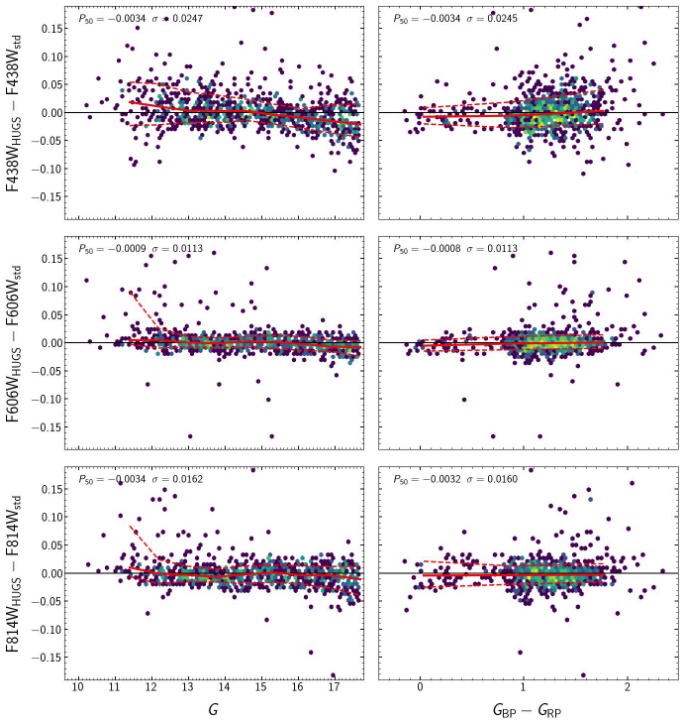


**Fig. 14.** Performances of standardised XPSP in the HST WFC3/UVIS ($F438W$, VEGAMAG) and ACS/WFC systems ($F606W$, $F814W$, VEGAMAG). We show Δmag as a function of $G$ magnitude (*left panels*) and $G_{BP} - G_{RP}$ colour (*right panels*) for the subsample of reference stars whose XP spectra has been released in *Gaia* DR3. The arrangement and the meaning of the symbols is the same as in Fig. 7 but the percentiles are computed in bins of 0.8 mag in width.

adopting the same criterion for the release of XP spectra, that is, `bp_num_of_transits` > 15 and `rp_num_of_transits` > 15, and $|C^\star| < 0.05$ in order to minimise the impact of contamination of the spectra. After cross-correlation with the *Gaia* EDR3 source catalogue, we end up with a sample of 1113 stars in the range $10.0 \leq G \leq 19.0$ mag, 968 of which have $G < 17.65$ mag. We used 'method 1' magnitudes from the HUGS catalogue, as they are presented as the best choice for the bright magnitude range we are considering (Nardiello et al. 2018). Here we compare magnitudes in the VEGAMAG systems, but the comparison has general validity, as transforming into STMAG or ABMAG would imply a simple zero-point shift.

The comparison of HUGS magnitudes with raw XPSP is shown in the left rows of the two panels of Fig. G.5. The typical scatter about the median is larger than in the cases described above. We verified that, in spite of the severe selection in $C^\star$, residuals still correlate with this parameter, showing that this extra scatter is due to the fact that we are not dealing with a sample dominated by fully isolated stars, as is the case for the reference samples considered above, but with many sources whose spectrophotometry (and possibly also HUGS magnitudes, to a lesser extent) suffer from some degree of contamination, effectively limiting the achievable precision. However, at least for $F606W$ and $F814W$, the original photometry is reproduced within ≃1% over the entire range of colour covered by the reference sample, the only systematic deviations being attributable to the hockey-stick effect. For this reason, we decided to limit the process of standardisation to the correction of this effect, avoiding any modification of the original passbands. The final results for the subsample of stars with $G < 17.65$ mag are shown in Fig. 14. The performances in $F606W$ and $F814W$, possibly the most widely used HST passbands, are satisfactory given the non-ideal conditions. The higher scatter and the lower accuracy of the $F438W$ XPSP are attributable to the same kind of problems affecting passbands sampling the blue end of the *BP* spectra.

While validation is limited to the passbands and the colour and magnitude ranges considered above, the results presented here and those of a limited set of additional tests we performed may suggest that HST photometry for $Rf > 1.4$ passbands should be reasonably well reproduced by XPSP, while the issues related to the blue and UV end of *BP* spectra remain valid also in this case.

## 4. Narrow-band photometry

In this section we explore the performance of XPSP in the realm of medium- and narrow-band photometry using a few widely used systems as test cases. Standardisation is attempted only for the version of the Strömgren system considered here. The J-PAS and J-Plus systems sample the performances of narrow-band XPSP over the entire range covered by XP spectra. General guidelines to use narrow band XPSP to calibrate surveys aimed at measuring emission line fluxes are also provided. Finally, we show an example of how XPSP can be used to take the design of a photometric system, and bring it into real life, measuring fluxes and magnitudes of real sources through its wide and medium-width passbands (the *Gaia* C1 system, Jordi et al. 2006).

### 4.1. Strömgren photometry and its standardisation

According to Sterken et al. (2011) the Strömgren system (Strömgren 1956) was originally designed to investigate the astrophysical properties of low-reddening main sequence stars. However, colour indices obtained from its *uvby* bands have been widely used to estimate the stellar effective temperature and surface gravity, as well as other parameters such as reddening and metallicity, over a wide range of stellar types and classes. For instance, see the use of the $(b - y)$ temperature sensitive colour in the Alonso et al. (1999) relations, or of $m_1 = (v - b) - (b - y)$ to derive metallicity (e.g. in stellar clusters, as done by Frank et al. 2015; Piatti et al. 2019, among others), or the correlation between $c_1 = (u - v) - (v - b)$ and nitrogen abundance (Grundahl et al. 2002).

In contrast to the cases discussed in Sect. 3, the Strömgren system lacks a generally accepted standard version, with its set of TCs and, especially, a large set of reliable standard stars. Among the many available versions of Strömgren TCs (see e.g. Bessell 2011), here we adopt those provided by the Spanish Virtual Observatory[17], which describe the filters mounted on the Wide Field Camera (WFC) of the *Isaac Newton* Telescope (INT) at El Roque de los Muchachos in the Canary Islands, as we have had some previous successful experience in using them (Massari et al. 2016).

Strömgren bands are entirely located in the *BP* realm. We limit our analysis to *bvy* bands here, because XPSP $u$ magnitudes suffer from the problems described in Sect. 2.2.2, affecting all the TCs in the range $\lambda < 400$ nm; moreover, its blue edge exceeds the blue limit of *BP* (see Fig. 5). While $b$ and $v$ passbands have $Rf = 2.0$ and $Rf = 1.8$, respectively, $y$ is slightly below the nominal $Rf = 1.4$ threshold for reproducible photometry established in Appendix B, with $Rf = 1.3$. This should

[17] http://svo2.cab.inta-csic.es/svo/theory/fps3/

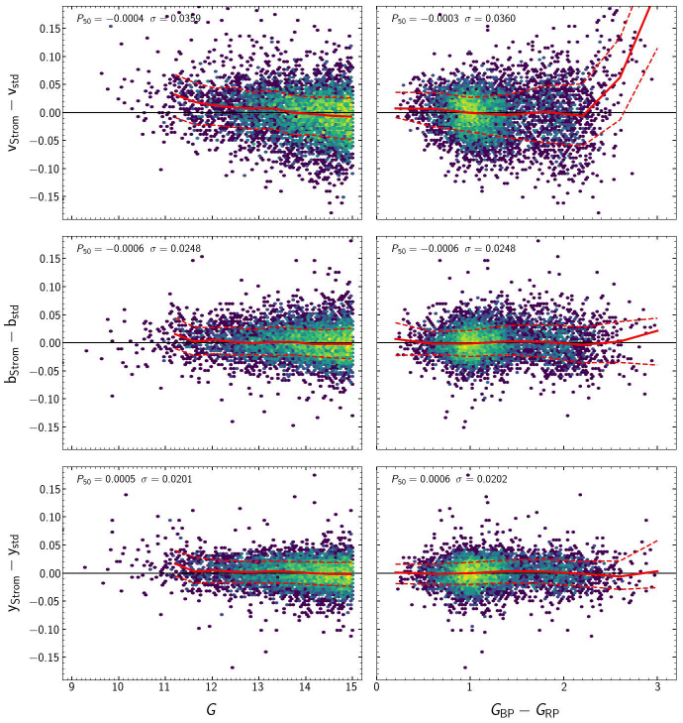


**Fig. 15.** Performances of standardised XPSP in the Strömgren system. The arrangement and the meaning of the symbols is the same as in Fig. 7.

not be considered as a serious issue, because the adopted limit is conservative and also because $Rf$ values below that threshold may only be troublesome in the presence of strong spectral features.

As a reference set for comparison and standardisation we chose the largest sample available in the literature of stars observed with the aforementioned TCs, which is the sample of stars located around the Galactic anticentre analysed by Monguió et al. (2013). This sample contains 23 992 stars, covering many spectral types and a wide colour range, $0 \leq G_{BP}-G_{RP} \leq 3$ mag, and, in turn, is calibrated on the set of standards defined in Crawford & Barnes (1970).

We performed a quality check in order to select only well-measured stars. We first considered the photometric error in each of the three considered Strömgren bands as a function of $G$-band magnitude and traced the median trend in steps of $\Delta G = 0.5$ mag. Only the sources with an error smaller than the 95th percentile of the distribution for each bin were kept in the sample. To be conservative, we further rejected all the sources with a photometric error of >0.1 in any band, as their calibrating power would be poor in any case. This effectively limits the reference sample to $G < 15.0$ mag stars. In turn, this implies that no correction for the hockey-stick effect is possible in the standardisation process. After this first selection based on the Strömgren photometry, we applied two other quality cuts based on *Gaia* EDR3 parameters by requesting `ruwe` < 1.4 (from `gaia_source`) and $-0.03 < C^* < 0.03$. After all these selections, the final reference sample includes 6158 stars.

The median difference between reference and XPSP raw magnitudes amounts to ≃0.14 mag in $v$, ≃0.00 mag in $b$, and ≃0.03 mag in $y$, with colour trends of amplitude ≲0.05 mag over the colour range covered by the reference sample (see Fig. G.6). The performance is worse for the bluer TCs, in line with the already mentioned issues with the blue part of the BP spectra, which are probably exacerbated by the lower S/N unavoidably associated with passband widths smaller than those discussed in Sect. 3. The median deviation of $v$ magnitudes is the largest among the non-UV passbands considered here. Magnitudes in J-PAS passbands of similarly narrow width and covering the same wavelength range (397−427 nm) have median deviations ≃0.08−0.10 mag (see Sect. 4.2). The relatively large deviation of the raw synthetic $v$ magnitudes might be ascribed to a combination of the systematic errors of EC XP spectra in the relevant wavelength range and the intrinsic problems historically affecting photometric calibration of the Strömgren system (see e.g. Bessell 2005; SMY11).

In Fig. 15 we show the usual Δ mag plots for the standardised magnitudes (see also Table G.3, for the corresponding $P_{50}$, $P_{16}$, and $P_{84}$ values). A small residual trend with magnitude remains in the bluest passband ($v$), as well as a strong colour term for $G_{BP}-G_{RP} \gtrsim 2.2$ mag, where the number of reference stars is low. On the other hand, the accuracy of the $b$ and $y$ magnitudes is very good, for $G > 11.5$ mag. The typical scatter at $G \leq 15.0$ mag is $\sigma \sim 40$ mmag in $v$, $\sigma \sim 30$ mmag in $b$, and $\sigma \sim 20$ mmag in $y$.

An immediate demonstration of the high photometric performance achieved by the standardised XPSP comes from a direct comparison between the Color Magnitude Diagrams (CMDs) of two Galactic GCs, namely NGC 5272 and NGC 6205 obtained from XPSP and from direct ground-based photometry taken from Massari et al. (2016) and Savino et al. (2018), respectively. Figure 16 shows that, once a strong selection to mitigate the effect of contamination of XP spectra is adopted ($-0.03 < C^* < 0.03$) for the stars in common between *Gaia* and the INT observations, the overall quality of the CMDs from synthetic photometry is clearly higher than their ground-based 'observed' counterparts. All sequences are significantly tighter in general, especially towards the AGB and the tip of the red giant branch (RGB), where the ground-based photometry may suffer from saturation and non-linearity effects. Also, the horizontal branch (HB), particularly that of NGC 5272, appears cleaner and better defined in the XPSP diagrams.

Finally, we stress again that this is a particular realisation of the Strömgren system. For instance, our standardised XPSP *vby* photometry fails to reproduce the colour indices provided by Hauck & Mermilliod (1998). The size of the mismatch depends on the stellar type and on the colour index, and ranges from ∼0.05 mag in $b - y$ for red giant stars, up to ∼0.1 mag in $m_1$ for blue giants. On the other hand, an indirect validation of the adopted standardised magnitudes is provided in Sect. 5.2, where metallicity estimates matching their spectroscopic counterparts within the uncertainties are obtained from standardised XPSP Strömgren colour indices.

### 4.2. Javalambre surveys

In this section, we test XPSP performance against the medium- and narrow-band photometry from the two surveys obtained at the Javalambre Observatory in Teruel, Spain (see Cenarro et al. 2014). The first is the J-PAS survey (Benitez et al. 2014), which includes a photometric system of 54 narrow and contiguous passbands and 6 wider passbands (including SDSS filters), covering a similar wavelength range to *Gaia*. In preparation for the full J-PAS catalogue, recently a small region in the sky was observed and released (the mini-JPAS catalogue, Bonoli et al. 2021), covering 1 deg$^2$ towards the Galactic halo (RA, Dec) = (+215°, +53°), up to magnitude 22−23 in the narrow bands and 24 in the broader passbands, with an absolute error of smaller than ∼0.04 mag.

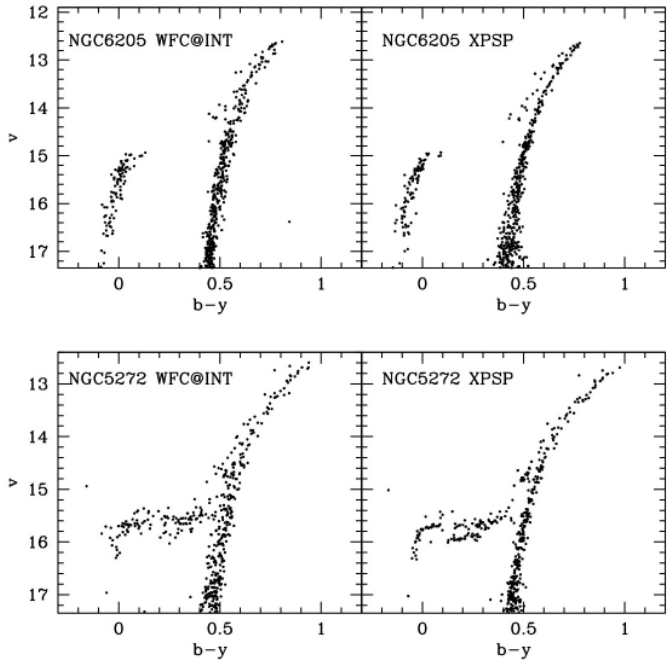


**Fig. 16.** Comparison between the observed (*left columns*) and synthetic (*right columns*) Strömgren CMDs of the GC NGC 5272 (*bottom panels*) and NGC 6205 (*upper panels*). The same stars are plotted in the *left* and *right panels*.

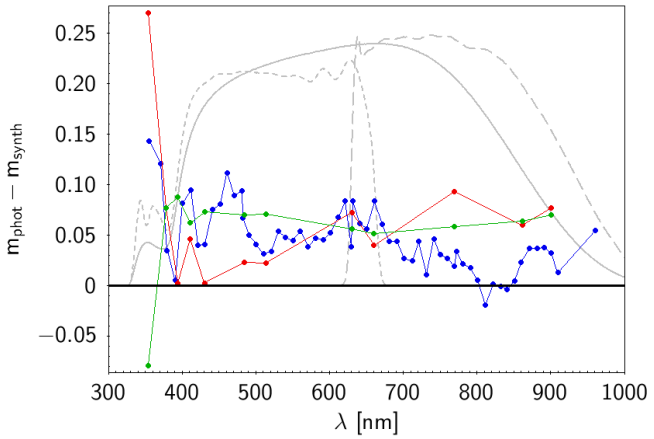


**Fig. 17.** Median residuals between the observed and synthetic magnitudes as a function of the mean wavelengths of the passbands for mini-JPAS (blue), J-PLUS in APOGEE DR16 (red), and J-PLUS WDs (green). In grey lines, we plot the *Gaia* passband transmissivity as in *Gaia* EDR3 (Riello et al. 2021) divided by a factor three to fit in the same scale of the residuals. The solid grey line represents *G* band, the short dashed line *BP* band (at shorter wavelengths), and the long dashed line *RP* band (at longer wavelengths).

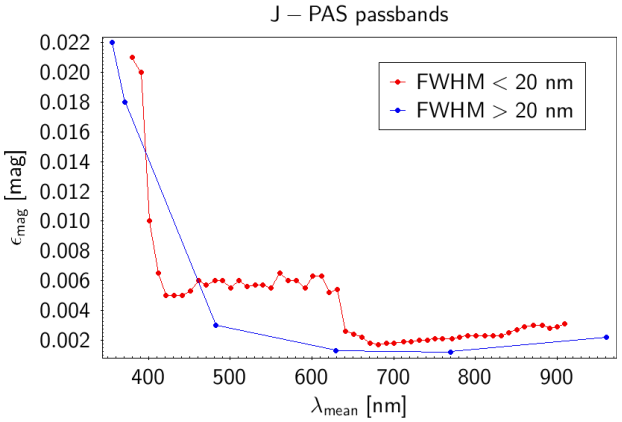


**Fig. 18.** Median magnitude uncertainties at the bright regime derived for J-PAS passbands as a function of wavelength. The red and blue lines show the behaviour for the narrow and wide passbands ($u_{JAVA}$ and SDSS passbands), respectively.

The cross-match with mini-JPAS yields 636 sources in common with the XP spectra in *Gaia* DR3. Although this is a small number of sources, they offer very detailed wavelength information when compared with XPSP results, which is very useful for estimating the level of detail that *BP* and *RP* spectra can provide, as XP spectra and J-PAS are of similar spectral resolution.

The second Javalambre photometric catalogue used here is the J-PLUS survey (Cenarro et al. 2019). The J-PLUS set of passbands includes five broad (similar to SDSS) and seven medium passbands (similar to some of the C1 passbands originally designed for *Gaia* purposes; see Sect. 4.4). The J-PLUS project made its DR2 catalogue available in November 2020, including 31.5 million sources with $r < 21$ mag with absolute calibration errors of 8−19 mmag (depending on the passband; López-Sanjuan et al. 2019). Among all these sources, we used only sources in common with APOGEE DR16 (Ahumada et al. 2020). We also considered a set of white dwarfs (WDs) from Gentile Fusillo et al. (2019) based on *Gaia* DR2.

We applied extra quality filters to avoid the most obvious problems in the XP spectra (blending, contamination, multiple sources, gating, issues in the astrometric solution, and large photometric excess flux). After all this filtering, the remaining set of sources for our analysis comprises 17 465 APOGEE sources and 337 WDs in J-PLUS DR2, and 583 sources in mini-JPAS. No standardisation of any kind has been attempted with the derived synthetic photometry in either J-PAS or J-PLUS systems.

The median Δmag (reference minus synthetic magnitudes) for these samples as a function of mean wavelength of the passband can be seen in Fig. 17. In addition to a general ZP (of about 0.05 mag) in the comparison, Fig. 17 indicates larger discrepancies at short wavelengths ($\lambda < 400$ nm), which are due to the known issues of externally calibrated XP spectra in this range (Sects. 2 and 3). Moreover, the bluest passband, $u_{JAVA}$, has a blue edge located beyond the blue edge of XP spectra (see Sect. 2 and Fig. 5). The lower accuracy of XPSP for filters in this spectral region is therefore not surprising, and neither is the fact that passbands whose SP comes from *BP* spectrophotometry have larger median residuals than those from *RP*, on average.

The uncertainty on the XPSP does not decrease when the magnitude decreases for the bright regime ($G \lesssim 12$ mag). This is due to the fact that, in this range, the calibration errors dominate the estimated uncertainty. Therefore, we analyse here the J-PAS median uncertainty at the bright end as derived for the synthetic photometry. The reader should note that this does not depend in any way on J-PAS data, only on XP spectra. The resulting calibration errors as a function of the central wavelength of the filter are shown in Fig. 18. We can see that passbands with short central wavelengths suffer an increase in systematic effects and also that the minimum uncertainties from the *BP* and *RP* instrument wavelength range are at different levels (being systematically larger for *BP*). For the reddest passbands in *RP*, the calibration error also increases progressively.

The homogeneous wavelength coverage in the J-PAS system allows us to evaluate the variation of the S/N obtained

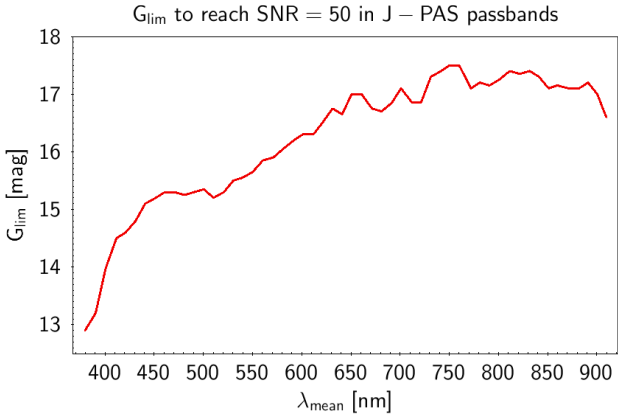


**Fig. 19.** $G$-band magnitude ($G_{\rm lim}$) needed to reach a median $S/N = 50$ in every one of the narrow ($FWHM \sim 15$ nm) J-PAS passbands with mean wavelength equal to $\lambda_{\rm mean}$.

whilst performing synthetic photometry on the *BP* and *RP* spectra. Figure 19 shows the $G$ magnitude needed in order to get a median $S/N = 50$ for every one of the narrow J-PAS passbands (the width of these passbands is about 14 nm).

### 4.3. Emission line photometry: the IPHAS system

The INT Photometric H$\alpha$ Survey of the Northern Galactic Plane (IPHAS, Drew et al. 2005) is designed to identify and characterise emission line stars and extended objects such as planetary nebulae. It uses passbands similar to SDSS $r$ and $i$ together with a narrow H$\alpha$ filter, leading to a H$\alpha$ TC with a full width at half maximum (FWHM) of 9.5 nm and $Rf \approx 1.13$. Broadband data are calibrated based on PanSTARRS, while the narrowband H$\alpha$ data rely on a fixed offset from the enclosing Sloan $r$ band, with further refinement based on overlapping fields anchored to fields with the best photometry (i.e. taken under stable photometric conditions). The final data release is presented in Monguió et al. (2020) as part of the INT Galactic Plane Survey (IGAPS).

We select two comparison sets of stars within the survey footprint ($-5° < b < +5°$, $30° < \ell < 215°$): a control sample consisting of a 1 in 40 sample of *Gaia* DR3 sources with relative parallax errors of better than 20%, and the set of emission line objects identified by Monguió et al. (2020) in the IGAPS catalogue through a linear cut in the ($r - i$, $r$−H$\alpha$) colour plane. The former sample contains approximately 500 000 sources, and the latter just over 8000. Sources are matched between the two catalogues using a 1″ positional cross-match. We remove sources that are saturated, have error flags set in the IGAPS catalogue, or have broadband uncertainties > 0.02 mag or H$\alpha$ uncertainties > 0.05 mag. We also require $G \leq 17.65$ mag, `xp_summary.bp_n_transits` ≥ 15 and `xp_summary.rp_n_transits` ≥ 15, and `phot_g_flux_over_error` > 50, and `phot_x_flux_over_error` > 10, for `x = bp,rp`. We do not filter based on $C_\star$ as a larger value of this parameter can also reflect the presence of emission lines (Riello et al. 2021). This does allow some extended sources to pass the selection, but such sources are rare and thus unlikely to affect the comparison (the vast majority of sources in both samples are classified in IGAPS as stellar). About one-third of the control sample and one quarter of the emitters sample pass these cuts, yielding 168 688 and 2165 sources respectively. These reduced samples are shown in Fig. 20.

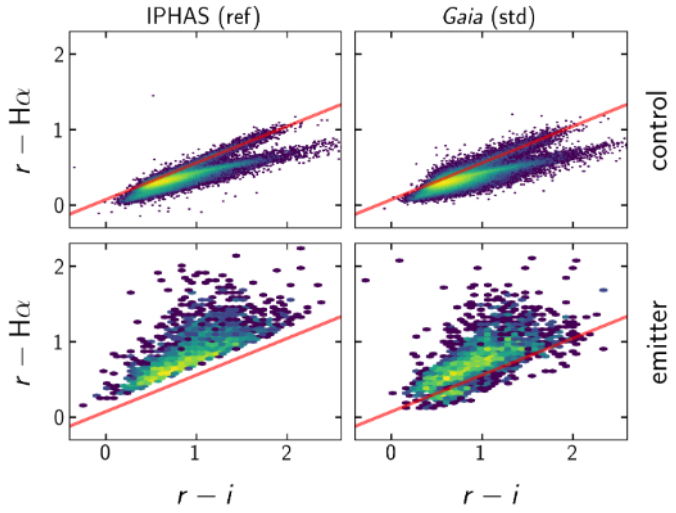


**Fig. 20.** IPHAS ($r-i$, $r$−H$\alpha$) for control (*top*) and emitter (*bottom*) samples after applying the cuts discussed in the text. The original IPHAS magnitudes are on the left and the *Gaia* synthetic magnitudes are on the right. The diagonal red line is the cut used by Monguió et al. (2020) to select emitting objects.

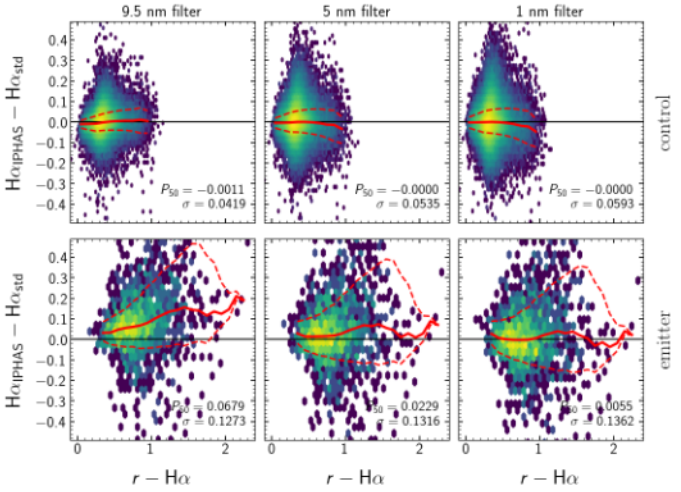


**Fig. 21.** Difference between IPHAS and synthetic H$\alpha$ magnitudes versus $r$−H$\alpha$ colour for the control (*top*) and emitter (*bottom*) samples, using synthetic H$\alpha$ filters of different widths. Potentially variable sources have been removed.

The broadband magnitudes of the control sample are well reproduced after standard corrections, which for simplicity we apply using polynomials rather than tweaking the shapes of the filters (we also note that we do not provide means to obtain this photometry with GaiaXPy). Here we do not intend to provide standardised XPSP, but instead to show that *Gaia* XPSP can be used to calibrate narrow-band photometry aimed at tracing line emission, especially in view of future surveys. We add 0.02 and 0.01 to the $r$ and H$\alpha$ ZP; respectively, while for $i$ we also apply a linear correction, with $i_{\rm std} = i_{\rm synth} + 0.01 + 0.11(r_{\rm synth} - i_{\rm synth})$. With these corrections, the IPHAS ($r-i$, $r-$H$\alpha$) colour plane is qualitatively reproduced.

However; after calibration using the control sample, the H$\alpha$ magnitudes in the emitter sample show both a dependency on the $r$−H$\alpha$ colour (which corresponds to the strength of the emission line) and a median offset of 0.07 mag (Fig. 21, left column). The H$\alpha$ fluxes in the synthetic photometry are generally lower

than those from IPHAS, with the discrepancy being greater for sources with stronger H$\alpha$ emission. We believe that this is due to flux in the emission line being lost outside the edges of the IPHAS H$\alpha$ passband (see Appendix B). For the vast majority of stars that do not have strong spectral features at this wavelength, such as these mentioned above, the flux lost is not a problem as it is replaced by flux bleeding in from outside the nominal filter wavelength range.

We considered other explanations for this discrepancy, such as variability or selection effects, but these can be ruled out by checking for consistency with photometry from second detections in the IPHAS data. Much of the IPHAS footprint is observed multiple times due to field overlaps, offsets to fill CCD gaps, and repeated observations to improve upon data taken in poor conditions. The emitter selection in Monguió et al. (2020) was based only on the primary detection. A small set of objects indeed show no emission in their second detections (largely corresponding to *Gaia* detections showing no emission in Fig. 20); aside from these, there is no systematic bias when comparing different IPHAS detections, ruling out selection effects or small-scale variability as the source of the discrepancy. In Fig. 21 we only include sources that have second detections, with the $r-$H$\alpha$ of the two detections consistent within 0.1 mag, which leaves 1682 sources in the emitter sample. This eliminates most of the sources lying below the selection line in the lower right plot of Fig. 20. As an additional check, we also compared a small set of emitters selected from the IPHAS sample with spectra published in Rodríguez-Flores et al. (2014). The high-resolution spectroscopy in that work is consistent with the IPHAS photometry for those sources, but the magnitude discrepancy was present for them in both simulated and actual *Gaia* photometry.

The sensitivity of the H$\alpha$ filter is related to its width, with narrower filters producing a greater magnitude difference between emitting and non-emitting sources. One way to better match the behaviour of the original filter is to use a narrower synthetic H$\alpha$ filter on the *Gaia* data. Doing so reduces the overall shift as well as the colour dependency, at the expense of greater scatter, particularly in the non-emitting sources (Fig. 21, centre and left columns). Nevertheless, we do not necessarily expect the narrowest filter to completely reproduce the range of colours from IPHAS (see Appendix B).

Despite the limitations discussed here, the practical functionality of the narrow-band filter – separating out emitters from non-emitting stars – is well reproduced by both the original IPHAS passband and the narrower versions, and moreover, the consistent performance in the H$\alpha$ passband of stars without strong H$\alpha$ emission enables flux calibration of survey fields, which in turn allows the selection of emission line stars even fainter than the publishing limit of the *Gaia* DR3 spectrophotometry. This is true despite the filter violating the *Rf* limitations discussed in Appendix B. Indeed, it should be possible to calibrate almost any narrow-band imagery taken with a well-characterised filter provided there are enough sufficiently bright, well-exposed stars in the field.

### 4.4. The project of a photometric system brought to life: C1

The original design for *Gaia* included a set of photometric passbands (Jordi et al. 2006), called C1B and C1M systems for broad and medium band photometry, respectively. The C1 system was especially thought to maximise the scientific return in terms of stellar astrophysical parameters. The spectral resolution requirements on the alternative prisms finally flying with the mission were made based on those passbands.

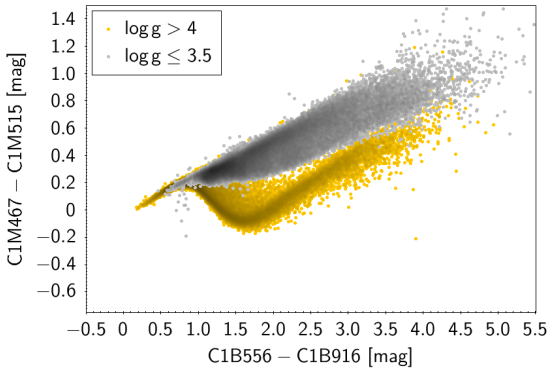


**Fig. 22.** Colour–colour diagram using the C1 photometric system, able to separate giant (grey) and main sequence stars (orange). log $g$ values are from GSP-Phot (Andrae et al. 2023).

Although some of the passbands in the C1 photometric system were finally implemented in the J-PLUS survey (Sect. 4.2), the synthetic photometry study in this paper provides the perfect opportunity to test the performance of the full C1 system. This illustrates the investigations that can be done even with future sets of passbands using EC XP spectra, which should be more accurate than only relying on simulated spectra from synthetic spectral libraries. Moreover, it serves as a good example of a photometric system that is realised in practice using only *Gaia* DR3 data. In principle, a general user of the *Gaia* Archive may conceive her/his own set of passbands designed for a specific science goal and get XPSP in that system for all the stars with XP spectra released in DR3. The example of applications shown here and in Sect. 5.3 for C1 showcases the performance that can be achieved for a well-designed system.

We can use the C1 synthetic photometry to learn about the performance of XP spectra, checking if they are able to trace the astrophysical information (see also Sect. 5.3). The aim here is not to repeat the work done by the *Gaia* DPAC, deriving again the astrophysical parameters of the sources (already available in the *Gaia* catalogue; Andrae et al. 2023; Fouesneau et al. 2023; Creevey et al. 2023), but is simply to evaluate whether or not the synthetic photometry derived with the C1 system is able to keep this information.

Using the C1 synthetic colour indices, we can perform for example a rough classification between giants and main sequence stars. For example, Fig. 22 shows the C1M467−C1M515 colour (sensitive to surface gravity) plotted against C1B556−C1B996 (sensitive to the effective temperature). Giants (in grey) and main sequence stars (in orange), which have different ranges of surface gravity (log $g$) values, as derived by DPAC/CU8 with the GSP-Phot module[18] (Andrae et al. 2023), are found at different positions in this diagram.

The C1 system can also be useful to estimate the metallicity of the studied sources, as we show in Sect. 5.3. On the other hand, Fig. 23 is intended to demonstrate the precision attainable with XPSP in this system. Two distance-corrected CMDs for the subset of stars from the *Gaia* Catalogue of Nearby Stars (GCNS; Gaia Collaboration 2021a) with $G < 17.65$ and the selection criteria listed in Sect. 6.2 are shown here. The CMD in the left panel is based on a combination of three C1B passbands,

[18] GSP-phot is the DPAC/CU8 module of the astrophysical parameters inference system (Apsis; Fouesneau et al. 2023) aimed at deriving the astrophysical parameters of the stars from XP spectra, parallaxes and $G$ apparent magnitudes, using a Bayesian full-forward modelling (Andrae et al. 2023).

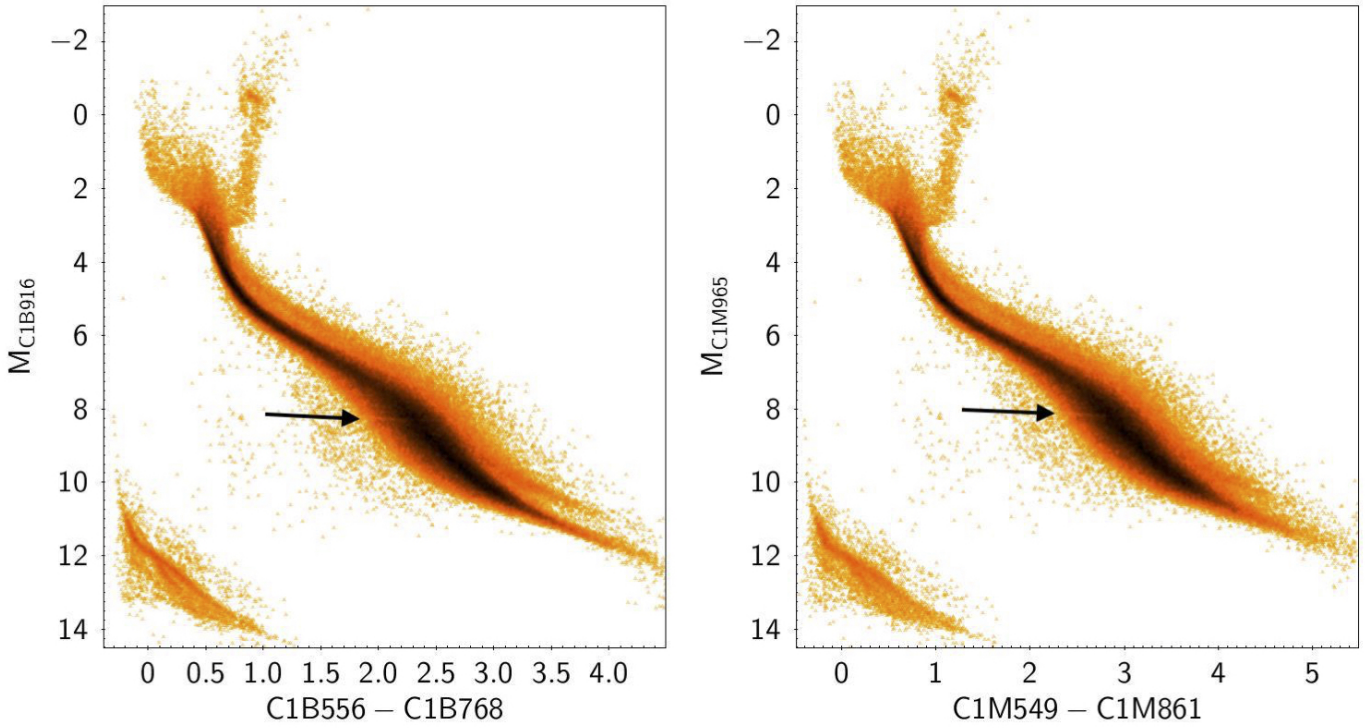


**Fig. 23.** Two examples of distance-corrected colour magnitude diagrams of the subset of the GCNS catalogue for which XPSP can be obtained, using combinations of broad (*left panel*) and medium width (*right panel*) passbands of the C1 system. Black arrows highlight the location of the Jao gap (Jao et al. 2018), which is clearly visible in both diagrams, showing the high precision of the XPSP in these bands.

while the one in the right panel is based on a combination of three C1M passbands. In both diagrams all the sequences typical of CMDs of the solar neighbourhood are very well defined, including the various WD subsequences (see Sect. 6.3, and references therein). The extremely subtle feature on the lower main sequence known as the Jao gap is clearly visible in both diagrams (Jao et al. 2018; Jao & Feiden 2020; Gaia Collaboration 2021a). This suggests that the precision of C1 XPSP is comparable to that achieved with the *Gaia* broadband system, as defined by Riello et al. (2021).

## 5. Performances verification experiments

In this section, we show a few examples of performance verification of XPSP against real science goals. In particular, we show that, in some cases, XPSP can be used to trace multiple populations (MPs) in GCs, or to estimate metallicity (also for extremely metal-poor stars) and even the abundance of $\alpha$ elements. It is also shown that XPSP can be used to identify emission line sources (ELS) with accuracy similar to that achieved from direct analysis of XP spectra. We discuss a further example of application in Sect. 6.3, namely classification of WDs.

### 5.1. Multiple populations in globular clusters

In the last four decades, the concept of GCs as a coeval and homogeneous simple stellar population has dramatically changed thanks to the discovery of star-to-star abundance variations in almost all GCs, which produce multiple photometric evolutionary sequences in the CMD (Bastian & Lardo 2018; Gratton et al. 2019, and references therein). These MPs can therefore be studied not only with spectroscopy, but also with high-quality photometry (see, e.g. Piotto et al. 2007; Milone et al. 2012, 2013, 2015; Lee 2019). In particular, *UV* passbands are sensitive to the deep CN molecular bands at 388 nm (Pancino et al. 2010; Sbordone et al. 2011), and therefore stars with normal and enhanced N can be efficiently separated photometrically (e.g. Yong et al. 2008; Lardo et al. 2011; Carretta et al. 2011, and references therein). About 20% of GCs also show multiple photometric sequences in CMDs not involving the *U* band, which are the result of different He and C+N+O abundances (Pancino et al. 2000; Sbordone et al. 2011; Milone et al. 2015; Monelli et al. 2013).

Here we use MPs in GCs to demonstrate the performance of the JKC synthetic standardised photometry presented in Sect. 3.2. For this purpose, we tested all the GCs selected for other performance verification cases (Sect. 5.2), complemented with a selection of half a dozen GCs hosting spectacular and well-studied MPs. In each GC, we selected the sample stars using the membership probability by Vasiliev & Baumgardt (2021, >0.9). In doing so, we implicitly adopted their careful and complex selection on the quality of the *Gaia* astrometry based on `RUWE`, the IPD parameters, and other indicators (see their Sect. 2). This selection appears in grey in Fig. 24. We further selected stars according to the following criteria (see also Sect. 3.2): (i) $|C*| < \sigma_{C*}$ (Riello et al. 2021, Sect. 9.4); (ii) `RUWE` < 1.4; (iii) `IPD_frac_multi_peak` < 7; (iv) `IPD_frac_odd_win` < 7; and (v) $\beta < 0.2$ (Riello et al. 2021, Sect. 9.3). This selection appears in red in Fig. 24. We note that of all the applied selections, only the one on $\beta$ really makes a difference, because selections on the other parameters were already explicitly or implicitly applied by Vasiliev & Baumgardt (2021).

In the top panels of Fig. 24, we compared the $V$, $B - V$ synthetic standardised photometry with the ground-based photometry by Stetson et al. (2019) for M 2 (NGC 7089), a GC well-known for hosting an anomalous RGB, containing a few percent of the stars redder than the main RGB

(Lardo et al. 2012, 2013). As can be noted, the sample selected with the Vasiliev & Baumgardt (2021) membership and quality criteria (in grey in the figure) displays a 'wind' of stars that are bluer than the red giant branch that is not present in the Stetson et al. (2019) photometry. This is caused by the fact that the typical seeing in the Stetson et al. (2019) data was of about 1.0″, while the typical aperture of XP spectra is of 3.5″ × 2.1″. The *Gaia* XP synthetic photometry therefore suffers more from crowding and blending effects. Additionally, the Stetson et al. (2019) photometry is obtained by PSF fitting and with sophisticated deblending routines, while in DR3 no detailed deblending has been performed. Future *Gaia* releases will tackle blending and contamination with ad hoc processing pipelines, but in *Gaia* DR3 we can use several indicators of crowding, such as the $\beta$ parameter defined by Riello et al. (2021). If we further select the sample as described above (red stars in Fig. 24), a cleaner RGB is obtained, but at the expense of completeness. In any case, in both the ground-based and the *Gaia* XPSP CMDs the anomalous branch is clearly visible as a sparsely populated sequence ≃0.2 mag redder than the main RGB, which demonstrates the very high performance of the synthetic photometry presented here.

In order to investigate the case of the $U$ band, we adopt the colour index $C_{UBI}$ (defined as $(U-B)-(B-I)$, Monelli et al. 2013) that combines and amplifies the effect of the variations in both N and He. The bottom panels of Fig. 24 show the case of NGC 6752. In the ground-based photometry, the presence of MPs is indicated by the well-separated RGBs, while in the synthetic photometry the separation is not so clearly visible, but the MP presence is clear because of the large width of the RGB (>0.1 mag) compared to the typical photometric errors (<0.03 mag, see Sects. 3.2 and 3.3). As in the case of M 2, a further selection including $\beta$ is necessary (red points in Fig. 24) in the synthetic photometry in order to clean the sample of untreated blends, also at the expense of completeness. We would like to highlight that NGC 6752 is among the closest GCs (less than 4 kpc) and is one of the cases in which MPs can be more clearly identified using XPSP. Another iconic GC, NGC 1851, which is more distant, very compact, and displays a rather complex RGB substructure in the ground-based $U$-band CMD (see Fig. 10 by Stetson et al. 2019), does not clearly reveal any substructure in the synthetic *Gaia* $U$-band and $C_{UBI}$ CMDs. This is likely because (i) the *Gaia* wavelength range does not fully include the $U$-band (as discussed in Sects. 2.2.2 and 3.3); (ii) the cluster is more compact than NGC 6752, with a half-light radius of $r_h = 0.51'$, to be compared with $r_h = 1.91'$ for NGC 6752 (Harris 1996), thus crowding effects are necessarily expected to produce a stronger effect on NGC 1851(see also Pancino et al. 2017); and (iii) the treatment of blending and contamination introduced in *Gaia* EDR3 (Riello et al. 2021) is still not the complete treatment planned for *Gaia* DR4. We also note that there is a zero-point offset between the XPSP and the ground-based photometry in the $U$ band that varies from GC to GC (0−0.2 mag). This is also likely caused by the above effects and is also due in part to the fact that the ground-based photometry Stetson et al. (2019, see in particular their Fig. 4; see also Appendix E, below) is based on a collection of images taken with different facilities and filters, and that the $U$-band is notoriously difficult to standardise (Altavilla et al. 2021).

In conclusion, the *BVRI* XPSP is of very high quality already, even for relatively distant GCs, in spite of the fact that the treatment of blends is not yet fully implemented in *Gaia* DR3, provided that one carefully selects the stars whilst considering parameters such as $\beta$ (Riello et al. 2021). In the case of the $U$-band, the photometric performance is unavoidably lower than what is needed to study these fine details. Further improvements are eagerly awaited in the next *Gaia* releases.

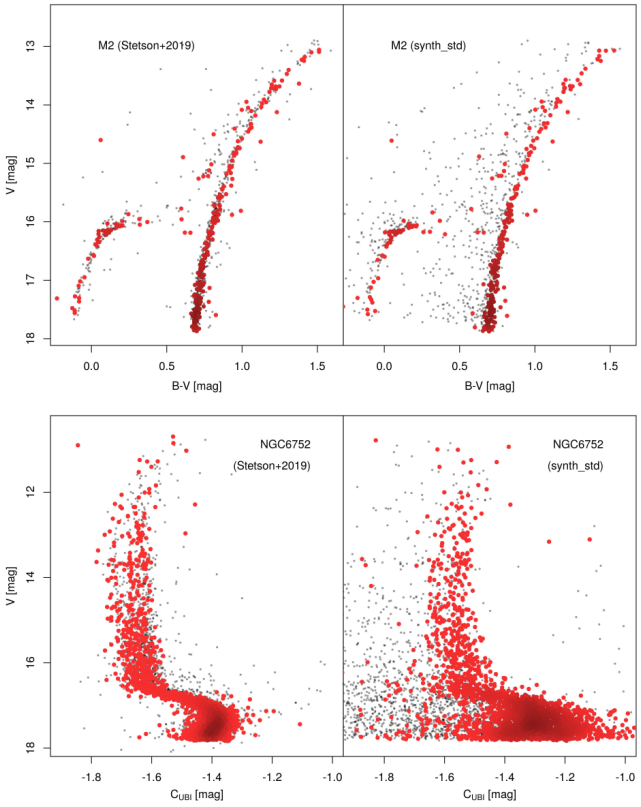


**Fig. 24.** *Top panels*: ground-based $V$, $B-V$ photometry of M 2 (*left*, Stetson et al. 2019) and the corresponding synthetic *Gaia* photometry (*right*). The samples selected using the Vasiliev & Baumgardt (2021) criteria are plotted in grey, while the ones further selected with the criteria described in Sect. 5.1 are plotted in red. *Bottom panels*: similar to the *top panels*, but for the case of the $V$, $C_{UBI}$ CMDs of NGC 6752.

### 5.2. Metallicity from the Strömgren system

The Strömgren index $m_1 = (v-b)-(b-y)$ has been widely used as a tool to infer the metal abundance of giant stars (see e.g. Richter et al. 1999; Anthony-Twarog & Twarog 2000). To explore the efficacy of our synthetic Strömgren photometry in recovering this parameter, we selected a sample of Galactic globular and open clusters (OCs), for which high-resolution spectroscopic [Fe/H] estimates exist. We determined their mean metallicity from *Gaia* XPSP by adopting the $m_0-(v-y)_0$–[Fe/H] relation provided by Calamida et al. (2007, based on GCs) for RGB stars, where $m_0$ and $(v-y)_0$ are the de-reddened version of the $m_1$ index and the $(v-y)$ colour, respectively. It worth noting that such a relation is based on photometry that is calibrated on the same list of standards upon which our standardised XPSP ultimately relies, namely that provided by Crawford & Barnes (1970, see Sect. 4.1). The extinction law by Cardelli et al. (1989), which provides $A_v/A_V = 1.397$, $A_b/A_V = 1.240$ and $A_y/A_V = 1.005$ as extinction coefficients, has been adopted to correct the Strömgren magnitudes for reddening. The 12 selected GCs span a metallicity range from [Fe/H] ∼ −2.5 to [Fe/H] ∼ −0.7. To extend this range towards higher values, we included four metal-rich OCs with metallicities in the range −0.2 < [Fe/H] < +0.4. In order to homogeneously select red giant

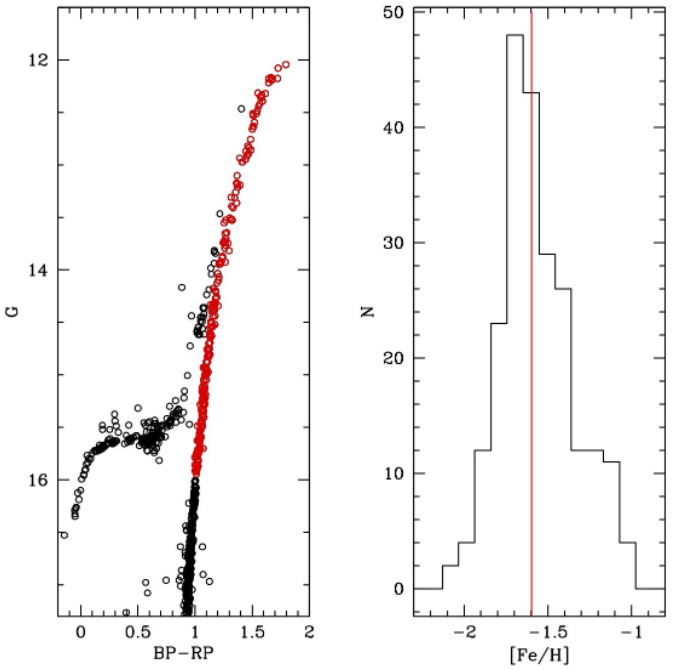


**Fig. 25.** *Left-hand panel*: *Gaia* CMD for the globular cluster NGC 5272. Red symbols indicate the giants for which a metallicity estimate was derived by means of the synthetic Strömgren photometry. *Right-hand panel*: metallicity distribution inferred from the synthetic $m_0$ index for the selected stars. The vertical red line marks the 2.5$\sigma$-clipped mean value of [Fe/H] = −1.59 ($\sigma$ = 0.14). The cluster spectroscopic metallicity as quoted in Carretta et al. (2009) is [Fe/H] = −1.50 ± 0.05.

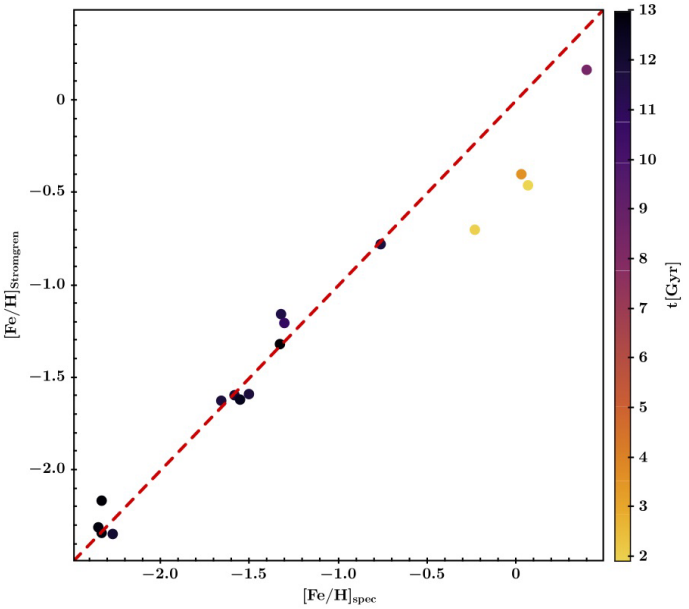


**Fig. 26.** Spectroscopic metallicity vs. photometric metallicity derived from the synthetic $m_0$ Strömgren index. The colour-coding indicates the age of each star cluster. Error bars are smaller than the symbol size.

stars in the analysed stellar clusters, we focused our analysis on all the stars from the red giant branch tip down to about 4 mag fainter, and manually excluded obvious asymptotic giant branch (AGB) stars. Quality cuts were applied to the selection in order to select stars with $-0.03 < C^* < 0.03$. Finally, the stellar membership to the GCs was ensured by setting a minimum membership probability as determined by Vasiliev & Baumgardt (2021) of 90%. This typically led to samples of several tens, or hundreds, of stars per cluster. Figure 25 summarises an example of such a procedure, by showing the *Gaia* CMD of the globular cluster NGC 5272 in the left-hand panel, by highlighting the selected targets with red symbols, and by reconstructing their metallicity distribution derived from the synthetic Strömgren index $m_0$ in the right-hand panel.

The list of 12 GCs includes NGC 104, NGC 288, NGC 362, NGC 4590, NGC 5272, NGC 6205, NGC 6218, NGC 6341, NGC 6752, NGC 7078, and NGC 7099. Their spectroscopic metallicity is taken from the homogeneous scale provided by Carretta et al. (2009), while we adopt the values provided in Harris (1996, with the 2010 update, available at `https://physics.mcmaster.ca/~harris/mwgc.dat`) for the extinction. The clusters have in any case been purposely selected among the low-extinction ones. The additional four open stellar clusters are NGC 2506, NGC 6791, NGC 6819, and M 67. Their spectroscopic metallicity and reddening are taken from Carretta et al. (2004), Bragaglia et al. (2014, 2001), and Zhang et al. (2021), respectively. The membership probability of the stars of these clusters is instead taken from Cantat-Gaudin et al. (2020). Figure 26 shows the one-to-one comparison between the spectroscopic metallicity and the 2.5$\sigma$-clipped mean metallicity determined from our Strömgren photometry for the 16 star clusters, which are colour-coded according to their age. The error bars correspond to the error on the mean of each cluster metallicity distribution. Clusters ages are taken from VandenBerg et al. (2013) for the GCs, and from Bossini et al. (2019) for the OCs.

From Fig. 26, it is clear that while the comparison for the GCs is good, the OCs are systematically offset by about 0.4 dex. Among these, the best behaved is NGC 6791, which is the oldest ($t$ = 8.4 Gyr). When interpreting these results, we should bear in mind that the colour of a giant star depends primarily on metallicity, but also on age. The sample of OCs has been chosen in such a way as to sample the high-metallicity part of the [Fe/H] distribution, and consists of systems that are much younger than the GCs used by Calamida et al. (2007) to calibrate their relation, that is, younger by between 4 and 10 Gyr. Therefore, the systematic offset of the OCs is likely due to this intrinsic age difference rather than to a poor sensitivity of the Calamida et al. (2007) relation at this high [Fe/H] (which is an effect that could nevertheless still contribute). As supporting evidence, we note that the observed offset goes in the direction of our interpretation, in the sense that young OCs have intrinsically bluer RGBs, thus mimicking an old and more metal-poor population.

To assess the precision and accuracy of the metallicity estimates obtained via synthetic photometry, we therefore restricted our analysis to the GCs sample, finding rather good results. The mean difference between our photometric estimates and the spectroscopic values is 0.02 dex, with a dispersion of 0.08 dex. Such a dispersion closely matches the findings by Calamida et al. (2007), who quote a precision for their relations of ∼0.1 dex. As a last remark, the nominal error on the mean metallicity obtained from the synthetic Strömgren photometry is quite small (because of the large number of available stars), and ranges from 0.01 dex in the case of NGC 104 up to 0.04 dex for NGC 7099 (i.e. from the more metal-rich to the more metal-poor GCs). This in turn means that, in the considered case, the dominant contribution to the observed dispersion comes from the precision of the $m_0 - (v - y)_0$–[Fe/H] relation itself.

Another way of testing the performance of our synthetic Strömgren photometry in determining the metallicty of giants is to directly compare with the spectroscopic measurements for nearby stars from large surveys such as GALAH (Buder et al. 2021) and APOGEE (Ahumada et al. 2020). To do so, after cross-matching *Gaia* DR3 sources with GALAH DR3 and APOGEE DR16, we select giants by requiring the spectroscopic log $g$ measurements of our sample to be smaller than 2.5 (this is a strict selection, excluding lower RGB stars). After inspecting the *Gaia* HR diagram of these giants (which we obtained by correcting the observed magnitudes for reddening and distance using *Gaia* DR3 parameters), a further cut at $M_G < 4.5$ was applied to exclude obvious dwarfs with likely uncertain log $g$. To avoid the inclusion of highly reddened sources, we also imposed $0 < E(G_{BP}-G_{RP}) < 0.1$. Finally, a quality cut at $-0.1 < C^* < 0.1$ was required to exclude low-quality *Gaia* measurements. We are then left with a sample of 3202 giants in common with GALAH DR3, and 5573 giants in common with APOGEE DR16 (all of these are located at a mean distance of ∼2 kpc, with $\sigma \sim 1$ kpc).

The upper panel of Fig. 27 shows the difference between the GALAH spectroscopic metallicity and that derived from the Strömgren $m_0$ index, as a function of the former. Overall, the agreement looks reasonable. The mean value of the distribution is $\Delta$[Fe/H] = 0.33, with a dispersion of $\sigma = 0.25$. The distribution itself shows a positive trend for higher metallicity and flattens for $[Fe/H]_{GALAH} < -0.5$. Such a trend is consistent with the age effect that has already been observed and discussed for the star clusters (see Fig. 26). More metal-rich stars are likely among the youngest of the sample, and the Calamida relation tends to underestimate their metallicity, while still performing reasonably well for the stars that are older and more metal-poor. Unfortunately, the GALAH sample is intrinsically lacking in metal-deficient stars, meaning that we cannot robustly test the behaviour at lower metallicity. For the sake of cross-validation, the second-row panel of Fig. 27 shows the difference between the photometric metallicity estimated from XP spectra by *Gaia* GSP-Phot (Andrae et al. 2023) and that coming from the synthetic Strömgren photometry, as a function of $[Fe/H]_{GALAH}$. In this case, the agreement is even better, with a mean difference of $\Delta$[Fe/H] = 0.16 and a somewhat tighter dispersion $\sigma = 0.20$.

The lower panels of Fig. 27 show the same kind of comparison, with the metallicity measurements coming from APOGEE. The behaviour is very similar to that described for GALAH. The mean difference between the spectroscopic and the Strömgren metallicity is $\Delta$[Fe/H] = 0.27, with a dispersion of $\sigma = 0.25$ and a similar positive trend. As in the previous case, the comparison improves when using *Gaia* GSP-Phot measurements as a reference, as the mean difference drops to $\Delta$[Fe/H] = 0.14 with a dispersion of $\sigma = 0.30$. Consistency within $\simeq \pm 0.2-0.3$ dex is also achieved in comparison to *Gaia* GSP-Phot metallicity.

Our analysis is particularly relevant in the case of distant sources, especially in the metal-poor regime. These are the cases where our primary source of metallicity from XP spectra encounters some limitations, while otherwise, GSP-Phot provides astrophysical parameters with good accuracy for the large majority of stars (Andrae et al. 2023; Fouesneau et al. 2023). For example, when estimating the mean metallicity of the GCs analysed above, GSP-Phot provides values that tend to significantly overestimate the metal content of these stellar systems. Hence, when robust estimates of the extinction exist, metallicity-sensitive distance-independent colour indices obtained from synthetic photometry from XP spectra, like that presented here and in the following sections, can provide an useful alternative solution that is highly complementary to GSP-Phot.

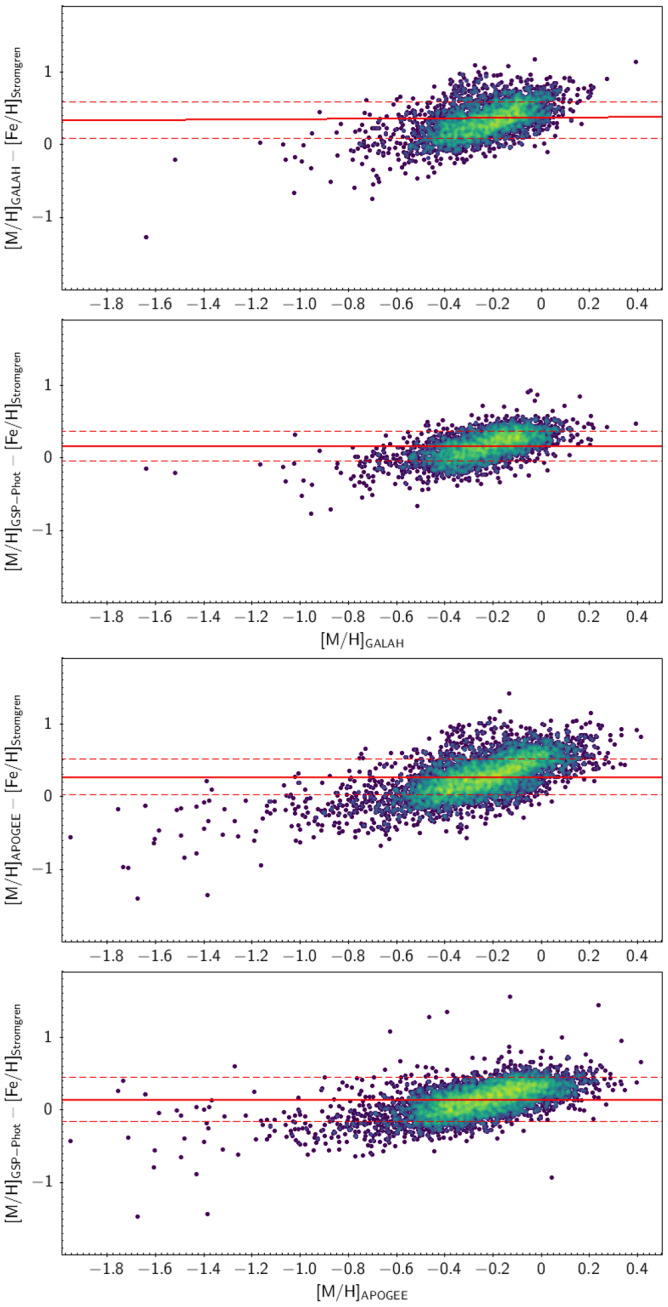


**Fig. 27.** *Top panel*: difference between the spectroscopic metallicity from GALAH DR3 and that coming from our synthetic Strömgren $m_0$ index. The red lines mark the mean value and the $1\sigma$ dispersion. *Second panel*: same but with *Gaia* GSP-Phot metallicty instead of that from GALAH. *Third panel*: difference between the spectroscopic metallicity from APOGEE DR16 and that coming from our synthetic Strömgren $m_0$ index. The red lines mark the mean value and the $1\sigma$ dispersion. *Bottom panel*: same but with *Gaia* GSP-Phot metallicity instead of that from APOGEE.

### 5.3. Metallicity from the C1 system

Metallicity and $\alpha$-element abundance information is more difficult to retrieve than temperature and surface gravity (see Sect. 4.4). Abundances leave an imprint in the spectra in narrow ranges of wavelength, and narrow passbands are more sensitive to fluctuations in the spectra. Nevertheless, metallicity and even the $\alpha$-element abundance can be studied with the C1 system.

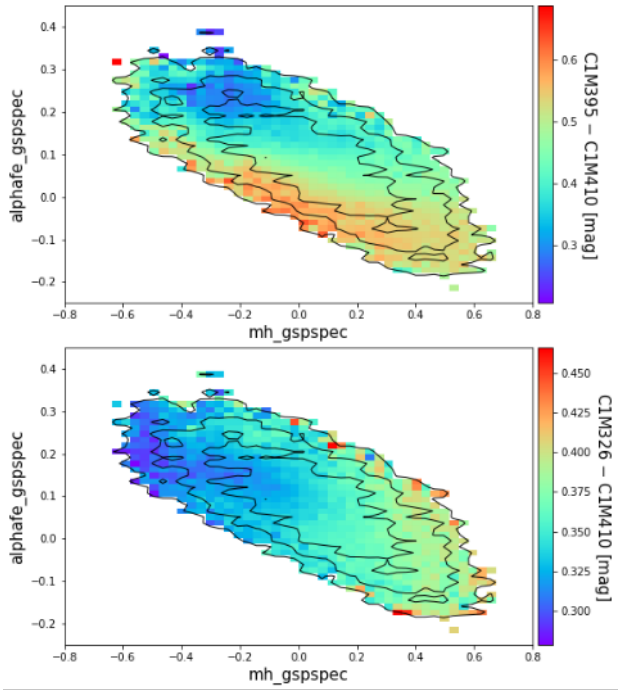


**Fig. 28.** $\alpha$-element abundance ([$\alpha$/Fe]) as a function of global metallicity ([M/H]) of the main sequence stars ($\log g > 4.0$ dex) and good quality flags (the first 13 digits in `flags_gspspec` equal to zero). All parameters were derived by GSP-Spec (Recio-Blanco et al. 2023c). Colour indices show the values of the colour C1M395−C1M410 (*top panel*), which is changing due to the $\alpha$ abundance and C1M326−C1M410 (*bottom panel*), which depends more on the global metallicity. Contours indicate density dropping by factors of 5.

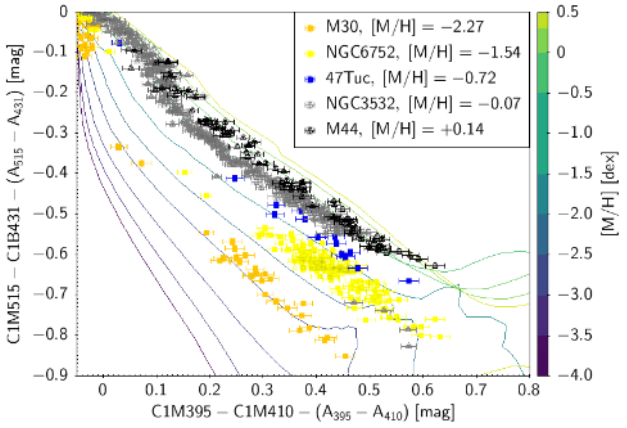


**Fig. 29.** C1 colour–colour diagram sensitive to global metallicity for a set of clusters corrected for reddening using $A_X$ absorption values in that passband derived as indicated in Appendix H. Lines represent the simulations performed using the BTSettl library (Allard et al. 2013; with a line colour depending on the global metallicity, [M/H]). Solid squares represent the stars in GCs and empty triangles the stars in OCs, all of them with their error bars. BTSettl models with $\log g = 2.0$ are plotted for GCs, and $\log g = 3.0$ for OCs.

For example, Fig. 28, where spectroscopic metallicities and [$\alpha$/Fe] values are taken from GSP-spec[19] (Recio-Blanco et al. 2023c), shows that when the pseudo-continuum at C1M410 is compared with C1M395 affected by CaII HK lines, we see a dependence with [$\alpha$/Fe] abundance. On the other hand, when the same C1M410 pseudo-continuum passband is compared with C1M326 (measuring the UV Balmer jump), this colour index has a stronger variation with metallicity ([M/H]) than with [$\alpha$/Fe] abundance. It is interesting to note that these passbands are able to trace chemical composition in spite of the fact that they sample a critical region of the XP spectra.

Using well-studied open and globular clusters we can test the relationship between chemical abundances and C1M colour indices. Figure 29 shows the colour C1M515−C1B431 plotted against C1M395−C1M410, which is able to separate different metallicities. Only sources with total uncertainty $\sigma_C < 0.02$ mag were plotted in the figure, where

$$\sigma_C \equiv \sqrt{\sigma^2_{\rm C1M395} + \sigma^2_{\rm C1M410} + \sigma^2_{\rm C1M515} + \sigma^2_{\rm C1B431}}. \quad (19)$$

The C1 synthetic photometry in Fig. 29 has been corrected from reddening effects using the relationships included in Appendix H. The absorption values used to perform this correction were obtained from literature estimates (Harris 1996 for M 30, NGC 6752, and NGC 104 GCs, Fritzewski et al. 2019 for NGC 3532, and Taylor 2006a for M 44 OCs).

The lines in Fig. 29 show the iso-metallicity lines derived from the BTSettl library (Allard et al. 2013). In the low metallicity range, the BTSettl lines were derived using a surface gravity value equal to $\log g = 2.0$, as only the giant stars in the GCs can be observed with enough accuracy in this colour–colour diagram. For higher metallicity, a value equal to $\log g = 3.0$ was used. As can be seen in Fig. 29, each cluster follows a metallicity track. Therefore, we can conclude that, as we have also seen for temperature and surface gravity, the XP spectra (and therefore the synthetic photometry) allow us to discriminate between the abundance effects present in the spectra.

The same diagram used for clusters in Fig. 29 can also be used for field stars. In Fig. 30 we show an example of this with a set of sources selected to compare with the results obtained with GSP-Phot (Andrae et al. 2023). We include sources with $\log g > 4$ dex, $A_G < 0.005$ mag, and $\sigma_C < 0.02$ mag from Eq. (19). The results of this cross-validation test are satisfactory.

### 5.4. Very metal-poor stars

In this section, we push the metallicity analysis one step further into the lower metallicity star regimes, that is, [Fe/H] < −2 dex. The more metal-poor the star, the more pristine it is. Ultra-metal poor (UMP) stars ([Fe/H] < −4 dex) belong to the earliest generations of stars formed in the Universe. Because of their very low abundance in metal elements, they are critical anchors to address questions on the formation of the first generation of stars, the (non-)universality of the initial mass function (IMF), the early formation stages of galaxies, and the first supernovae (e.g. Beers & Christlieb 2005). However, the minimum metallicity at which low-mass stars can form is still an open question (see Greif 2015, and references therein). Only 42 UMPs are known to date in our Galaxy despite simulations predicting multiple thousands of them (Karlsson et al. 2013). These stars are scarce objects, and are relatively faint sources because of their low masses. Finding them is therefore a challenge, and we are limited to mostly finding them in our Galaxy.

[19] GSP-spec is the DPAC/CU8 Apsis module designed to derive chemical abundances from *Gaia* RVS spectra (Recio-Blanco et al. 2023c). Here, to select well-measured abundances we considered only stars with the first 13 digits in `flags_gspspec` equal to zero. We note that very similar results as those shown in Fig. 28 are obtained if APOGEE DR16 abundances are used instead of GSP-spec ones.