Gaia Data Release 3

External calibration of *BP/RP* low-resolution spectroscopic data

P. Montegriffo[1],[*], F. De Angeli[2], R. Andrae[3], M. Riello[2], E. Pancino[4,5], N. Sanna[4], M. Bellazzini[1],
D. W. Evans[2], J. M. Carrasco[6], R. Sordo[7], G. Busso[2], C. Cacciari[1], C. Jordi[6], F. van Leeuwen[2],
A. Vallenari[7], G. Altavilla[8,5], M. A. Barstow[9], A. G. A. Brown[10], P. W. Burgess[2], M. Castellani[8], S. Cowell[2],
M. Davidson[11], F. De Luise[12], L. Delchambre[13], C. Diener[2], C. Fabricius[6], Y. Frémat[14], M. Fouesneau[3],
G. Gilmore[2], G. Giuffrida[8], N. C. Hambly[11], D. L. Harrison[2,15], S. Hidalgo[16], S. T. Hodgkin[2], G. Holland[2],
S. Marinoni[8,5], P. J. Osborne[2], C. Pagani[9], L. Palaversa[17,2], A. M. Piersimoni[12], L. Pulone[8], S. Ragaini[1],
M. Rainer[4], P. J. Richards[18], N. Rowell[11], D. Ruz-Mieres[2], L. M. Sarro[19], N. A. Walton[2], and A. Yoldas[2]

1. INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, 40129 Bologna, Italy
2. Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
3. Max-Planck-Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
4. INAF – Osservatorio Astrofisico di Arcetri, Largo E. Fermi, 5, 50125 Firenze, Italy
5. Space Science Data Center – ASI, Via del Politecnico SNC, 00133 Roma, Italy
6. Institut de Ciències del Cosmos (ICC), Universitat de Barcelona (IEEC-UB), c/ Martí i Franquès, 1, 08028 Barcelona, Spain
7. INAF – Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122 Padova, Italy
8. INAF – Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monte Porzio Catone, Roma, Italy
9. School of Physics & Astronomy, University of Leicester, Leicester LE9 1UP, UK
10. Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
11. Institute for Astronomy, School of Physics and Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
12. INAF – Osservatorio Astronomico d'Abruzzo, Via Mentore Maggini, 64100 Teramo, Italy
13. Institut d'Astrophysique et de Géophysique, Université de Liège 19c, Allée du 6 Août, 4000 Liège, Belgium
14. Royal Observatory of Belgium, 3 Avenue Circulaire, 1180 Brussels, Belgium
15. Kavli Institute for Cosmology, Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
16. IAC – Instituto de Astrofisica de Canarias, Via Láctea s/n, 38200 La Laguna SC, Tenerife, Spain
17. Ruđer Bošković Institute, Bijenička Cesta 54, Zagreb, Croatia
18. STFC, Rutherford Appleton Laboratory, Harwell, Didcot OX11 0QX, UK
19. Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia, c/ Juan del Rosal 16, 28040 Madrid, Spain

ABSTRACT

Context. Gaia Data Release 3 contains astrometry and photometry results for about 1.8 billion sources based on observations collected by the European Space Agency (ESA) *Gaia* satellite during the first 34 months of its operational phase (the same period covered by *Gaia* early Data Release 3; *Gaia* EDR3). Low-resolution spectra for 220 million sources are one of the important new data products included in this release.

Aims. In this paper, we focus on the external calibration of low-resolution spectroscopic content, describing the input data, algorithms, data processing, and the validation of the results. Particular attention is given to the quality of the data and to a number of features that users may need to take into account to make the best use of the catalogue.

Methods. We calibrated an instrument model to relate mean *Gaia* spectra to the corresponding spectral energy distributions (SEDs) using an extended set of calibrators: this includes modelling of the instrument dispersion relation, transmission, and line spread functions. Optimisation of the model is achieved through total least-squares regression, accounting for errors in *Gaia* and external spectra.

Results. The resulting instrument model can be used for forward modelling of *Gaia* spectra or for inverse modelling of externally calibrated spectra in absolute flux units.

Conclusions. The absolute calibration derived in this paper provides an essential ingredient for users of *BP/RP* spectra. It allows users to connect *BP/RP* spectra to absolute fluxes and physical wavelengths.

Key words. catalogs – surveys – instrumentation: photometers – instrumentation: spectrographs – techniques: photometric – techniques: spectroscopic

---

[*] Corresponding author: P. Montegriffo, e-mail: paolo.montegriffo@inaf.it

1. Introduction

The European Space Agency (ESA) mission *Gaia* (Gaia Collaboration 2016) is designed to be self-calibrating for the large majority of its data products. For example, the core product of the mission, namely exquisitely accurate and precise astrometry for ≃1.8 billion celestial sources, is entirely based on observations obtained by the mission itself (relative positions at different epochs and the relative colours of the sources), while external data are only used for validation (Lindegren et al. 2021). Analogously, the removal of any instrumental imprint and/or space- or time-dependent inhomogeneities from the mission all-sky photometry and spectrophotometry is achieved using repeated measurements of large sets of internal calibrators (Riello et al. 2021; Carrasco et al. 2021; De Angeli et al. 2023).

However, within the production chain of photometry and spectrophotometry, there are two remarkable exceptions to this generally adopted approach:

(1) The physical flux scale, the main ingredient in the conversion of internally calibrated fluxes (expressed in $e^- \, s^{-1}$) into physical units ($W \, m^{-2} \, nm^{-1}$), which is determined using an external set of spectrophotometric standard stars, the *Gaia* spectrophotometric standard stars (SPSSs; Pancino et al. 2021).

(2) The physical wavelength scale, required to convert internal pseudo-wavelength labels (pixels; see Sect. 2) associated to fluxes in *BP* and *RP* spectra into wavelengths in physical units (nm), achieved (mainly) thanks to a set of external spectra of sources with strong emission lines at known wavelength.

As there is no way to infer these scales from *Gaia* data alone, it is necessary to make use of external calibration data.

The *BP* and *RP* Instrument Models (IMs), which include these two fundamental components, depend on a number of factors (the dispersion relation, the instrument response, and the line spread function (LSF); see Sects. 2 and 4), which are derived using external data in the process known as *absolute calibration*. The IM is the fundamental tool for forward modelling of *BP* and *RP* observations, starting from a theoretical model spectrum, with the main goal being to facilitate the inference of astrophysical parameters from their *BP*/*RP* spectra by comparison on the plane of observations (Creevey et al. 2023). Once the parameters of the IM have been estimated (see Sect. 6), the model can also be used in the opposite direction, that is, to transform an internally calibrated mean *BP*/*RP* spectrum (De Angeli et al. 2023) into a wavelength- and flux-calibrated spectrum that we call an externally calibrated spectrum (ECS). As the IM also includes the modelling of the LSF at any wavelength, its application significantly reduces the effect of photon mixing inherent to the slit-less spectra produced by the *BP* and *RP* spectrophotometers, enhancing the effective spectral resolution of ECS. It is important to realise that the IM solves for all the relevant factors (e.g. calibrations of flux and wavelength, and LSF) simultaneously, as they are deeply and inseparably entangled in *BP*/*RP* spectra.

While *BP* and *RP* spectrophotometry has already been used for internal processing in previous releases (Riello et al. 2021), with *Gaia* Data Release 3 (*Gaia* DR3, Gaia Collaboration 2023a) the *BP*/*RP* spectra of about 220 million sources are released for the first time. These can be retrieved from the *Gaia* Archive[1] as internally calibrated mean spectra in a continuous representation (see De Angeli et al. 2023 for details) while for a subset of sources with $G < 15$ they will also be provided as ECS sampled on a standard wavelength grid (see Appendix B

for more details on data formats). The Python package `GaiaXPy` (De Angeli et al., in prep.) has been developed to help users to convert spectra from continuous to sampled representation in the internal or absolute flux scale (ECS). The tool also implements the IM presented here to allow for the simulation of *Gaia*-like mean spectra from a given spectral energy distribution (SED); for example a synthetic stellar spectrum or an absolute-flux-calibrated measured spectrum.

In this paper, we illustrate how the *BP*/*RP* IM is derived and how internally calibrated mean *BP* and *RP* spectra are converted into ECS, discussing the performances and the limitations of the final products. After giving an overview of the external calibration approach (Sect. 2) and a description of the external calibrators (Sect. 3), we describe the implementation of the IM (Sect. 4) and the method implemented to reconstruct the ECS (Sect. 5). Section 6 is dedicated to the description of the processing scheme and the main results are shown in Sect. 7, while Sect. 8 is dedicated to the validation of the calibrations. Finally, in Sect. 9 we discuss a few known problems.

2. Overview of the problem

The instrument model (described in detail in Sect. 4) allows us to relate a mean *BP* or *RP* spectrum to the corresponding SED via an integral equation of the following kind:

$$n_e(u) = \int I(u, \lambda) \cdot n_p(\lambda) \, d\lambda, \tag{1}$$

where the observed spectrum $n_e(u)$ is the internally calibrated mean spectrum in units of $e^- \, s^{-1}$ (u denotes a coordinate in data space, often referred to as a *pseudo-wavelength*) and the kernel I is a combination of few components:

(i) the LSF, that is, the instantaneous one-dimensional intensity distribution in the spectrum of a monochromatic point source;

(ii) the dispersion model, that is, the relation that links absolute wavelengths to pseudo-wavelength coordinates; and

(iii) the response model, which represents the ratio between the number of detected photons for wavelength interval and the number of photons per wavelength interval entering the telescope aperture.

As the detectors are photon-counting devices, all the relations are expressed in terms of $n_p(\lambda)$, the spectral photon flux distribution (SPD). The SPD (hereafter expressed in units of photon $m^{-2} \, s^{-1} \, nm^{-1}$) is related to the source SED via the equation

$$n_p(\lambda) = \frac{10^{-8} \lambda}{hc} f(\lambda), \tag{2}$$

where hc is the product of the Planck constant and the vacuum speed of light, and the SED $f(\lambda)$ is expressed in units of $W \, m^{-2} \, nm^{-1}$ (the factor 10^{-8} compensates for per-nanometre flux units; all other measurements and constants are expressed in S.I. units).

The external calibration concept is based on two key assumptions: the first is that all differential effects that impact raw observations (spatial variations of the instrument across the focal plane, different observing configurations, time evolution of instrument characteristics, etc.) have been removed by the internal calibration chain (De Angeli et al. 2023) that precedes the external calibration; the second is that the common reference system defined by the internal calibration is similar to the physical instrument in some unspecified point of the focal plane.

These assumptions underlie the design of the external calibration strategy:

(a) The external calibration model is unique for all sources and in particular does not depend on source luminosity or colour. The first assumption above, although plausible, cannot be a priori guaranteed and any infringement of it will manifest as systematic differences within different classes of sources.

(b) Also, the IM can be parametrised in terms of corrections to the *nominal* instrument, which is defined by our pre-launch knowledge and laboratory measurements made on the satellite components.

This second point helps to cope with degeneracies that are present in possible IM solutions: in principle the optimisation of the IM parameters could be derived from an arbitrarily large number and variety of standard stars (sources with known SED from independent ground- or space-based observations) by matching the predictions of the model with the corresponding observed *BP/RP* mean spectra. As discussed by Weiler et al. (2020), the traditional approach of deriving a simple response of the instrument as a function of wavelength – by computing the ratio between the observed spectrum and the SED for a limited set of (possibly featureless) calibrators – will not work for a *Gaia*-like instrument because of the rather large width of the LSF compared to the wavelength scale of the response variations. As a consequence, the derived response changes with the spectral type of the calibrator: the LSF must be taken into account, creating the need for a much larger set of calibrators.

3. Calibrators

A reliable calibrator must satisfy several stringent requirements: it must be an isolated and point-like source with stable flux and high signal-to-noise ratio (S/N), and it should not be subject to strong interstellar extinction in order to avoid polarisation that could cause variations in the measured flux with the observing geometry and so on (see Pancino et al. 2012, 2021, and references therein).

The data set of spectrophotometric standard stars (SPSSs)[2], expressly built over the years for the calibration of *Gaia* photometric and spectroscopic data, is composed of 111 stellar sources, calibrated to the CALSPEC[3] scale (Bohlin et al. 2014, 2020) with flux accuracy of about 1% (Pancino et al. 2012, 2021; Altavilla et al. 2015, 2021). The SPSSs have been monitored for short-term constancy at the 0.005 mag level over a few hours (Marinoni et al. 2016). The current SPSS release is based on about 25% of the spectra collected for the project; a more complete release, which will contain about 200 SPSSs, will be used to calibrate future *Gaia* releases. The SPSS data set was recently complemented (Pancino et al. 2021) by a second set of 60 stars with looser requirements on the absolute accuracy (up to 5%) and flux stability (up to variations of about 0.05 mag) but including stellar types not contained in the SPSS set (bright O, B and late M stars), the passband validation library (PVL). These were originally intended to be used for validation purposes only, but a subset of these were eventually included in some phases of the actual calibration of the IM. The consequence of the severe criteria applied to the selection of primary calibrators is that the resulting stellar spectra are not independent from a mathematical point of view: their principal components span only a subspace of all possible spectral shapes and consequently not all the necessary instrument components would be constrained by
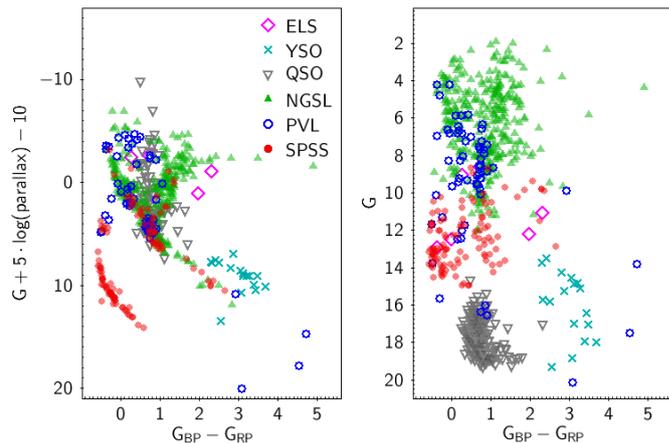
Fig. 1. Colour–magnitude diagrams for the whole set of calibrators used in external calibration processing and validations. On the vertical axes are absolute (*left*) and apparent (*right*) *G* magnitudes.

these, allowing for degeneracies in the solutions. Implementing the IM as a perturbation of the nominal model confers the advantage that unconstrained components have a reliable a priori estimation. However, to enforce more observational constraints to the IM, the set of primary calibrators has been extended by adding a set of secondary calibrators, including a wide variety of sources featuring strong emission lines over the entire wavelength range (mostly quasi-stellar objects (QSOs) and Wolf Rayet stars). Although these objects often show variability and require a special treatment in the processing (see Sect. 6), they are essential to provide strong constraints to the wavelength and LSF calibrations. For the *Gaia* DR3, a total of 211 peculiar sources have been selected from the literature, including 188 QSOs from the Sloan Digital Sky Survey (SDSS; see Lyke et al. 2020, and references therein), 17 young stellar objects (YSOs) from the X-shooter spectral library (Verro et al. 2022), and six emission line sources (ELSs) from STELIB (Le Borgne et al. 2003). Most of these objects have SEDs only partially covering the *Gaia* wavelength range: this limitation had consequences on the processing strategy, as described in Sect. 6. Finally, we also used the catalogue of the Next Generation Spectral Library (NGSL, Heap & Lindler 2016) for validation purposes, which consists of 348 bright sources with magnitude ranging from *G* = 1.97 and *G* = 12.0. Figure 1 shows the whole pool of calibrators and validation sources in a colour–magnitude diagram either in terms of absolute (left) or apparent (right) *G* magnitudes as function of $G_{BP}-G_{RP}$ colour.

4. Instrument model

The *Gaia* satellite observes the sky spinning around its axis (Gaia Collaboration 2016): the light collected by its telescopes is projected onto the focal plane where an array of charge-coupled device (CCD) detectors make measurements while operating in time-delay integration (TDI) mode. In the case of *Gaia* spectrophotometers, two slit-less prisms disperse the light onto two separate rows of seven CCDs each, both covering the focal plane in the *across scan* (AC) direction. The dispersion direction, referred as *along scan* (AL) being aligned with the transit direction of the projected light, is perpendicular to the AC direction. Each spectrophotometer covers part of the spectral wavelength interval, and the two ranges partially overlap: the blue photometer (*BP*) covers the nominal range [330, 680] nm

while the red photometer (*RP*) covers the range [640, 1050] nm. For each observed source, to limit the telemetry of the satellite, only a small window around the position of the source is actually read and transmitted to Earth during each transit. Windows are 60×12 pixels wide in each of the AL and AC directions. The maximum exposure time of an observation (~4.4 s) is fixed by the velocity by which the image transits on a CCD, but, to avoid saturation, it can be reduced according to the magnitude of the source by limiting the activation of the reading window to only a portion of the CCD with *gates* (De Angeli et al. 2023). Sources brighter than $G \simeq 11.5$ are transmitted as 2D windows (called window class 0, WC0) while for fainter sources, data are binned in the AC direction producing 1D spectra of 60 samples (WC1). Each source is observed several times during the lifetime of the mission under many different observing configurations and conditions (see Carrasco et al. 2021, for an exhaustive description). The internal calibration (De Angeli et al. 2023) has the complex task of calibrating all these configurations to reduce spectra to a common reference system, the *mean instrument*.

If we assume that the dispersion of the prism is perfectly aligned with the AL direction, then we can model the dispersed image of a point-like source in the data space as

$$\Im(u,w) = P_\tau \int_0^\infty n_\mathrm{p}(\lambda)\, P_\lambda\,(u - u_\mathrm{d}(\lambda), w)\, R(\lambda)\, \mathrm{d}\lambda, \qquad (3)$$

where u is the continuous coordinate in data space in the AL direction, w is the continuous coordinate in data space in the AC direction, P_τ is the telescope pupil area, $n_\mathrm{p}(\lambda)$ is the SPD of the source, $u_\mathrm{d}(\lambda)$ is the dispersion function, $P_\lambda(u,w)$ is the effective monochromatic point spread function (PSF) at wavelength λ, and $R(\lambda)$ is the overall instrument response function. This relation assumes indirectly that non-linear effects, such as those produced by charge transfer inefficiency (CTI) effects[4], are not important and can be neglected. As all internally calibrated spectra are binned to 1D windows, we can integrate the previous equation in the AC direction, obtaining the following model, which is suitable for describing a mean spectrum:

$$n_\mathrm{e}(u) = P_\tau \int_0^\infty n_\mathrm{p}(\lambda)\, L_\lambda\,(u - u_\mathrm{d}(\lambda))\, R(\lambda)\, \mathrm{d}\lambda, \qquad (4)$$

where n_e is given in units of $\mathrm{e}^- \mathrm{s}^{-1}$ and $L_\lambda(u)$ is the effective monochromatic LSF at wavelength λ obtained by integrating P_λ in the AC direction. This is the explicit form of Eq. (1). A detailed description of each factor of the model is given in the following subsections.

4.1. Dispersion model

Airbus Defence and Space (DS), the company in charge of developing and building the *Gaia* satellite, provided nominal dispersion functions based on chief-ray analysis for the *BP* and *RP* prisms in units of millimetres as a function of wavelength, by fitting a sixth-degree polynomial to the unperturbed *Gaia* optical design. For each field of view (FoV), dispersion functions are

---

[4] CTI, by delaying the release of the charge by a physical pixel as CCD charges scroll in the AL direction, would result in a deformation of the source spectrum that depends not only on the source SED but also on the *scene* of the observation, i.e. the temporal sequence of sources observed by that particular pixel immediately before the considered transit.
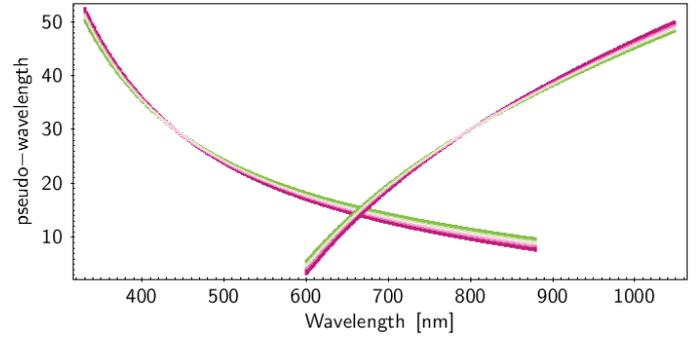
Fig. 2. Pre-launch nominal dispersion relations for *BP* and *RP* instruments. Different curves refer to all FoV/CCD row combinations.

provided for the centre of each CCD (the dispersion varies in the AC direction) in the form of the coefficients A_i of the expansion

$$\mathrm{AL}(\omega) - \mathrm{AL}(\omega_\mathrm{ref}) = \sum_i A_i \omega^i, \qquad (5)$$

where
- $\mathrm{AL}(\omega)$ denotes the AL image position in mm,
- $\omega = 1/\lambda$ in nm^{-1} denotes the wavenumber, and
- $\omega_\mathrm{ref} = 1/440\,\mathrm{nm}^{-1}$ for *BP* and $1/800\,\mathrm{nm}^{-1}$ for *RP*, corresponding roughly to the central wavelength of each instrument.

The AL position in pixel units u is obtained by dividing Eq. (5) by the pixel size of a CCD in the AL direction P_AL. Nominal dispersion curves for all FoV/CCD row combinations are shown in Fig. 2. The dispersion curve varies across the focal plane due to the tilt of the prisms with respect to the focal plane assembly: the comparison of the dispersion functions for different AC positions shows that the functions are related through a linear scaling to a high degree and are virtually independent from the FoV. We can therefore arbitrarily assume the coefficients for any of the CCD rows or FoV and model the generic dispersion function as:

$$u_\mathrm{d}(\lambda) = \sum_{k=0}^{N_u - 1} d_k \cdot \left[\frac{1}{P_\mathrm{AL}} \sum_{i=0}^{N} A_i \frac{1}{\lambda^i} \right]^k, \qquad (6)$$

where $u_\mathrm{d}(\lambda)$ denotes the AL image position in pixel units, coefficients A_i are arbitrarily assumed as those of CCD row 4, and FoV 1 and d_k are the IM parameters to be optimised in the calibration process. For *Gaia* DR3, we assume a number of parameters $N_u = 3$. Lower order parameters can be interpreted as follows: parameter d_0 represents the zero point of the dispersion relation and by construction is the reference AL position u_ref corresponding to the reference wavenumber ω_ref; and parameter d_1 is the scale of the dispersion relation. The nominal values are $d_0 = 30$, $d_1 = 1.0$, and $d_2 = 0.0$ for both *BP* and *RP* instruments.

4.2. The line spread function model

The LSF model is the only component of the IM that cannot be implemented as a simple perturbation of a nominal model, as in the cases of the dispersion and response models. Numerical monochromatic LSFs were provided by Airbus DS for testing purposes for each combination of FoV and CCD row; Fig. 3 shows two sets of these LSFs computed at wavelength $\lambda = 440\,\mathrm{nm}$ for the *BP* instrument (left) and $\lambda = 800\,\mathrm{nm}$ for the *RP* instrument (right). These models were built upon the optical PSF model of the telescope, which included optical aberrations based on laboratory measurements of the wavefront error
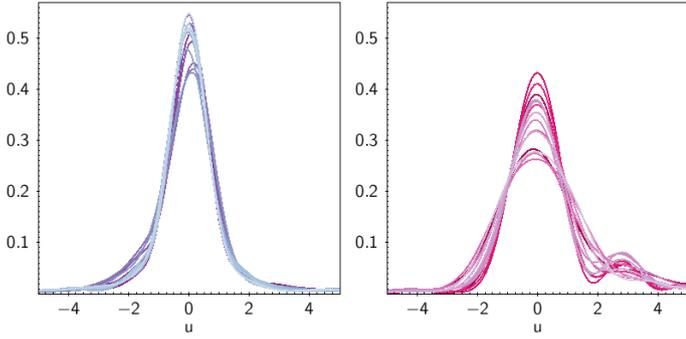
Fig. 3. Example of pre-launch nominal monochromatic LSF computed at wavelength $\lambda = 440$ nm (*left*) and $\lambda = 800$ nm (*right*). Different curves represent the LSF for each FoV/CCD row combination.
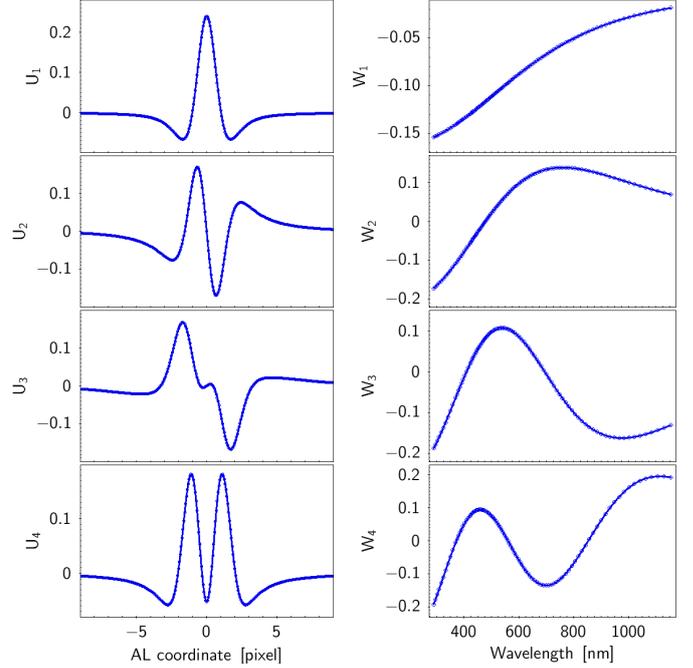

Fig. 4. LSF basis functions. *Left*: first four basis functions of matrix U as a function of the AL coordinate. *Right*: first four basis functions of matrix W as a function of wavelength.

(WFE) maps made on the telescope mirrors. The great variations in the shapes of the LSF shown in the figure are essentially due to variations in the WFE map from one FoV/CCD pair to the other. The problem is that these WFE maps are not applicable to the flying instrument because several factors (changes in physical conditions, mechanical stress of the launch, defocusing of the instrument, etc.) lead to them changing in an unpredictable way.

The strategy adopted for the current model implementation, which is explained in more detail in Appendix C, is to create a large sample of theoretical PSFs based on the optical design of *Gaia*, including randomly generated realistic WFE maps: these optical PSFs are then converted to effective PSFs that include a number of effects (charge diffusion, smearing introduced by TDI, pixel integration) to account for the discretised nature of the data. Effective PSFs are then marginalised in the AC direction to obtain a set of numerical LSFs sampled on a two-dimensional grid in spatial and wavelength coordinates. These numerical LSFs are then used to build a set of two-dimensional basis functions to allow the LSF to be modelled with a minimum number of free parameters for a given accuracy. The reduction to the 2D basis functions is achieved by means of generalised principal component analysis (GPCA, Ye et al. 2004), a fast and efficient algorithm for 2D image compression used to concentrate relevant information of a given data set in a small number of dimensions. Unlike the usual principal component analysis, GPCA is able to preserve the spatial locality of pixels in an image by projecting the images to a vector space that is the tensor product of two lower dimensional vector spaces. These two vector spaces are designed by matrices U and W whose columns represent the basis functions with which the dependencies are modelled along the spatial coordinate u and the wavelength coordinate λ, respectively. Figure 4 shows the first four bases for vector space U (left) and W (right). The U and W bases are interpolated to continuous variables (u, λ) by 1D interpolation. To ensure that the interpolation for the U bases satisfies the 'shift invariant sum' condition, which preserves the underlying function normalisation independently from the sub-pixel position of the sampling grid, these bases were then fitted with an S-spline model (Lindegren 2009). The interpolation for W bases is achieved by a cubic spline. The model for the LSF is finally given by

$$L(u, \lambda) = \overline{L}(u, \lambda) + \sum_{m=1}^{\ell_1} \sum_{n=1}^{\ell_2} d_{m,n} \cdot U_m(u) \cdot W_n(\lambda), \quad (7)$$

where $\overline{L}(u, \lambda)$ is the numerical mean LSF of the theoretical set and $d_{m,n}$ are the IM parameters that are fitted during the external calibration processing. As explained in Appendix C, the LSF wavelength modelling requires roughly half the dimensions needed for AL dependency modelling, hence $\ell_1 \simeq 2\ell_2$. Reiterating the caveat expressed at the beginning of this section regarding the absence of a proper *nominal* LSF model, when all parameters $d_{m,n}$ are set to zero the LSF coincides with the mean numerical model $\overline{L}(u, \lambda)$.

Equation (7) represents an undispersed LSF centred on the origin of the U bases. The dispersed LSF in the data space can be obtained by shifting the origin by an amount given by the dispersion relation as seen in Eq. (4): $L(u - u_{\rm d}(\lambda), \lambda)$. However, it is worth noting that the LSF origin does not necessarily coincide with its centroid: the centroid is an intrinsic property of the LSF, and in the case of a symmetric LSF is naturally given by the point of symmetry ξ, that is $L(\xi - u) = L(\xi + u) \; \forall u$, but in general the LSF is not symmetric and its shape can change with wavelength, introducing some degeneracy between the chromaticity and the dispersion. As sketched in Fig. 5, the dispersion model of Eq. (6) provides the location of the LSF origin: as far as the instrument model is concerned, the degeneracy between dispersion and chromaticity has little significance because the dispersion model and the LSF origin are consistently defined; however, physical interpretation of the data requires a dispersion relation that gives the centroid of the monochromatic LSF as a function of wavelength. This dispersion relation is provided as a lookup table where the centroid u_0 is computed at each wavelength λ by solving the non-linear equation:

$$\int_{-\infty}^{\infty} L(u_0 + u, \lambda) \, w(u/s) \, {\rm d}u = 0, \quad (8)$$
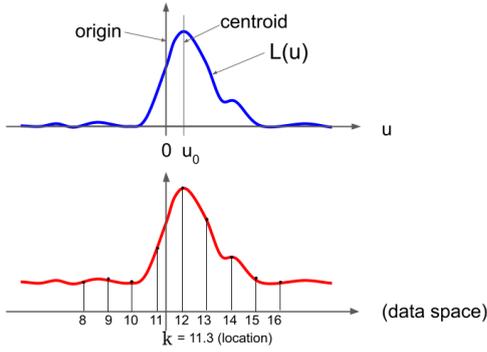
Fig. 5. Definition of centroid, origin, and location. *Top*: a schematic monochromatic LSF at a given wavelength λ_0 with origin $u = 0$ and centroid $u = u_0$. *Bottom*: location κ of the LSF in the data stream of sample values: a narrow emission line at wavelength λ_0 is overlying a continuous signal at $k = \kappa$.

Fig. 6. Pre-launch nominal responses for *BP* and *RP* instruments.

where the weighting function w is the Tukey's bi-weight:

$$w(z) = \begin{cases} z(1 - z^2)^2 & \text{if } |z| < 1 \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

and the scale parameter $s = 2.7$ is a value suitable for the *Gaia* case (Lindegren 2006). This dispersion function is provided for both *BP* and *RP* instruments as a single CSV file tabulated for wavelengths ranging from 320 nm to 1100 nm in steps of 0.5 nm[5].

4.3. Response model

The response $R(\lambda)$ defined in Sect. 2, as the ratio between the number of detected photons and the number of photons entering the telescope aperture per wavelength interval, is modelled as the product of the individual responses of each physical element (e.g. primary mirror, secondary mirrors) hit along the optical path. It changes across the focal plane, and depends on the observing configuration of each source and transit (gates, window class) and on time (contamination and decontamination issues). Assuming that all these dependencies have been accounted for by the internal calibration, we are left with a function of the wavelength alone.

The nominal pre-launch response curve for the mean *BP* and *RP* instruments can be described (Jordi et al. 2006) as the product of the following elements:

$$R_N(\lambda) = T_0(\lambda)\rho_{\text{att}}(\lambda)Q(\lambda)T_{\text{p}}(\lambda), \tag{10}$$

where
- $T_0(\lambda)$ is the telescope (mirrors) reflectivity;
- $\rho_{\text{att}}(\lambda)$ is the attenuation due to rugosity (small-scale variations in smoothness of the surface) and molecular contamination of the mirrors;
- $Q(\lambda)$ is the typical CCD quantum efficiency curve;
- $T_{\text{p}}(\lambda)$ is the prism (fused silica) transmittance curve including the filter coating.

These quantities were initially measured by Airbus DS during on-ground laboratory test campaigns and are plotted in Fig. 6. As can be seen, the steepest features of these curves are the *BP* and *RP* cut-offs produced by the prism transmittance curves, and
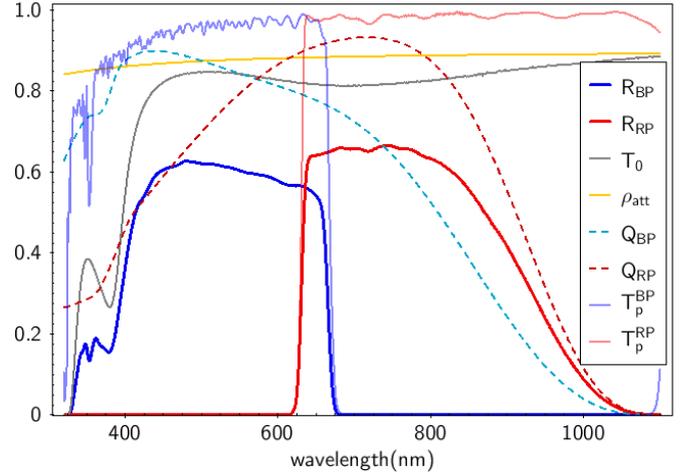
the steep BP drop around $\simeq 400$ nm which is mainly due to mirror reflectivity. Laboratory measurements showed that the precise location, that is, in wavelength, of the cut-off varies across the focal plane due to the uneven thickness of the prism coating (which is a few nm in both instruments). Moreover, combining measurements taken all over the focal plane results in a further smearing of the nominal curve. Therefore, a suitable modelling of the actual response cut-offs has been achieved by assuming the nominal curves, degrading the wavelength resolution by convolution with a rectangular window of width varying with the spectral dispersion per pixel, and re-shaping the cut-off mathematically with a two-parameter Gauss error function for *RP* and a two-parameter complementary error function for *BP* to control the wavelength position λ_C and the slope σ_C of these features (the tabulated transmittance curve is in practice truncated just before the cut-off and multiplied by the error function to mimic the cut-off shape). Nominal values for (λ_C, σ_C) are (667.9, 4.71) and (631.0, 4.0) for *BP* and *RP*, respectively. The corresponding curves are represented in Fig. 6 as blue (*BP*) and red (*RP*) thick lines. It is well known that the actual on-board overall response was heavily affected by rapid and discontinuous variations due to water vapour contamination of the satellite instrument components (Gaia Collaboration 2016) and to the various decontamination campaigns. The internal calibration initialises the internal reference system using only high-quality data collected in periods of low and slowly varying contamination (De Angeli et al. 2023), which ensures that the mean instrument response is not too different from the nominal $R_N^*(\lambda; \lambda_C, \sigma_C)$. Nevertheless, to be able to model any deviation from the nominal curve, the response is multiplied by a parametric function:

$$R(\lambda) = R_N^*(\lambda; \lambda_C, \sigma_C) \cdot R_{\text{d}}(\lambda; r_i). \tag{11}$$

Parameters r_i are referred to as response *shape* parameters to distinguish them from the two response *cut-off* parameters λ_C and σ_C; their nominal values are all zeros. The distortion model R_{d} is implemented as the exponential of a linear combination of a set of basis functions $\mathcal{R}$ in the AL sampling space u:

$$R_{\text{d}}(u) = e^{\sum r_i \mathcal{R}_i(u)}, \tag{12}$$

which is transformed to wavelength space λ through the dispersion relation:

$$R_{\text{d}}(\lambda) = R_{\text{d}}(u_{\text{d}}(\lambda)) \frac{du}{d\lambda}. \tag{13}$$

---

[5] Available at https://www.cosmos.esa.int/web/gaia/dr3-bprp-instrument-model

The exponential form guarantees the non-negativity of the overall model, while modelling in the sample space ensures the natural instrument spectral resolution, avoiding over-fitting where the spectral resolution is lower. The basis functions $\mathcal{R}_i$ used for the *Gaia* DR3 models are spline functions of second order with an initial uniform knot spacing in u that becomes non-uniform in later processing stages (see Sect. 6 for details). The full set of IM response parameters to be optimised is therefore $(r_i, \lambda_C, \sigma_C)$: when they are set to their nominal values, the response model closely resembles the nominal model.

5. Basis inversion and SED reconstruction

Once the instrument model is defined, Eq. (4) can be used to estimate the spectral photon flux distribution $n_p(\lambda)$ corresponding to an observed spectrum $n_e(u)$. Obtaining the SPD/SED allows the user to inspect *Gaia* spectra in a format that is more intuitive and of common usage. It may be convenient to define an *effective* spectral photon distribution as:

$$n_p^*(\lambda) = n_p(\lambda) \cdot R(\lambda), \tag{14}$$

so that

$$n_e(u) = P_\tau \int_0^\infty n_p^*(\lambda) \, L_\lambda \, (u - u_d(\lambda)) \, \mathrm{d}\lambda. \tag{15}$$

The effective spectrum is the observed spectrum deconvolved by the LSF function and transformed to the wavelength space through the dispersion relation (and scaled by some factor). Therefore, its shape will preserve the basic features of the observed spectrum. However, Eq. (15) is a Fredholm integral equation of the first kind, which is difficult to solve for the unknown $n_p^*(\lambda)$ because such integral equations are often ill-posed problems: large variations in the solution $n_p^*(\lambda)$ can occur for a slightly perturbed observable $n_e(u)$ (as is the case here, as n_e is affected by noise). However, as *BP/RP* mean spectra are modelled as a linear combination of basis functions (Carrasco et al. 2021):

$$n_e(u) = \sum_n b_n \varphi_n(u), \tag{16}$$

an interesting solution can be found by modelling the effective spectral photon distribution as a linear combination of the same spectral coefficients b_n with a particular set of bases:

$$n_p^*(\lambda) = \sum_{n=1}^N b_n \phi_n(\lambda), \tag{17}$$

where the ϕ_n bases satisfy the following condition:

$$\varphi_n(u) = P_\tau \int_0^\infty \phi_n(\lambda) \, L_\lambda \, (u - u_d(\lambda)) \, \mathrm{d}\lambda. \tag{18}$$

In practice, the externally calibrated spectral photon distribution (and related SED) corresponding to each pair of observed *BP* and *RP* spectra can be reconstructed by finding a set of proper functions whose images through the dispersed LSF model are the bases of the internal representation: the great advantage of this approach with respect to solving Eq. (15) directly is that it requires inverting the integral equation for a set of analytic functions that are by definition noise-free. We refer to these functions ϕ_n as the inverse bases hereafter.

To choose the most suitable representation for the inverse bases, it is useful to review the representation used for internally calibrated mean spectra described in Carrasco et al. (2021) and De Angeli et al. (2023) and summarised here for convenience. The basis functions implemented for *Gaia* DR3 are orthonormal Gauss-Hermite functions $\varphi_n(\theta)$ where a linear transformation is set between the pseudo-wavelengths axis u and the argument of the Hermite functions θ as:

$$\theta = \frac{u - \Delta\theta}{\Theta}. \tag{19}$$

The number of bases for *Gaia* DR3 has been set to $N = 55$ for both *BP* and *RP* spectra. An optimisation post-process is applied to mean spectra basis functions to concentrate most of the information in the lower order spectral coefficients: this optimisation takes the form of a rotation of the bases specified by a square matrix V_C. This rotation has no consequence for the basis inversion algorithm described here because, whenever Eq. (18) is satisfied, the inverse bases for the optimised bases are simply obtained by applying the same rotation matrix to the inverse basis set ϕ_n. The model chosen to represent each ϕ_n function is a linear combination of the same bases that model their image through the instrument model, that is, a linear combination of Hermite functions:

$$\phi_n(\theta) = \sum_{k=1}^K h_{k,n} \cdot \varphi_k(\theta), \tag{20}$$

with the same mapping between axes u and θ as adopted for the bases used to represent internally calibrated spectra:

$$\varphi_n\left(\frac{u - \Delta\theta}{\Theta}\right) = P_\tau \int_0^\infty \sum_{k=1}^K h_{k,n}\varphi_k\left(\frac{u_d(\lambda) - \Delta\theta}{\Theta}\right) L_\lambda\,(u - u_d(\lambda))\,\mathrm{d}\lambda. \tag{21}$$

The reason for choosing such a model is that, as in the case of the effective photon distribution, left and right bases should share the same basic features, as the function on the left hand side of the equation is a smeared version of that on the right hand side once mapped to the same axis u.

We could solve Eq. (21) for coefficients $h_{k,n}$ in a least-squares sense by sampling φ_n on a sufficiently dense and extended grid on the θ axis, but in this case the optimal number of bases K of the model would be undefined and it would not be clear whether or not a limit to this number were set by some hidden condition. A more appealing possibility is to project Eq. (21) into the coefficient space b_n of mean spectra, where the nth function φ_n is represented by definition by a vector of coefficients that are all null except the nth one equal to unity. In matrix notation, let **b** denote the array of coefficients, **s** a mean spectrum sampled on a given grid **u** of U points, and $D \in \mathbb{R}^{U \times N}$ the design matrix whose element D_{nu} is the value of the nth Hermite function evaluated at the uth pixel grid point. Consequently,

$$\mathbf{s} = D \cdot \mathbf{b}, \tag{22}$$

and

$$\mathbf{b} = D^\dagger \cdot \mathbf{s}, \tag{23}$$

where $D^\dagger \in \mathbb{R}^{N \times U}$ is the pseudo-inverse of the design matrix (see Appendix D for details). The integrals of Eq. (21) can be computed numerically by trapezoidal integration over a fine regular wavelength grid with Λ points and step $\delta\lambda$ extending over the

wavelength interval where the response is not null: let $\mathcal{L} \in \mathbb{R}^{U \times \Lambda}$ represent the instrument dispersed LSF model kernel sampled over that discrete grid, with the (i, j)th element being

$$\mathcal{L}_{i,j} = P_\tau \cdot L_{\lambda_j}\left(u_i - u_d(\lambda_j)\right) \delta\lambda. \tag{24}$$

We can write an equation like Eq. (21) for each of the left N Hermite basis functions: if the left term is interpreted as a column of the design matrix D, then all the N relations can be condensed into one single equation:

$$D = \mathcal{L} \cdot D_\varphi \cdot H, \tag{25}$$

where $D_\varphi \in \mathbb{R}^{\Lambda \times K}$ is the design matrix for the right φ_k bases sampled on the wavelength integration grid, while matrix $H \in \mathbb{R}^{K \times N}$ contains in its columns the set of coefficients $h_{k,n}$ that define the shape of each inverse basis. By multiplying Eq. (25) by $D^\dagger$ from the left we finally obtain:

$$I_N = D^\dagger \cdot \mathcal{L} \cdot D_\varphi \cdot H, \tag{26}$$

where the left hand term is the $(N \times N)$ identity matrix. By setting

$$B = D^\dagger \cdot \mathcal{L} \cdot D_\phi, \tag{27}$$

it is evident that the problem can be solved if B is a square matrix with $K = N$, that is, the number of inverse bases is equal to the number of bases for mean spectra representation. In this case, the matrix H that defines the basis functions for the externally calibrated spectra is simply given by:

$$H = B^{-1}. \tag{28}$$

The model for inverse bases consists of a matrix of coefficients H for each of the BP and RP instruments. Let λ be the wavelength grid over which we sample the externally calibrated spectrum corresponding to a pair of BP and RP observed spectra and Λ be the dimension of vector λ. We can build a design matrix $D_\varphi \in \mathbb{R}^{\Lambda \times N}$ by sampling the N Hermite functions on the grid:

$$\theta = \frac{u_d(\lambda) - \Delta\theta}{\Theta}. \tag{29}$$

If V_C is the orthogonal rotation matrix that defines the optimisation of the basis functions for the internally calibrated mean spectra, we obtain that the sampled effective spectral photon distribution is defined as

$$\mathbf{n}_p^* = \left(D_\varphi \cdot H \cdot V_C^T\right) \cdot \mathbf{b}. \tag{30}$$

If we build two more design matrices D^P_φ and D^E_φ whose elements are respectively defined as

$$D^P_{\varphi\,i,j} = D_{\varphi\,i,j} \frac{1}{R(\lambda_i)}, \tag{31}$$

and

$$D^E_{\varphi\,i,j} = D_{\varphi\,i,j} \frac{1}{R(\lambda_i)} \frac{10^8\,hc}{\lambda_i}, \tag{32}$$

where $R(\lambda)$ is the instrument response and hc the product of the Planck constant and the vacuum speed of light, then we get

$$\mathbf{n}_p = \left(D^P_\varphi \cdot H \cdot V_C^T\right) \cdot \mathbf{b}, \tag{33}$$

which is the SPD in units of photons s^{-1} m^{-2} nm^{-1}, and

$$\mathbf{f}_\lambda = \left(D^E_\varphi \cdot H \cdot V_C^T\right) \cdot \mathbf{b}, \tag{34}$$

which is the SED in units of W m^{-2} nm^{-1}.

As the design matrices of Eqs. (31) and (32) depend on the inverse of the instrument response function, the sampling wavelength grid must be limited to the range [330, 650] nm for BP and [635, 1050] nm for RP in order to avoid large errors in the reconstructed spectra.

The BP and RP instruments produce two partially overlapping SPDs and SEDs: these are combined into a single distribution by computing a weighted mean in the overlapping region $[\lambda_{lo}, \lambda_{hi}]$, where the weight varies linearly with wavelength:

$$w_{BP}(\lambda) = 1 - \frac{\lambda - \lambda_{lo}}{\lambda_{hi} - \lambda_{lo}}, \tag{35}$$

and

$$w_{RP}(\lambda) = 1 - w_{BP}(\lambda), \tag{36}$$

for $\lambda_{lo} < \lambda < \lambda_{hi}$.

BP and RP spectra coefficients are accompanied by the covariance matrix K_{bb}: this is used to calculate the covariance matrix for the sampled spectrum, which in the case of the SED is computed as

$$K_{ff} = \left(D^E_\varphi \cdot H \cdot V_C^T\right) \cdot K_{bb} \cdot \left(D^E_\varphi \cdot H \cdot V_C^T\right)^T. \tag{37}$$

The square roots of the diagonal elements of K_{ff} are the errors associated with the sampled SED.

6. Processing

The instrument model has been designed to reproduce the *nominal* instrument model when all parameters are set to their nominal values: these parameters are d_i for the dispersion (Eq. (6)), $d_{m,n}$ for the LSF (Eq. (7)), and r_i, λ_C, σ_C for the response model (Eqs. (11) and (12)). The concept of *nominal* refers to the dispersion function and to the overall response curve, for which pre-launch laboratory measurements are available. The LSF model instead is initialised as the mean of a large number of theoretical LSF models: for this reason it corresponds to a flat WFE map and is symmetric. Model parameter optimisation can be fulfilled with a sufficient number of calibrators by minimising in a least squares sense a χ^2-based cost function

$$\chi^2 = \mathbf{r}^T \cdot W \cdot \mathbf{r}, \tag{38}$$

where the array of residuals $\mathbf{r}$ and the weight matrix W can be evaluated in two different spaces:

1. Sample space: the observed BP/RP mean spectrum is sampled according to Eq. (22):

$$\mathbf{r} = D \cdot \mathbf{b} - \mathcal{I}_{u,\lambda} \cdot \mathbf{n}_p, \tag{39}$$

where $\mathcal{I}_{u,\lambda}$ represents the instrument matrix sampled on the wavelength array of the source SPD and on the same AL grid as the mean spectrum; $(\mathcal{I}_{u,\lambda}\mathbf{n}_p)$ represents the discretised version of Eq. (1), and the weight matrix is computed as

$$W = \left(D \cdot K_{bb} \cdot D^T + \mathcal{I}_{u,\lambda} \cdot K_{pp} \cdot \mathcal{I}_{u,\lambda}^T\right)^{-1}, \tag{40}$$

where K_{bb} is the covariance matrix of the spectra coefficients, the product $DK_{bb}D^T$ is the pixel covariance matrix of the sampled BP/RP spectrum, and K_{pp} represents the covariance matrix of the source SPD.
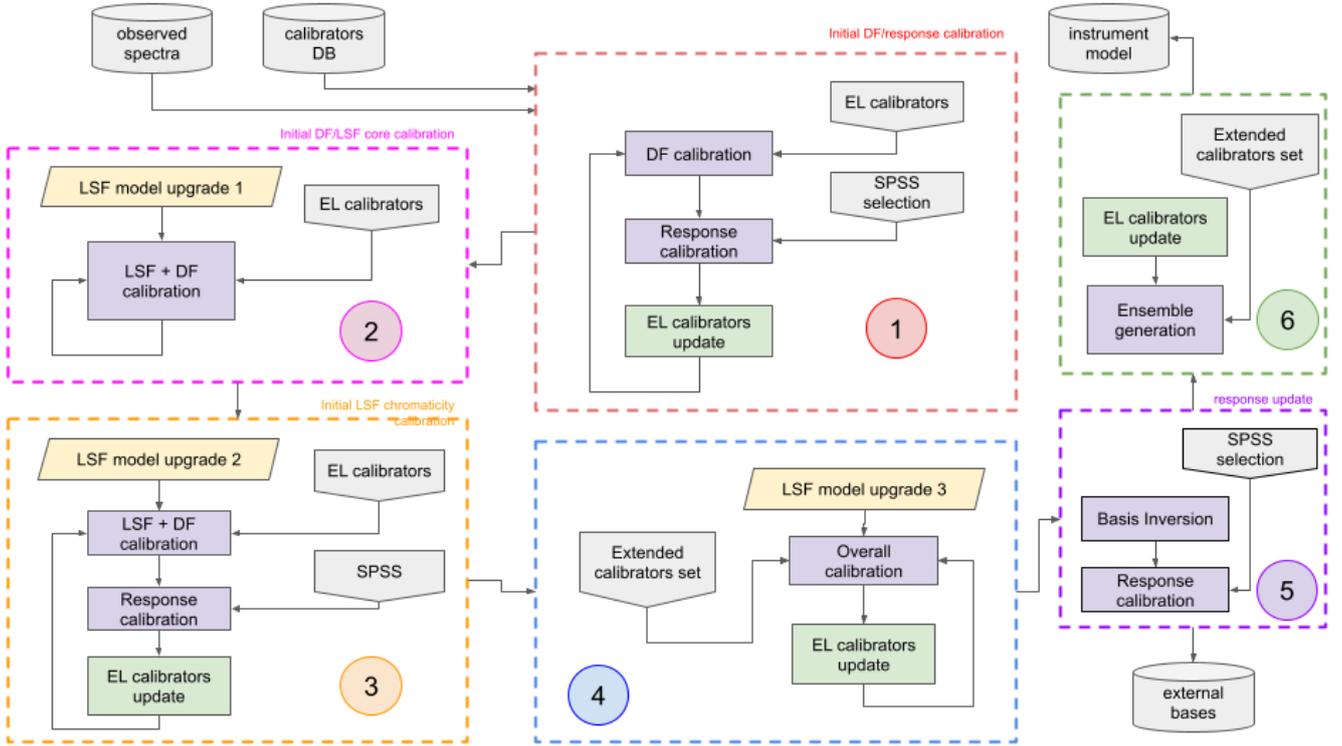
Fig. 7. Processing scheme.

2. Coefficient space: the model prediction is projected in the coefficient space according to Eq. (23)

$$\mathbf{r} = \mathbf{b} - D^{\dagger} \cdot \mathcal{I}_{u,\lambda} \cdot \mathbf{n}_{\mathrm{p}}, \tag{41}$$

and the weight matrix is the inverse of the sum between coefficients and projected SPD covariances:

$$W = \left(K_{\mathrm{bb}} + \left(D^{\dagger} \cdot \mathcal{I}_{u,\lambda} \right) \cdot K_{\mathrm{pp}} \cdot \left(D^{\dagger} \cdot \mathcal{I}_{u,\lambda} \right)^{T} \right)^{-1}. \tag{42}$$

In all scenarios, we include the usage of the covariance matrices for both the observed *BP/RP* spectra and the calibrator SPDs to allow for a total least-squares regression analysis of the data (Huffel & Vandewalle 1991). Although the SPD covariance matrix would be extremely important to properly account for systematic effects of wavelength calibration errors, especially in regions where the SPD changes steeply with the wavelength for certain spectral types, only errors on fluxes are usually available in the literature. Therefore, only *BP/RP* covariances are full matrices while SPD covariances are simply diagonal matrices populated with the corresponding variances on the sampled flux. Moreover, total regression is generally highly demanding in terms of computational cost because it requires an evaluation and inversion of the covariance matrix at each step of the solver, and has therefore only effectively been taken into account in the final stages of the calibration, as is clarified below.

When optimisation is carried out in sample space, the AL grid used to compare sampled spectra is usually oversampled by some factor with respect to the *Gaia* pixel, and therefore the coefficients covariance matrix transformed to sample space $(D \cdot K_{\mathrm{bb}} \cdot D^{T})$ (hereafter samples covariance) does not have full rank and cannot generally be inverted. A first obvious solution is to take into account only the diagonal elements of the samples covariance, even if the off-diagonal elements are not negligible because mean spectra are continuous functions and hence

random noise will manifest as random wiggles in the form of long-range correlations between pixels. A formal solution for a correct weighting scheme in sample space is to compute the weight matrix as:

$$W = D^{\dagger T} \cdot K_{\mathrm{bb}}^{-1} \cdot D^{\dagger}, \tag{43}$$

where the pseudo-inverse of the samples covariance matrix $D^{\dagger}$ has been defined in Eq. (D.4). This approach was tested but some occasional numerical instabilities discouraged us from using it for the present calibrations. Cost evaluation in coefficient space would solve the problem of the invertibility of the matrix, allowing for a full exploitation of spectra covariances. However, this approach could not be followed for the current release because of the incomplete wavelength coverage of many emission line calibrators used in the optimisation process (Eq. (23) implicitly requires the sampled spectrum to extend over the entire AL range). For calibrators with incomplete wavelength coverage, χ^2 computation was limited to the available section of the model spectrum, excluding a safety margin of a couple of pixels where truncation occurs because the redistribution of light produced by LSF will cause some systematic difference between the partial spectrum and the one that would have been obtained by a complete SPD. For this reason, we decided to carry out the cost evaluation in sample space.

Provided that the cost function is not linear with respect to the model parameters, the optimisation process can be carried out using an implementation of the differential evolution algorithm (DEA) as described by Storn & Price (1997): although this class of algorithms has been shown to achieve global optimisation with a natural ability to escape local minima traps in the χ^2 space, we find it convenient to proceed with the bootstrapping of the model by limiting the number of free parameters at the first stages of the processing, and gradually enhancing the model complexity (i.e. increasing the number of parameters) only when

convergence is progressively achieved. This led to a complex processing scheme that is sketched in Fig. 7.

Calibrators are divided into two groups, absolute flux or primary calibrators (labelled SPSS) and secondary calibrators, sources with emission line features (labelled EL calibrators). As explained in Sect. 3, the latter group contains potentially variable sources. To overcome this problem, an update process equivalent to a grey flux calibration is implemented for these calibrators: the input SPD of each EL source is scaled by a parameter that is evaluated at each calibration cycle to minimise the squared residuals between the current model prediction and the corresponding observed mean spectrum. In the first stage of the processing, the response model is initialised with a low number of shape parameters (see Eq. (11), 8 parameters for *BP*, 5 for *RP*), while cut-off parameters are set to their nominal values, the dispersion model degree is set to 1, and the LSF model is symmetric. This stage is designed to provide a reliable initialisation of the dispersion relation and the response shape: the cost evaluation is limited to the central region of the spectra, avoiding the wings and the cut-off regions; the optimisation of the dispersion parameters, based on the EL calibrators, is alternated with the optimisation of the response shape, which is achieved by using a selection of featureless SPSSs. Each optimisation cycle typically consists of approximately 3000−4000 DEA iterations involving about 50 walkers (different realisations of model parameters, initially distributed randomly around the starting set of parameters). The iterations stop when the individual costs from all the walkers converge to a common value. The set of parameters with the lowest cost is used to initialise the subsequent optimisation cycle. Once convergence is reached, the second stage begins, which entails modelling the shape of the central part of the LSF: the LSF model is initialised with $(\ell_1, \ell_2) = (4, 1)$ bases, and LSF and dispersion parameters are optimised together to model any possible asymmetry of the LSF core. A second upgrade to the LSF model is made in stage 3 where the number of bases is increased to $(\ell_1, \ell_2) = (4, 2)$ to allow modelling of any chromaticity effect: an LSF and dispersion parameter fit is alternated with response model adjustment and EL source update. In stage 4, the LSF and dispersion models are set to their final configurations, the number of response shape parameters are doubled while the cut-off parameters are left free to change; all model parameters are optimised together using an extended set of calibrators (EL + SPSS). The number of bases is set to $(\ell_1, \ell_2) = (7, 3)$ for *BP* and is left unchanged for *RP*; in both instruments the dispersion model degree is set to 2. In stage 5, the basis inversion process is performed to allow the reconstruction of the effective SPD for the SPSS: these are divided by the corresponding source SPDs to obtain the data shown in Fig. 8. This plot was obtained using a subset of 41 SPSSs selected to be as featureless as possible and, given the definition of effective spectra in Eq. (14), it traces the overall instrument response curves. The top plot shows data plotted as a function of wavelength. The bottom plot shows the residuals between data and model represented as function of the AL coordinate (or pseudo-wavelength); given that the dispersion direction of the *BP* instrument is inverted with respect to that of the *RP* instrument, *BP* data have been mirrored horizontally in order to sort the displayed data according to wavelength. *RP* data are offset by 60 samples in pseudo-wavelength to avoid superposition with *BP*. It is possible to recognise the signs left in the data by the first Balmer absorption lines ($H\alpha$ through $H\epsilon$) as peaks highlighted in the plots. Interestingly, a wavy regular pattern is visible at all wavelengths with a nearly constant frequency in the pseudo-wavelength space. The origin of this pattern is not fully understood; it could be related to wiggles in
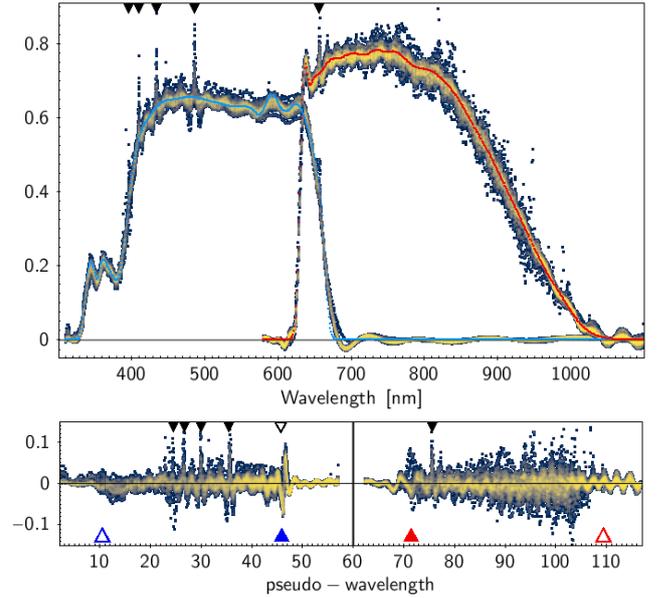
Fig. 8. *BP* and *RP* response curves traced by the ratio between effective SPD computed from *BP* and *RP* spectra and source SPDs from ground-based observations. *Top*: data are plotted against wavelength; black triangles mark the position of Balmer lines. *Bottom*: residuals between data and the response models are plotted against pseudo-wavelengths. *BP* pseudo-wavelengths have been swapped left-to-right, while *RP* data are shifted by 60 samples. Blue- and red-filled triangles mark the position of *BP* and *RP* cut-of, *f* respectively, while blue and red open triangles show where the response drops to zero.

the mean spectra (De Angeli et al. 2023). Using this data, we upgraded the instrument response distortion model by increasing the number of spline knots to model these wiggles, especially in the range [500, 800] nm, carefully excluding the signature left by the $H\alpha$ line (the response model must not be a function of any source astrophysical parameter). The total number of response parameters is 26 for *BP* and 23 for *RP*. The model curves represented in the top plot were computed after this upgrade. The last stage of the processing is dedicated to the creation of the *ensemble* of instrument models: the DEA solver is run on the last instrument models until parameter relaxation is achieved (typically after about 400 iterations), and then all the 50 walkers are saved into the database providing the 50 instrument models. The walker with the lowest chi-square is chosen to represent the instrument model, which enables forward modelling to simulate mean *BP* and *RP* spectra or inverse modelling to provide SEDs through the inverse basis representation. The ensemble instead is used to derive the uncertainties in the simulated spectra. The instrument model ensemble and the inverse bases are used in the `GaiaXPy` tool (De Angeli et al., in prep.) to simulate mean spectra and to generate sampled *Gaia BP* and *RP* calibrated spectra in the absolute system.

7. Results

The final *BP* and *RP* instrument models were used to create Fig. 9 where the corresponding instrument matrices $\mathcal{I}_{u,\lambda}$ are represented: this plot shows at a glance how *BP* and *RP* mean spectra are simulated for a given SPD. The instrument matrix can be used to express Eq. (1) in a discretised form as already done in Eq. (39): a mean spectrum is the row-by-column product
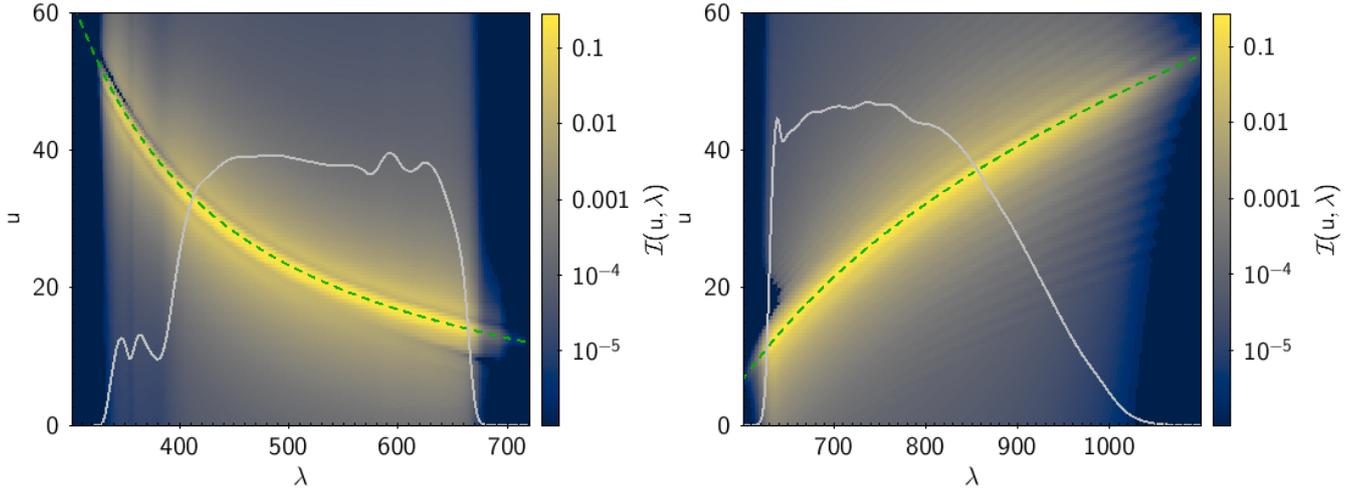
Fig. 9. Visual representation of instrument matrix for *BP* (*left panel*) and *RP* (*right panel*) instruments. Green dashed curves are the dispersion relations, and white curves sum up the matrix columns and represent the response curves.
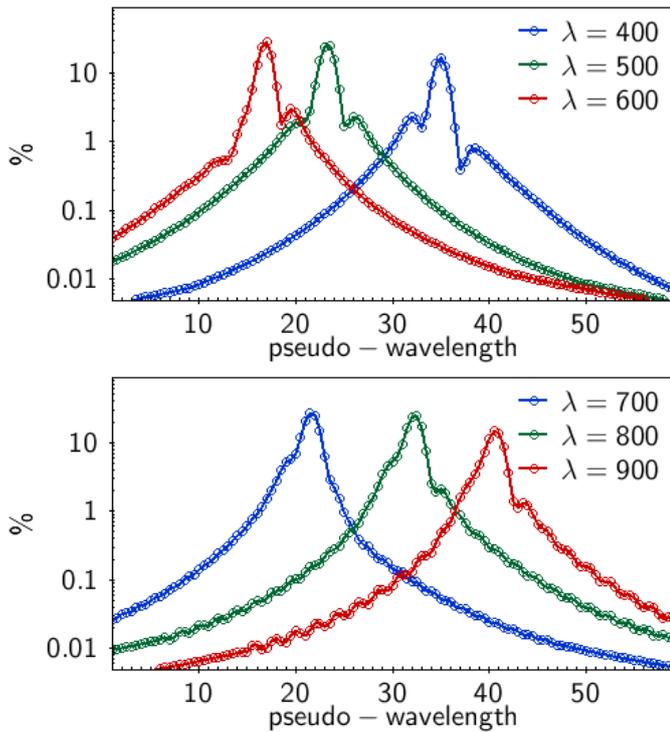

Fig. 10. Three selected columns of the instrument matrix for *BP* (*top panel*) and *RP* (*bottom panel*) in logarithmic scale. The curves represent the dispersed monochromatic LSFs at three different wavelengths rescaled by 100 to represent a percentage distribution.

matrix column profile, and the final mean spectrum is given by the accumulation on the *u* axis of each dispersed packet. One of the elements that catches the eye is that the dispersion direction is opposite for the two instruments, and hence the wavelength decreases from left to right in the *BP* spectra plots. Furthermore, in Fig. 9, the *RP* instrument exhibits a pattern of ripples that run roughly parallel to the dispersion relation. Intensity and distance from the LSF crest grow with wavelength: these ripples are the signature left by diffraction patterns of theoretical LSF in the *U* bases of the LSF model (Montegriffo 2017); the amplitude of the diffraction pattern scales linearly with λ and this is the reason why it is virtually invisible below 700 nm. Figure 10 shows the dispersed monochromatic LSF in logarithmic scale for wavelengths $\lambda = (400, 500, 600)$ nm for *BP* (top panel) and $\lambda = (700, 800, 900)$ nm for *RP* (bottom panel). The structure of the *RP* ripples is clearly visible in the bottom plot. On the other hand, we can figure out which wavelength ranges of the SPD contribute more to the photon budget of a particular sample of the mean spectrum by looking at the corresponding matrix row: this contribution for samples at $u = (20, 30, 40)$ is visualised in Fig. 11 for both instruments as the percentage of the incoming photons that are detected in each data sample. The width of the distribution sets the level of smearing of the spectrum at that wavelength. As can be seen, there are sections of mean spectra (as the case shown for *BP* at $u = 20$) that receive photons from a very large wavelength range: this explains why it is so important to also resolve for the LSF while calibrating mean spectra.

7.1. Inverse basis functions

Figure 12 shows the first few inverse bases ϕ_n^P for the *BP* and *RP* instruments (in units of photons s^{-1} m^{-2} nm^{-1}) as a function of wavelength. The top panels show the inverse bases of the canonical Hermite functions while the bottom panels represent the inverse bases for the optimised basis functions. The amplification of the bases below $\lambda \lesssim 400$ nm and above $\lambda \gtrsim 900$ nm is due to the normalisation of the bases by the response function $R(\lambda)$ (see Eq. (31)). For the same reason, it is pointless to look at the bases outside the represented wavelength range because they diverge as the response goes to zero.

It is interesting to compare the image through the instrument model of the inverse bases $\varphi_n^\dagger$ (hereafter referred to as the

of the instrument matrix with the SPD (sampled on the same wavelength grid of $\mathcal{I}_{u,\lambda}$). Provided that: (i) columns and rows represent small intervals of wavelengths and AL sample coordinates respectively, (ii) each column represents a monochromatic dispersed LSF scaled by the response (the response curve is over-plotted in white) at that wavelength and the telescope pupil area, and (iii) the loci of LSF maxima, highlighted by the dashed green lines, correspond to the dispersion functions, then we can visualise the mean spectrum formation by splitting the incoming SPD into packets of photons according to their wavelength. These are then distributed following the corresponding

reconstructed bases),

$$\varphi_n^\dagger(u) = P_\tau \int_0^\infty \phi_n^P(\lambda)\, L_\lambda\left(u - u_{\mathrm{d}}(\lambda)\right) R(\lambda)\, \mathrm{d}\lambda, \qquad (44)$$

against the original bases φ_n. This provides a means to evaluate the accuracy of the inverse representation: in principle the reconstructed bases should be equal to the original ones. This comparison is made in the top panels of Fig. 13 for the canonical Hermite functions and in Fig. 14 for the optimised versions. In both cases, the lines represent the original basis functions φ_n, while open squares represent the reconstructed bases $\varphi_n^\dagger$. The bottom panels of the figures show the difference

$$\Delta\varphi = \varphi_n(u) - \varphi_n^\dagger(u). \qquad (45)$$

In the case of non-optimised basis functions, the residuals for *BP* gradually increase at shorter wavelengths (from left to right) for higher order terms, becoming as high as 1%; in the *RP* case the residuals are much smaller, about 0.1%, over the entire wavelength range. In the case of optimised *BP* functions, the situation seems to worsen with residuals of ~2% over the entire pseudo-wavelength range, while for *RP* the quality of the comparison remains good. To understand this behaviour, it is convenient to quantify the quality of a reconstructed basis as a function of the order of the basis itself. This is done in Fig. 15 where, for each basis, we compute the quantity

$$\sigma_n = \frac{\int_{u_0}^{u_1} \left|\varphi_n^\dagger(u) - \varphi_n(u)\right|\,\mathrm{d}u}{\int_{u_0}^{u_1} \left|\varphi_n(u)\right|\,\mathrm{d}u}, \qquad (46)$$

which is used as a proxy for the percentage error on the pseudo-wavelength integral of the basis function. The reader must be aware that this is only an indirect way of assessing the accuracy of each inverse basis: we are evaluating the accuracy of external bases by forward modelling in the sampled spectra space where they are never used, and use them only in the inverse modelling of ECS. The plot in the top panel shows the reconstruction error for the non-optimised case, and one can see that the quality of the reconstruction is very high for both instruments (with numerical precision for *RP* outperforming the *BP* case by a factor of ten) for approximately the first 47 lower order bases, and worsens very quickly for the higher order terms. We understand this behaviour as an intrinsic limit of the current inverse basis model implementation (set in Eq. (20) with $K = 55$): an inverse basis represents by definition a given Hermite function deconvolved by the LSF function, hence its model should have the ability to reproduce features with higher spatial frequencies with respect to its smeared representation. If we consider the ϕ_{54} case, despite the fact that it should have higher spatial frequencies than the φ_{54} Hermite function, we are effectively modelling it as a linear combination of the $\varphi_0 \ldots \varphi_{54}$ Hermite functions and therefore we are possibly missing the higher frequencies. This problem will be addressed in future releases; for now we note that this inadequacy of the model will mainly affect the accuracy in reproducing features with high spatial frequencies, such as narrow emission lines (introducing some systematic errors into the shape of the ECS), but it will not have consequences for the vast majority of the sources. As the optimisation process consists in a rotation of the bases, this determines a redistribution of the error budget of the higher terms to lower order components (and this explains the behaviour of BP residuals in Fig. 14): however, thanks to the efficiency of the optimised bases in concentrating most of the information in the lower order coefficients, this will
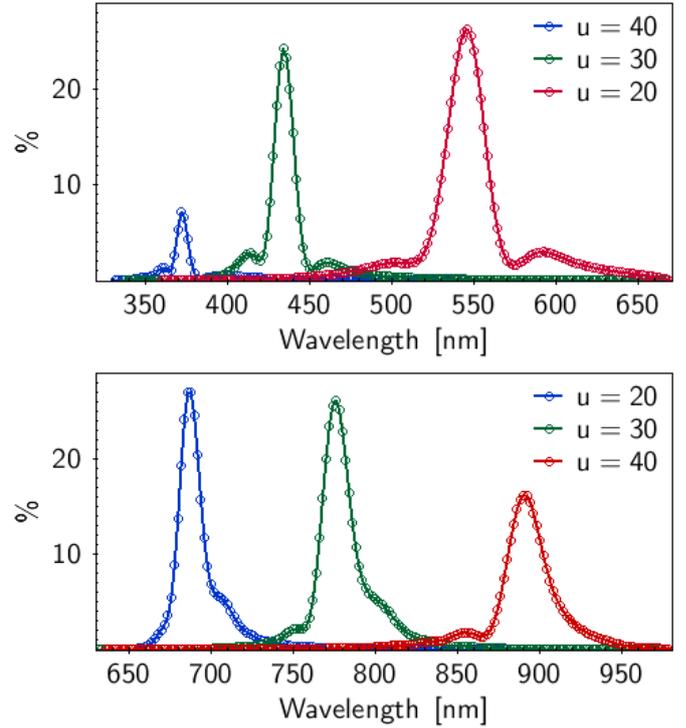
Fig. 11. Three selected rows of the instrument matrix for *BP* (*top panel*) and *RP* (*bottom panel*). The curves represent the percentage as function of wavelength of the incoming photons that are accumulated in the corresponding data sample.

have no effect on the overall quality of the ECS reconstruction. In Sect. 8.3, we perform a test to evaluate the impact of this reconstruction error on real data.

7.2. Spectral resolution

The spectral resolution of a *Gaia* spectrum changes considerably across the wavelength range. This is even more evident in ECS where the resolution changes abruptly near the boundary between the *BP* and *RP* wavelength ranges. The full width at half maximum (FWHM) of the instrument LSF model provides a direct measurement of the spectral resolution of mean *BP* and *RP* spectra. However, the resolution of ECS should be increased by the deconvolution from the LSF smearing effect obtained by the bases inversion process. To evaluate the spectral resolution of ECS, we simulate the response of the instrument model to a monochromatic signal at a well-defined wavelength as follows:

(1) we produce a synthetic spectrum with a flat null continuum and a single emission line of zero intrinsic width (the Dirac delta function) centred at a given wavelength;

(2) we forward model the corresponding mean spectrum into the coefficients space;

(3) we reconstruct the corresponding ECS;

(4) we measure the FWHM of the line: this is the measurement of the 'instrumental width' at the considered wavelength, that is the size of the spectral resolution element;

(5) and then we repeat with a synthetic spectrum whose emission line is displaced by 10 nm until the whole *BP* and *RP* spectral range is covered.

As we use the same instrument model in the forward and reverse directions, we are testing the theoretical resolution that would be obtained given 'perfect' knowledge of the instrument model. The presence of undetected systematic differences between the
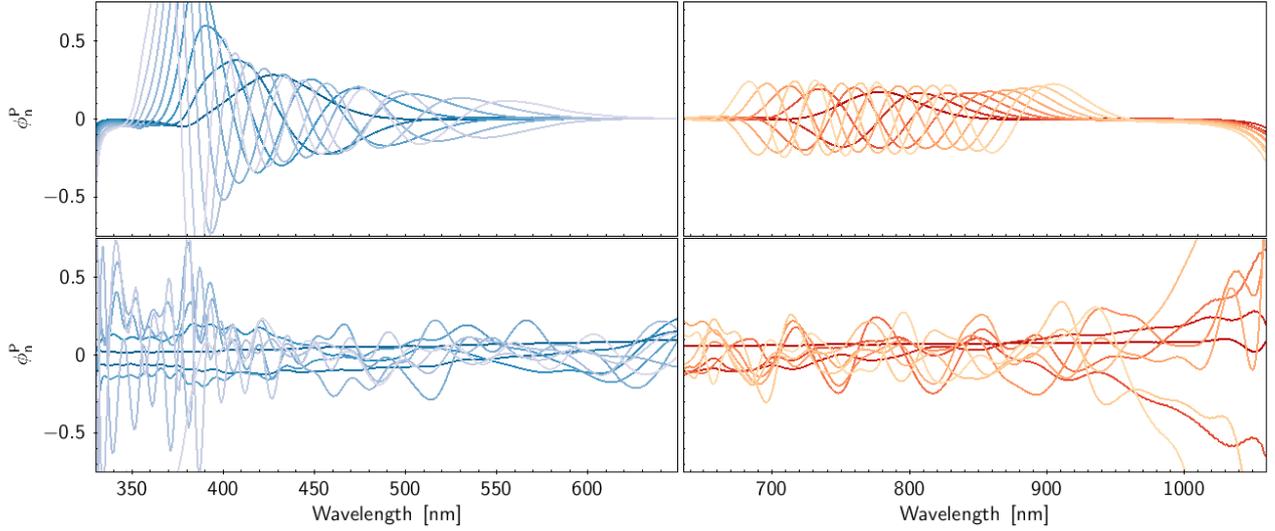
Fig. 12. First nine inverse basis functions for *BP* (*left*) and *RP* (*right*) instruments. *Top panels*: inverse bases corresponding to the canonical Hermite functions, *bottom panels*: inverse bases corresponding to the optimised bases.
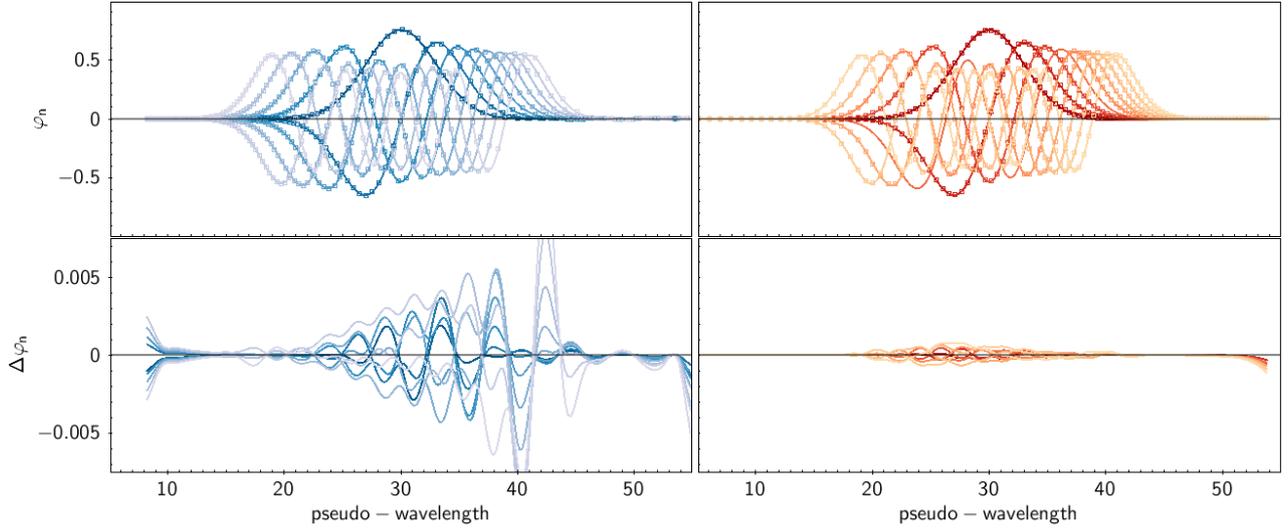
Fig. 13. Reconstruction error of the Hermite functions. *Top panels*: comparison between the first nine canonical Hermite functions (continuous lines) and the image of the inverse, non-optimised basis functions (open squares) for *BP* (*left*) and *RP* (*right*) instruments as a function of the pseudo-wavelength *u*. *Bottom panels*: residuals between the two families of curves shown in the *top panels*.

instrument model and the actual instrument will possibly worsen the resolution of the current ECS with respect to this ideal case. Furthermore, the total regression results in Gaussian noise on the elements of the instrument matrix. In the inversion of the noisy instrument matrix, this well-behaved Gaussian random noise can become non-trivial noise and even create systematic errors due to the non-linearity of inversion. Figure 16 shows the simulated ECS for a subset of the Dirac delta spectra. The reconstructed spectra show wiggles due to the inversion process. These will create artefacts in real ECS around small-scale features such as emission lines. As can be noticed, these wiggles tend to become very large at wavelengths greater than 1000 nm due to the amplification due to the response normalisation (see Eqs. (31) and (32)). We show that the presence of such large fluctuations is common in ECS at wavelengths $\lambda \gtrsim 1030$ nm.

The FWHM for internally calibrated *BP*/*RP* mean spectra (hereafter ICS) and for ECS are represented in Fig. 17 as a function of wavelength together with the CCD pixel scale that should

be regarded as an upper limit to the resolution achievable by the basis inversion process (it is not possible to reconstruct a signal smaller than the pixel integration scale with the current approach). The plot also reproduces the scale length corresponding to 2 pixels, which is the scale on which Nyquist sampling can take place: interestingly, the *BP* instrument suffers from a small under-sampling at all wavelengths (the FWHM being equal to ~1.7 pixel) and therefore does not satisfy the Nyquist criterion. This fact could play a relevant role in causing the rather high number of wiggles seen in *BP* mean spectra compared to the *RP* ones (see De Angeli et al. 2023). Values for the FWHM of a monochromatic signal measured on ICS, ECS, and the corresponding spectral resolution are provided in Table 1 as a function of wavelength.

The spectral resolution is finally shown in Fig. 18: the *RP* instrument exhibits a more uniform variation with respect to wavelength compared to the *BP* case. The anomalous drop in *RP* resolution just below 650 nm in the ECS curve as well
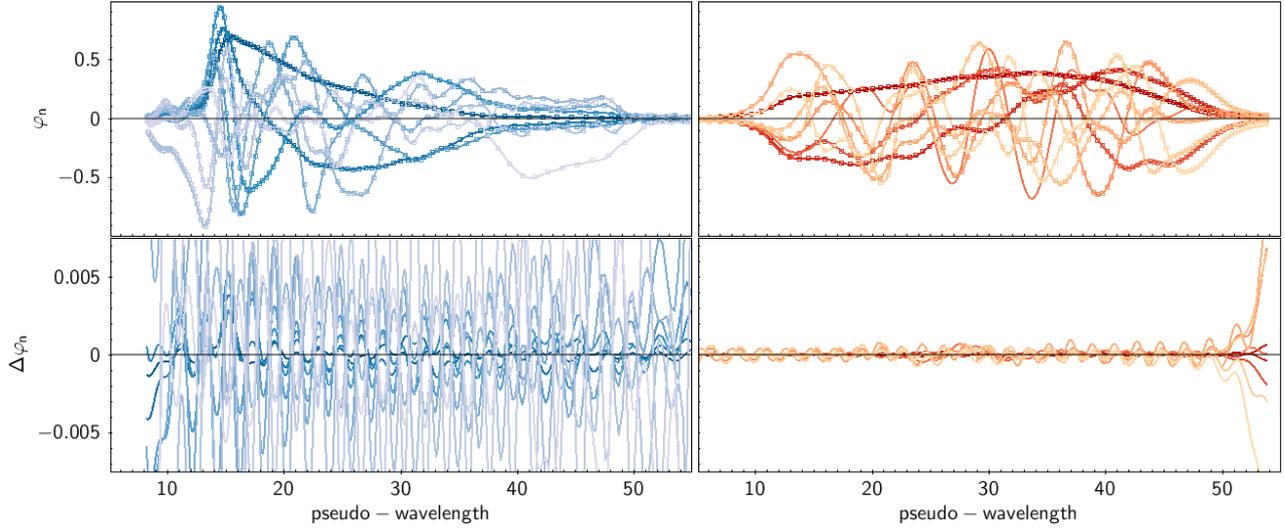
Fig. 14. Same as Fig. 13 but for the optimised basis functions.
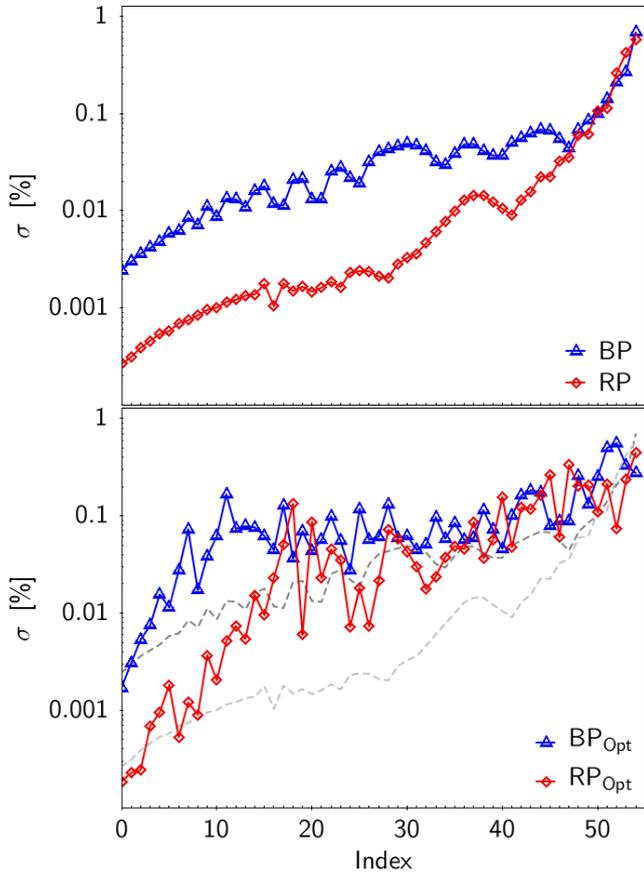
Fig. 15. Numerical reconstruction error of the inverse bases as a function of basis index for canonical Hermite functions (*top panel*) and for optimised basis functions (*bottom panel*); in the *bottom panel*, the grey dashed lines represent the curves of the *top panel* and are plotted to help in the comparison.

Fig. 16. Response of the instrument model to a simulated monochromatic stimulus at different wavelengths.
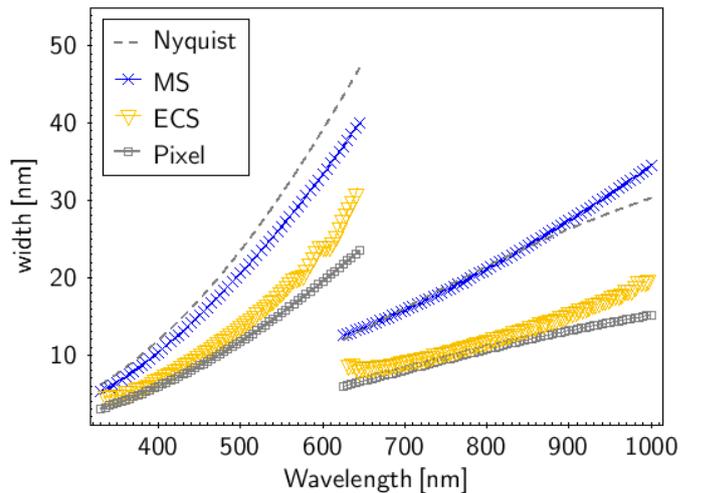
Fig. 17. Full width at half maximum measured in nm as a function of wavelength for a monochromatic signal in *BP* and *RP* mean spectra (blue crosses) and in the externally calibrated spectra (yellow triangles) compared to the width of one CCD pixel (grey squares) and the width of two pixels to represent the Nyquist sampling limit (dashed grey line).

as small fluctuations visible in the *BP* curve around 600 nm are possibly due to ripples in the response profiles affecting the basis inversion process. The final resolution of a typical *Gaia* ECS drops from $R_\lambda = 70$ to $R_\lambda \sim 22$ in the wavelength range $\lambda \in [330, 640]$ nm, and then suddenly rises to

Table 1. Full width at half maximum of a monochromatic signal measured on ICS, ECS, and spectral resolution for ECS.

λ	ICS	ECS	R_λ	λ	ICS	ECS	R_λ	λ	ICS	ECS	R_λ
350.0	6.483	4.893	71.5	570.0	29.25	20.11	28.3	780.0	19.99	11.33	68.8
370.0	7.900	5.508	67.2	590.0	31.98	23.12	25.5	800.0	21.14	11.95	66.9
390.0	9.490	6.355	61.4	610.0	34.84	24.21	25.2	820.0	22.33	12.60	65.1
410.0	11.24	7.417	55.3	630.0	37.81	28.63	22.0	840.0	23.55	13.24	63.4
430.0	13.12	8.611	49.9	640.0	13.13	8.576	74.6	860.0	24.80	13.85	62.1
450.0	15.13	9.994	45.0	660.0	13.98	8.478	77.8	880.0	26.09	14.50	60.7
470.0	17.25	11.36	41.4	680.0	14.87	8.741	77.8	900.0	27.40	15.47	58.2
490.0	19.46	12.88	38.0	700.0	15.81	9.116	76.8	920.0	28.76	16.07	57.2
510.0	21.75	14.56	35.0	720.0	16.79	9.662	74.5	940.0	30.15	17.13	54.9
530.0	24.15	16.31	32.5	740.0	17.81	10.11	73.2	960.0	31.57	17.86	53.8
550.0	26.63	18.31	30.0	760.0	18.88	10.69	71.1	980.0	33.01	18.83	52.0

$R_\lambda \sim 78$ before decreasing smoothly to $R_\lambda \sim 55$ at longer wavelengths.

7.3. Wavelength calibration accuracy

To assess the accuracy of the wavelength calibration, we compared the wavelength position of emission lines measured on the *BP* and *RP* mean spectra of a selected sample of QSO extracted from our calibrator set with the expected positions measured on simulated spectra computed with the instrument model and the corresponding source SEDs known from ground-based observations. We created a starting list of sources by selecting a sample of 102 SDSS QSOs showing strong emission lines in their *BP* and *RP* spectra and compiling a list of estimated wavelength positions for 263 lines based on the QSO SDSS redshift values and the wavelength at rest of each line. The differences between the AL position corresponding to the peak of each line in the mean spectra and in the simulated spectra are shown in the upper panel of Fig. 19. Open symbols refer to spectra where the SDSS SED does not cover the full *BP/RP* wavelength range, and therefore we do not have a reliable simulation of the expected ICS: in these cases, the expected AL position is based on the initial estimate of the line wavelength. The red curve represents a smoothed median of the data while the black lines delimit the region within ±10% of the LSF FWHM. As can be seen, at wavelengths below 400 nm there is an indication of a lower accuracy of the dispersion relation due to the lack of a reliable number of QSOs with emission lines in this range employed in the instrument calibration process. We hope to resolve this issue in future releases. The other interesting element in this plot is the presence of a small systematic offset in the *RP* spectra (roughly equal to one-tenth of a pixel), possibly due to systematic error in the LSF model, which is unable to correctly reproduce the instrument chromaticity as can be seen in Fig. 20 where three examples are shown (the LSF core is more asymmetric in *RP* than in *BP*). As the resolution of ECS is higher than that of ICS, we also checked the accuracy of wavelength calibration for the ECS case: starting from the estimated initial wavelength positions, we searched for local maxima in the ECS and in the corresponding simulated ECS (which is computed as described at beginning of Sect. 8). The wavelength positions of the peaks are then refined by computing the centroid of the lines accordingly to Eq. (8). The residuals between observed and expected wavelength positions are plotted in the lower panel of Fig. 19. Most of the points are enclosed between ±10% of the expected FWHM for ECS, represented by the dashed lines (the dark lines represent ±10% FWHM for ICS). A systematic difference of about 1 nm is
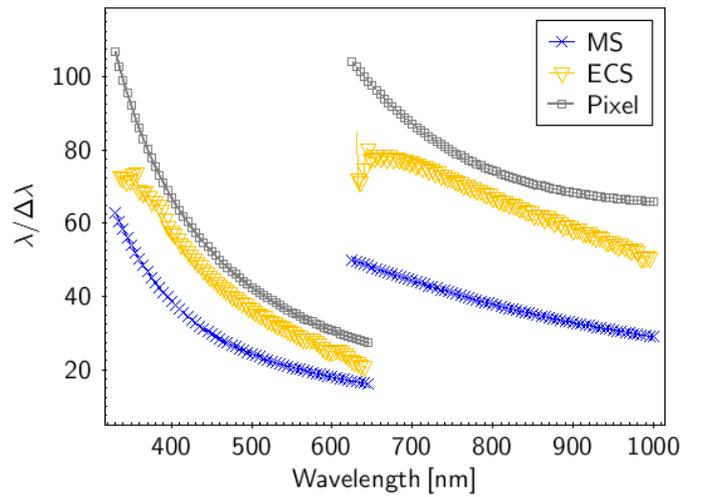
Fig. 18. Spectral resolution as a function of wavelength for *BP* and *RP* mean spectra (blue crosses) and externally calibrated spectra (yellow triangles) compared to the resolution corresponding to the width of one CCD pixel (grey squares).

confirmed in the *RP* wavelength range. Finally, we notice a lower precision in *BP* calibration that may be a consequence of LSF under-sampling seen in Sect. 7.2. We note that the number of measured lines in ECS is greater than that in ICS because several lines in ICS were lost because of the intrinsic shape of ICS (typically faint lines falling in regions with large variations of the instrument response).

8. Validation

Validation of the instrument models and externally calibrated spectra is mainly based on comparisons between observations (i.e. *Gaia* mean *BP/RP* spectra) and model predictions for sources with known SED. A test of this kind was also carried out by Andrae et al. (2023) where observed spectra of known solar twins were compared both to model spectra of the Sun and to IM predictions based on X-shooter SEDs (Verro et al. 2022) of the same sources, showing remarkably good agreement. Hereafter, we use the term *observation* to refer to *Gaia* spectra and *simulation* to refer to external data simulated via the instrument model. The comparison data always rely on external data sets. In the following sections, we use sources from the SPSS, PVL, and NGSL sets for the comparisons. Sources with magnitude $G < 4$ were filtered out from the NGSL sample to avoid saturation in *BP/RP*
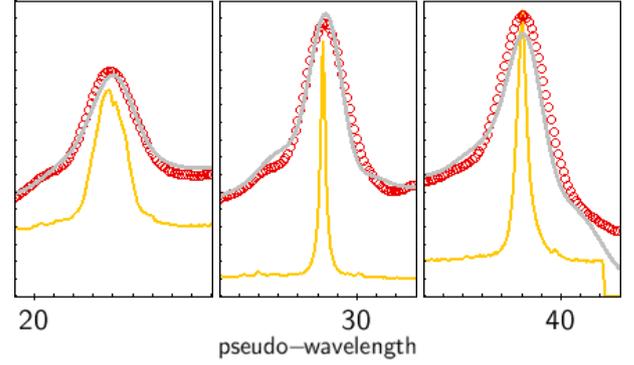
Fig. 20. Examples of emission lines in mean RP spectra. Red symbols are observed spectra, the grey line is the model prediction, and the yellow curve is the source high-resolution SPD from SDSS (in arbitrary units). The asymmetries in the line profile are not accurately reproduced by the model predictions.

where $u(\lambda)$ is the BP/RP dispersion function and u^{-1} the inverse dispersion function. The two degraded SEDs $f_{\mathrm{BP}}^{\dagger}$ and $f_{\mathrm{RP}}^{\dagger}$ are then combined into a unique SED $f^{\dagger}(\lambda)$ following the same rule described in Sect. 5 for the combination of BP/RP ECS. Finally, a further validation comes from computing synthetic photometry on ECS and comparing the resulting magnitudes to photometric standards, as described in Sect. 8.4.

8.1. Sampled mean spectra comparisons

Figure 21 shows the comparison between observed and simulated spectra for two *Gaia* sources chosen from the PVL set as an example: source Gaia DR3 1435896975388228224 (top) and source Gaia DR3 4339417394313320192 (bottom). *BP* spectra have wavelengths decreasing from left to right due to the dispersion direction of the prism. These two sources have magnitudes of $G = 9.68$ and $G = 13.81$ and colour indices of $G_{\mathrm{BP}}-G_{\mathrm{RP}} = 0.03$ and $G_{\mathrm{BP}}-G_{\mathrm{RP}} = 4.73$, respectively. Because of the low resolution, the appearance of *Gaia* mean spectra is generally quite smooth and only a few conspicuous spectral features are visible as in the shown examples. The agreement between model (grey lines) and observed (blue and red open circles) is very good in this well-behaved example. The complete set of plots for SPSS, PVL, and NGSL samples can be found in Montegriffo et al. (2022).

To make a more comprehensive comparison, for each source we compute the percentage difference between the observed spectrum and the model prediction for the SPSS, PVL, and NGSL sets: these curves are plotted in Fig. 22 as a function of wavelength, colour coded by $G_{\mathrm{BP}}-G_{\mathrm{RP}}$. The vertical grey line highlights the boundary between *BP* and *RP* data. The two yellow curves represent the P_{16} and P_{84} percentile distributions that are used as a proxy for the $\pm 1\sigma$ distributions: for wavelengths higher than $\lambda \simeq 400$ nm the accuracy of the calibration is mostly enclosed in the $\pm 2\%$ level marked by the two horizontal dashed blue lines, with some sources showing systematic offsets. Below 400 nm there is a clear systematic error correlated to source colour: this is the most prominent systematic error known in current *Gaia BP* and *RP* spectra and is analysed in detail in Sect. 8.1.1. To eliminate sources with systematic errors in the absolute flux scale (which can be present at the 1% level), we apply the same grey calibration described in Sect. 6.

Figure 23 shows residuals after such an additional step: precision is at the $\pm 1\%$ level except for $\lambda \lesssim 400$ nm. The three

Fig. 19. Wavelength calibration accuracy evaluated on a sample of QSOs with emission lines in their spectra. *Top panel*: difference in AL sample position of emission line peaks measured on *BP* and *RP* mean spectra and the expected position measured on simulated mean spectra (filled symbols). Open symbols refer to SEDs with incomplete wavelength coverage for which the expected position has been estimated based on the line wavelength at rest and the QSO redshift value. The vertical grey line separates *BP* from *RP* data. The black lines delimit the region within $\pm 10\%\,FWHM$ of the LSF for mean spectra. *Bottom panel*: observed and expected emission line wavelengths as measured on the ECS. Dashed grey lines represents $\pm 10\%$ of the expected LSF FWHM for calibrated spectra, and black lines delimit the region within $\pm 10\%\,FWHM$ of the LSF for mean spectra.

spectra (see Sect. 9.1). Unless we specify otherwise, plots combine data from all three data sets. The comparison can be made in terms of sampled internally calibrated mean BP/RP spectra or in terms of externally calibrated spectra. In the first case, the observations and predictions are sampled on a common AL grid. However, results are often shown as a function of the more familiar absolute wavelength, with 640 nm being used as the boundary between BP and RP. Here, ECS are represented as SEDs: in this case, the expected distribution of a source is not the SED at its original high resolution, but rather its spectral resolution is degraded at the resolution equivalent to one *Gaia* pixel with the following procedure: first we compute a degraded SED $f_X^{\dagger}(\lambda)$ for each BP and RP instrument:

$$f_X^{\dagger}(\lambda) = \int_{\lambda_{\mathrm{lo}}(\lambda)}^{\lambda_{\mathrm{hi}}(\lambda)} f(\lambda)\,\mathrm{d}\lambda, \tag{47}$$

where $f(\lambda)$ is the original SED and the integration limits are computed as:

$$\lambda_{\mathrm{lo}}(\lambda) = \lambda - u^{-1}\left(u(\lambda) - 0.5\right), \tag{48}$$

$$\lambda_{\mathrm{hi}}(\lambda) = \lambda + u^{-1}\left(u(\lambda) + 0.5\right), \tag{49}$$

Fig. 21. Comparison between mean and simulated spectra for source `Gaia DR3 1435896975388228224` (*top*) and source `Gaia DR3 4339417394313320192` (*bottom*), both belonging to the PVL dataset. Blue and red points represent the *Gaia BP* and *RP* mean spectra, while the grey curves represent the model predictions for the corresponding instrument. The superimposed yellow curves represent the source SPD arbitrarily rescaled in flux. We note that the wavelength in *BP* decreases from left to right.
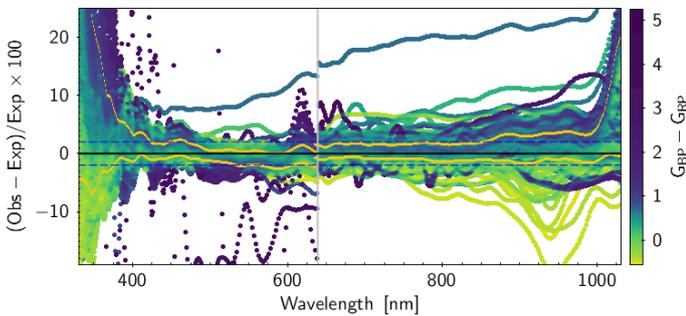
Fig. 22. Difference between observed and expected mean flux as a function of wavelength for *BP* and *RP* sampled mean spectra for the whole SPSS, PVL, and NGSL sets. Yellow curves represent the P_{16} and P_{84} percentiles. The colour map encodes the $G_{BP}-G_{RP}$ colour index.

Fig. 23. Difference between observed and expected mean flux as a function of wavelength for *BP* and *RP* sampled mean spectra. A grey calibration has been applied to spectra. The three curves represent the smoothed medians for SPSS (red), PVL (orange), and NGSL (yellow). The colourmap encodes the $G_{BP}-G_{RP}$ colour index.
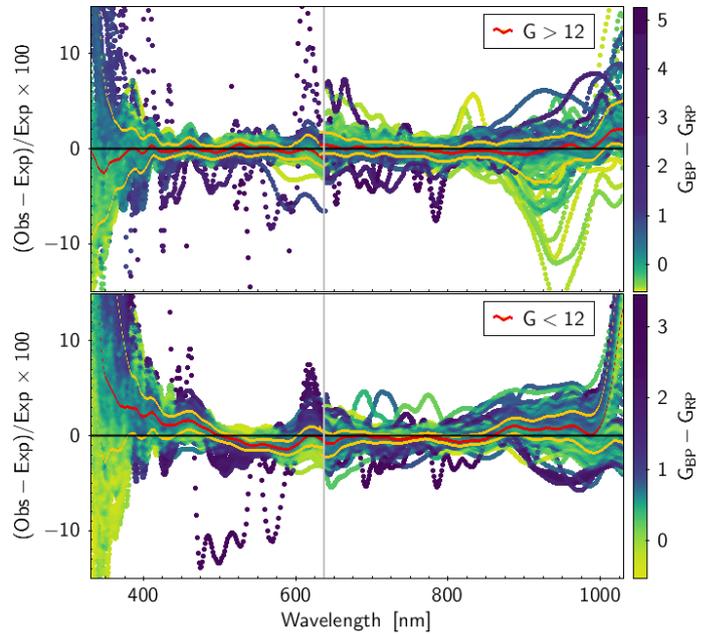
Fig. 24. Quantities are the same as Fig. 23 but data have been divided into two luminosity classes: sources with $G > 12$ (*top*) and $G < 12$ (*bottom*). Red and orange curves represent P_{50}, P_{16}, and P_{84} percentiles distributions, respectively.

curves in the plot represent the median distribution for SPSS (red), PVL (orange), and NGSL (yellow) sources. As can be seen, while the SPSS curve is almost completely flat, the median distributions for the other two control groups show significant deviations in the BP data as well as in the redder part of the RP spectra. As PVL and NGSL sources are generally brighter than SPSS, with the quality of PVL and NGSL being at least similar to SPSS, this suggests that the problem is unlikely due to a problem in PVL and NGSL data, but is rather related to magnitude-related systematic errors in *Gaia* mean spectra. This is confirmed by the two plots in Fig. 24 where sources with $G > 12$ are shown in the top panel while the bottom panel includes only sources with $G < 12$. Residuals from the first group are very flat and the

precision is well below the $\pm 1\%$ level, while for brighter sources the situation is more complex.

Figure 1 from De Angeli et al. (2023) clearly illustrates the extreme complexity faced by the internal calibration when dealing with *BP* and *RP* data: sources brighter than $G = 11.5$ are observed as 2D windows and, depending on their magnitude, under a high number of different observing configurations (so-called gates, which effectively limit the exposure time and the area of the CCD over which integration occurs, are activated to prevent chip saturation of bright sources). These results may suggest that the internal calibration of bright sources is not fully converged on the common internal reference system, which is mostly dominated by faint sources observed mainly as 1D windows and under more uniform conditions. Bearing in mind that one of the assumptions of the external calibration is that the instrument model does not depend on the source magnitude, by definition it cannot account for internal inhomogeneities in the data related to different observing conditions. It is expected that the external calibration model shows a better fit to the faint
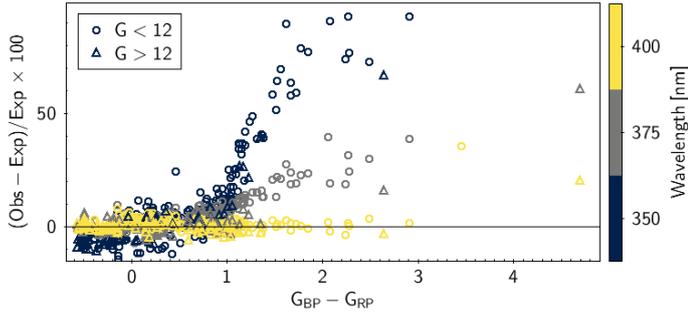
Fig. 25. Percentage residuals measured at wavelength $\lambda = \{350, 375, 400\}$ nm plotted as a function of $G_{BP}-G_{RP}$ colour index. Open circles represent sources with $G < 12$ while open triangles represent sources with $G > 12$.
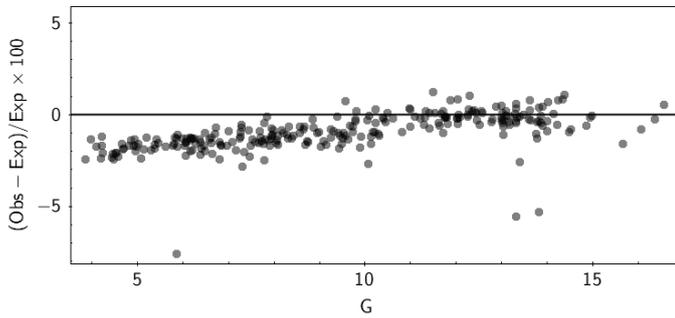
Fig. 26. Percentage residuals as a function of G magnitude measured on BP spectra at wavelength $\lambda = 580$ nm.

8.1.1. Systematic effects

The most prominent systematic effect visible in previous plots is a colour term affecting residuals at wavelengths $\lambda \lesssim 400$ nm. Figure 25 shows the normalised residuals as a function of $G_{BP}-G_{RP}$ colour index; residuals are calculated at wavelengths 350, 375, and 400 nm. As can be seen, the colour term, which is virtually absent at $\lambda = 400$ nm, rapidly grows when wavelengths decrease, with observed fluxes for redder sources being higher than expected and fluxes of blue sources being slightly lower than expected. There is some evidence that the colour term is magnitude dependent, because it is smaller for faint sources at the same wavelength. Although there may be issues in this wavelength range related to the wavelength calibration accuracy (see Sect. 7.3), as well as possible systematic effects in the LSF model, we believe that the presence of non-linear effects left by the internal calibration chain might not be excluded.

A second magnitude-dependent effect is visible in BP spectra for wavelengths between 560 and 600 nm and is represented in Fig. 26: the plot shows residuals measured at $\lambda = 580$ nm as a function of G magnitude of the source. While residual are flat for sources fainter than $G \simeq 11$, a linear trend is clearly visible with fluxes becoming progressively lower than expected, up to -2% at the bright end. This systematic effect is located immediately before the BP/RP interface and could be related to internal LSF calibration issues that induce small deformations in the shape of the mean spectrum (as already seen in the residuals in the bottom panel of Fig. 24). A further example is shown in Fig. 27
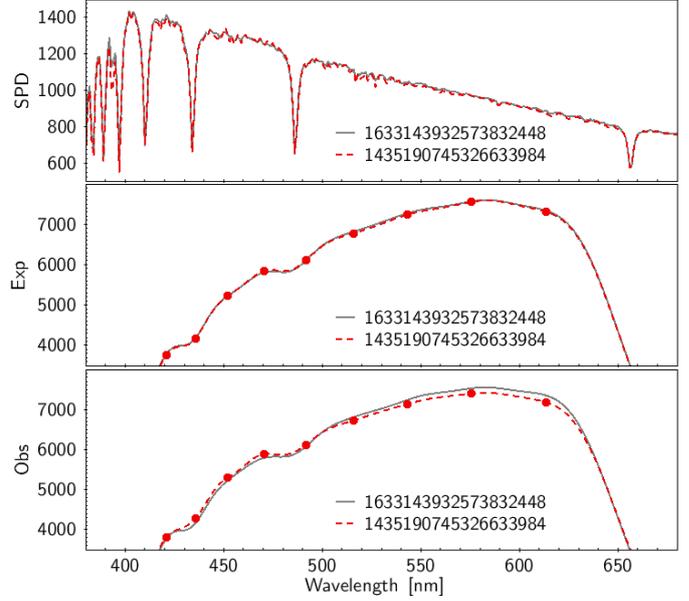
Fig. 27. Comparison between flux distributions of two sources with the same SPD shape but different magnitude (source ID number shown in the legend). Fluxes for the second source have been normalised to match the same level of the first. The comparison is shown between high-resolution SPDs (*top panel*), expected BP mean spectra (*middle panel*), and observed mean BP spectra (*bottom panel*).
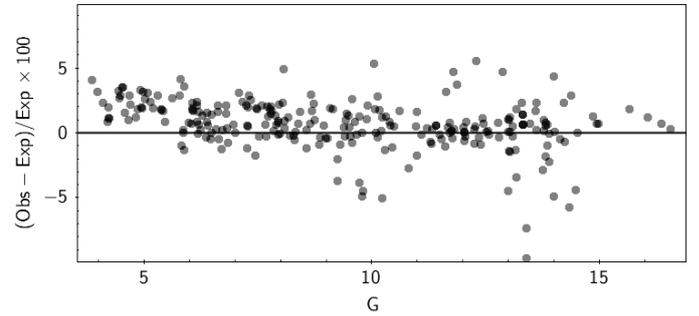
Fig. 28. Percentage residuals as a function of G magnitude measured at wavelength $\lambda = 950$ nm.

with the comparison between spectra of two PVL sources whose SPD shape is extremely similar but with different luminosity levels: source `Gaia DR3 1633143932573832448` at magnitude $G_A = 12.46$ and source `Gaia DR3 1435190745326633984` with magnitude $G_B = 6.82$. In all the panels, the flux of the second source has been rescaled by a factor of $10^{-0.4(G_A-G_B)}$ to put both fluxes on the same scale. As can be seen, while the expected BP mean spectra are almost indistinguishable, in the bottom panel the observed spectrum for the bright source is below the expected level.

A similar magnitude-dependent systematic effect is also visible in the red part of RP spectra, even if data are more noisy in this second case: the plot in Fig. 28 shows residuals measured at $\lambda = 950$ nm as a function of G magnitude. In this case, residuals are almost flat for $G \gtrsim 9$, but at the brighter magnitude end, a linear trend can be seen as a rise with fluxes becoming higher than expected by $\sim 3\%$ at $G \simeq 4$.

end because of the magnitude distribution of the SPSSs, most of which fall in the well-behaved faint magnitude range.
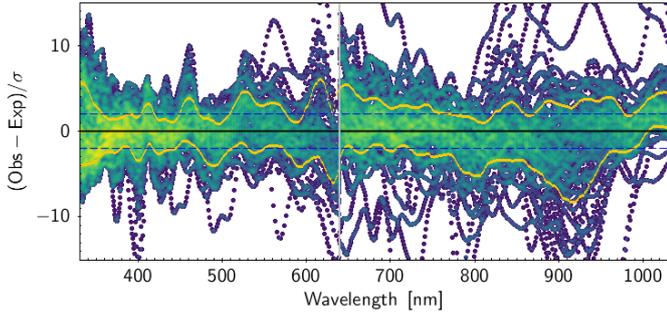
Fig. 29. Difference between observed and expected mean flux normalised by the squared sum of errors as a function of wavelength for *BP* and *RP* sampled mean spectra. Yellow curves are the P_{16} and P_{84} quantile distributions while horizontal dashed blue lines set the ±2 level.
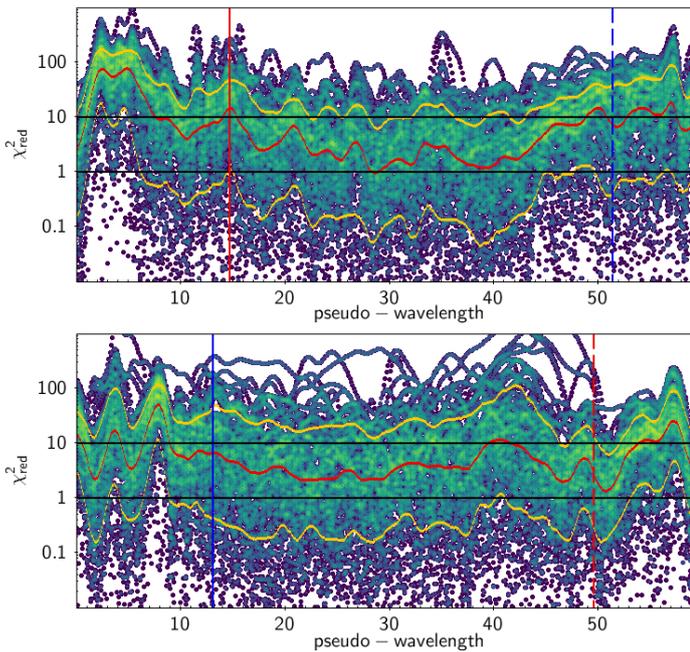
Fig. 30. Contribution per sample to χ^2 values for *BP* (*top*) and *RP* (*bottom*) computed for sources with $G > 12$.

8.1.2. Errors

De Angeli et al. (2023) observed that mean spectra errors are generally underestimated for spectral coefficients with low index and are slightly overestimated for higher order coefficients. For most of the sources, this translates to a general underestimation of errors because most of the information is contained in the lower order coefficients. Moreover, these errors are only reliable for fainter sources. To assess the validity of error on fluxes, we normalise the residuals between observed and expected mean fluxes by the error on fluxes obtained by summing in quadrature the error from the *BP* and *RP* mean spectra, with the error on the model prediction coming from the SPD error budget and a calibration error obtained by the instrument models ensemble (see Sect. 6). These normalised residuals are shown in Fig. 29 for sources with $G > 12$. The yellow curves represent P_{16} and P_{84} percentiles used as a proxy for the ±1σ level. As can be seen, $\sigma \gtrsim 2$ for most of the wavelength range (dashed blue horizontal lines are set at ±2 to guide the eye): this means that differences are larger than the estimated uncertainties. It is worth pointing out that the considered error does not contain the con-

tribution of covariances between samples: however, these covariances are not negligible because the mean spectra are represented as a continuous model, and therefore random noise takes the form of random wiggles that give rise to long-range correlations across several samples. The solution is to evaluate a χ^2 by projecting sampled spectra into coefficient space as described by Eq. (23) and to use the full coefficient covariance matrix defined by Eq. (42). However, before doing so, it is useful to look at the full distribution of normalised residuals as a function of the AL sampling coordinate. In Fig. 30, we plot these residuals in the form

$$\frac{(\text{Obs} - \text{Exp})^2}{\sigma^2}, \tag{50}$$

interpreting this quantity as the contribution of each sample to the χ^2_{red} computation. The top panel shows residuals for the BP case: the vertical red line at $u \simeq 14.7$ corresponds to the wavelength $\lambda = 640$ nm and roughly identifies the position of the filter cut-off while the vertical blue dashed line at $u \simeq 51.4$ corresponds to wavelength $\lambda = 330$ nm and roughly corresponds to the filter cut-off: outside this range the mean spectrum contains only the contribution coming from other parts of the spectrum due to the effect of LSF smearing. Red and yellow curves represent the P_{50} and the P_{16} and P_{84} percentile distributions. A relevant piece of information coming from this plot is that while the χ^2 contribution coming from the spectrum is reliable, with the median confined within the 1−10 range, we observe unrealistic values in the spectra wings that would make the global χ^2 useless. The situation is similar for the *RP* case shown in the bottom plot (the dashed blue line identifies the filter cut-off at $\lambda = 640$ nm, the red line the cut-off at $\lambda = 1030$). The discrepancies in the wings of spectra are driven by small systematic differences between observations and model predictions that are mostly due to the LSF model: the transmissivity of the instrument in that sample range is almost null and all the detected photons are due indeed to the smearing action of the LSF (they originate from different wavelengths); a small systematic effect in the wings of the LSF model can justify the observed discrepancies. To evaluate the χ^2 in the coefficient space, we then safely excluded a few pixels in the wings of sampled spectra from the computation, projecting only the sample range $u = [9, 53]$, and obtaining the distributions visualised in Fig. 31 for both instruments. We plot the distributions for sources with $G > 12$ as blue and red histograms for *BP* (left) and *RP* (right), respectively, while brighter sources are represented as green distributions in both plots. The peak for the faint sources is located roughly at $\chi^2 \simeq 6$ for *BP* and $\chi^2 \simeq 3$ for *RP*: these rather large values may be in part due to an underestimation of spectra covariances, but there is evidence that the calibration error budget coming from the instrument models ensemble may also be underestimated because of the suboptimal procedure followed for its construction (differential evolution algorithm instead of a more appropriate Markov chain Monte Carlo method). Values for bright sources are meaningless because of the combined effect of systematic differences in the shapes of spectra seen in Sect. 8.1 and unreliable coefficient covariances.

8.2. Comparing mean spectra in wavelength space

Figure 32 shows the comparison between *Gaia* ECS and the corresponding reference SED from external data for the same sources as in Fig. 21. The ECS is compared to the degraded version of the SED computed as described at the beginning of Sect. 8. Although the agreement between the distribution is
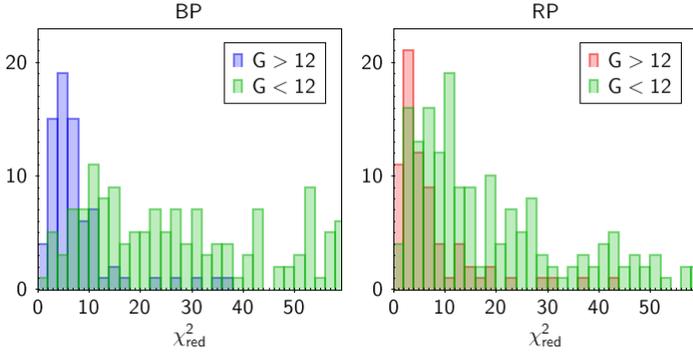
Fig. 31. Reduced χ^2 distributions for *BP* (*left*) and *RP* (*right*) computed on the set of SPSS, PVL, and NGSL spectra.

Fig. 32. Comparison between ECS and model SED for source `Gaia DR3 1435896975388228224` (*top*) and source `Gaia DR3 4339417394313320192` (*bottom*). The black curve represents the *Gaia* ECS, the yellow curve represents the corresponding high-resolution SED from external data, and aquamarine open circles are the SED degraded at a spectral resolution corresponding to one *Gaia* pixel.

Fig. 33. Comparison between ECS and model SED for source `Gaia DR3 1435190745326633984` (*top*) and source `Gaia DR3 1633143932573832448` (*bottom*), the same sources represented in Fig. 27. The colour coding and symbols are the same as in Fig. 32.

both belonging to the SPSS catalogue. Finally, an example of ECS comparison for two sources with emission lines is shown in Fig. 35: source `Gaia DR3 4007020769942990080` (top) and source `Gaia DR3 4467076569812517248` (bottom) with magnitudes of $G = 16.45$ and $G = 17.31$, respectively, both contained in the SDSS catalogue, exhibit an extremely different wiggle intensity in the blue part of their spectra, with the fainter one showing a better-behaved ECS (therefore wiggle intensity is not directly related to S/N).

A more global analysis is presented in Fig. 36, which shows the percentage residuals computed on the whole set of SPSS, PVL, and NGSL samples divided into two luminosity classes, sources with $G > 12$ in the top panel and sources with $G < 12$ in the bottom panel. If we compare these residuals with the equivalent quantities measured on ICS (Fig. 24), the effect of wiggles is evident. Nevertheless, the red lines representing the median of the distributions, once smoothed by a Gaussian with $\sigma = 10$ nm, show a substantially equivalent behaviour to that seen in the ICS residuals. Finally, Fig. 37 shows the normalised residuals for the faint sample. The error is computed here as the sum in quadrature of the contribution of the ECS error and the error on the model SED computed by propagating the errors on the full-resolution SED following Eq. (47) under the hypothesis that errors are independent. Percentiles P_{16} and P_{84} (red curves), smoothed by a Gaussian with $\sigma = 10$ nm, are used as a proxy to the actual standard deviation of the residuals which proves to be wider than expected. This result may in part be due to the lack of covariance terms for the high-resolution SED samples, which results in underestimation of the model error; however, the result is consistent with what is seen in Fig. 29 for the ICS comparison. The systematic colour-dependent difference seen in ICS at wavelengths $\lambda < 400$ nm is still present and is illustrated in Fig. 38: a sample of SPSS ECS with colour ranging from $G_{BP}-G_{RP} = 0.06$ (top left panel) to $G_{BP}-G_{RP} = 2.65$ (bottom right panel) is compared against the corresponding model SED. While the ECS is

excellent, some wiggles are present, especially in the top plot in the blue part of the spectrum and around the Balmer absorption lines. The wiggles originate in the basis inversion process and their effect can vary a lot in ECS as can be seen by looking at the other examples in this section.

Figure 33 shows the ECS SED comparison for the same sources represented in Fig. 27. These sources have reference SEDs with almost the same shape, but different magnitudes ($G = 6.82$ in the top panel vs. $G = 12.46$ in the bottom panel). The intensity of wiggles is much higher in the bright source with respect to the fainter one, especially for $\lambda \lesssim 500$ nm, indicating that neither the shape of the spectrum nor the noise level necessarily determines the amount of wiggling. Another example of two sources with similar spectral type and comparable apparent magnitude but different wiggle intensity is given in Fig. 34: sources `Gaia DR3 1399559249961569792` ($G = 13.32$) and `Gaia DR3 2323394345824851584` ($G = 13.62$),

Fig. 34. Comparison between ECS and model SED for source Gaia DR3 1399559249961569792 (*top*) and source Gaia DR3 2323394345824851584 (*bottom*), two SPSSs of similar spectral type and magnitude. The colour coding and symbols are the same as in Fig. 32.
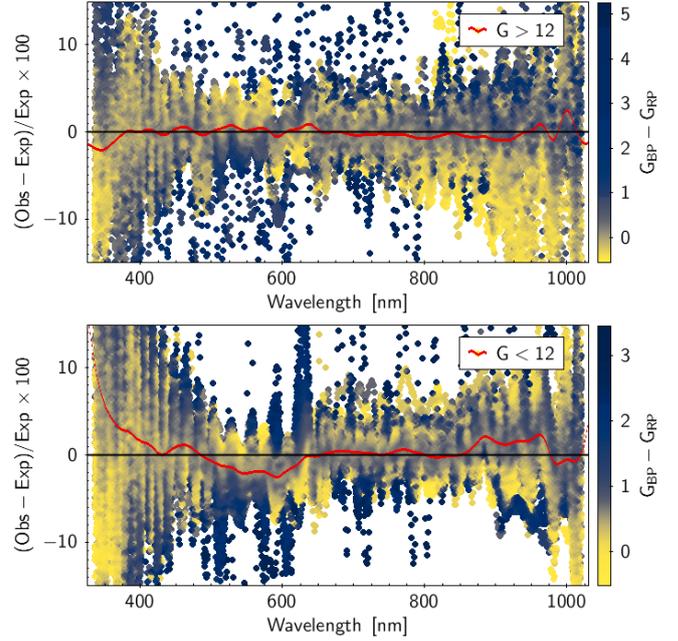
Fig. 36. Percentage residuals between ECS and reference fluxes for sources with $G > 12$ (*top*) and $G < 12$ (*bottom*), for the full set of SPSS, PVL, and NGSL samples. The colour map encodes the $G_{BP} - G_{RP}$ colour index. The red curve represents the median distribution smoothed by a Gaussian with $\sigma = 10\,\mathrm{nm}$.
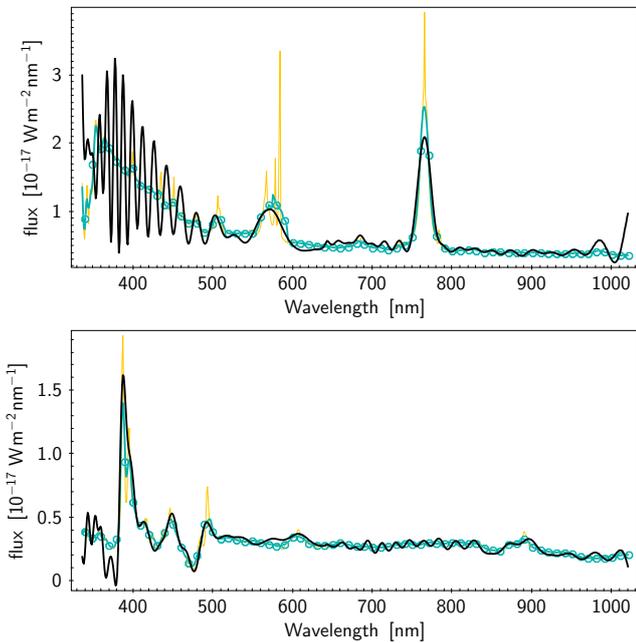
Fig. 35. Comparison between ECS and model SED for source Gaia DR3 4007020769942990080 (*top*) and source Gaia DR3 4467076569812517248 (*bottom*), two QSOs from the SDSS catalogue. The colour coding and symbols are the same as in Fig. 32.

Fig. 37. Normalised residuals between ECS and reference fluxes for sources with $G > 12$. The two red curves represent the P_{16} and P_{84} percentile distributions smoothed by a Gaussian with $\sigma = 10\,\mathrm{nm}$. The ± 2 levels are also plotted as horizontal dashed black lines.

8.3. Effect of the reconstruction error on ECS

In Sect. 7.1 we evaluate the numerical reconstruction error on inverse bases by forward modelling their image through the instrument model and comparing the resulting functions with the original Hermite functions. As discussed in Sect. 7.2, well-behaved Gaussian noise on the instrument model can badly propagate to inverse bases, creating systematic errors in ECS. To look for such systematic effects, we performed the following test: using measured *BP/RP* mean spectra coefficients of SPSS, PVL, and NGSL sources, we computed the corresponding ECS, and then we forward modelled the *BP/RP* mean spectra corresponding to these ECS in sample space and compared the results with the original *Gaia BP/RP* sampled mean spectra. This comparison is shown in Fig. 39 as percentage residuals (top panel) and as normalised residuals (bottom panel) as a function of wavelength. In the second case, the normalisation is done by

fainter than expected in the bluest source, it becomes systematically brighter than expected for redder sources. Moreover, the difference is higher at smaller wavelengths, following the same behaviour shown in Fig. 25 for ICS.
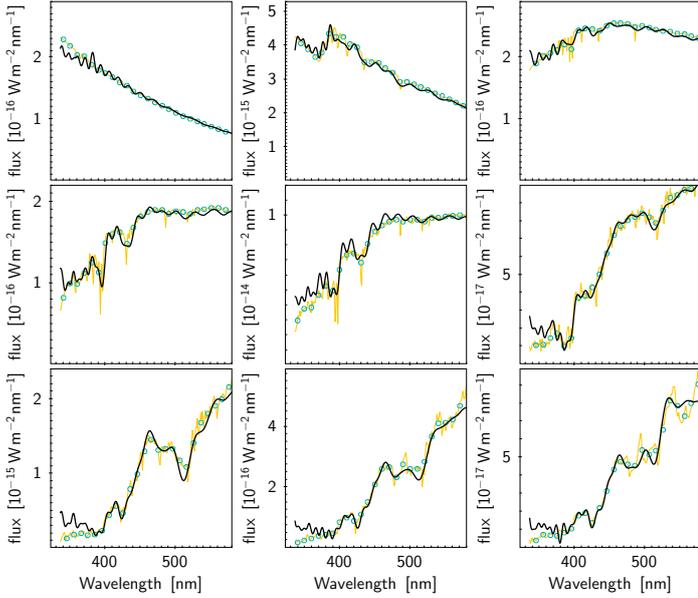
Fig. 38. Comparison between ECS and model SED for SPSS with colour index $G_{\rm BP}-G_{\rm RP}$ respectively equal to (from *left* to *right*, from *top* to *bottom*) 0.06, 0.27, 0.83, 1.03, 1.07, 1.35, 1.68, 2.26, and 2.65. ECS flux at wavelengths $\lambda < 400$ nm is systematically lower than expected for bluer sources and higher than expected for the redder ones. The colour convention and symbols are the same as in Fig. 32.

dividing the residuals by the error on the model spectra[6]. In both cases, residuals are computed as observation minus model and the colour-coding indicates source colour. The percentage residuals show a strong systematic difference in *RP* for wavelengths $\lambda \gtrsim 950$ nm, while in the remaining *RP* wavelength range, residuals are extremely well behaved, and are generally confined to the $\pm 0.1\%$ region. In comparison, BP shows more important ripples confined to the range $\pm 0.2\%$ for wavelengths in the interval $[480, 600]$ nm, progressively worsening outside this interval. It is interesting to notice the colour dependence of the amplitude of the ripples from the $G_{\rm BP}-G_{\rm RP}$ colour index, with redder sources showing more extended ripples than bluer ones. The mean behaviour of the residuals at the shortest wavelength is more clear in the bottom panel, where the normalisation by the error has the effect of smoothing the ripples in the blue part of the *BP* spectrum, particularly for the reddest sources: residuals reveal a significant increasing systematic difference for wavelengths $\lambda \lesssim 350$ nm. Residuals for *RP* at longer wavelengths show a trend with colour index; this is expected considering the larger flux errors at these wavelengths for blue sources than for redder ones. Finally, it is worth noting that these systematic effects have a much smaller amplitude than those seen in the ICS comparison in Fig. 24 and consequently their amplitude in wavelength space could possibly be overwhelmed by other systematic effects, such as those mentioned in relation to Figs. 36 and 37.

8.4. Synthetic photometry on ECS

The availability of calibrated low-resolution SEDs means various applications are possible, such as performing wide- and medium-band synthetic photometry in any photometric system whose passbands are fully enclosed in the $330-1050$ nm wavelength range covered by *Gaia BP* and *RP* spectra. The sci-

---

[6] Errors on the measured *BP/RP* spectra are irrelevant here because we start from the noisy spectrum and compare back to it.
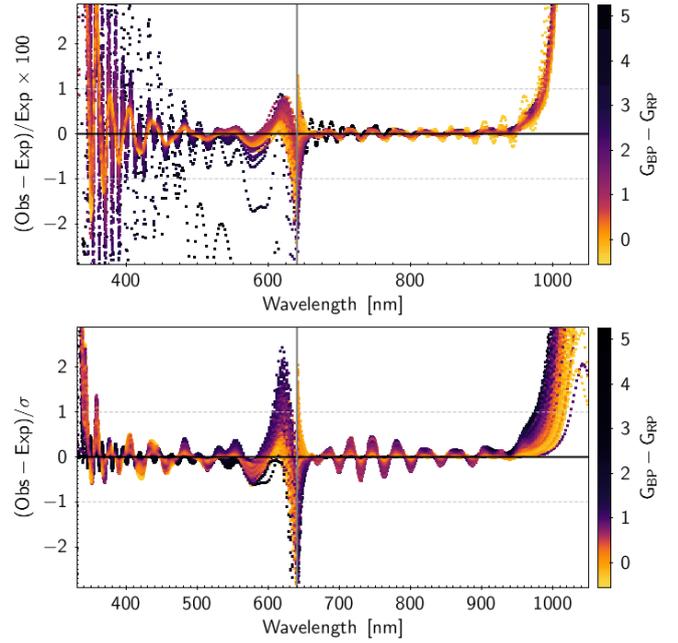
Fig. 39. Effect of the reconstruction error on ECS. *Top*: percentage residuals between mean *BP/RP* fluxes and corresponding forward modelled mean spectra computed from the ECS and instrument models as a function of wavelength; the plot uses data for the full set of SPSS, PVL, and NGSL samples. *Bottom*: residuals are normalised by the model error. In both cases, the colour map encodes the $G_{\rm BP}-G_{\rm RP}$ colour index, while horizontal dashed lines placed at ± 1 provide a reference. The vertical grey line at $\lambda = 640$ sets the separation between *BP* and *RP* data.

ence verification paper by Gaia Collaboration (2023b) is dedicated to this specific application. Here, we use this technique for the validation of our calibrations because it allows comparisons with completely independent reference data. In the following, we illustrate two different test cases: a comparison with Landolt standard photometry in the Johnson-Kron-Cousins (JKC hereafter, Bessell 2005) system and a comparison with HIPPARCOS photometry in the B_T, V_T, and H_p bands (van Leeuwen et al. 1997). In both cases, we first show a comparison between synthetic photometry realised on external SED data and synthetic photometry on ECS data computed for the SPSS, PVL, and NGSL samples to verify the consistency between the two scales. We then test the ECS synthetic photometry against external photometric standards to allow for independent validation of the ECS flux scale. The acronym HRS is used to distinguish external high-resolution spectra from ECS data.

8.4.1. Landolt photometry

A first comparison is made by computing synthetic photometry in the JKC system (Johnson & Morgan 1953; Johnson 1963; Kron et al. 1953; Cousins 1973, 1983, 1984) for a set of nearly 32 800 sources belonging to the Landolt collection of standard stars (Landolt & Uomoto 2007; Landolt 1992, 2007, 2009, 2013; Clem & Landolt 2013, 2016); for a detailed description of this reference sample and the selection criteria applied to the original collection please refer to Pancino et al. (2022) and references therein. The comparison is performed in the *B*, *V*, *R*, and *I* passbands only because *U* band does not fall entirely within the wavelength range covered by *Gaia BP* and *RP* spectra: passband response curves are taken from Bessell & Murphy (2012). Before comparing synthetic ECS photometry with Landolt data,
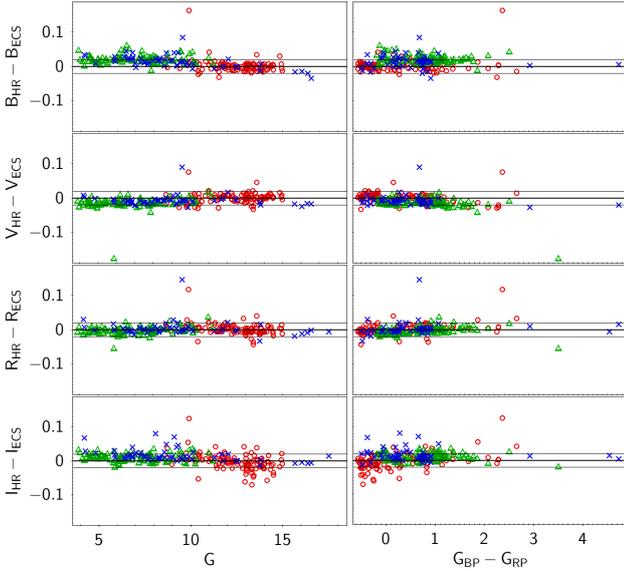
Fig. 40. Comparison between synthetic photometry computed on external high-resolution spectra and that computed on ECS in the Johnson-Kron-Cousins photometric system. Data are from SPSS (open red circles), PVL (blue crosses), and NGSL (open green triangles) samples. The comparison is shown as a function of G magnitude (*left*) and colour index $G_{BP}-G_{RP}$ (*right*). Two horizontal lines at ±0.02 are shown for reference.

it is useful to check the impact of systematic effects in ECS on the synthetic photometry itself. To do this, we compare synthetic photometry computed on external reference SEDs for SPSS, PVL, and NGSL with the corresponding synthetic photometry computed on ECS (Fig. 40). Residuals are shown as a function of G magnitude and colour index $G_{BP}-G_{RP}$ on the left and right panels, respectively. As expected, the most relevant systematic differences are seen in the $G \lesssim 11$ bright range (covered mostly by NGSL sources, plotted as green open triangles, and PVL data, plotted as blue crosses): synthetic ECS magnitudes are brighter than their HRS counterparts by nearly 0.02 mag in B, are fainter than expected by about 0.01 mag in V and R bands, and are again brighter by $\simeq 0.01$ mag in I band. Plots against colour index do not show any relevant colour trend, except for the blue part of the I band where SPSS data (red open circles) create a cluster of points with a mean residual around $\simeq -0.01$ mag (however, we note that SPSS data are more noisy in I band compared to other filters).

Figure 41 shows residuals between the standard photometry and ECS synthetic photometry for the Landolt standards as a function of G magnitude (left panels) and $G_{BP}-G_{RP}$ colour index (right panels) for the B, V, R, and I bands of the JKC system. Red curves represent the median (P_{50}) of the distributions. Residuals as a function of G magnitude reveal a magnitude-dependent term at the faint end (at $G \gtrsim 16$): as discussed by Gaia Collaboration (2023b), this trend in residuals is interpreted as the result of a systematic overestimation of background in *Gaia BP* and *RP* spectra, which causes the actual *BP* and *RP* fluxes to be slightly underestimated. This effect was already revealed by Evans et al. (2018) by their comparison between *Gaia* DR2 photometry and external data, where the *hockey stick* term was first used to describe the observed feature, and by Riello et al. (2021), who compared G-band photometry and synthetic G photometry computed on ECS (see their Fig. 23). At magnitudes $G \lesssim 11$, residuals show offsets analogous to those seen in Fig. 40. However
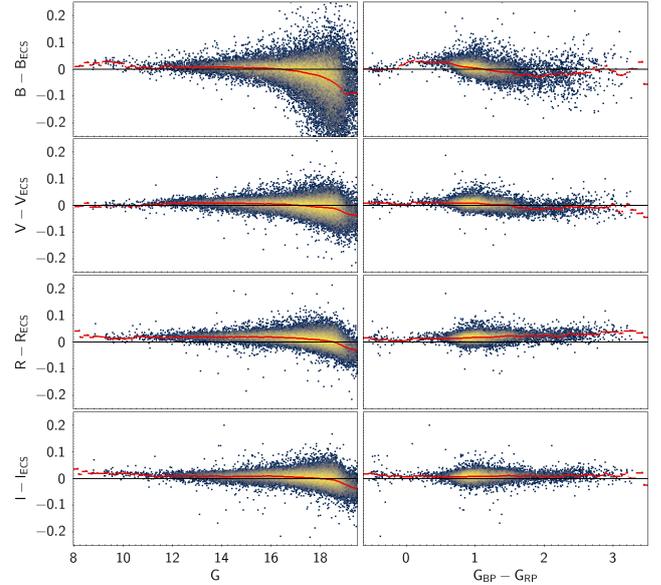
Fig. 41. Comparison between Landolt standard *BVRI* magnitudes and synthetic *BVRI* photometry in the Johnson-Kron-Cousins system computed on externally calibrated *BP* and *RP* spectra for a sample of 32 781 sources selected from the Landolt collection of standards. *Left*: residuals as a function of magnitude. The red lines represent the smoothed median distributions. *Right*: residuals as a function of $G_{BP}-G_{RP}$ colour for sources with $G < 17$.

Table 2. Offsets measured in the *BVRI* bands between Landolt standards and ECS synthetic photometry flux scale.

	B	V	R	I
Δ mag	0.005	0.005	0.018	0.009
St. Dev.	0.027	0.017	0.015	0.015

the overall accuracy estimated through the comparison with Landolt standards is excellent and ranges from 5 mmag in B band to 18 mmag in R band, while the standard deviations for sources with $G < 17$ range from 27 mmag for B to 15 mmag for R band (values are reported in Table 2). To minimise the noise due to the hockey stick effect and better trace the genuine colour terms, plots as a function of $G_{BP}-G_{RP}$ colour index refer only to sources with $G < 17$. These residuals reveal a significant colour term in the B filter causing differences of up to 0.06 mag in the considered colour range. This colour term could be related to systematic effects impacting the blue part of ECS, considering that the B transmission curve extends to wavelengths $\lambda < 400$ nm. A smaller colour term is still visible in the V and R bands while residuals in I band are almost flat with differences within 0.01 mag.

8.4.2. HIPPARCOS photometry

The HIPPARCOS mission[7], in addition to state-of-the art astrometry for the time, provided space-based photometry in three wide bands for the $\simeq 10^5$ stars included in the catalogue (Perryman et al. 1997; van Leeuwen et al. 1997). The HIPPARCOS B_T and V_T passbands are similar to JKC B and V, respectively, while the H_p band ranges from $\simeq 340$ nm to

---
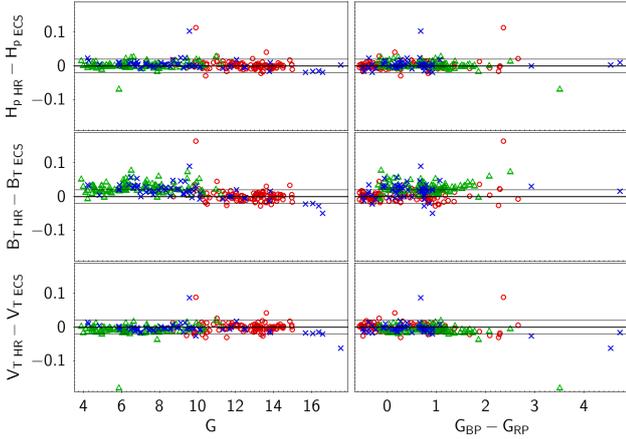
Fig. 42. Comparison between synthetic photometry computed on external high-resolution spectra and on ECS in the HIPPARCOS H_p and *Tycho* B_T, V_T photometric system. SPSSs are represented as open red circles, PVLs as blue crosses, and NGSLs as open green triangles. The difference between the two sets of magnitudes is shown as a function of G magnitude (*left*) and colour index $G_{BP}-G_{RP}$ (*right*). Two horizontal lines at ±0.02 are shown for reference.

$\simeq$900 nm, peaking around $\simeq$450 nm. HIPPARCOS photometry is generally recognised as a benchmark of excellent precision (see e.g. Bessell 2005), and 95% of the stars have uncertainties of <0.01 mag in H_p. The comparison with *Gaia* photometry is limited by the relatively bright magnitude limit of the HIPPARCOS catalogue, where the vast majority of stars have $G < 11.0$. The original passband curves (van Leeuwen et al. 1997) were revised by Bessell & Murphy (2012) who provide improved transmissivity curves that we adopt for the comparisons shown in this section.

Figure 42 shows the comparison between HRS and ECS synthetic photometry computed on SPSS, PVL, and NGSL spectra. Symbols are the same as Fig. 40. As can be seen, thanks to its large width, the H_p photometry shows excellent agreement between the two sets, with no magnitude or colour trends. On the other hand, the B_T filter, covering the more problematic *BP* wavelength region at $\lambda < 400$ nm, shows differences of up to 0.02 mag for sources brighter than $G \lesssim 10.5$ – the magnitude range covered by HIPPARCOS catalogue – while the agreement is still excellent at fainter magnitudes. V_T photometry shows better agreement, with synthetic ECS magnitudes being slightly fainter ($\sim$0.01 mag) than their HRS counterparts at the bright end.

The original HIPPARCOS catalogue, once cross-correlated with the *Gaia* DR3 catalogue, gives a list of 99 525 matches that reduce to a list of 88 662 sources when the selection criteria for published *BP* and *RP* spectra are applied (Fouesneau et al. 2023; De Angeli et al. 2023). According to van Leeuwen et al. (1997), the magnitude scales for HIPPARCOS passbands were chosen such that $H_p = V_T = V$ and $B_T = B$ at $B - V = 0$, and therefore in order to compare these magnitude scales to synthetic ECS magnitude, we need to adjust our zero points accordingly, with the following relations (see details in Appendix E):

$$H_p = H_{p\,ECS} + 0.022 \tag{51}$$

$$B_T = B_{T\,ECS} + 0.054 \tag{52}$$

$$V_T = V_{T\,ECS} + 0.011. \tag{53}$$

Residuals between standard HIPPARCOS/*Tycho* magnitudes and synthetic ECS magnitudes computed through Bessell passbands are shown in Fig. 43 as a function of $G_{BP}-G_{RP}$ colour index. The H_p panel (top) reveals a small colour term spanning an
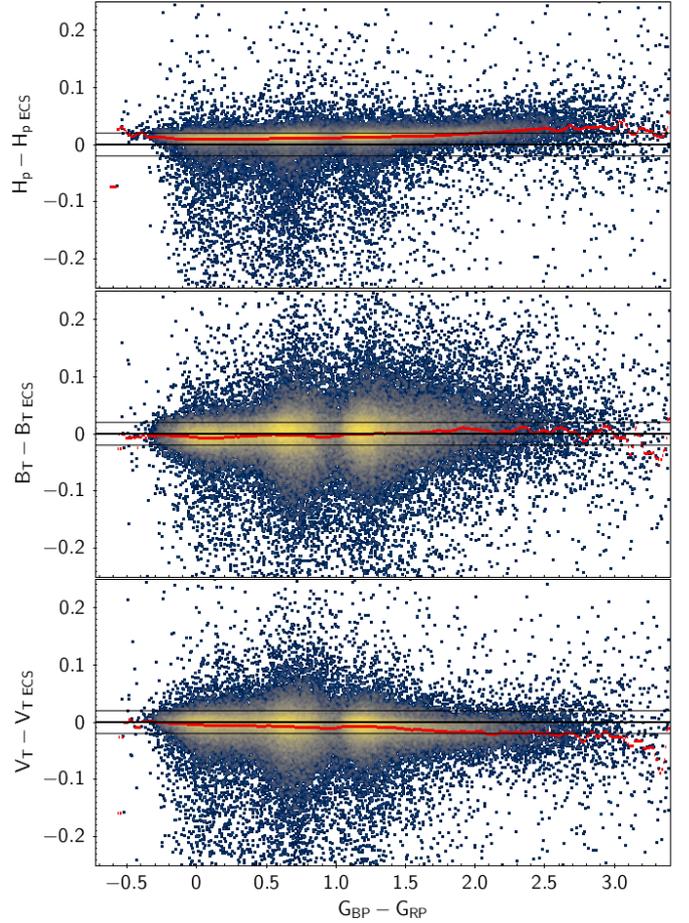
Fig. 43. Residuals between HIPPARCOS/*Tycho* standard magnitudes and synthetic photometry from ECS computed assuming revised Bessell & Murphy (2012) passbands. Red curves are the median distributions; horizontal lines at ±0.02 are shown for reference.

excursion of less than 2 mmag along the whole colour range. At $G_{BP}-G_{RP} = 0.0$, the median distribution (red curve) shows an offset of roughly 10 mmag. Due to the rather large width of the H_p band, the collected flux is larger than that from the other two bands, resulting in a very narrow sequence of residuals with a standard deviation of $\sim$12 mmag. B_T band residuals are almost flat along the whole colour range, with a standard deviation of $\sim$34 mmag. Good agreement is also found in V_T band, with a small colour term spanning an excursion similar to that of H_p band but opposite in sign; the standard deviation of the residuals is $\sim$26 mmag. The comparison with HIPPARCOS photometry also provides a unique opportunity to test the spatial homogeneity of *BP* and *RP* spectrophotometry over the entire sky with space-based data. In order to show the residuals between the HIPPARCOS and *Gaia* synthetic magnitude scales, excluding colour systematic effects, we fitted each median distribution of Fig. 43 with a fourth degree polynomial function in $G_{BP}-G_{RP}$ colour and subtracted this term from Δmag: Figure 44 shows the HEALPiX level 3 maps in Galactic coordinates (Hammer Aitoff projections) of residuals in the three bands, limiting the representation to sources with $|C^\star| < 0.05$[8], $-0.5 < G_{BP}-G_{RP} < 3.0$ and $|\Delta$mag$| < 0.2$ to avoid spurious signals due to outliers. The maps show that the synthetic photometry from *Gaia BP* and *RP* spectra matches the original HIPPARCOS photometry to within

---

[8] See Riello et al. (2021) for $C^\star$ definition.
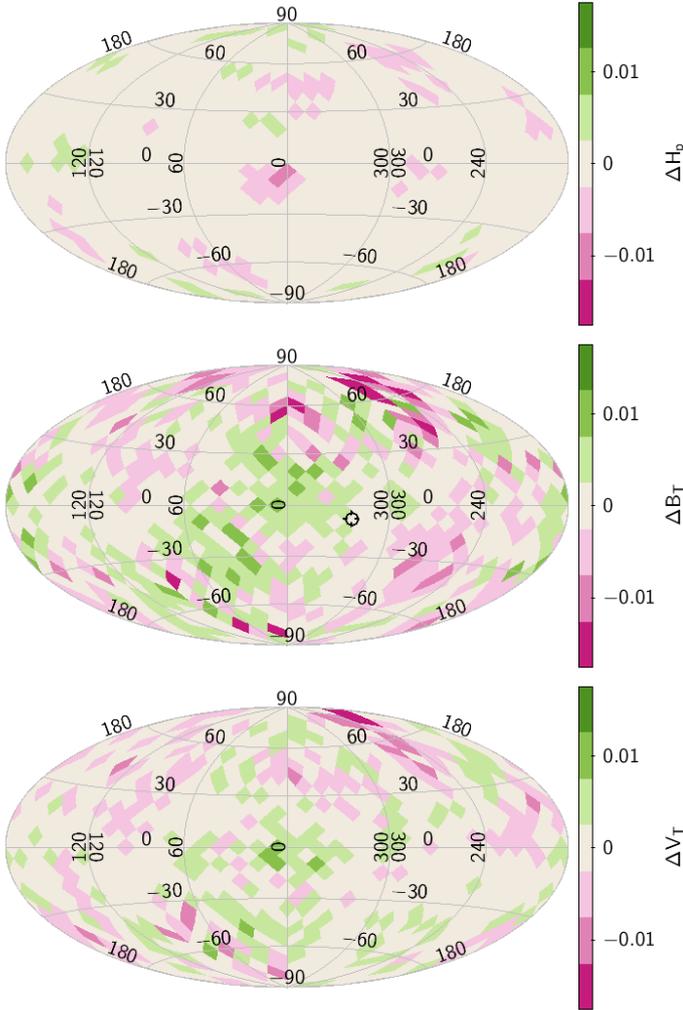
Fig. 44. HEALPiX level 3 maps in Galactic coordinates (Hammer Aitoff projection) of the median residuals between standard Hipparcos/*Tycho* magnitudes and synthetic magnitudes from ECS, after correction for the small colour trends shown in Fig. 43. To avoid spurious signals, we limited the comparison to stars with $-0.5 < G_{BP}-G_{RP} < 3.0$, $|\Delta mag| < 0.2$, and $|C^\star| < 0.05$. The $\simeq 80\,000$ sources involved are not evenly distributed but cover the entire sky at this resolution: beige pixels correspond to absolute median differences $\lesssim 2.5$ mmag and not to empty pixels.

Fig. 45. Examples of heavily saturated spectra taken from the NGSL set: sources `Gaia DR3 2067518817314952576` (*top*) with $G = 2.11$ and $G_{BP}-G_{RP} = 0.99$ and `Gaia DR3 2106630885454013184` (*bottom*) with $G = 2.36$ and $G_{BP}-G_{RP} = 2.42$. Colour coding and symbols are the same as in Fig. 21.

± 5.0 mmag in the vast majority of cases, in all three Hipparcos bands. Fluctuations of amplitude $\gtrsim 10.0$ mmag are quite rare. It is interesting to compare these maps with the sky density of objects with BP/RP spectral data represented in the top left panel of Fig. 28 of De Angeli et al. (2023): sky regions with higher residuals in the B_T map correspond to regions with a lower number of BP/RP observations.

9. Known problems and caveats

9.1. Saturation

Saturation can affect observations of bright sources. Even though an ad hoc gating scheme has been designed and configured to reduce the effective exposure time and avoid saturation, large uncertainties in the on-board magnitude estimates at the bright end imply that gate activation is not always optimal. Moreover, the saturation level is different between different CCDs and dis-

persion differences across the focal plane will also play a role. A bright source can therefore have a mixture of saturated and non-saturated BP/RP observed spectra. The size of the effect on combined mean spectra will depend on the fraction of saturated versus non-saturated epoch spectra and on the size of the effects on single observations. As a consequence, the shape of a saturated spectrum may not always be recognisable as a spectrum with a cropped, sharp, flat section at the top of it, corresponding to some fixed flux threshold, but it can exhibit a very complicated behaviour, as can be seen in Fig. 45 where a couple of examples are shown.

The NGSL sample (containing very bright sources up to $G = 1.97$) represents an ideal dataset for assessing the effect of saturation in *Gaia* spectra. We use synthetic photometry to measure the saturation level present in spectra by comparing the synthetic flux in BP and RP measured on the *Gaia* ECS with the synthetic flux obtained on NGSL SEDs. The difference between the corresponding synthetic magnitudes is represented in Fig. 46 as a function of the source magnitude in the same filter: results for BP and RP are shown respectively in left and right top panels. Saturation in BP spectra starts around $BP \simeq 4.5$ depending on the colour, with redder sources being increasingly affected. This behaviour is due to the properties of the BP dispersion relation: in the case of a red source as in the example shown in Fig. 45 (lower panel), most of the source flux is concentrated in a few pixels. These are the ones that will saturate in the case of a sufficiently bright source. At the same flux level, a bluer/hotter source has less chance of saturating because photons are distributed over a wider range of samples. In the case of the RP, saturation is found at magnitudes $RP \lesssim 3$ and affects mostly red sources. In the bottom left panel of Fig. 46, data for both BP and RP are plotted against the G magnitude. Sources showing signs of saturation in their BP or RP spectra (differences in shape
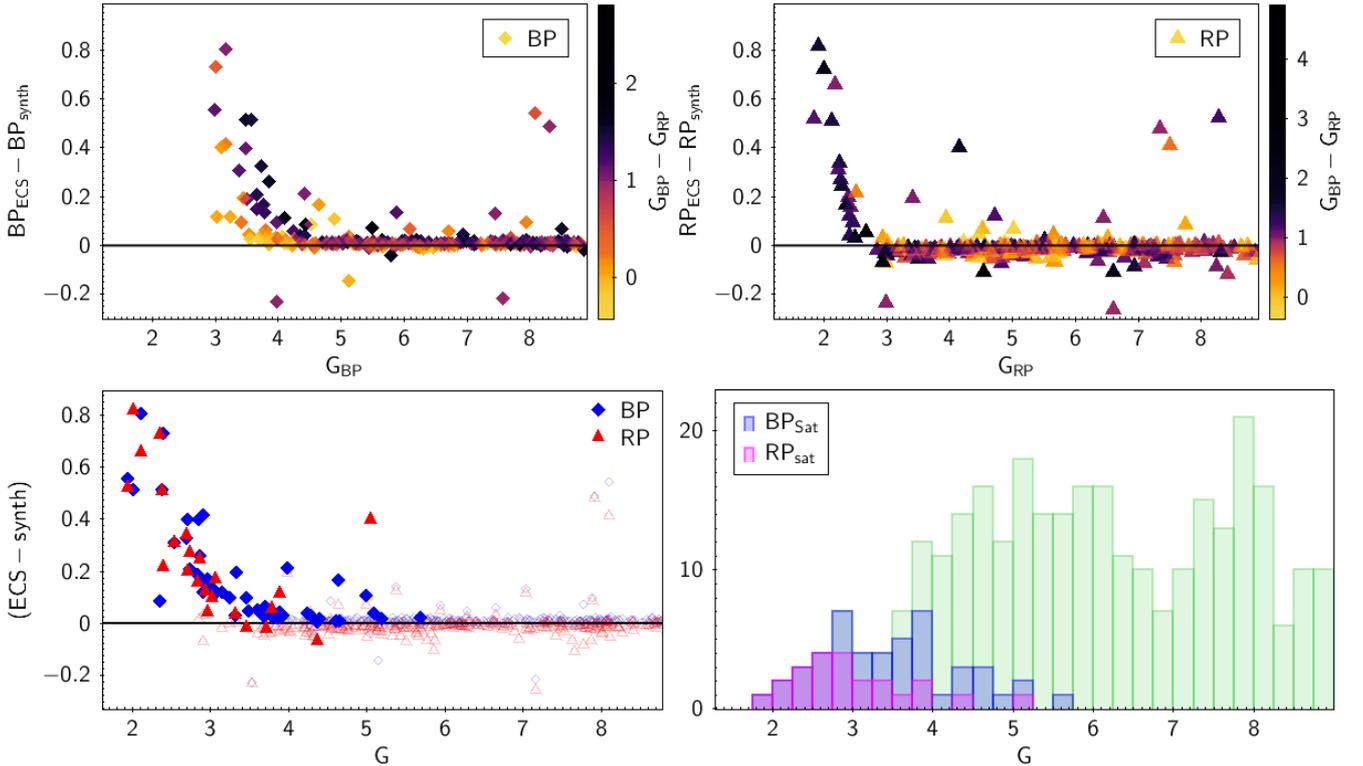
Fig. 46. Saturation in *BP* and *RP* spectra. *Top*: difference between synthetic magnitudes computed on NGSL SEDs and the corresponding synthetic magnitude computed on *Gaia* ECS, used as an indicator of saturation in *Gaia* spectra. Differences are computed for the *BP* band (*left*) and *RP* (*right*). *Bottom*: delta synthetic magnitudes measured on both *BP* and *RP* spectra against the *G* magnitude of the source (*left*); filled symbols represent spectra with clear signs of saturation (selected manually) in their spectra shape. *Right panel*: distributions of *BP* and *RP* saturated spectra against the distribution for unsaturated sources.
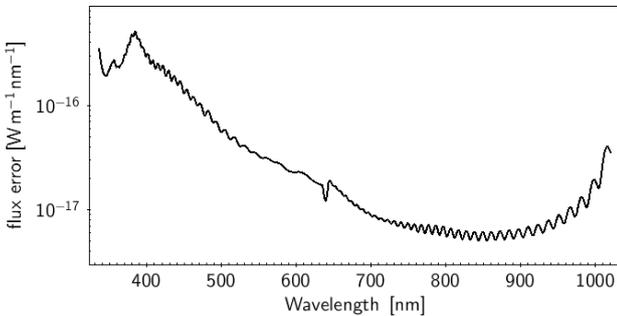
Fig. 47. Flux error distribution for the externally calibrated spectrum of source `Gaia DR3 1435896975388228224`. Errors are characterised by the presence of wiggles.

effects such as CCD non-linearities and larger uncertainties in the source astrometry could play a role.

9.2. Wiggles in ECS errors

The ECS flux error distribution against wavelength shows wiggles, as seen in Fig. 47 for source `Gaia DR3 1435896 975388228224`. This is visible for all sources. Considering that the flux error is computed as the squared root of the diagonal elements of the covariance matrix on the ECS (see Eq. (37)) and that the flux error distribution of *BP*/*RP* ICS does not show such wiggles, we deduce that the effect is a numerical problem introduced by the inversion process and may be due to poor conditioning of the inverse basis design matrices.

10. Summary and conclusions

This paper describes the process leading to the external calibration of *Gaia BP* and *RP* low-resolution spectra. We derived an instrument model for the *BP* and *RP* spectrophotometers that allows us to forward-model observed *Gaia* mean spectra starting from empirical or theoretical SEDs. The model has been optimised using a large set of primary flux calibrators (SPSS) and secondary calibrators (PVL and emission lines sources).

Through inverse modelling, we generated a set of inverse bases allowing the reconstruction of externally calibrated low-resolution SEDs for the 220 million sources with published *BP*/*RP* spectra within *Gaia* DR3. These reconstructed SEDs are flux- and wavelength calibrated and have an increased resolution with respect to the original mean spectra due to mitigation of

between observed ICS and model prediction) were selected manually and are visualised in the plot as filled symbols while well behaved sources are represented as open transparent ones. The bottom right panel of Fig. 46 shows the distribution with *G* magnitude of all NGSL objects and those that show saturation in *BP* or *RP*: from the histogram, we can conclude that saturation is not significant at *G* > 5, and that most sources will have saturated *BP* spectra at *G* ≲ 4, and it is only *G* ≲ 3.5 that saturation will also occur in most *RP* spectra. A final word of caution should be added here to say that given the very small number of sources brighter than the thresholds indicated in this section and the fact that *Gaia* was not designed to cover such bright magnitudes, further investigation is needed to completely understand the origin of the discrepancies observed in Fig. 46. Other

the LSF smearing effect. Errors associated to calibrated spectra appear to be underestimated, especially for sources brighter than magnitude $G \simeq 12$.

Comparison with external data reveals the presence of systematic effects in the wavelength region below $\lambda < 400$ nm. These systematic effects are a function of the colour of the sources and, to a lower degree, their magnitude. Furthermore, smaller magnitude-dependent systematic effects impacting sources brighter than $G \simeq 12$ are present in the wavelength range $560-600$ nm and in the red part of the spectra at $\lambda \simeq 950$ nm. Externally calibrated spectra are occasionally affected by the presence of numerical artefacts in the form of ripples produced by the basis inversion process.

Calibrated spectra were also validated by comparing synthetic photometry in the Johnson-Kron-Cousins system with Landolt standard sources and in the HIPPARCOS H_p and *Tycho*-2 B_T, V_T passbands with the HIPPARCOS catalogue. In both cases, we find excellent agreement overall, with typical accuracy ranging from 5 to 20 mmag and precision between 15 and ~30 mmag, as measured by the standard deviation of the residuals over wide colour and magnitude ranges. The comparison with HIPPARCOS photometry demonstrates the excellent degree of spatial homogeneity of the synthetic photometry with a relative accuracy of less than 2.7 mmag over 90% of the entire sky. Additional validation by comparison with a large data set of reliable external photometry is provided in Gaia Collaboration (2023b). Finally, caution must be exercised when using the spectra of very bright sources because of the presence of saturation, especially for sources brighter than $G \simeq 4$.

References

Altavilla, G., Marinoni, S., Pancino, E., et al. 2015, Astron. Nachr., 336, 515

Altavilla, G., Marinoni, S., Pancino, E., et al. 2021, MNRAS, 501, 2848

Andrae, R., Fouesneau, M., Sordo, R., et al. 2023, A&A, 674, A27 (*Gaia* DR3 SI)

Astropy Collaboration (Price-Whelan, A., et al.) 2018, AJ, 156, 123

Bessell, M. S. 2005, ARA&A, 43, 293

Bessell, M., & Murphy, S. 2012, PASP, 124, 140

Boch, T., & Fernique, P. 2014, in Astronomical Data Analysis Software and Systems XXIII, eds. N. Manset, & P. Forshay, ASP Conf. Ser., 485, 277

Bohlin, R. C., Gordon, K. D., & Tremblay, P. E. 2014, PASP, 126, 711

Bohlin, R. C., Hubeny, I., & Rauch, T. 2020, AJ, 160, 21

Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, A&AS, 143, 33

Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, A&A, 652, A86

Clem, J. L., & Landolt, A. U. 2013, AJ, 146, 88

Clem, J. L., & Landolt, A. U. 2016, AJ, 152, 91

Cousins, A. W. J. 1973, Mem. R. Astron. Soc., 77, 223

Cousins, A. W. J. 1983, S. Afr. Astron. Obs. Circ., 7, 47

Cousins, A. W. J. 1984, S. Afr. Astron. Obs. Circ., 8, 59

Creevey, O. L., Sordo, R., Pailler, F., et al. 2023, A&A, 674, A26 (*Gaia* DR3 SI)

De Angeli, F., Weiler, M., Montegriffo, P., et al. 2023, A&A, 674, A2 (*Gaia* DR3 SI)

Evans, D. W., Riello, M., De Angeli, F., et al. 2018, A&A, 616, A4

Fabricius, C., Høg, E., Makarov, V. V., et al. 2002, A&A, 384, 180

Fouesneau, M., Frémat, Y., Andrae, R., et al. 2023, A&A, 674, A28 (*Gaia* DR3 SI)

Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1

Gaia Collaboration (Vallenari, A., et al.) 2023a, A&A, 674, A1 (*Gaia* DR3 SI)

Gaia Collaboration (Montegriffo, P., et al.) 2023b, A&A, 674, A33 (*Gaia* DR3 SI)

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357

Heap, S. R., & Lindler, D. 2016, in The Science of Calibration, eds. S. Deustua, S. Allam, D. Tucker, & J. A. Smith, ASP Conf. Ser., 503, 211

Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27

Huffel, S., & Vandewalle, J. 1991, The Total Least Squares Problem: Computational Aspects and Analysis (Philadelphia: Society for Industrial and Applied Mathematics)

Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90

Johnson, H. L. 1963, in Basic Astronomical Data: Stars and Stellar Systems, ed. K. A. Strand, 204

Johnson, H. L., & Morgan, W. W. 1953, ApJ, 117, 313

Jordi, C., Høg, E., Brown, A. G. A., et al. 2006, MNRAS, 367, 290

Kron, G. E., White, H. S., & Gascoigne, S. C. B. 1953, ApJ, 118, 502

Landolt, A. U. 1992, AJ, 104, 340

Landolt, A. U. 2007, AJ, 133, 2502

Landolt, A. U. 2009, AJ, 137, 4186

Landolt, A. U. 2013, AJ, 146, 131

Landolt, A. U., & Uomoto, A. K. 2007, AJ, 133, 768

Le Borgne, J. F., Bruzual, G., Pelló, R., et al. 2003, A&A, 402, 433

Lindegren, L. 2006, Gaia Data Processing and Analysis Consortium (DPAC) Technical Note GAIA-C3-TN-LU-LL-068, http://www.cosmos.esa.int/web/gaia/public-dpac-documents

Lindegren, L. 2009, Gaia Data Processing and Analysis Consortium (DPAC) Technical Note GAIA-C3-TN-LU-LL-084, http://www.cosmos.esa.int/web/gaia/public-dpac-documents

Lindegren, L. 2010, Gaia Data Processing and Analysis Consortium (DPAC) Technical Note GAIA-C3-TN-LU-LL-084, http://www.cosmos.esa.int/web/gaia/public-dpac-documents

Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, A&A, 649, A2

Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, ApJS, 250, 8

Marinoni, S., Pancino, E., Altavilla, G., et al. 2016, MNRAS, 462, 3616

MATLAB 2018, Version 9.4.0 (R2018a) (Natick, Massachusetts: The MathWorks Inc.)

Montegriffo, P. 2017, Gaia Data Processing and Analysis Consortium (DPAC) Technical Note GAIA-C5-OABO-PMN-012, http://www.cosmos.esa.int/web/gaia/public-dpac-documents

Montegriffo, P., Pancino, E., & Sanna, N. 2022, Gaia Data Processing and Analysis Consortium (DPAC) Technical Note GAIA-C5-OABO-PMN-016, http://www.cosmos.esa.int/web/gaia/public-dpac-documents

Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23

Pancino, E., Altavilla, G., Marinoni, S., et al. 2012, MNRAS, 426, 1767

Pancino, E., Sanna, N., Altavilla, G., et al. 2021, MNRAS, 503, 3660

Pancino, E., Marrese, P. M., Marinoni, S., et al. 2022, A&A, 664, A109

Pérez, F., & Granger, B. E. 2007, Comput. Sci. Eng., 9, 21

Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49

Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3

Rodrigo, C., & Solano, E. 2020, XIV.0 Scientific Meeting (virtual) of the Spanish Astronomical Society, 182

Storn, R., & Price, K. 1997, J. Glob. Optim., 11, 341

Taylor, M. B. 2005, in Astronomical Data Analysis Software and Systems XIV, eds. P. Shopbell, M. Britton, & R. Ebert, ASP Conf. Ser., 347, 29

Taylor, M. B. 2006, in Astronomical Data Analysis Software and Systems XV, eds. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, ASP Conf. Ser., 351, 666

van Leeuwen, F. 2007, A&A, 474, 653

van Leeuwen, F., Evans, D. W., Grenon, M., et al. 1997, A&A, 323, L61

Verro, K., Trager, S. C., Peletier, R. F., et al. 2022, A&A, 660, A34

Weiler, M., Carrasco, J. M., Fabricius, C., & Jordi, C. 2020, A&A, 637, A85

Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9

Ye, J., Janardan, R., & Li, Q. 2004, Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, KDD04, Seattle, Washington, USA, 354

Appendix A: Acknowledgements

The *Gaia* project and data processing have made use of:

– the Set of Identifications, Measurements, and Bibliography for Astronomical Data (SIMBAD, Wenger et al. 2000), the 'Aladin sky atlas' (Bonnarel et al. 2000; Boch & Fernique 2014), and the VizieR catalogue access tool (Ochsenbein et al. 2000), all operated at the Centre de Données astronomiques de Strasbourg (CDS);

– the National Aeronautics and Space Administration (NASA) Astrophysics Data System (ADS);

– the SPace ENVironment Information System (SPENVIS), initiated by the Space Environment and Effects Section (TEC-EES) of ESA and developed by the Belgian Institute for Space Aeronomy (BIRA-IASB) under ESA contract through ESA's General Support Technologies Programme (GSTP), administered by the BELgian federal Science Policy Office (BELSPO);

– the Spanish Virtual Observatory (https://svo.cab.inta-csic.es) project funded by MCIN/AEI/10.13039/501100011033/ through grant PID2020-112949GB-I00 (Rodrigo & Solano 2020);

– the software products TOPCAT, STIL, and STILTS (Taylor 2005, 2006);

– MATLAB (MATLAB 2018);

– Matplotlib (Hunter 2007);

– IPython (Pérez & Granger 2007);

- Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration 2018);
- NumPy (Harris et al. 2020);
- pyphot (http://github.com/mfouesneau/pyphot);
- the HIPPARCOS-2 catalogue (van Leeuwen 2007). The HIPPARCOS and Tycho catalogues were constructed under the responsibility of large scientific teams collaborating with ESA. The Consortia Leaders were Lennart Lindegren (Lund, Sweden: NDAC) and Jean Kovalevsky (Grasse, France: FAST), together responsible for the HIPPARCOS Catalogue; Erik Høg (Copenhagen, Denmark: TDAC) responsible for the Tycho Catalogue; and Catherine Turon (Meudon, France: INCA) responsible for the HIPPARCOS Input Catalogue (HIC);
- the Tycho 2 catalogue (Høg et al. 2000), the construction of which was supported by the Velux Foundation of 1981 and the Danish Space Board;
- The Tycho double star catalogue (TDSC, Fabricius et al. 2002), based on observations made with the ESA HIPPARCOS astrometry satellite, as supported by the Danish Space Board and the United States Naval Observatory through their double-star programme;
- the VizieR catalogue access tool (CDS, Strasbourg, France).

This publication made extensive use of the online authoring Overleaf platform (https://www.overleaf.com/).

Appendix B: Data products

The BP and RP data are made available via the *Gaia* archive[9]. BP and RP mean spectra are published for 219,197,643 **sources**. This list includes mostly sources with *G*-band magnitude brighter than 17.65 mag and more than 15 CCD transits contributing to the generation of the mean spectra for both BP and RP. Other selection criteria were also applied; see De Angeli et al. (2023) for more details.

All BP and RP spectra are provided in their continuous representation. See Sect. 4 and Appendices A and B in De Angeli et al. (2023) for a description of the data format and for download instructions. For a subset of sources with BP and RP spectral data including only sources brighter than *G* = 15 mag (the exact list can be obtained by selecting entries with gaia_source.has_xp_sampled='t') BP and RP spectra are also provided in the sampled representation on a default grid as externally calibrated spectra in absolute flux and wavelength systems. These can be retrieved via DataLink from the archive interface or programmatically selecting the sampled type (XP mean sampled spectra on the web interface and retrieval_type='XP_SAMPLED' when using astroquery from Python).

Externally calibrated sampled spectra for all sources, optionally on a user-defined wavelength grid, can be obtained via GaiaXPy (De Angeli et al., in prep.), a Python package developed to help users of the BP and RP spectral data.

Inverse bases computed for a default wavelength sampling (XpMerge/XpSampling tables) are distributed in a dedicated Cosmos page: https://www.cosmos.esa.int/web/gaia/dr3-xpmergexpsampling GaiaXPy provides the possibility to sample ECS on a user-defined wavelength grid.

GaiaXPy provides the possibility to forward model mean BP and RP sampled spectra via the IM on a user-defined pseudo-wavelength grid: optionally GaiaXPy is capable to project these spectra into coefficients space allowing for the simulation of spectra in their continuous representation. GaiaXPy is avail-

able on a dedicated Cosmos page: https://www.cosmos.esa.int/web/gaia/gaiaxpy.

Appendix C: LSF model implementation

The current model for the PSF/LSF representation is based on the preliminary studies and developments described in several technical notes by Lindegren (2006, 2009, 2010). Here we sketch the basic ideas while a complete description of the model is available in Montegriffo (2017). The optical PSF P_λ^O describes the instantaneous two dimensional intensity distribution in the image of a point source on the CCD: it depends on the optical diffraction of the instrument and the optical aberrations and it is computed as the Fourier transform of the complex amplitude of the incident wavefront in the pupil plane A. Let (x, y) be linear coordinates in the pupil plane in meters, and (u, v) the corresponding angular coordinates in the image plane in radians, then for a given WFE map $w(x, y)$ we get

$$A(x, y) = \begin{cases} e^{i\frac{2\pi}{\lambda} w(x,y)} & \text{for } (x, y) \in \text{ pupil} \\ 0 & \text{otherwise,} \end{cases} \quad \text{(C.1)}$$

and

$$P_\lambda^O(u, v) = \frac{1}{\lambda^2 D_x D_y} \left[\int \int_{-\infty}^{\infty} A(x, y) e^{i\frac{2\pi}{\lambda}(xu+yv)} \, dx \, dy \right]^2, \quad \text{(C.2)}$$

where D_x and D_y are the pupil dimensions along x and y: *Gaia* telescopes have two rectangular primary mirrors with dimensions $D_x = 1.4510$ m and $D_y = 0.5016$ m.

The WFE map depends on the FoV and field angles, but is independent from wavelength and should be relatively stable over time. However, it is important to understand that the optical PSF is never observed directly: what is observed is the effective PSF, which takes into account the discretised nature of the data as well as the additional smearing introduced by the TDI technique and the effects due to charge diffusion. The effective PSF P_m can be computed as the inverse Fourier transform of the product between the modulation transfer function (MTF) M and the optical transfer function O_m (by definition O_m is the Fourier Transform of the optical PSF, Eq. C.2). The total MTF must combine the effect of several spatial response functions, i.e.:

- the AL pixel integration (a rectangle of width p_u equal to the CCD AL pixel size expressed in radians);
- the AL four-phase TDI charge transfer (a rectangle of width p_u/n_u, where $n_p = 4$ is the number of phases per pixel);
- the AC pixel integration (a rectangle of width p_v equal to the CCD AC pixel size expressed in radians);
- the charge diffusion (modelled as a bivariate normal distribution with diffusion width $\sigma_u = \sigma_v = 4\mu m$);

$$M(f_x, f_y) = \text{sinc}(\pi f_x p_u) \, \text{sinc}\left(\pi f_x \frac{p_u}{n_p}\right) \text{sinc}(\pi f_y p_v) \, e^{-2\pi^2(\sigma_u^2 f_x^2 + \sigma_v^2 f_y^2)}, \quad \text{(C.3)}$$

where f_x and f_y are the spatial frequencies AL and AC. The effective PSF can then finally computed as:

$$P_m(u, v) = \frac{1}{4\pi^2} \int \int_{-\infty}^{\infty} O_m(f_x, f_y) \, M(f_x, f_y) \, e^{-i2\pi(f_x u + f_y v)} \, df_x \, df_y. \quad \text{(C.4)}$$

By definition, the PSF is normalised to unity. The effective LSF $L_m(u)$ is obtained by summing Eq. C.4 over the AC direction.

Among all the effects cited here, the WFE maps modelling optical aberrations are the most uncertain and as explained in Sect. 4.2 numerical maps measured by Airbus are unusable. Instead we model optical WFE maps as a linear combination of normalised Legendre polynomials up to the 6th order: these maps are randomly generated and scaled to have a root mean square (RMS) uniformly distributed between 40 and 60 nm. Using this setup we generated up to 5000 random numerical WFE maps to compute the corresponding numerical LSFs: in practice each LSF is sampled on a discrete wavelength grid of c points unevenly distributed between 288 nm and 1150 nm with a sampling scheme chosen to guarantee the same relative accuracy of the PSF/LSF computation over the entire wavelength range. Moreover these monochromatic LSFs are sampled over a regular grid of r points centred on $u = 0$ and with a spatial resolution high enough to sample the images at least to the Nyquist frequency. Each numerical LSF is included twice by reversing the AL axis in order to preserve the symmetry of the problem: LSFs are arranged into a stack of matrices $L_i \in \mathbb{R}^{r \times c}$ with $i = 1, \ldots, 10000$. From this stack we calculate the mean LSF $\overline{L}$:

$$\overline{L} = \frac{1}{n} \sum_{i=1}^{n} L_i, \tag{C.5}$$

and then we compute a stack of residuals

$$\widetilde{L}_i = L_i - \overline{L}, \tag{C.6}$$

that is reduced to a set of two-dimensional basis functions by means of GPCA. In practice, an optimal (ℓ_1, ℓ_2)-dimensional axis system, with $\ell_1 < r$, $\ell_2 < c$, defined by two matrices $U \in \mathbb{R}^{r \times \ell_1}$ and $W \in \mathbb{R}^{c \times \ell_2}$ with orthonormal columns, is derived such that the projections of the data points $\widetilde{L}_i$ onto this axis system have the maximum variance over all the possible (ℓ_1, ℓ_2)-dimensional axis systems, where the variance is defined as:

$$var(U, W) = \frac{1}{n-1} \sum_{i=1}^{n} \|U^T \cdot \widetilde{L}_i \cdot W\|_F. \tag{C.7}$$

The symbol $\|.\|_F$ denotes the Frobenius norm of a matrix. For given U and W matrices, the projection of $\widetilde{L}_i$ can be computed as

$$D_i = U^T \cdot \widetilde{L}_i \cdot W \quad \text{with} \quad D_i \in \mathbb{R}^{\ell_1 \times \ell_2}. \tag{C.8}$$

From D_i we can reconstruct $\widetilde{L}_i$ by setting

$$\widetilde{L}_i \approx U \cdot D_i \cdot W^T, \tag{C.9}$$

then

$$L_i \approx U \cdot D_i \cdot W^T + \overline{L}, \tag{C.10}$$

or, in extended notation:

$$L(u_i, \lambda_j) = \overline{L}_{u_i, \lambda_j} + \sum_{m=1}^{\ell_1} \sum_{n=1}^{\ell_2} d_{m,n} \cdot U_{i,m} \cdot W_{j,n}. \tag{C.11}$$

The numerical mean LSF $\overline{L}$ is represented in Fig. C.1 in logarithmic scale; the second term of Eq. C.11 is the component of the model that is fitted during the optimisation process and represents the deviation of the current model from the mean LSF.

The reconstruction error for L_i is then

$$E_i = \|\widetilde{L}_i - U \cdot D_i \cdot W^T\|_F = \|\widetilde{L}_i - U \cdot U^T \cdot \widetilde{L}_i \cdot W \cdot W^T\|_F. \tag{C.12}$$
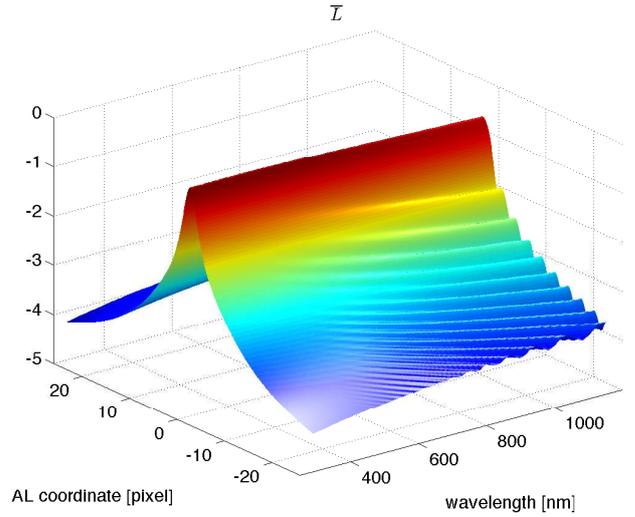

Fig. C.1. Mean LSF model $\overline{L}$ in logarithmic scale.
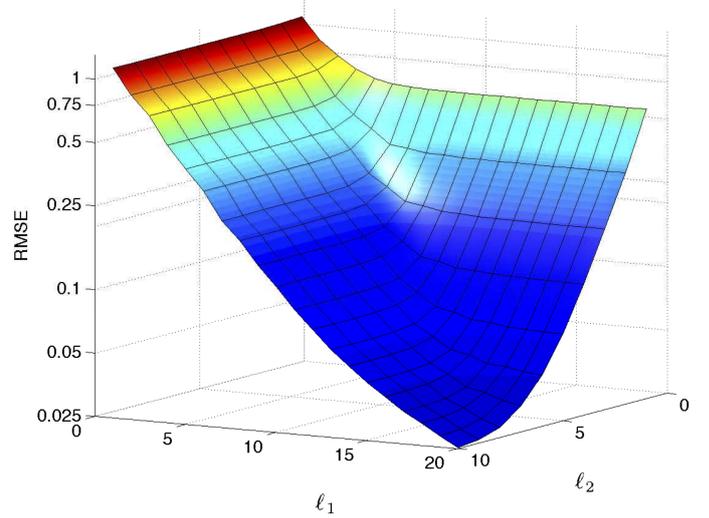

Fig. C.2. Empirical RMS error for the set of 10000 simulated LSFs measured as function of order (ℓ_1, ℓ_2).

The RMS error is defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} E_i^2}, \tag{C.13}$$

and it measures the average reconstruction error.

Figure C.2 shows the RMS error that has been measured on the complete set of 10000 LSF as a function of (ℓ_1, ℓ_2): the shape of the surface suggests that the LSF wavelength modelling should require about half dimensions with respect to the AL dependency modelling. The optimal subspace dimensions for GPCA has been set to $(\ell_1, \ell_2) = (20, 10)$. Columns of matrices U and W can be interpreted as basis functions to model the dependencies along the spatial coordinate u and the wavelength coordinate λ respectively. The numerical LSF given by Eq. C.11 can be easily 2D-interpolated to continuous variables (u, λ) by 1D-interpolation of U and W bases separately. To ensure that the interpolation for the U bases satisfies the 'shift invariant sum' condition –i.e. preserves the underlying function normalisation independently from the sub-pixel position of the sampling

grid– these bases were then fitted with an S-spline model (Lindegren 2009). The interpolation for W bases is less problematic and is achieved by a cubic spline. The numerical bases are extrapolated in the AL direction by a smooth transition to a Lorentzian profile in the wings, while in the wavelength domain extrapolation is not needed, being the wavelength range of BP and RP photometers fully covered by the W bases. Finally, an efficient interpolation scheme based on GPCA decomposition has been implemented also for the mean $\overline{L}$, thus enabling to exploit a continuous LSF model as:

$$L(u, \lambda) = \overline{L}(u, \lambda) + \sum_{m=1}^{\ell_1} \sum_{n=1}^{\ell_2} d_{m,n} \cdot U_m(u) \cdot W_n(\lambda). \tag{C.14}$$

Appendix D: Projection of sampled spectra into coefficient space

To exploit the projection of spectra from sample space to coefficients space it is convenient to adopt matrix formalism. Let

$$\mathbf{b}^T = b_1, b_2, \ldots, b_N \tag{D.1}$$

represent the vector of $(1 \times N)$ coefficients; then a spectrum $\mathbf{s}$ sampled on a given grid $\mathbf{u}$ of M points will be given by

$$\mathbf{s} = D \cdot \mathbf{b}, \tag{D.2}$$

where $D \in \mathbb{R}^{M \times N}$ is the design matrix of the Hermite functions sampled on the $\mathbf{u}$ grid. In principle the design matrix D could be rectangular, and hence not invertible; however we can always compute a pseudo-inverse matrix $D^\dagger \in \mathbb{R}^{N \times U}$ such that

$$\mathbf{b} = D^\dagger \cdot \mathbf{s}, \tag{D.3}$$

by first multiplying both sides of Eq. 22 by D^T from the left and then by multiplying the results by the inverse of matrix $D^T \cdot D$ thus obtaining:

$$D^\dagger = \left(D^T \cdot D\right)^{-1} \cdot D^T. \tag{D.4}$$

The pseudo-inverse can be computed in a numerically stable way through a singular value decomposition of matrix D itself:

$$D = U \cdot S \cdot V^T, \tag{D.5}$$

where

$$U^T \cdot U = I, \tag{D.6}$$

and

$$V^T \cdot V = I. \tag{D.7}$$

The pseudo-inverse is then given by

$$D^\dagger = V \cdot S^{-1} \cdot U^T. \tag{D.8}$$

Appendix E: Hipparcos/Tycho synthetic photometry zero points

Magnitude scales for Hipparcos passbands were chosen such that $H_p = V_T = V$ and $B_T = B$ at $B - V = 0$ (van Leeuwen et al. 1997). In order to reproduce these scales we need to adjust the zero points of the synthetic ECS photometry accordingly. In order to do so we have evaluated the offsets δ_i to let the median of the distributions $V - (H_p + \delta_1)$, $B - (B_T + \delta_2)$ and $V - (V_T + \delta_3)$ to be equal to zero at $B - V = 0$. Residuals with the zero point correction applied are shown on Fig. E.1 while the assumed zero point values can be retrieved in Sect. 8.4.2.
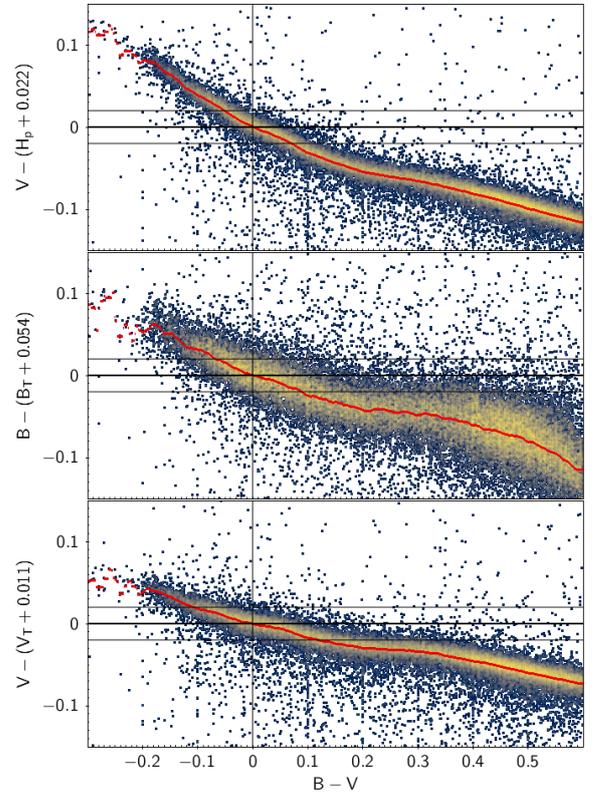

Fig. E.1. Calibration of the zero point for the Hipparcos/Tycho passbands.

Appendix F: *Gaia*-related acronyms

Table F.1. *Gaia*-related acronyms used in the paper. Each acronym is also defined at its first occurrence in the paper.

Acronym	Description	See
AC	ACross scan direction	Sect. 4
AL	ALong scan direction	Sect. 4
BP	Blue Photometer	Sect. 1
CCD(s)	Charge Coupled Device(s)	Sect. 4
CTI	Charge Transfer Inefficiency	Sect. 4
DEA	Differential Evolution Algorithm	Sect. 6
DR	Data Release	Sect. 1
DS	Defence and Space	Sect. 4.1
ECS	Externally Calibrated Spectrum	Sect. 1
ELS	Emission Line Sources	Sect. 3
ESA	European Space Mission	Sect. 1
FoV(s)	Field(s) of View	Sect. 4.1
FWHM	Full Width at Half Maximum	Sect. 7.2
G, G_{BP}, G_{BP}	Integrated *Gaia* magnitudes	Sect. 3
GPCA	Generalised Principal Component Analysis	Sect. 4.2
HRS	High Resolution Spectrum	Sect. 8.4
ICS	Internally Calibrated BP/RP mean Spectra	Sect. 7.2
IM	Instrument model	Sect. 1
JKC	Johnson-Kron-Cousins	Sect. 8.4
LSF	Line Spread Function	Sect. 1
MTF	Modulation Transfer Function	Sect. C
NGSL	Next Generation Spectral Library	Sect. 3
PSF	Point Spread Function	Sect. 4
PVL	Passband Validation Library	Sect. 3
QSO	Quasi-stellar object	Sect. 3
RMS	Root Mean Square	Sect. C
SDSS	Sloan Digital Sky Survey	Sect. 3
SED	Spectral Energy Distribution	Sect. 1
SPD	Spectral Photon Distribution	Sect. 2
SPSS	Spectro-Photometric Standard Stars	Sect. 3
RP	Red Photometer	Sect. 1
TDI	Time Delayed Integration	Sect. 4
WC(s)	Window Class or strategy	Sect. 4
WFE	WaveFront Error	Sect. 4.2