A&A, 714, A344 (2026)
https://doi.org/10.1051/0004-6361/202661086
© The Authors 2026

Astronomy & Astrophysics

Stellar masses and ages in *Gaia* Data Release 4 from the Final Luminosity Age Mass Estimator algorithm

O. L. Creevey[1,⋆], L. Casamiquela[2], Y. Lebreton[2,3], C. Ordenovic[1], F. Thévenin[1], M. Fouesneau[4], R. Andrae[4,18], C. A. L. Bailer-Jones[4], F. Pailler[5], B. Pichon[1], C. Babusiaux[6], A. Barbier[5], N. Baudeau[7], N. Brouillet[8], S. Cassisi[9,10], B. Edvardsson[11], O. Kochukhov[11], G. Kordopatis[1], A. C. Lanzafame[12,13], A. J. Korn[11], S. Meszaros[14,15,16], C. Navarrete[1,5], C. Robin[5], R. Sordo[17], C. Soubiran[8], and D. R. Reese[2]

[1] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France
[2] LIRA, Observatoire de Paris, Université PSL, Sorbonne Université, Université Paris Cité, CY Cergy Paris Université, CNRS, 92195 Meudon, France
[3] Université Rennes, CNRS, IPR (Institut de Physique de Rennes) – UMR 6251, 35000 Rennes, France
[4] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[5] Centre national d'études spatiales (CNES), 2 Place Maurice Quentin, 75039 Paris, France
[6] Université Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France
[7] Telespazio France for CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[8] Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[9] INAF-Osservatorio Astronomico d'Abruzzo, via M. Maggini sn., 64100 Teramo, Italy
[10] INFN, Sezione di Pisa, Largo Pontecorvo 3, 56127 Pisa, Italy
[11] Department of Physics and Astronomy, Uppsala University, Box 524, 751 20 Uppsala, Sweden
[12] Dipartimento di Fisica e Astronomia "Ettore Majorana", Università di Catania, Via S. Sofia 64, 95123, Catania, Italy
[13] INAF – Osservatorio Astrofisico di Catania, Via S. Sofia 78, 95123 Catania, Italy
[14] ELTE Eötvös Loránd University, Gothard Astrophysical Observatory, 9700 Szombathely, Szent Imre H. st. 112, Hungary
[15] MTA-ELTE Lendület "Momentum" Milky Way Research Group, Hungary
[16] HUN-REN CSFK, Konkoly Observatory, Konkoly Thege Miklós út 15–17, Budapest 1121, Hungary
[17] INAF-Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122 Padova, Italy
[18] Deutsches Zentrum für Astrophysik, Postplatz 1, 02826 Görlitz, Germany

Received 23 May 2026 / Accepted 1 August 2026

ABSTRACT

Context. The masses and ages of stars are key quantities for understanding exoplanetary, stellar, and galactic evolution. In the context of *Gaia*, these parameters provide insights into the stellar populations and help trace the formation and history of the Galaxy.
Aims. As part of the *Gaia* Data Processing and Analysis Consortium, the Final Luminosity Age Mass Estimator (FLAME) pipeline processes *Gaia* data to derive stellar parameters comprising luminosities, radii, masses, and ages. This paper discusses the methods and data used in FLAME for *Gaia* data releases and the expected performances of FLAME for the fourth *Gaia* data release.
Methods. Two main components comprise FLAME. The first component is analytical and is used to estimate luminosity, radius, and radial velocity correction due to gravitational redshift by exploiting the atmospheric, astrometric, and photometric parameters produced within *Gaia*. The second is a model inference based on two main approaches: a classical minimisation approach and a Bayesian framework. The aim is to derive the mass, age, and evolutionary stage. The two-step implementation offers flexibility in handling photometric properties that are prone to systematic errors.
Results. Tests with simulated data, the Sun, and well-characterised samples of stars including some clusters have shown that the methods in FLAME perform as expected, producing results in statistical agreement with the literature. The mass and age of giant stars are extremely sensitive to the input atmospheric parameters, and good agreement with external data for these stars is only reached when the same input data are used. This emphasises the difficulty in validating masses and ages of giants in any catalogue. We provide new stellar fundamental parameters for some high-velocity stars, stars with very low mass companions, and a selection of stars in the PLATO field of view. We also discuss the expected uncertainties and show that typical mass uncertainties are in relatively good agreement with analytical predictions from a mass–luminosity relation. Typical relative uncertainties on age for solar-like stars vary between 20% and 40%, and for masses between 1.2 and 1.3 $M_\odot$, they are around 20%. For giants, the relative uncertainties in age are typically 10% for lower masses, increasing to 15–20% for higher masses.
Conclusions. We conclude that FLAME produces valid results in *Gaia* data releases. In *Gaia* Data Release 4, approximately 500 million sources have results from the pipeline, thus providing large samples of stars with high precision and accuracy in their stellar parameters. As the results depend on the quality of the input data, samples much decreased in quality are also expected. Users should consult the *Gaia* online documentation and flags for guidelines regarding the exploitation of the catalogue.

Key words. methods: data analysis – catalogs – stars: fundamental parameters – stars: statistics – Galaxy: stellar content

⋆ Corresponding author: orlagh.creevey@oca.eu

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Subscribe to A&A to support open access publication.

1. Introduction

The age of a star is a fundamental parameter for understanding the formation and evolution of astrophysical systems. Reconstructing the history of our Milky Way, which has been shaped by the accretion of smaller galaxies (Searle & Zinn 1978; Helmi et al. 2018), requires determining the ages of stars belonging to its various populations (Helmi 2020; Gallart et al. 2019), as revealed by *Gaia* (Gaia Collaboration 2016a). Furthermore, for individual stars, estimating the age of a star that hosts an exoplanet is essential for constraining the evolutionary history of exoplanetary systems, a goal that will be further advanced by the upcoming PLATO mission (Rauer et al. 2025). Concerning the scientific return of the two space missions, stellar mass is a fundamental parameter, as it governs the evolution of a star over time.

The *Gaia* mission (Gaia Collaboration 2016b), launched in 2013 by the European Space Agency (ESA), has the aim of producing a comprehensive six-dimensional map of the Milky Way through measurement of the precise positions and velocities of approximately 1% of the Galaxy's stars (Gaia Collaboration 2016a, 2018, 2023b). In addition to mapping the Galaxy (Rowell et al. 2021; Torra et al. 2021), *Gaia* provides detailed physical characterisation of its observed sources using high-precision astrometry (Lindegren et al. 2021), photometry (Riello et al. 2021), and low- (De Angeli et al. 2023) and medium-resolution (Sartoretti et al. 2018, 2023; Katz et al. 2023) spectroscopy. The determination of stellar masses and ages is part of this characterisation effort carried out by the *Gaia* Data Processing and Analysis Consortium (DPAC), specifically through the Coordination Unit 8 (CU8 – Astrophysical Parameters) pipeline named Apsis (Bailer-Jones et al. 2013; Andrae et al. 2018; Creevey et al. 2023; Fouesneau et al. 2023; Delchambre et al. 2023).

The mass and age of stars are both derived by the module Flame within the Apsis workflow. The Flame module is designed to compute the radii, luminosities, radial velocity correction due to gravitational redshift, masses, ages, and evolutionary stages of all processed sources. In *Gaia* Data Release 3 (GDR3), June 2022, the Flame results are provided in the `astrophysical_parameters` and `astrophysical_parameters_supp` tables. The goal of this article is to describe the Flame module and its resulting products as well as to evaluate its performance in preparation for *Gaia* Data Release 4 (GDR4), December 2026.

The article is organised as follows. The architecture of the module and the methodology are described in Sect. 2 together with the models implemented in the analysis. The *Gaia* data used by Flame are presented in Sect. 3. The validation of the methodology based on simulated data and the Sun is discussed in Sect. 4, while Sect. 5 extends this validation by comparing Flame results with those from other catalogues. In Sect. 6, we apply Flame to infer the stellar properties – in particular masses and ages – for several samples of stars of interest. In Sect. 7, we finalise the work by discussing the expected uncertainties in Flame parameters in GDR4 before concluding in Sect. 8.

2. Method and models

The Flame code[1] (Creevey & Lebreton 2022; Ulla et al. 2022; Creevey et al. 2023) derives stellar evolutionary parameters and mass by exploiting atmospheric parameters, astrometry, photometry, and stellar models. It comprises two components: Algo1 which estimates stellar luminosity L, radius R, and radial velocity correction for gravitational redshift $V_{\rm GR}$, and Algo2 which infers model parameters using the derived L and complementary atmospheric parameters. For both components it exploits atmospheric parameters produced within Apsis (see Sect. 3). The choice to separate the Flame algorithm into an observational module and a model-based inference component is twofold. Firstly, it enables the encapsulation of all complexities of the photometric bands into a single parameter – the bolometric correction (BC). Secondly, it allows for a direct comparison of the derived physical properties of the stars with the true theoretical ones predicted by the stellar structure equations, namely L, $T_{\rm eff}$, and $\log g$. Additionally, this separation facilitates both technical and scientific validation of the method based on observational data obtained from other observatories, enabling direct comparisons between the resulting parameters.

2.1. Algo1

The first algorithm in Flame, Algo1, computes L, R, and $V_{\rm GR}$. The luminosity is calculated using the equation

$$M_{\rm bol} - M_{\rm bol,\odot} = -2.5 \log_{10}\left(\frac{L}{L_\odot}\right), \tag{1}$$

where

$$M_{\rm bol} = G - 5\log_{10} d + 5 - A_G + BC. \tag{2}$$

Here, G is the *Gaia* G-band magnitude, A_G is the extinction in the G-band, d is the distance in parsecs, and BC is the bolometric correction in G band, which depends on the atmospheric parameters $T_{\rm eff}$, $\log g$, [Fe/H], and $[\alpha/{\rm Fe}]$[2]. We adopt $M_{\rm bol,\odot}$ = 4.74 mag (IAU 2015 resolution; Mamajek et al. 2015). Within Flame, G is calculated from the observed flux f_G, its uncertainty, and the photometric zeropoint of the passband, allowing for accurate propagation of uncertainties.

Once L is determined, R is calculated from the Stefan-Boltzmann relation with the same $T_{\rm eff}$ used for the determination of the BC:

$$R/R_\odot = \frac{(L_\star/L_\odot)^{1/2}}{(T_{\rm eff}/T_{\rm eff\odot})^2}. \tag{3}$$

The radial velocity correction to the gravitational redshift is calculated from the input $\log g$ and the derived R,

$$V_{\rm GR} = \frac{GM}{Rc} = \frac{GM}{R^2}\frac{R}{c} = g\frac{R}{c}, \tag{4}$$

where c is the speed of light in vacuum.

Parameter values and their uncertainties are estimated using two methods: the first a bootstrap method, where the input observations $\{T_{\rm eff}, \log g, [{\rm Fe/H}], [\alpha/{\rm Fe}], f_G, d, A_G\}$ are perturbed N times to generate distributions of $\{L, R, V_{\rm GR}\}$. The second method uses the Markov chain Monte Carlo (MCMC) samples from the preceding module (GSP-Phot). The resulting median of each distribution defines the parameter value, while the 0.16 and 0.84 quantiles provide the lower and upper bounds of the 68% confidence interval.

[1] https://indico.ict.inaf.it/event/2345/

[2] [Fe/H] is the logarithm of iron-to-hydrogen abundance ratio relative to the Sun, and $[\alpha/{\rm Fe}]$ is the α-element-to-iron ratio.

Bolometric correction

The BC, in the G-band, is evaluated on a grid of stellar atmosphere models using the relevant passbands and zeropoints. For GDR3 the passbands[3] and the BCs[4] can be retrieved from the ESA cosmos website. The grids of models and methods used for calculating the BC within *Gaia*-DPAC APSIS modules are described in detail in Korn et al. (in prep.). Each model requires the input parameters $T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe] and returns BC values in the G, G_{BP}, and G_{RP} bands. For $T_{\rm eff}$ in the range 2500–8000 K, we use the MARCS models (Gustafsson et al. 2008), while for $T_{\rm eff}$ above 8000 K, Line-by-Line opacity stellar models (typically referred to as LL or A models) are adopted (Shulyak et al. 2004). The transition at 8000 K is handled by matching each set of models defined by [α/Fe], [Fe/H], $\log g$ in $T_{\rm eff}$, ensuring a smooth connection between the two grids.

The models include a zeropoint offset that must be corrected, effectively defining $\mathrm{BC}_{G,\odot}$. In GDR3, this was achieved by defining $M_{G,\odot}$ = 4.66 mag by using several independent methods (models, external datasets, and solar analogues), which then yielded $\mathrm{BC}_{G,\odot}$ = 0.08 mag (Creevey et al. 2023), with an estimated uncertainty of ±0.015 mag. In GDR4, these values are revised based on the GDR4 photometric passbands and will be published in the online documentation and APSIS papers.

2.2. ALGO2

The second component of FLAME performs model inference by comparing the observed stellar properties with their model-based counterparts. This allows the inference of the stellar mass (M), age (τ), and evolutionary stage (ϵ) where ϵ is a number typically between 100 and 1600, with values < 420 indicating main sequence, and > 490 from the base of the red giant; see Hidalgo et al. (2018) for details, or the data model here[5]. Unlike in GDR3, GDR4 additionally provides the corresponding model parameters for $T_{\rm eff}$, L, $\log g$, and [M/H], where [M/H] is the ratio of total metal to hydrogen abundance observed in the atmosphere.

2.2.1. Parameter inference

The model-based inference, ALGO2, comprises two methods. The results published in GDR4 are obtained by exploiting the method modelled on the Stellar Parameters INferred Systematically code (SPINS, Lebreton & Reese 2020), a publicly available Python pipeline[6]. The code employs a Bayesian framework to provide the posterior probability distribution function of the inferred stellar parameters from a set of observational constraints, a grid of stellar models, and a set of priors. The probability distribution function is sampled using a MCMC solver built on the emcee Python package (Foreman-Mackey et al. 2013), coupled with an interpolation scheme for the stellar models. FLAME uses the version of the SPINS code described in detail in Casamiquela et al. (2024), which includes several improvements, such as the on-the-fly computation of the autocorrelation time and the automatic convergence evaluation. SPINS allows the use of priors on the grid parameters, such as the initial mass function, the metallicity distribution function, or the star formation rate; see Lebreton & Reese (2020) for details. We used a flat prior on the age with range [0, 13.8] Ga and the initial mass function of Salpeter (1955). Further details on the priors for GDR4 are discussed in the online documentation. SPInS has been adapted in Java to be implemented in the DPAC/CU8 pipeline at CNES.

The observational input to the model-parameter inference comprises the set $\{L, \mathrm{[M/H]}, T_{\rm eff}, \log g\}$ and their uncertainties. The uncertainties are taken as half of the difference between the upper (84%) and lower (16%) percentiles for each input parameter. The results of the model-parameter inference are MCMC samples, and we use the median, 16$^{\rm th}$ and 84$^{\rm th}$ percentiles, after elimination of the burn in (50% of the total model calculations), to define the model property and confidence interval. The model properties that are published are M, τ, ϵ, and the associated model parameters (L, $T_{\rm eff}$, [M/H], $\log g$), along with their confidence intervals. We additionally publish the median of the log-posterior value, which can be used in conjunction with the model $T_{\rm eff}$, L, $\log g$, and [M/H], to judge the quality of the fit.

A second inference method was also implemented in FLAME and is based on a Levenberg–Marquardt approach (Creevey & Lebreton 2022). In GDR3 the model-based parameters were derived using this implementation, which provides similar results to SPINS when the dataset are of high quality. Additionally in GDR3 a narrow prior based on solar metallicity was implemented due to some known issues on the input metallicity, and empirical corrections to the masses for a non-solar metallicity star are available in Creevey & Lebreton (2022)[7].

2.2.2. Stellar models

The stellar models employed in this work are taken from the BaSTI[8] grid of stellar models (Hidalgo et al. 2018). These models are computed assuming a solar-scaled heavy-element distribution and include key input physics such as atomic diffusion of helium and metals, convective-core overshooting, and mass loss. The adopted solar mixture follows Caffau et al. (2011), complemented by the element abundances from Lodders (2010). Evolutionary tracks are available for initial stellar masses in the range $M_0 \in [0.1\,M_\odot, 15\,M_\odot]$ and metallicities [M/H] ∈ [−3.197, +0.45]. We additionally include models at [M/H] = +0.45 without diffusion, since this metallicity is absent in the BaSTI diffusive grid due to the lack of suitable radiative opacity tables. However, at a so high metallicity, the extension of the outer convection zone is so large, that the effect of atomic diffusion is hugely reduced – if any – with respect to more metal-poor stars. Moreover, only a small fraction of stars in our sample is affected, making this approximation quite reasonable.

The initial helium abundance follows a helium-to-metal enrichment ratio of $\Delta Y/\Delta Z$ = 1.31 (see Hidalgo et al. 2018, for details). For each stellar model defined by a given age, initial mass, and chemical composition, the grid provides the current stellar mass, evolutionary stage, luminosity, and effective temperature. From these quantities, the radius, surface gravity, and $V_{\rm GR}$ can be derived. The likelihood function used for model inference is based on the observables $\{L, T_{\rm eff}, \log g, \mathrm{[M/H]}\}$.

The BaSTI grid spans all evolutionary phases, from the pre-main sequence up to the first thermal pulses on the asymptotic

[3] https://www.cosmos.esa.int/web/gaia/dr3-passbands
[4] https://www.cosmos.esa.int/web/gaia/dr3-bolometric-correction-tool
[5] https://gea.esac.esa.int/archive/documentation/GDR3/Gaia_archive/chap_datamodel/sec_dm_astrophysical_parameter_tables/ssec_dm_astrophysical_parameters.html
[6] https://gitlab.obspm.fr/dreese/spins
[7] *Gaia* Public Documents are found https://www.cosmos.esa.int/web/gaia/public-dpac-documents
[8] http://basti-iac.oa-abruzzo.inaf.it

Table 1. Parameters published in the *Gaia* archive from FLAME.

Parameter	description	XP	RVS
L	luminosity	`lum_flame`	`lum_flame_spec`
R	radius	`radius_flame`	`radius_flame_spec`
V_{GR}	radial velocity correction due to the gravitational redshift	`gravredshift_flame`	`gravredshift_flame_spec`
BC	bolometric correction	`bc_flame`	`bc_flame_spec`
M	mass	`mass_flame`	`mass_flame_spec`
τ	age	`age_flame`	`age_flame_spec`
ϵ	evolution stage	`evolstage_flame`	`evolstage_flame_spec`
[M/H]†	model metallicity	`mh_flame_model`	`mh_flame_model_spec`
$\log g^\dagger$	model $\log g$	`logg_flame_model`	`logg_flame_model_spec`
$T_{\rm eff}{}^\dagger$	model $T_{\rm eff}$	`teff_flame_model`	`teff_flame_model_spec`
$L^\dagger$	model L	`lum_flame_model`	`lum_flame_model_spec`

Notes. Here, XP and RVS refer to the output parameters based on the input atmospheric parameters using XP and RVS spectra, respectively. In GDR3 the FLAME XP-based results are available in the `astrophysical_parameters` table, while the FLAME RVS-based results are found in the `astrophysical_parameters_supp` table. In GDR4 the table structure of the astrophysical parameters has been updated, and the FLAME results will be available in the `ap_xp` and `ap_rvs`, respectively. All parameters are accompanied by an upper and lower confidence interval. †Available from GDR4 onwards.

giant branch, the carbon ignition, or the age of the Universe, depending on stellar mass. For use in FLAME, we constructed a dedicated sub-grid by excluding high-mass models ($M > 10\,M_\odot$) and retaining only the evolutionary stages from the zero age main sequence to the tip of the red giant branch (RGB). This sub-grid comprises a total of 16 622 293 stellar models.

For most stars, the global metallicity derived in the *Gaia* analysis is from MARCS model atmospheres (Gustafsson et al. 2008). The MARCS grid adopts the solar abundance mixture of Grevesse et al. (2007), whereas the BaSTI models are based on the solar composition of Caffau et al. (2011), complemented by Lodders (2010). These two mixtures differ in their overall metallicity scale, with the Caffau et al. (2011) mixture being higher by 0.0976 dex relative to that of Grevesse et al. (2007). To ensure consistency between the observational inputs and the BaSTI model grid, we therefore applied a correction of this value to the estimated [M/H] values as input to ALGO2.

The metallicity [M/H] is the decimal logarithm of the ratio of the total number abundances of metals to hydrogen with respect to the Sun. If the assumption of a universal solar-scaled mixture is valid, then [M/H] ≈ [Fe/H]. However, when a star's mixture is enriched in α-elements (O, Ne, Mg, Si, S, Ar, Ca, Ti) with respect to the Sun, that is [α/Fe] ≠ 0, this approximation is no longer valid. The correct way to infer the parameters of such a star would then be to interpolate within several sets of grids of stellar evolutionary tracks based on different α-element enhancements. BaSTI grids are also available for an enrichment of [α/Fe] = +0.4. However, using it in FLAME would require [α/Fe] as an additional parameter in the SPInS optimisation, which would increase computation time. Salaris et al. (1993) showed that α-enhanced isochrones are well-mimicked by standard solar-scaled ones if metallicity is rescaled according to α-element enrichment. For the Caffau et al. (2011) solar mixture used in the BaSTI grids, we inferred a correction of the form

$$[\mathrm{M/H}]_\alpha \approx [\mathrm{Fe/H}] + 0.76[\alpha/\mathrm{Fe}], \quad (5)$$

or in quadratic form

$$[\mathrm{M/H}]_\alpha \approx [\mathrm{Fe/H}] + 0.208[\alpha/\mathrm{Fe}]^2 + 0.655[\alpha/\mathrm{Fe}]. \quad (6)$$

For the 34 million stars with non-zero [α/Fe], we corrected the metallicity using Eq. (6).

3. *Gaia* data

As part of the *Gaia*-DPAC system, FLAME can only use *Gaia* data and *Gaia*-derived parameters as input. The module produces two sets of results, one based on the atmospheric parameters derived from the APSIS module GSP-PHOT (Andrae et al. 2023), which analyses the *Gaia* BP and RP spectra (together known as XP) and are available for the vast majority of *Gaia* sources, and one based on the atmospheric parameters from the APSIS module GSP-SPEC (Recio-Blanco et al. 2023), which analyses the high resolution RVS spectra for a set of several million sources with a high enough signal-to-noise ratio (S/N) to derive reliable parameters. In GDR3 results for approximately 300 million sources with $G < 18.25$ mag were produced by processing the results from GSP-PHOT, and these are found in the `astrophysical_parameters` table, while results for approximately 6 million sources with $G < 13$ were produced by processing the input atmospheric parameters from GSP-SPEC, and these are found in the `astrophysical_parameters_supp` table. All results derived using FLAME contain `flame` in their parameter name, for example, `mass_flame` in the former table or `radius_flame_spec_lower` in the latter, where 'lower' means the 16th percentile value; see Table 1. For GDR4 we expect results on approximately 500 million sources based on the GSP-PHOT atmospheric parameters and 30 million sources based on the GSP-SPEC ones, and these will be available in tables called `ap_xp` and `ap_rvs`, respectively.

The processing by FLAME is done on a source-by-source basis. This implies that it has no knowledge if the object is a member of a binary system or a stellar cluster. Indeed it can also be a non-stellar source, and while we could publish a meaningless radius and luminosity for such non-stellar sources, it is not expected that the model parameters converge and may therefore be absent. As additional constraints cannot be placed on the processing of the sources, there are evidently parameter spaces that may not perform well, such as the ages of very low mass stars or

binaries with non-negligible secondary mass components. We chose to not remove these parameters from the archive, as they may in any case provide some important information to the user in GDR4. For examples of identifying such sources, the user should consult the online documentation. Examples of use cases through a Jupyter notebook will be made available with GDR4, and some examples using both GDR3 and GDR4 are given in Appendix B.

3.1. Parameters from GSP-PHOT

GSP-PHOT processes the low resolution XP spectra (De Angeli et al. 2023) to derive the atmospheric parameters ($T_{\rm eff}$, $\log g$, and [M/H]) and interstellar extinction in the G-band, A_G. The low-resolution spectra span approximately 300–1000 nm and have a typical resolving power of between 30 and 70 (De Angeli et al. 2023). For not very high S/N the $T_{\rm eff}$ and A_G can become degenerate, as they both produce similar changes in the spectral shape of the source[9]. GSP-PHOT also exploits the parallax and estimates a distance using priors, and this along with the amplitude of the XP spectra provides an independent estimate of the stellar radius. We note that GSP-PHOT employs stellar isochrones to help constrain the solution for the atmospheric parameters, and therefore a mass and age is associated with the optimal parameters. While this introduces some dependence on the priors adopted by GSP-PHOT, this dependence is not very different from one using parameters derived from empirical training, whose values will be restricted to the training sample. As FLAME is not only dependent on these atmospheric parameters, it is likely that the other constraints will result in large likelihoods and hence poorer logPosterior values. Full details of GSP-PHOT processing can be found in Andrae et al. (2023).

The APs from GSP-PHOT are produced by several stellar libraries (four libraries in GDR3, three in GDR4: MARCS, PHOENIX, A (LL), and OB, see Korn et al., in prep.), and while all of the results are available in the *Gaia* releases in the `astrophysical_parameters_supp` table, FLAME only processes the best model for each source. These are found in the `astrophysical_parameters` table and the best models are typically MARCS for FGK stars, and A for intermediate temperature stars (OB stars are outside of the FLAME validity range and are therefore not processed). In GDR4 the GSP-PHOT best and FLAME results based on XP spectra will be in a table called `ap_xp`.

GSP-PHOT is based on a MCMC approach. In ALGO1 FLAME processes the samples provided in the MCMC by GSP-PHOT. Table 2 summarises the input data to FLAME.

3.2. Parameters from GSP-SPEC

GSP-SPEC processes the high resolution RVS spectra to produce atmospheric parameters and chemical abundances. The RVS spectra span wavelengths in the range 840–870 nm and have a typical resolving power of $R \sim 11\,000$ (Sartoretti et al. 2018). As the RVS spectra are not sensitive to interstellar extinction and therefore do not provide this information, we use A_G provided by GSP-PHOT in Eq. (2), while we rely uniquely on the parallax as the distance estimator.

[9] As an example of the degeneracy between $T_{\rm eff}$ and A_G; see the illustration in Fig. 1 on the *Gaia* stories web page: https://www.cosmos.esa.int/web/gaia/dr3-what-is-in-between-the-stars

Table 2. *Gaia* data input into FLAME.

XP	RVS
d	ϖ
G	G
$T_{\rm eff\,XP}$	$T_{\rm eff\,RVS}$
$\log g_{\rm XP}$	$\log g_{\rm RVS}$
$[{\rm M/H}]_{\rm XP}$	$[{\rm M/H}]_{\rm RVS}$
–	$[\alpha/{\rm Fe}]_{\rm RVS}$
$A_{G\,\rm XP}$	$A_{G\,\rm XP}$

Notes. Here, XP refers to the parameters provided by GSP-PHOT, which processes the Gaia XP spectra, while RVS refers to those provided by GSP-SPEC, which processes the Gaia RVS spectra.

Two sets of results from GSP-SPEC are available in the *Gaia* archive: one based on the Matisse-Gauguin algorithm (Bijaoui et al. 2010; Recio-Blanco et al. 2016, 2023) and a second one based on a neural network. FLAME processes the results produced by the Matisse-Gauguin algorithm which are found in the `astrophysical_parameters` table in GDR3, but the FLAME results are made available in the `astrophysical_parameters_supp` table. For GDR4 the GSP-SPEC and RVS-based FLAME results are available in a table called `ap_rvs`.

3.3. Parameters from astrometry and photometry

In addition to the atmospheric parameters, A_G, and d provided by the APSIS modules, FLAME also uses the parallax, ϖ, from the astrometric solution and the G-band flux from the photometric-processing system. Within APSIS, the parallax bias (Lindegren et al. 2021) is also applied.

4. Validation of the methodology

In this section we focus on validating the methodology used in FLAME. While ALGO1 is uniquely concerned with the application of analytical equations, our attention here is on the performance of ALGO2 which handles model-based parameter inference. To achieve this, we assume an input observational dataset consisting of $\{L,\ T_{\rm eff},\ \log g,\ [{\rm M/H}]\}$, with various expected errors. We then apply FLAME ALGO2 to recover the stellar model parameters (M, τ, ϵ), along with the corresponding model properties (L, $T_{\rm eff}$, $\log g$, [M/H]).

4.1. Simulated data from the stellar grid

We began with tests using simulated data from the model grid that is used for the inference. This allowed us to perform a direct comparison of the input-output. We selected 372 stars across the Hertzsprung-Russell (HR) diagram, ensuring coverage along the main sequence, sub-giant phase, and RGB, at all metallicities. A first set was selected using points directly on the model grid, and then a second set was selected using data points between model grid points. Figure 1 shows the distribution of the simulated dataset (similar for both cases) across the HR diagram, colour-coded by [M/H], with a few solar-metallicity evolution tracks highlighted in the background. By simulated dataset we imply that we perturbed the observables using a random Gaussian distribution scaled by uncertainties varying between 1–5% errors on $T_{\rm eff}$ and L, while the errors in $\log g$ and [M/H] were scaled by 0.1 dex. Here we focus on the set of results corresponding to the 2% errors in $T_{\rm eff}$ and L because it is a typical

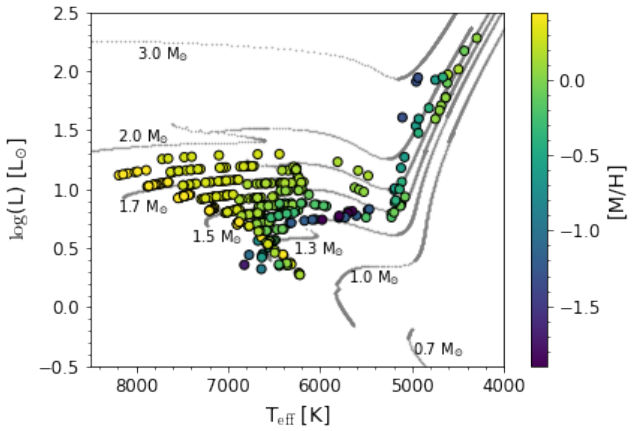


Fig. 1. Hertzsprung–Russell diagram showing the sample of simulated data used for validating the methodology. The background grey tracks are solar-metallicity evolution tracks for different initial mass values.

uncertainty on the input parameters for bright Gaia sources (e.g. $G < 14$).

The results of the analysis are shown in Fig. 2. In the top panel we compare the input and retrieved masses, while the bottom panels compare the ages. The top left compares the input mass, $M_{\rm input}$, with $M_{\rm FLAME}$ colour-coded by $\log g$. The identity line is also shown to help guide the eye and it can be seen that FLAME retrieves the expected results, with no particular offsets or trends with other parameters. We also show linear fits with black lines. In the right panel, we show the histogram of the differences in mass but scaled by the FLAME uncertainties. As expected, the distribution peaks around zero, with roughly 68% of the values lying within $\pm 1\sigma$ (0.005 ± 0.045 M$_\odot$), where σ is the median absolute deviation (MAD).

The lower panel shows the same for the age, this time colour-coded by M. Again no significant offsets or trends are visible, demonstrating that FLAME successfully retrieves the ages. The corresponding histogram shows the differences scaled by their uncertainties. The mean offset is 0.0 Ga with a MAD of 0.3 Ga.

4.2. Validation using solar parameters and impact of observational errors

The second set of tests used to validate our method focuses on the star we know best: the Sun. Our aim here is to assess the accuracy of both the models and the method while also investigating the impact of each observational uncertainty on the resulting stellar parameters.

We simulated datasets representing the Sun by considering various scenarios of data quality. For each scenario we made 500 simulated datasets. The best-case scenario is denoted as Σ_0, which assumes a 1% error in L, 60 K in $T_{\rm eff}$, 0.05 dex in $\log g$ and [M/H]. We then reduced the precision in each of the observables one by one or in sets of two, as shown in the upper half of Table 3. The left column is the name of the dataset, and the subscript to Σ indicates which parameter uncertainty is increased; for example, Σ_{LT} indicates that the uncertainties in both L and $T_{\rm eff}$ were increased.

The median M and τ of the set of 500 simulations along with the median of their uncertainties are presented in Table 3 in the last two columns. We kept only a precision to two decimal places. Here one can easily see for high precision measurements that the mass is retrieved accurately within the uncertainty. The age is also retrieved correctly within the uncertainty. The largest impact happens when we decrease the precision in all measurements, i.e. from Σ_0 to Σ_{LTGM} and Σ_{LGTM2}, and while the actual parameters do not change much (because there are no systematic errors), it can be seen that the uncertainties increase as expected.

Table 3. Impact of observational and systematic errors in L, $T_{\rm eff}$, $\log g$, and [M/H] on the retrieval of the solar mass and age.

	σ_L (%)	$\sigma_{T_{\rm eff}}$ (K)	$\sigma_{\log g}$ (dex)	$\sigma_{\rm [M/H]}$ (dex)	$M \pm \sigma_M$ (M$_\odot$)	$\tau \pm \sigma_\tau$ (Ga)
Σ_0	1	60	0.05	0.05	0.99 ± 0.05	4.9 ± 3.1
Σ_L	3	60	0.05	0.05	0.99 ± 0.04	4.9 ± 3.1
Σ_T	1	120	0.05	0.05	0.99 ± 0.04	5.1 ± 3.1
Σ_{LT}	3	120	0.05	0.05	0.99 ± 0.04	5.1 ± 3.1
Σ_{GM}	1	60	0.15	0.15	0.98 ± 0.03	5.9 ± 4.3
Σ_{LTGM}	3	120	0.15	0.15	0.98 ± 0.06	6.1 ± 4.3
Σ_{LTGM2}	10	250	0.25	0.25	0.97 ± 0.06	6.0 ± 4.4
$\Sigma_{0,\rm Lp}$	+5				1.01 ± 0.04	4.3 ± 2.9
$\Sigma_{0,\rm Lm}$	–5				0.98 ± 0.04	5.7 ± 3.2
$\Sigma_{0,\rm Tp}$		+80			0.99 ± 0.04	5.0 ± 3.1
$\Sigma_{0,\rm Tm}$		–80			0.99 ± 0.05	5.1 ± 3.1
$\Sigma_{0,\rm Gp}$				+0.1	1.04 ± 0.03	1.7 ± 1.7
$\Sigma_{0,\rm Gm}$				–0.1	0.93 ± 0.03	10.5 ± 2.6
$\Sigma_{0,\rm Mp}$			+0.1		1.01 ± 0.04	4.8 ± 3.1
$\Sigma_{0,\rm Mm}$			–0.1		0.97 ± 0.04	5.4 ± 3.1
$\Sigma_{LTGM,\rm Gp}$				+0.1	0.99 ± 0.06	4.8 ± 4.2
$\Sigma_{LTGM,\rm Gm}$				–0.1	0.96 ± 0.06	7.4 ± 4.3

4.3. Impact of systematic errors on the solar parameters

We continued our sequence of validation tests by investigating how systematic errors in the observational data affect the inferred model properties. We simulated datasets in the same manner as above by modifying one observational property at a time, specifically 5%, 80 K, 0.1 dex, and 0.1 dex for L, $T_{\rm eff}$, $\log g$, and [M/H]. We took as reference a set of Σ_0 but while acknowledging that the systematic errors will always be larger than the uncertainty.

The results of these tests are also given in Table 3 in the lower half of the table. The subscripts L, T, G, M, denote L, $T_{\rm eff}$, $\log g$, and [M/H], respectively, with 'p' and 'm' being a 'plus' or 'minus' systematic error by the amount specified in the table. As shown by the resulting parameters, the largest impact comes from the systematic errors in $\log g$ where both M and τ no longer agree with the solar parameters. The next parameter with most impact is [M/H] which changes primarily the mass.

To clarify if such a systematic error had the same impact in the case of worse precision parameters, we repeated the exercise using Σ_{LTGM} for the systematic in $\log g$. We observed that with more freedom to explore the parameter space due to the lower precision, the systematic does not play such an important role.

The analysis described here was restricted to some very specific test cases for a solar-like star. It is not possible to extrapolate the results to other parameter spaces or another combination of uncertainties and systematics. To do so, a full-scale analysis would be needed, and this is beyond the scope of the paper.

5. Comparison with external datasets

To further validate our mass and age inference, we apply FLAME to a set of existing catalogues and compare our results. We begin

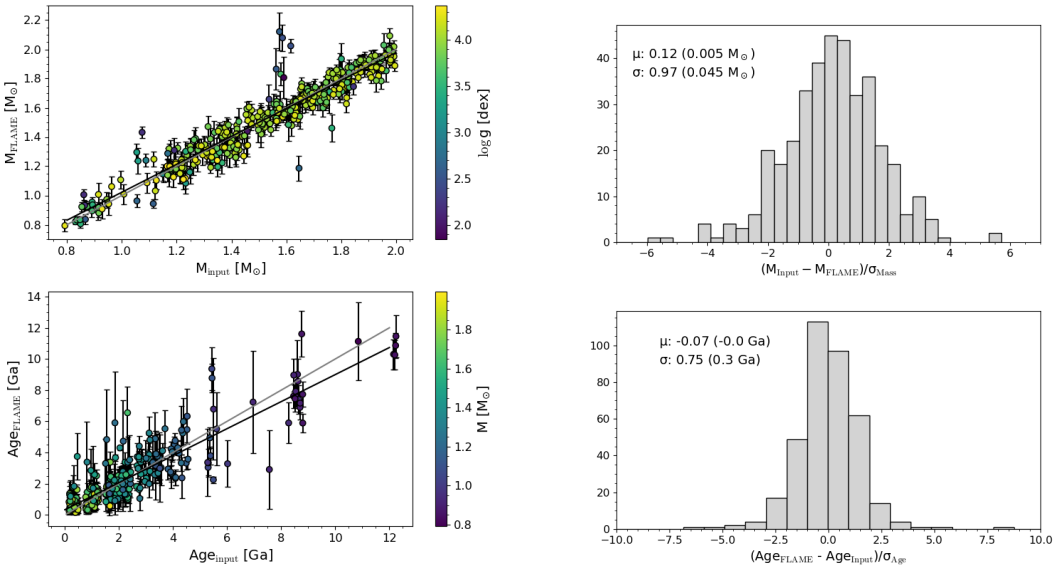


Fig. 2. Comparison between the true input M (*top*) and age (*bottom*) with the output from FLAME. The left panels illustrate a direct one-to-one comparison, colour-coded by $\log g$ for mass and M for age, with the bisector in grey to guide the eye and a linear fit in black. The right panels show histograms of the differences scaled by the uncertainties.

with a comparison of FLAME results with the *Gaia* benchmark stars (GBSs; Heiter et al. 2015; Soubiran et al. 2024). We then compare the mass and age with results from catalogues that exploit asteroseismology. Finally we illustrate a comparison with larger catalogue datasets using pre-operational GDR4 data.

5.1. Validation test with the Gaia benchmark stars

The GBSs are stars for which their $T_{\rm eff}$ and $\log g$ have been derived through fundamental relations, in particular using interferometry for measuring angular diameters. The primary goal of the GBS is to serve as benchmarks for testing stellar physics in both atmospheres and interiors (e.g. Collet et al. 2009; Creevey et al. 2015, 2024). In addition they help calibrate stellar spectroscopic surveys, in particular *Gaia*, and allow one to derive high precision metallicities and abundances (Jofré et al. 2014, 2015, 2017; Casamiquela et al. 2026). They constitute a valuable reference for testing performances of pipelines, validating large-scale catalogues of stellar $T_{\rm eff}$, radii, and masses, as well as for calibrating stellar system such as globular clusters (e.g. VandenBerg et al. 2014).

The most recent release of the GBSs, version 3 (v3; Soubiran et al. 2024), comprises over 200 stars of which 190 have also been observed by *Gaia* and are present in GDR3. In addition to the fundamental parameters $T_{\rm eff}$ and $\log g$, the catalogue provides two sets of stellar masses, derived using different methodologies based on distinct stellar evolution models – the BASTI and STAREVOL (Lagarde et al. 2012, 2017) evolutionary tracks. That paper focusses on the fundamental observed properties, and therefore stellar ages have not been published.

As our stellar models extend only to the tip of the RGB, we selected approximately 140 stars that are unlikely to be red clump stars. The published values of L, $T_{\rm eff}$, $\log g$, [Fe/H], together with their associated uncertainties, were then used as input parameters for processing with FLAME.

We first compared our derived radii with those from the reference study to verify consistency. Given that the methodology involves a degree of randomness, exact agreement is not expected. A comparison of the radii relative to their uncertainties yields a median difference of -0.08σ, while the median relative radius difference is -0.002σ, where $\sigma = 0.003$ R$_\odot$. These results indicate that our derived radii are in high agreement with the published values. The good agreement is not surprising as we use similar stellar models and some of the same input data.

We compared our derived masses with their masses from the BASTI evolution tracks ('MassB' from their catalogue) in Fig. 3. In the top panel we show a mass-mass plot focussing on the region with $M < 2.2$ M$_\odot$ (135 stars). An identity line is also drawn to help guide the eye. The colour code indicates the surface gravity and thus to a large extent the evolution status of the star. Most of the main-sequence stars (brighter colours) fall almost directly on the bisector, indicating strong agreement with their results. A subset of the very evolved stars show an offset of the order of 0.1–0.3 M$_\odot$; however, the offset is insignificant when considering the uncertainties. The lower panel displays a histogram of the mass differences scaled by the quadrature sum of their uncertainties, and it can be seen that most of the masses are within 2σ, indicating a very satisfactory agreement. The few outliers at more than 4σ are for very evolved stars and these differences could be due to the fact that we use different input constraints – they use $T_{\rm eff}$, [Fe/H], L, and R, and in this region of the HR diagram there is an important degeneracy between metallicity and mass. In Table C.1 we list the FLAME mass and age of the 140 GBS stars analysed here.

5.2. Comparison with asteroseismic data

Asteroseismology (e.g. Ulrich 1997; Aerts 2021) has the capacity to infer very accurate and precise stellar properties (surface gravity, radius, mass, and age; see e.g. Lebreton & Goupil (2014); Metcalfe et al. (2014); Hekker & Ball (2014); Creevey et al. (2017); Lund et al. (2025)) because the observed frequencies probe the inner structure of the star. Asteroseismic samples therefore provide an excellent dataset for which to validate our methodology. We focus on two main dataset comparisons: a set of main-sequence and sub-giant stars observed by both Kepler (Borucki 2016) and the Apache Point Observatory Galactic Evolution Experiment (APOGEE, Gunn et al. 2006; Majewski et al. 2017), as part of what is known as the APOKASC sample (Pinsonneault et al. 2014). We use the results from

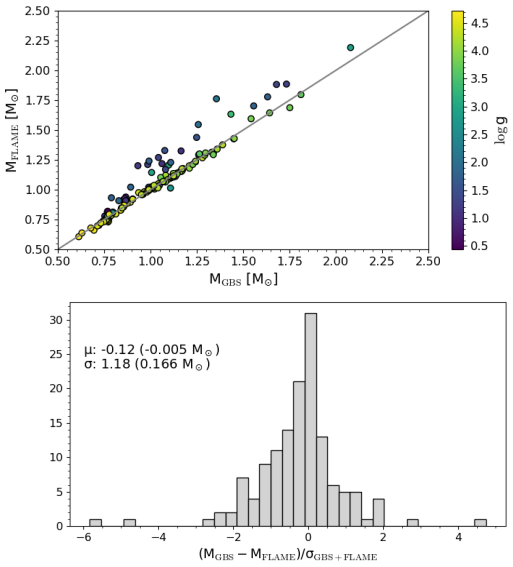


Fig. 3. Comparison of mass with the GBSs. The grey line in the top panel is a bisector to guide the eye.

Serenelli et al. (2017) who exploit the so-called global seismic quantities for a sample of approximately 450 stars. The global seismic quantities are the mean large frequency separation $\langle\Delta\nu\rangle$ and the frequency of maximum power ν_{max} which are sensitive to the global parameters of mass and radius. Then we used the results from Pinsonneault et al. (2018), who provide results on a sample of approximately 6000 giants. For these two datasets, we performed a new inference on the masses and ages and compared them with the seismic analysis.

For the two datasets, we retrieved the GSP-Spec atmospheric parameters from GDR3 and applied the published calibrations, along with the Flame luminosities. For the giants, we retrieved only the sources that correspond to an RGB status because Flame (for GDR3 and GDR4) exploits models that span from the zero age main sequence to the tip of the RGB only. We only include those sources that have all four input parameters ($T_{\rm eff}$, log g, [M/H], L).

We illustrate the comparison between the resulting Flame mass and the asteroseismic mass in the left panels of Fig. 4 by the background grey points. Top and bottom panels correspond to Serenelli et al. (2017) and Pinsonneault et al. (2018), respectively. For the non-evolved stars, there is a relatively good agreement between the two masses with a median offset of –3% but a large scatter giving a MAD of 11%. For the giants, the median offset in relative mass is –7% and the MAD is 13%. However, a linear trend with mass is also present, which gives an unsatisfactory agreement. This disagreement is not so surprising for stars in this evolution stage. All stellar evolution tracks of giants of different masses and metallicities occupy a very narrow region in the HR diagram, with the result that even minor changes in the input parameters can change the inferred masses and ages significantly. Similar trends were also seen for the mass and ages of giants published in GDR3. In this comparison case, it is likely that the Flame results are incorrect, because we expect asteroseismology to well constrain the solution. However, in the majority of comparisons with methods using similar input constraints as Flame it is not possible to know which results are the more accurate, and thus it is difficult to validate masses and ages of giant stars.

To fully validate Flame, we therefore made a new dataset comprising the same input $T_{\rm eff}$, [Fe/H], and log g reported in the asteroseismic catalogues. For the Pinsonneault et al. (2018) catalogue, we use the parameter 'logg_seis' which is the surface gravity derived from the asteroseismic analysis. As it is usually very well determined, we added 0.05 dex in quadrature to the errors given in that paper for our test in order to better reproduce typical uncertainties expected from atmospheric parameter analyses.

In addition to these atmospheric parameters, we calculate new luminosities. We exploit the GDR3 G mag BC tool in order to derive new BCs using the input catalogue atmospheric parameters. We apply the parallax bias correction to the GDR3 parallax (Lindegren et al. 2021)[10] and use the corrected parallax to estimate the distance. For the Serenelli et al. (2017) sample, as no extinction value is available in their catalogue, and `ag_gspphot` is not available for the full sample, we exploit the extinction map from Vergely et al. (2022) to derive extinctions for the full sample. We then use the tool made available through *Gaia*[11] which is based on the work of Danielski et al. (2018) to convert A_V to A_G. For the APOKASC sample, we use their published value of extinction in the V band A_V and convert it to A_G using an approximate conversion constant (the stars are all of similar log g, $T_{\rm eff}$, and extinction so the extinction coefficient will be very similar for the subset). With these new datasets, we then re-ran Flame for all of the stars in the two catalogues.

The resulting comparison of the masses is shown in Fig. 4 in the left panels, but this time represented by the coloured points, with the colour-code indicating log g. It is immediately clear that the trend with mass has now statistically disappeared. In addition the median offset between the relative masses is at $< +0.01\%$ with a MAD of 3% for the non-evolved stars, while it is 2% with a MAD of 7% for the giants, indicating very compatible results with the asteroseismic results.

The right panels in Fig. 4 show direct comparisons between the Flame age and the asteroseismic age. The results are in agreement and follow the bisector. The points deviating most from the bisector are also those points whose mass deviates most from the seismic mass.

5.3. Flame *GDR4 performance*

In this subsection, we present some results that are indicative of the GDR4 performance of Flame. We use data from a pre-operational GDR4 run to compare Flame masses and ages to those from a large external catalogue and to examine the distribution of ages for a few open clusters.

5.3.1. StarHorse comparison

As a further validation test we exploit the StarHorse catalogue (Queiroz et al. 2023) which produces stellar parameters for around 10 million sources which are in common between Gaia and other spectroscopic surveys. They additionally derive stellar ages for approximately 2.5 million main-sequence and sub-giant stars by exploiting the PARSEC isochrones (Bressan et al. 2012) and GDR3 data. This catalogue therefore provides a useful test set for Flame covering a large range of masses, ages, $T_{\rm eff}$, luminosities, and [M/H]. For this specific comparison test we use the GDR4 GSP-Phot data as input and process the data using Flame. As shown in the previous section, with different input data we expect to have different results. This particular set then

[10] https://gitlab.com/icc-ub/public/gaiadr3_zeropoint
[11] https://www.cosmos.esa.int/web/gaia/edr3-extinction-law

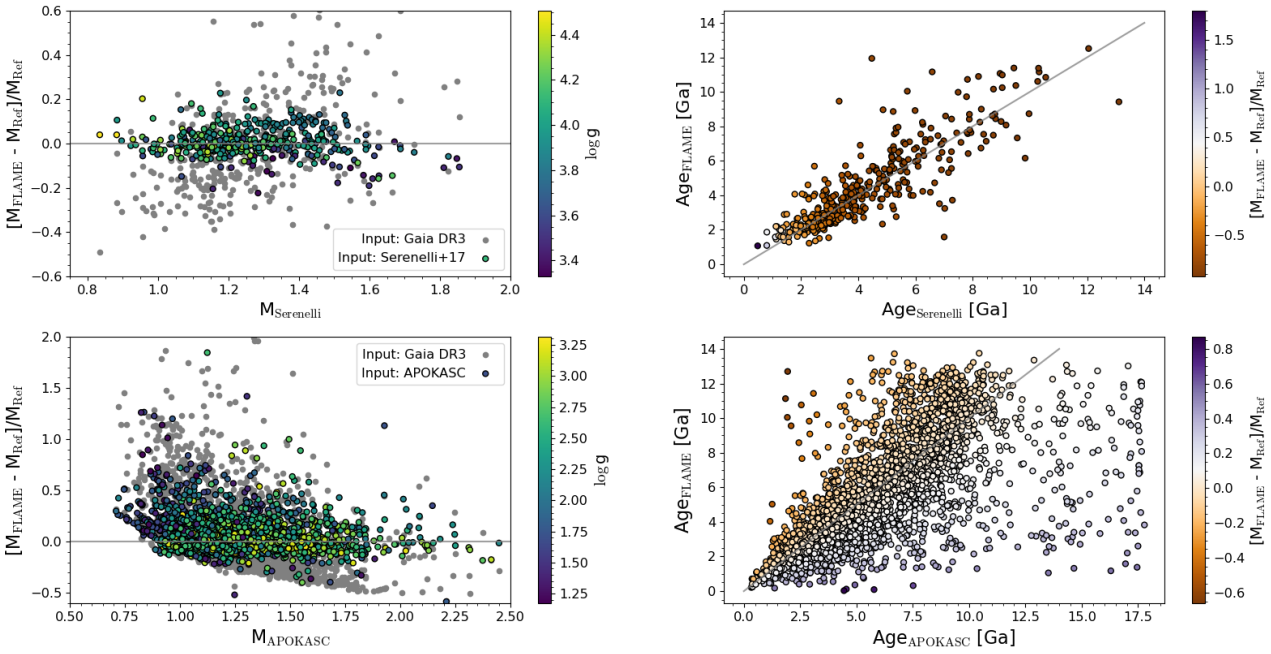


Fig. 4. *Left panels*: relative difference between FLAME mass and asteroseismic mass. The grey symbols show the results when we rerun FLAME using GDR3 atmospheric parameters for these sources, while the colour-coded symbols show the comparison when we use the atmospheric parameters from the reference catalogue as input to FLAME. *Right panels*: comparison of FLAME age and asteroseismic age when using the atmospheric parameters from the reference catalogue as input.

allowed us to provide an overall comparison of how the FLAME masses and ages in GDR4 may compare with other catalogues.

Several test runs on GDR4 data were performed during the development and pre-operational test phase. The aim of these tests are to validate the scientific and technical parts of the processing and post-processing code. They use as input a list of several million sources, comprising lists of known objects in catalogues that help to validate FLAME and the other APSIS modules, and a list of randomly chosen sources to cover a large range of magnitude, parallax, sky coverage, and precision in these parameters. A subset of the StarHorse catalogue is included in this test set. Matching with the output from FLAME we obtained a total of 381 254 stars in common all having $\log g > 3.2$.

To show a meaningful comparison, we selected those stars with input data that were in broad agreement with the GDR4 input data. Specifically we required that the StarHorse $\log g$ and the GDR4 input $\log g$ agreed within 0.5 dex, the StarHorse [M/H] and the GDR4 input [M/H] agreed to 0.5 dex, and the StarHorse $T_{\rm eff}$ agreed with the GDR4 $T_{\rm eff}$ to within 400 K. This resulted in a total of 344 767 stars or 90% of the original sample. Figure 5 shows the comparison between the two datasets for mass and age.

In the left panel, we show the mass comparison for the 90% sample. There is good overall agreement. The median offset in mass over the full range is 3%, with the FLAME masses being slightly higher, and the MAD is 5%.

To make a reasonable age comparison, we further restricted the sample so that the StarHorse mass matches the FLAME mass to within 10%. This results in a total of 285 325 stars or 75% of the original sample. We show the comparison of the ages for this subsample in the right panel of Fig. 5, colour-coded by the StarHorse $\log g$. We again observed a very good agreement between the ages with a median offset of –0.16 Ga and a MAD of 0.75 Ga, where the scatter is caused by the mass differences. If we relax the constraint on the agreement in masses we find a mean offset of –0.1 Ga with a MAD of 1.0 Ga. Overplotted in orange is a random sample consensus linear fit with slope 0.8667 and intercept –0.3934.

5.3.2. Clusters

As a final validation test we investigate the distribution of ages from a few well-known open clusters. We use the selection of cluster members from Casamiquela et al. (2024) who selected high probability main-sequence and the sub-giant members with $M_G < 4.5$; and as input data we again use the pre-operational GDR4 test run data as in Sect. 5.3.1. We crossmatched the list of source IDs with the FLAME dataset for the following clusters which represent a young, intermediate, and older cluster: NGC 3532, NGC 2682, and NGC 188.

We show the distribution of the ages of each of the clusters in Fig. 6 as the histogram, along with a kernel-density estimate (kde). The age of the cluster according to the literature is given in the label in each panel along with the median age value derived by FLAME, and it can be seen that the FLAME ages peak very close to the literature age, and only some outliers are seen. The kde maxima only vary slightly from the median age (6.2, 4.3, and 1.5 Ga, respectively). For the youngest cluster, there is the largest dispersion. The outliers in each of the panels are often stars with low masses for which ages are difficult to derive, or stars for which the input parameters may not be in agreement with the expected values. We remind readers that within the APSIS processing, all stars are processed on a star-by-star basis without any knowledge of it being associated with a cluster, so the input metallicities and extinction will be different for every star.

6. Application of FLAME to datasets of high astrophysical interest

Having validated the methodology (Sect. 4) and compared results with some well-established external datasets (Sect. 5),

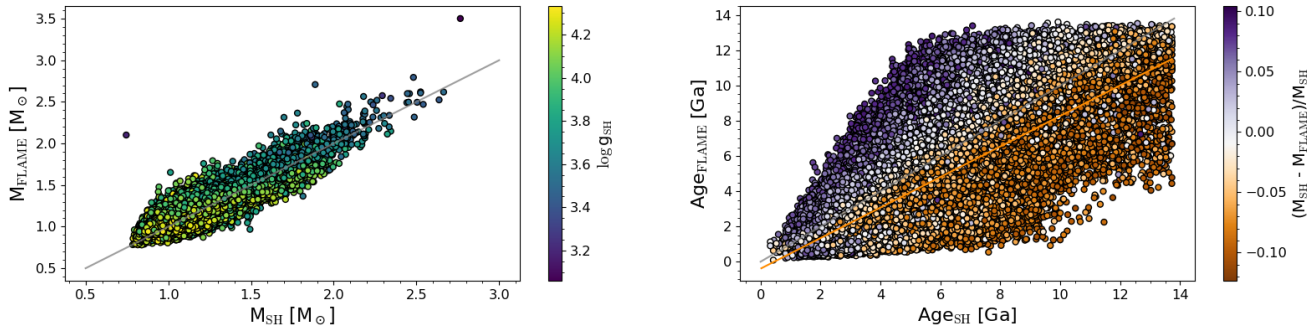


Fig. 5. Comparison of the masses and ages from FLAME using pre-operational GDR4 data as input and the results from the StarHorse catalogue. *Left*: comparison of masses after imposing some constraints on the input data (see text; 344 767 stars, or 90% of the original samples). *Right*: comparison of ages after imposing the same constraints and restricting the stars to those with masses that agree to 10%.

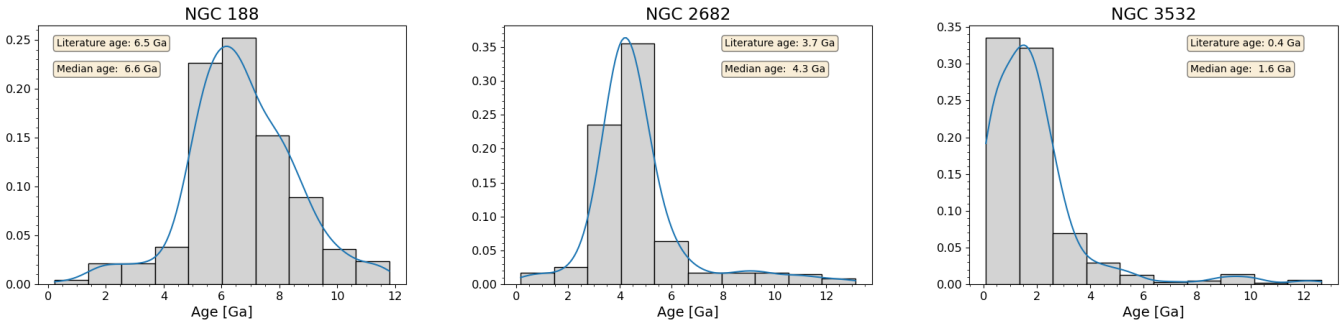


Fig. 6. Distribution of ages for three clusters as derived by FLAME using a dataset representative of GDR4 data.

we then used FLAME to derive new masses and ages for various astrophysically interesting targets. We examine a sample of high-velocity stars, stars with very low mass companions, and a subsample of targets from the science calibration and validation targets of the PLATO Input Catalogue.

6.1. High-velocity stars

High-velocity stars (Katz et al. 2025) likely originate in different regions of the Galaxy and travel through diverse stellar environments. They are most commonly thought to be ejected by interactions with the supermassive black hole at the Galactic centre, though alternatives such as stellar explosions or cluster dynamics have been proposed. Determining their masses and ages can provide key insights into their origins. With this in mind, we aim to derive the ages of the sample of high-velocity and metal-poor stars presented in Katz et al. (2025).

As input data to FLAME we use the $T_{\rm eff}$, $\log g$, [Fe/H], and [α/Fe] as given in Katz et al. 2025. As there are no observational errors published with the $T_{\rm eff}$ and $\log g$, we adopt a classical 100 K and 0.1 dex as an error for all of the stars. We use Eq. (6) to calculate [M/H] from [Fe/H] and [α/Fe], and we adopt the published error on [Fe/H] as the error on [M/H]. We use these spectroscopic parameters to derive the BC.

To derive the luminosities, we extracted complementary data from GDR3; we use the astrometric information and derive the parallax bias to correct the published parallaxes. In addition, as only 35 of the 51 stars have published extinction values, we follow the procedure mentioned in Sect. 5.2 to derive new extinction values. The resulting extinction values are broadly in agreement with the existing A_G in GDR3.

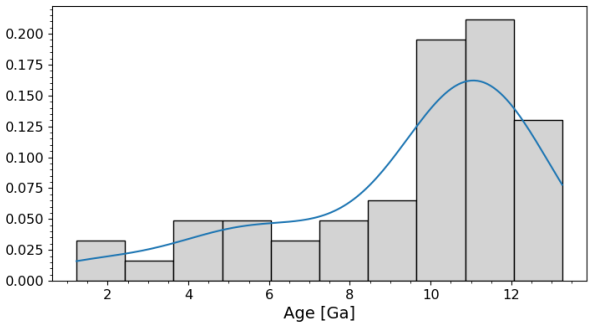


Fig. 7. Distribution of ages of high-velocity stars from Katz et al. (2025) derived with FLAME.

Using the corrected parallax, BC, A_G, and GDR3 G, we calculate the luminosities using Eqs. (1) and (2). The input dataset to FLAME therefore comprises {L, $T_{\rm eff}$, [M/H], $\log g$}, and we derive their masses and ages for the full sample.

We show the distributions of the ages of the high-velocity sample in Fig. 7 and we publish their luminosities, radii, masses, and ages in Table E.1. The distribution of the ages peak around 11 Ga, corresponding to the era of the halo phase of the Milky Way. It is not unexpected to find some younger stars in the sample because we do not know the exact origin of their star formation.

6.2. Solar-like stars with very low mass companions

Low-mass stars burn nuclear fuel extremely slowly, and can live for tens to hundreds of billions of years. Their observable surface properties evolve only minimally over time, and therefore

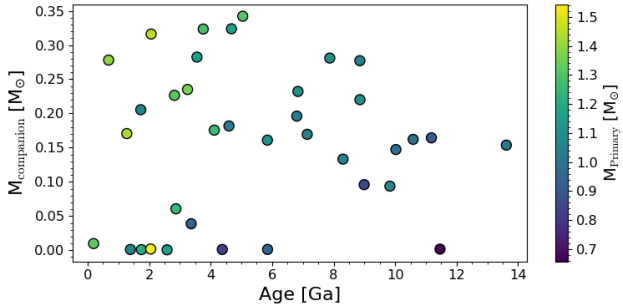


Fig. 8. Companion mass as a function of stellar age for the sample of systems with low-mass companions in Maxted (2025).

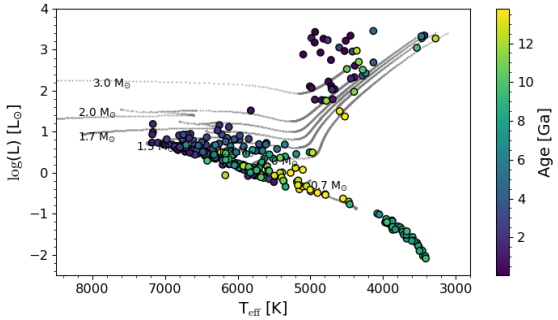


Fig. 9. Hertzsprung–Russell diagram of a subset of the benchmark scvPIC stars, colour-coded by age.

their masses and ages remain poorly constrained. In particular, if we consider objects on the boundary of sub-stellar and stellar objects, knowing their masses and ages can lead to a better understanding of their formation and evolution.

When such objects are members of binary systems with more massive primaries, the primary stars can often be characterised much more accurately using independent techniques such as spectroscopy, stellar modelling, and activity or rotation diagnostics. By deriving reliable ages and masses for these primary stars, we can indirectly constrain the physical properties of their low-mass companions.

With this in mind, we consider the sample of 36 stars with low-mass companions studied by Maxted (2025); of which eight are in fact transiting exoplanets. The low-mass companions contribute less than 0.1% of the total flux in the V band. This means that we can use assumptions of an isolated star to make a meaningful determination of the properties of the primary component.

To construct our dataset of $\{L, T_{\rm eff}, \log g, [{\rm Fe/H}]\}$ we follow the same approach as described in Sect. 6.1. We adopt the published $T_{\rm eff}$, $\log g$, and [Fe/H] from Maxted (2025), and we calculate the primary star's luminosity using G, A_G, BC, and a corrected parallax (Eqs. (1) and (2)). We then process this dataset with FLAME.

The resulting masses and ages of these systems are given in Table F.1 where we use the published mass ratio (q) as given by Maxted (2025) to calculate the companion masses. In Fig. 8 we also illustrate the mass of the companion star as a function of the stellar age. We note that q has a very small dependence on the mass estimate, but as our mass estimates are in close agreement with those of Maxted (2025) (a median relative difference of 2% and MAD of 3%), q should not change significantly.

6.3. PLATO input catalogue

The ESA mission PLATO (Rauer et al. 2025) is currently planned to launch in January 2027. Its objective is to search for and characterise exoplanets, in particular those similar to Earth. Unlike *Gaia* which scans the full sky, PLATO will stare at a fixed field of view for at least two full years, allowing the collection of very long and high precision datasets capable of detecting Earth-like planets around Sun-like stars. As part of the preparation of the PLATO data validation, a selection of stars known as the science, calibration, and validation (scv) PLATO Input Catalogue (PIC) stars has been made to help validate the PLATO data and pipelines. A subset of the scvPIC has been prepared by the PLATO Benchmark Stars work package (Merle et al., submitted) and is presented in Zwintz et al. (2026).

About 200 of the targets in the scvPIC from the PLATO benchmark stars comprise high quality samples of data from GDR3 known as the golden sample of astrophysical parameters (Gaia Collaboration 2023a), members of open clusters, and targets that can be measured with interferometry. We refer readers to Zwintz et al. (2026) for details on the target selection and the description of the targets that made it to the scvPIC. In this section we aim to provide new masses and ages of this dataset by exploiting the astrophysical parameters from GDR3, as produced by GSP-PHOT and FLAME. We use as input data the L, $T_{\rm eff}$, [M/H], and $\log g$, as given directly in the *Gaia* archive, and we process these datasets with the GDR4 version of FLAME.

The new masses and ages of these stars are given in Table D.1, and in Fig. 9 we illustrate the HR diagram colour-coded by the resulting age. These homogeneous parameters could be used for validating some of the first stellar parameters from PLATO.

7. Specifics of *Gaia* data

Sections 4 and 5 focussed primarily on the validation of the model inference method ALGO2, while Sect. 6 applied the model-based inference to a set of sources of astrophysical interest. In this section we revert to the complete methodology and examine the expected performance of FLAME in the context of *Gaia* data.

7.1. Inflation of input errors from GSP-PHOT and GSP-SPEC

As observed in GDR3, the formal uncertainties in atmospheric parameters can be of the order of 10 K in $T_{\rm eff}$ and less than 0.03 dex in $\log g$ and [M/H]. As explained in Andrae et al. (2023) and Recio-Blanco et al. (2023) these formal uncertainties reflect the S/N of the data and even if systematic errors may be present, these cannot be captured by the uncertainties in the fitting method. Andrae et al. (2023) explain that by comparing with external data, typical errors are of the order of 110 K for $T_{\rm eff}$, and 0.1–0.25 for $\log g$. While it is essential to report formal uncertainties for the atmospheric parameters that are derived directly from the data, these small uncertainties can have a negative impact in particular on the model-based FLAME parameters. Firstly, the uncertainty in the BC is often of the order of 0.001 mag, which means it contributes minimally to the total error budget for luminosity, leading to unrealistically small uncertainties. Secondly, in model inference, if the uncertainties are overly constrained, the resulting mass and age estimates from FLAME may be biased, as discussed in Sect. 4.3 for the solar case, and will likely have unrealistic error margins.

To address this, we have introduced an inflation parameter for the atmospheric parameter uncertainties in GDR4.

For GSP-Spec data, this has been introduced in a straightforward manner by adding in quadrature 60 K, 0.1 dex, and 0.1 dex in the $T_{\rm eff}$, $\log g$, and [M/H] parameters, respectively, before the BC is evaluated. This implies that the inflated uncertainties get propagated all the way to the mass and age inference.

For GSP-Phot, the approach differs, as Flame does not process the published values along with their confidence intervals. Instead, it directly processes the MCMC samples. As the MCMC samples retain correlations among the independent parameters, the inflation has to be done by maintaining these correlations. In practice this implies the multiplication of the distance of each sample point from its mean by a relative factor $T/T_{\rm eff}$. It, in effect, stretches the samples away from each other and maintains the correlations. The value of T has been determined empirically by using multiple data samples and comparing their mass and age parameters with those from Flame after scaling by the new Flame uncertainty. In the ideal case, the agreement between the datasets falls within $\pm 3\sigma$ for almost all of the targets. The chosen value of T corresponds to 50 K. The MCMC samples have five independent fitting parameters, $\{T_{\rm eff}$, [M/H], $\log g$, A_G, $d\}$, which are used in Flame. Thus, the inflation of the parameters has been applied by accounting for all these correlations.

7.2. Expected uncertainties

7.2.1. Luminosity

The main purpose of Algo1 is to derive the stellar luminosity. It depends on numerous observations, cf. Eqs. (1) and (2), namely $\{T_{\rm eff}, \log g$, [M/H], $f_G, \varpi\}$. The combination of the atmospheric parameter uncertainties are responsible for the uncertainty in BC, and then this along with f_G and ϖ determine $\sigma(L)$. The uncertainties $\sigma(f_G)$ are closely tied to the stellar magnitude, while $\sigma(\varpi)$ depends not only on magnitude, but also the spectral types and luminosity classes. For example, for solar-like stars the relative parallax error will increase with G, while for hot A stars we will also see a relative increase, but of a different order compared to the solar-like star case. In addition, the S/N of the input spectra from which the APs are derived also depend on G: the brighter the star, the higher the S/N, the smaller the uncertainties in $\{T_{\rm eff}, \log g$, [M/H]$\}$. However, the performance of the atmospheric parameter modules will also vary from star type to star type.

To show the performance of the determination of L, we considered a solar-like star, i.e. $|(L/L_\odot - 1)| < 0.15$ and `spectraltype_esphs` = 'G', and searched for a sample of these stars at $G = 9, 11, 13, 15, 17$ in GDR3. We used the mean value of the parameters `g_mean_flux_over_error`, `varpi`, and `varpi_error` as associated representative uncertainties at each of the G magnitudes. At the same time we also considered uncertainties on $T_{\rm eff}$, which vary between 10 and 200 K, and $\sigma_{\log g} = \sigma_{\rm [M/H]}$, which vary between 0.025 and 0.250 dex.

For each $T_{\rm eff}$ and $\log g$ error, we performed bootstrap simulations to derive the BC, and using the set of expected values for the astrometry and photometry, along with the Eqs. (1) and (2) we then calculated L. We used the standard deviation of L as an estimate of the expected uncertainty and then calculated the relative uncertainty, $\sigma(L)/L$.

This is illustrated in Fig. 10 where we show by the blue dashed lines $\sigma(L)/L$ as a function of $\sigma(T_{\rm eff})$ assuming $\sigma_{\log g} = \sigma$[M/H] = 0.05 dex for a solar-like star with $G = 11$. As $\sigma(T_{\rm eff})$ increases, $\sigma(L)/L$ varies from about 0.5% to 3%. The continuous line represents the results when we increase the $\sigma_{\log g}$ to 0.25 dex. As can be noted, the uncertainty in $\log g$ and [M/H] also play a small role in determining $\sigma(L)/L$.

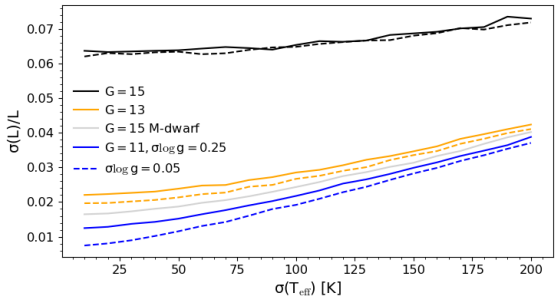


Fig. 10. Luminosity errors as a function of input $T_{\rm eff}$ uncertainty for different G.

The orange lines show results for an assumed solar-like star at $G = 13$. The difference between the impact of $\sigma_{\log g}$ going from 0.025 (dashed lines) to 0.250 dex (continuous lines) is much smaller than for the $G = 11$ case. Here, the parallax and photometry begin to have a more important role in $\sigma(L)/L$, and increase the minimum uncertainty to 2%. The black lines represent $G = 15$ and we can now see that there is virtually no impact due to varying $\sigma_{\log g}$ (no difference in dashed and continuous lines) while $\sigma_{T_{\rm eff}}$ still plays a small role by increasing $\sigma(L)/L$ from just over 6% to just over 7%. Doing the same exercise for a $G = 17$ star, we are entirely dominated by the parallax and photometry errors and $\sigma(L)/L$ remains at a constant 35% (not shown).

The results above are representative of predicted uncertainties for a solar-like star, using GDR3-like astrometry and photometric precision. To illustrate the predicted uncertainties for an M-dwarf for $G = 15$, we follow the same methodology and illustrate only the result for $\sigma_{\log g} = 0.25$ by the lightgrey solid line. Here $\sigma(L)/L$ varies from 1.5% to 4%. In this regime the parallaxes are quite large (these are nearby stars) and the parallax uncertainties therefore play a much smaller role. We expect all of these values to improve in GDR4 due to the expected increase in precision of the astrometric and photometric data.

7.2.2. Radius

As R is derived directly from L and $T_{\rm eff}$, through the Stefan–Boltzmann law, the uncertainties are straight forward to predict from a standard propagation of errors:

$$\sigma(R)/R = \sqrt{\frac{\sigma_L^2}{4L} + \frac{4\sigma_x^2}{x^2}}, \tag{7}$$

where $x = T_{\rm eff}/T_\odot$. By considering four of the limiting cases from the above shown in Fig. 10, we calculated for $G = 11$ with $\sigma_{T_{\rm eff}} = 50$ K and 200 K, $\sigma(R) = 0.02$ and 0.07 $R_\odot$, while for $G = 15$ it varies from 0.035 to 0.08 $R_\odot$; we considered that the star is 1$R_\odot$.

7.2.3. Mass

If we were to use a relation such as the $M - L$ relation for main-sequence stars (e.g. with an index of approximately three), we could approximate $\sigma(M) \sim \sqrt{\frac{\sigma_L^2}{4L}}$. For the cases illustrated above, where σ_L ranges from 0.014 to 0.070 $L_\odot$, $\sigma(M)$ would

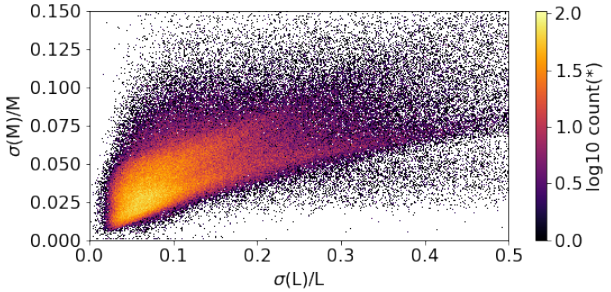


Fig. 11. Mass errors as a function of luminosity errors for a sample of approximately 10 million stars.

vary between 0.007 and 0.035 $M_\odot$. However, in FLAME the mass is constrained by a set of physical values set by the stellar evolution models. For a single metallicity, the mass (and age) are constrained additionally by $T_{\rm eff}$ and $\log g$. On top of that, the sensitivity of each of the observations vary with mass and evolution stage. To illustrate the typical uncertainties in mass that we obtain from FLAME, we show in Fig. 11 the results from one of the pre-operational test runs, described earlier in Sect. 5.3.1, on approximately 10 million stars spanning a wide range of masses, evolution stages, and metallicities with over 90% having $G < 17$. To make a comparable plot, we show how the relative uncertainty in M varies with relative uncertainty in L, with the colour code indicating the number of sources. Yellow indicates a bin (point-like) with approximately 100 stars while black indicates 1 source. The majority of the relative uncertainties from FLAME are typically less than 20–30% in L, while less than 10% in mass. We indeed observe similar σ_M/M to that from a $M - L$ relation, where typical 7% uncertainties result in between 2 and 4% relative mass uncertainties.

7.2.4. Age

Age uncertainties have a highly nonlinear dependence on the uncertainties in the other parameters. The uncertainty depends not only on the observational errors but also on the star's evolutionary stage and mass. Moreover, when the inferred age is near the boundaries of the model grid – whether very young or very old – the uncertainty limits are truncated at the upper or lower ends. To illustrate the expected performance of the age estimates, we use the same sample of approximately 10 million stars mentioned above and examine the age uncertainties within specific mass ranges. Additionally we focus on main-sequence stars and restrict the metallicity to [M/H] > −0.5, as metallicity significantly impacts a star's evolution and, consequently, its age.

Figure 12 illustrates the distribution of age uncertainties for mass ranges spanning the FGK range (different colours). The top panel shows normalised histograms of relative age uncertainty while the lower figure shows normalised histograms of the age error in Ga. As this figure illustrates, the lower mass stars tend to have very large relative uncertainties of the order of 70% which decreases slightly with mass. This is due to the relatively slow rate of change of the surface properties of these low-mass stars during the main sequence and thus the incapability of the observed properties to constrain the age well (given their unknown masses). The higher mass stars (red) will have relative age uncertainties of the order of 20% because $T_{\rm eff}$, L, and $\log g$ change relatively quickly as the star ages, and thus these properties allow us to efficiently narrow in on their ages. For

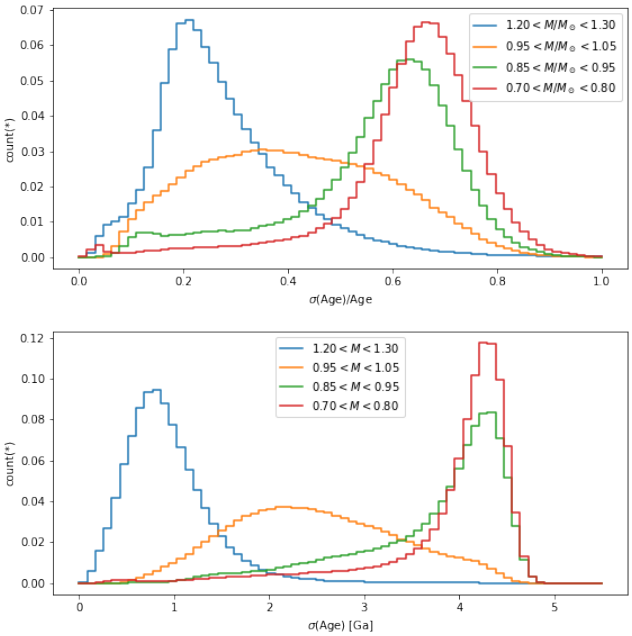


Fig. 12. Histograms illustrating the distributions of relative age uncertainty (*top*) and absolute age uncertainty (*bottom*) for main-sequence solar-metallicity stars in different mass ranges.

solar-like stars (green) we have a broad distribution of uncertainties spanning 15–60%, that depends not only on the input observational errors but the actual assigned age which statistically covers the full range of age. One can also note that despite the attempts to inflate the uncertainties, there are still some stars where the relative uncertainties are of the order of less than 5%, and this happens for approximately 1% of the 10 million sample.

If we generate the same figures for all stars beyond the main sequence (i.e. evolution stage >420), the distributions change significantly. For lower-mass stars (~0.8 $M_\odot$), the uncertainties peak around 10% and increase to 15–20% for higher masses.

8. Conclusions

The FLAME pipeline is one of the official software tools used within *Gaia*-DPAC to produce astrophysical parameters of sources that appear in the *Gaia* data releases. Specifically, the pipeline is designed to estimate luminosity, radius, radial velocity correction due to gravitational redshift, mass, age, and evolutionary status. In this paper, we laid out the architecture of the code, explained the methodology and models, and described the input *Gaia* data that is processed by the code. The code is separated into an analytical part and a model-inference part, which allows all the complexities of the model atmospheres and passbands to be encapsulated into a single value in the former – the BC. The model inference to estimate mass, age, and evolutionary status then compares the physical intrinsic properties of the star with those predicted from the BaSTI stellar evolution tracks, namely, the luminosity, metallicity, surface gravity, and effective temperature.

We then focussed our efforts on validating the model-based inference part of the code to estimate the mass and age. We did this by testing FLAME on simulated stars and the Sun in Sect. 4 and then by comparing our results with those of external catalogues in Sect. 5. We found that when we use the same input data as the external catalogue, we obtain masses and ages

that are in tight agreement with those of the external catalogues. However, when we use a source of input data different from that in the external catalogue, our results broadly agree in mass and age across main-sequence and sub-giant stars but the results for giants do not always agree. This is due to the fact that the evolved models all occupy a very tight region in the HR diagram during this phase, and therefore any small change in input data can lead to a very different optimal model. In addition, stars observed in the lower-right region of the HR diagram could be prone to large uncertainties on age due to the slow evolution of the observed surface properties.

In Sect. 6, we applied FLAME to several sets of stars in order to produce new luminosities, masses, and ages, including a high radial velocity metal-poor sample from Katz et al. (2025), a sample of stars with very low mass companions from Maxted (2025), and finally a sample of high-quality bright stars from the PLATO field of view. Their parameters are given in the appendix and made available at the CDS.

We discussed specifics of FLAME related to the GDR4 data, such as the addition of a metallicity offset to ensure consistency between the input and model metallicity (Sect. 2), the source of the atmospheric parameters (Sect. 3), the inflation of uncertainties, and the expected uncertainties in GDR4 (Sect. 7). Uncertainties in luminosity depend not only on the G magnitude, parallax, and parallax uncertainty but also on the spectral type and the quality of the input atmospheric parameters. The precision of the latter plays a very different role depending on the spectral type and magnitude of the star (which together intrinsically relate to the parallax). The precision in luminosity can reach as high as 1% for a $G = 11$ star G-type star (luminosity close to one), even when errors on the $T_{\rm eff}$ are around 50 K. For the same type of star with a magnitude of $G = 15$, the quality of the atmospheric parameters plays practically no role, showing a flat distribution around 6–7%. However, a $G = 15$ M-dwarf star can reach 1–2% uncertainties in luminosity. Uncertainties on the radius depend entirely on the precision in luminosity and $T_{\rm eff}$ because the Stefan-Boltzmann law is invoked. For *Gaia* data, the radius uncertainty can reach as low as 1–2%. Mass uncertainties roughly follow the expected uncertainty predicted by an analytical mass-luminosity relation, but stellar models help constrain the uncertainties better, and tests during the pre-operational development phase for GDR4 show typical mass uncertainties peak between 2% and 6%. For FGK stars, we expect to retrieve ages with no better than 50% for main-sequence stars with masses below 0.8 $M_\odot$. For masses between 1.2 and 1.3 $M_\odot$, expected precisions peak around 20%. For giants we find the opposite trend, with typical precisions of the order of 10% for stars with masses around 0.8 $M_\odot$ and the uncertainties increasing for higher masses.

FLAME is a powerful and efficient tool for deriving the masses and ages of stars from *Gaia* data. The results demonstrate that the derived masses are expected to be accurate to within 5%, while the ages can be estimated with a reliability of as good as 10%, provided the input data is of high quality. These findings underline the potential of FLAME to contribute significantly to the field of galactic archaeology with data from GDR4 and offer valuable insights into the history and evolution of our Galaxy and to the characterisation of exoplanetary systems.

Data availability

The full Tables C.1 and D.1 are available at the CDS via https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/714/A344

Acknowledgements. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. SC acknowledges financial support from PRIN-MIUR-22: CHRONOS: adjusting the clock(s) to unveil the CHRONO-chemo-dynamical Structure of the Galaxy" (PI: S. Cassisi) finanziato dall'Unione Europea – Next Generation EU, and Theory grant INAF 2023 (PI: S. Cassisi). OK acknowledges support by the Swedish Research Council (grant agreement no. 2023-03667) and by the Swedish National Space Agency. CN acknowledges financial support from the Centre national d'études spatiales (CNES), France (ROR: https://ror.org/04h1h0y33) within the framework of the *Gaia* mission. CBJ and RA acknowledge funding from the DLR (German space agency) via grant 50 QG 2102.

References

Aerts, C. 2021, Rev. Mod. Phys., 93, 015001
Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, A&A, 616, A8
Andrae, R., Fouesneau, M., Sordo, R., et al. 2023, A&A, 674, A27
Bailer-Jones, C. A. L., Andrae, R., Arcay, B., et al. 2013, A&A, 559, A74
Bijaoui, A., Recio-Blanco, A., de Laverny, P., & Ordenovic, C. 2010, in ADA 6 – Sixth Conference on Astronomical Data Analysis, eds. J. L. Starck, M. Saber Naceur, & R. Murtagh, 9
Borucki, W. J. 2016, Rep. Progr. Phys., 79, 036901
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Caffau, E., Ludwig, H. G., Steffen, M., Freytag, B., & Bonifacio, P. 2011, Sol. Phys., 268, 255
Casamiquela, L., Reese, D. R., Lebreton, Y., et al. 2024, A&A, 692, A243
Casamiquela, L., Soubiran, C., Jofré, P., et al. 2026, A&A, 705, A167
Collet, R., Nordlund, Å., Asplund, M., Hayek, W., & Trampedach, R. 2009, Mem. Soc. Astron. Italiana, 80, 719
Creevey, O. L., Cassisi, S., Thévenin, F., Salaris, M., & Pietrinferni, A. 2024, A&A, 689, A243
Creevey, O. L., & Lebreton, Y. 2022, Masses and ages from FLAME in DR3, Gaia DPAC Public Technical Notes. Online at GAIA-C8-TN-OCA-OLC-036
Creevey, O. L., Thévenin, F., Berio, P., et al. 2015, A&A, 575, A26
Creevey, O. L., Metcalfe, T. S., Schultheis, M., et al. 2017, A&A, 601, A67
Creevey, O. L., Sordo, R., Pailler, F., et al. 2023, A&A, 674, A26
Danielski, C., Babusiaux, C., Ruiz-Dern, L., Sartoretti, P., & Arenou, F. 2018, A&A, 614, A19
De Angeli, F., Weiler, M., Montegriffo, P., et al. 2023, A&A, 674, A2
Delchambre, L., Bailer-Jones, C. A. L., Bellas-Velidis, I., et al. 2023, A&A, 674, A31
Foreman-Mackey, D., Conley, A., Meierjurgen Farr, W., et al. 2013, emcee: The MCMC Hammer, Astrophysics Source Code Library [record ascl:1303.002]
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2023, A&A, 674, A28
Gaia Collaboration (Brown, A. G. A., et al.) 2016a, A&A, 595, A2
Gaia Collaboration (Prusti, T., et al.) 2016b, A&A, 595, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Gaia Collaboration (Creevey, O. L., et al.) 2023a, A&A, 674, A39
Gaia Collaboration (Vallenari, A., et al.) 2023b, A&A, 674, A1
Gallart, C., Bernard, E. J., Brook, C. B., et al. 2019, Nat. Astron., 3, 932
Grevesse, N., Asplund, M., & Sauval, A. J. 2007, Space Sci. Rev., 130, 105
Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, A&A, 582, A49
Hekker, S., & Ball, W. H. 2014, A&A, 564, A105
Helmi, A. 2020, ARA&A, 58, 205
Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, Nature, 563, 85
Hidalgo, S. L., Pietrinferni, A., Cassisi, S., et al. 2018, ApJ, 856, 125
Jofré, P., Heiter, U., Soubiran, C., et al. 2014, A&A, 564, A133
Jofré, P., Heiter, U., Soubiran, C., et al. 2015, A&A, 582, A81
Jofré, P., Heiter, U., Worley, C. C., et al. 2017, A&A, 601, A38
Katz, D., Sartoretti, P., Guerrier, A., et al. 2023, A&A, 674, A5
Katz, D., Gómez, A., Caffau, E., et al. 2025, A&A, 704, A294
Lagarde, N., Decressin, T., Charbonnel, C., et al. 2012, A&A, 543, A108
Lagarde, N., Robin, A. C., Reylé, C., & Nasello, G. 2017, A&A, 601, A27
Lebreton, Y., & Goupil, M. J. 2014, A&A, 569, A21
Lebreton, Y., & Reese, D. R. 2020, A&A, 642, A88
Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, A&A, 649, A2
Lodders, K. 2010, in Principles and Perspectives in Cosmochemistry, Astrophysics and Space Science Proceedings, 16, 379
Lund, M. N., Chontos, A., Grundahl, F., et al. 2025, A&A, 701, A285
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94

Mamajek, E. E., Torres, G., Prsa, A., et al. 2015, arXiv e-prints [arXiv:1510.06262]
Maxted, P. F. L. 2025, MNRAS, 539, 3741
Metcalfe, T. S., Creevey, O. L., Doğan, G., et al. 2014, ApJS, 214, 27
Pinsonneault, M. H., Elsworth, Y., Epstein, C., et al. 2014, ApJS, 215, 19
Pinsonneault, M. H., Elsworth, Y. P., Tayar, J., et al. 2018, ApJS, 239, 32
Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2023, A&A, 673, A155
Rauer, H., Aerts, C., Cabrera, J., et al. 2025, Exp. Astron., 59, 26
Recio-Blanco, A., de Laverny, P., Allende Prieto, C., et al. 2016, A&A, 585, A93
Recio-Blanco, A., de Laverny, P., Palicio, P. A., et al. 2023, A&A, 674, A29
Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3
Rowell, N., Davidson, M., Lindegren, L., et al. 2021, A&A, 649, A11
Salaris, M., Chieffi, A., & Straniero, O. 1993, ApJ, 414, 580
Salpeter, E. E. 1955, ApJ, 121, 161
Sartoretti, P., Katz, D., Cropper, M., et al. 2018, A&A, 616, A6
Sartoretti, P., Marchal, O., Babusiaux, C., et al. 2023, A&A, 674, A6
Searle, L., & Zinn, R. 1978, ApJ, 225, 357
Serenelli, A., Johnson, J., Huber, D., et al. 2017, ApJS, 233, 23
Shulyak, D., Tsymbal, V., Ryabchikova, T., Stütz, C., & Weiss, W. W. 2004, A&A, 428, 993
Soubiran, C., Creevey, O. L., Lagarde, N., et al. 2024, A&A, 682, A145
Torra, F., Castañeda, J., Fabricius, C., et al. 2021, A&A, 649, A10
Ulla, A., Creevey, O. L., Álvarez, M. A., et al. 2022, Gaia DR3 documentation Chapter 11: Astrophysical parameters, Gaia DR3 documentation, European Space Agency; Gaia Data Processing and Analysis Consortium. Online at https://gea.esac.esa.int/archive/documentation/GDR3/index.html, chap. 11
Ulrich, R. 1997, in KITP Conference: Solar Neutrinos: News About SNUs, eds. J. Bahcall, W. Haxton, P. Langacker, H. Robertson, & Y. Totsuka, 28
VandenBerg, D. A., Bond, H. E., Nelan, E. P., et al. 2014, ApJ, 792, 110
Vergely, J. L., Lallement, R., & Cox, N. L. J. 2022, A&A, 664, A174
Zwintz, K., Aerts, C., Tkachenko, A., et al. 2026, arXiv e-prints [arXiv:2604.04042]

Appendix A: Author contributions

The authors of this paper have contributed in many ways to the FLAME software and feedback on performance. The FLAME team comprises OLC (scientific lead), CO (main java developer), FT, YL, LC, BP, and have contributed significantly to all aspects mentioned below. Significant support in the development of its integration into the CNES framework and scientific validation framework has been provided by NBa, FP, CR, CBJ, MF, RA, and AB. BE, OK, SC, AK, RS have contributed to the provision of input models, bolometric corrections, and simulations. DR has provided software support for the python-java development of the SPInS code. NBr, AL, GK, CS, RS, AK, CN, RA, MF, CB have contributed to the scientific validation and feedback loop. All authors have contributed to writing this manuscript.

Appendix B: Example queries for good practice use of FLAME results

Here we provide some example uses of good practice of FLAME results using GDR3 and GDR4.
Select a high quality HR diagram using FLAME parameters. In GDR3 the query is:

```
select top 10000 lum_flame, gs.teff_gspphot
from gaiadr3.astrophysical_parameters as ap
inner join gaiadr3.gaia_source as gs
              on gs.source_id = ap.source_id
where lum_flame is not null and
random_index < 10000000 and
phot_g_mean_flux_over_error > 2000 and
parallax_over_error > 20 and
phot_bp_mean_flux_over_error > 1300 and
phot_rp_mean_flux_over_error > 1500
```

In GDR4 the query is:

```
select top 10000 lum_flame, teff_gspphot
from gaiadr4.ap_xp as ap
inner join gaiadr4.all_source_astrometry as ast
             on ast.source_id = ap.source_id
where random_index_as < 1000000  and
flags_flame like '0___'
```

Select specific examples of variable stars that are likely to have non-reliable FLAME results. In GDR3 the query is:

```
select top 10000 ap.source_id
from gaiadr3.astrophysical_parameters as ap
inner join gaiadr3.vari_summary as vs
                on vs.source_id = ap.source_id
where lum_flame is not null and
(in_vari_eclipsing_binary = 'True' OR
                in_vari_rrlyrae = 'True')
```

In GDR4 the query is:

```
select top 10000 ap.source_id
from gaiadr4.ap_xp as ap
inner join gaiadr4.vari_summary as vs
                on vs.source_id = ap.source_id
where lum_flame is not null and
(in_vari_eclipsing_binary = 'True' OR
                in_vari_rrlyrae = 'True')
```

Appendix C: FLAME stellar parameters of the GBS

Table C.1. New masses and ages of a subset of the GBSs (extract).

GDR3 source ID	$M_\odot$ [$M_\odot$]	Age [Ga]
6904703228801028864	1.096 ± 0.015	7.8 ± 0.6
2195115561168483712	0.987 ± 0.038	11.0 ± 1.6
2190891581091291520	1.201 ± 0.044	4.7 ± 0.6
1872046609345556480	0.661 ± 0.009	10.7 ± 2.3
1872046574983497216	0.606 ± 0.011	7.8 ± 4.0
1787990525934491776	0.754 ± 0.007	13.0 ± 0.7
2278918344568877056	1.595 ± 0.060	1.8 ± 0.1
6412595290592307840	0.702 ± 0.008	12.1 ± 1.6
1895198193816300544	1.234 ± 0.020	3.1 ± 0.8
1907131544341497600	4.097 ± 0.168	0.2 ± 0.1
1982252790090609152	1.328 ± 0.038	3.7 ± 0.3
1876331990259358976	4.143 ± 0.134	0.2 ± 0.0
2719475542667772416	1.163 ± 0.016	4.7 ± 0.3

Notes. The full table of 140 stars is available at the CDS. See Sect. 5.1 for details.

Appendix D: Mass and age of scvPIC stars

Table D.1. Newly derived masses and ages of a subset of the PLATO scvPIC (extract; Zwintz et al. 2026).

GDR3 source ID	$M_\odot$ [$M_\odot$]	Age [Ga]
5318186221414047104	0.977 ± 0.013	4.6 ± 0.9
5317884439832479872	0.927 ± 0.048	7.5 ± 3.8
5299155152615066752	1.234 ± 0.008	2.7 ± 0.2
5537152725129555712	0.804 ± 0.010	13.1 ± 0.7
5537149289155845888	0.959 ± 0.005	8.5 ± 0.3
4823123689293923968	0.903 ± 0.027	10.8 ± 2.3
4777119126354782592	0.766 ± 0.004	13.4 ± 0.4
4791496409119237888	0.741 ± 0.013	12.4 ± 1.2
5538812506654586624	1.230 ± 0.037	1.6 ± 0.7
5558725074543077632	1.260 ± 0.041	2.3 ± 0.5

Notes. The full table of 338 stars is available at the CDS. See Sect. 6.3 for details.

Appendix E: Mass and age of high-velocity stars

Appendix F: Mass and age of systems with low-mass companions

Table E.1. Newly derived luminosities, masses, and ages of the high-velocity stars from Katz et al. (2025).

Star	GDR3 source ID	[M/H]	$L_\star$ [$L_\odot$]	$M_\odot$ [$M_\odot$]	Age [Ga]
Gaia_2	1336408284224866432	-1.783	2.145 ± 0.036	0.807 ± 0.054	10.4 ± 2.9
Gaia_108	139417803820690048	-0.366	1.743 ± 0.029	0.959 ± 0.075	8.8 ± 3.4
Gaia_45	1839706467664901248	-1.126	124.907 ± 2.108	0.839 ± 0.032	12.1 ± 1.2
Gaia_41	2726872575862585856	-1.634	4.396 ± 0.074	0.836 ± 0.038	10.8 ± 1.3
Gaia_56	2618198324041826432	-1.926	1.162 ± 0.020	0.764 ± 0.036	10.5 ± 3.0
Gaia_6151	1619258337664073088	-2.475	1.972 ± 0.033	0.798 ± 0.042	9.7 ± 2.6
Gaia_54	1235633782930379776	-2.181	0.819 ± 0.014	0.765 ± 0.048	6.4 ± 4.6
Gaia_1120	1763032952014500096	-1.905	0.356 ± 0.006	0.674 ± 0.025	5.6 ± 4.4
Gaia_42	1992247007192756224	-1.572	31.110 ± 0.525	0.831 ± 0.031	11.2 ± 1.3
Gaia_42	1992247007192756224	-1.595	31.110 ± 0.525	0.831 ± 0.031	11.2 ± 1.3
Gaia_66	2033964849204648064	-0.648	0.977 ± 0.016	0.834 ± 0.057	9.8 ± 3.5
Gaia_53	2038062247987954048	-1.758	3.054 ± 0.052	0.788 ± 0.030	12.5 ± 1.6
Gaia_11	1953616147184247808	-1.558	113.432 ± 1.914	0.834 ± 0.034	11.1 ± 1.4
Gaia_646	4510483566596988416	-0.378	6.023 ± 0.102	1.268 ± 0.165	3.8 ± 1.0
Gaia_32	1413748207000103808	-2.004	0.399 ± 0.007	0.688 ± 0.028	5.5 ± 4.4
Gaia_55	1420093656106274176	-0.611	1.216 ± 0.021	0.863 ± 0.056	9.8 ± 3.6
Gaia_1	1353777750444932736	-1.294	71.310 ± 1.204	0.844 ± 0.033	11.5 ± 1.5
Gaia_47	2145301671471881984	-1.330	46.579 ± 0.786	0.935 ± 0.027	7.8 ± 0.6
Gaia_1220	4250930137323602944	-1.790	163.637 ± 2.762	0.838 ± 0.040	10.7 ± 1.8
Gaia_2799	946807746355535360	-0.771	3.396 ± 0.057	1.000 ± 0.094	7.3 ± 1.6
Gaia_2082	187566551874865664	-0.181	4.006 ± 0.068	1.093 ± 0.112	5.4 ± 1.4
Gaia_26	1814359288672674560	-0.962	409.346 ± 6.909	0.929 ± 0.027	10.8 ± 0.8
Gaia_2620	3069023032306067968	-1.597	55.183 ± 0.931	1.069 ± 0.035	4.7 ± 0.5
Gaia_8	1768154507240342528	-1.750	60.189 ± 1.016	0.830 ± 0.034	11.1 ± 1.4
Gaia_3444	340605204521472640	-0.944	2.144 ± 0.036	0.880 ± 0.076	10.6 ± 2.2
Gaia_1288	2095628244392589824	-1.708	284.018 ± 4.793	0.826 ± 0.026	11.5 ± 1.1
Gaia_651	1907386149999121920	0.015	0.639 ± 0.011	0.939 ± 0.039	4.5 ± 4.0
Gaia_4892	1977970978277811840	-1.041	398.908 ± 6.732	0.927 ± 0.024	12.8 ± 0.9
Gaia_234	2228292706060074112	-1.645	382.410 ± 6.454	0.813 ± 0.014	13.3 ± 0.7
Gaia_57	4288765848587649664	-0.589	1.407 ± 0.024	0.933 ± 0.075	6.5 ± 4.3
Gaia_360	3436807447014964 48	-1.068	1.475 ± 0.025	0.837 ± 0.067	10.3 ± 3.8
Gaia_50	4500737499823974656	-1.652	52.292 ± 0.883	0.856 ± 0.035	10.4 ± 1.4
Gaia_5857	3779863003378779392	-1.060	2.331 ± 0.039	0.879 ± 0.080	9.4 ± 2.8
Gaia_3	1335426489060934272	-1.888	1055.064 ± 17.806	0.800 ± 0.018	12.2 ± 0.9
Gaia_16	1778657836877668096	-1.352	37.086 ± 0.626	1.605 ± 0.054	1.2 ± 0.1
Gaia_14	1984025890028941184	-1.634	43.450 ± 0.733	0.819 ± 0.022	11.8 ± 1.0
Gaia_2150	3644341873063382272	-0.768	1.407 ± 0.024	0.855 ± 0.069	8.9 ± 3.2
Gaia_46	1839321466796403072	-1.142	230.435 ± 3.889	0.827 ± 0.023	12.9 ± 1.3
Gaia_1671	1863896960443202816	-1.656	1.440 ± 0.024	0.791 ± 0.048	9.5 ± 3.4
Gaia_379	4606414581726715264	-2.502	206.256 ± 3.481	0.822 ± 0.033	11.1 ± 1.5
Gaia_23	1345432216672749184	-1.575	121.314 ± 2.047	1.246 ± 0.038	2.8 ± 0.3
Gaia_4550	2699001399207617792	-1.079	29.349 ± 0.495	0.837 ± 0.030	12.1 ± 1.3
Gaia_51	1776283372797846272	-1.730	205.049 ± 3.461	0.818 ± 0.021	11.7 ± 0.9
Gaia_28	4608346805911470080	-2.861	49.595 ± 0.837	1.530 ± 0.053	1.4 ± 0.2
Gaia_3033	3910720318902016000	-2.602	4.912 ± 0.083	0.785 ± 0.018	12.4 ± 1.0
Gaia_64	4205559241036808320	-0.633	0.564 ± 0.010	0.794 ± 0.049	7.8 ± 4.5
Gaia_4960	2668731259979103872	-2.455	2.625 ± 0.044	0.798 ± 0.041	10.6 ± 2.3
Gaia_1319	2047058074180427904	-2.261	3.093 ± 0.052	0.778 ± 0.021	12.4 ± 1.3
Gaia_1024	1421352528201172608	-1.294	92.496 ± 1.561	0.840 ± 0.032	11.5 ± 1.3
Gaia_2616	3070924882479137408	-1.613	31.118 ± 0.525	0.844 ± 0.033	10.9 ± 1.4
Gaia_36	4236626594855062656	-1.586	261.964 ± 4.421	0.834 ± 0.041	11.4 ± 1.7

Notes. See Sect. 6.1 for details.

Table F.1. Newly derived individual masses and system age of stellar systems with low-mass companions using Maxted (2025).

Star	GDR3 source ID	q	$M_{\rm comp}$ [$M_\odot$]	$M_{\rm primary}$ [$M_\odot$]	Age [Ga]
EBLM_J0057-19	2356249333811520512	0.126	0.133	1.059 ± 0.027	8.3 ± 1.1
EBLM_J0113+31	313579414867698176	0.190	0.196	1.034 ± 0.018	6.8 ± 0.3
EBLM_J0123+38	323261744326472832	0.262	0.342	1.305 ± 0.007	5.0 ± 0.1
EBLM_J0228+05	2517108778307431168	0.119	0.171	1.435 ± 0.028	1.3 ± 0.2
EBLM_J0500-46	4810245453195049472	0.162	0.169	1.049 ± 0.027	7.1 ± 0.7
EBLM_J0526-34	4822911449189918208	0.252	0.323	1.284 ± 0.018	3.8 ± 0.1
EBLM_J0540-17	2970781118606872192	0.140	0.176	1.258 ± 0.024	4.1 ± 0.3
EBLM_J0608-59	5494443978353833088	0.282	0.324	1.148 ± 0.018	4.7 ± 0.3
EBLM_J0627-67	5280290694161913472	0.246	0.282	1.148 ± 0.041	3.6 ± 0.4
EBLM_J0719+25	870991952855398656	0.144	0.161	1.118 ± 0.018	5.8 ± 0.4
EBLM_J0941-31	5440085875822885888	0.172	0.235	1.363 ± 0.023	3.2 ± 0.1
EBLM_J0955-39	5419963580383767936	0.192	0.205	1.071 ± 0.026	1.7 ± 0.5
EBLM_J1013+01	3835492798481111168	0.180	0.164	0.915 ± 0.019	11.2 ± 0.9
EBLM_J1305-31	6180966394355426816	0.254	0.281	1.107 ± 0.013	7.9 ± 0.3
EBLM_J1522+42	1393971390134144640	0.165	0.162	0.983 ± 0.015	10.6 ± 0.6
EBLM_J1928-38	6739146911148825344	0.266	0.277	1.041 ± 0.015	8.9 ± 0.4
EBLM_J1934-42	6689082332805585280	0.177	0.182	1.024 ± 0.017	4.6 ± 1.2
EBLM_J2040-41	6678103674840655360	0.152	0.147	0.969 ± 0.015	10.0 ± 0.6
EBLM_J2046+06	1735977265094166912	0.200	0.278	1.394 ± 0.021	0.7 ± 0.2
EBLM_J2046-40	6678572680972611328	0.152	0.154	1.008 ± 0.008	13.6 ± 0.2
EBLM_J2217-04	2626910437568266240	0.199	0.220	1.105 ± 0.013	8.9 ± 0.3
EBLM_J2315+23	2839193907753424384	0.209	0.232	1.113 ± 0.015	6.8 ± 0.3
EBLM_J2343+29	2867806567563054720	0.114	0.096	0.842 ± 0.015	9.0 ± 1.8
EBLM_J2359+44	1923017735011519488	0.218	0.316	1.453 ± 0.018	2.1 ± 0.1
EPIC_219654213	4183945728602365568	0.172	0.227	1.319 ± 0.031	2.8 ± 0.1
NGTS-EB-7	5504617415848984320	0.088	0.094	1.067 ± 0.022	9.8 ± 0.6
WASP-30	2434845586060285824	0.049	0.061	1.249 ± 0.026	2.9 ± 0.5
CoRoT-1	3105507886130792448	0.001	0.001	1.206 ± 0.012	1.7 ± 0.2
HAT-P-7	2129256395211984000	0.001	0.002	1.543 ± 0.009	2.1 ± 0.1
HD_189733	1827242816201846144	0.001	0.001	0.825 ± 0.009	4.4 ± 0.9
HD_209458	1779546757669063552	0.001	0.001	1.145 ± 0.009	2.6 ± 0.3
Kepler-1	2131314401800665344	0.001	0.001	1.072 ± 0.008	1.4 ± 0.3
WASP-4	6535499658122055552	0.001	0.001	0.941 ± 0.010	5.9 ± 0.8
WASP-18	4955371367334610048	0.007	0.010	1.326 ± 0.009	0.2 ± 0.1
WASP-43	3767805209112436736	0.003	0.002	0.658 ± 0.014	11.5 ± 2.3
KOI-205	2077896798333728640	0.042	0.039	0.921 ± 0.028	3.4 ± 2.1

Notes. q is the mass ratio from Maxted (2025). See Sect. 6.2 for details.